به نام خدا

راهنما و مجموعه خلاصه مقالات و سخنرانی‌های

چهارمین کنفرانس ملی انفورماتیک ایران

پژوهشکده علوم کامپیوتر
پژوهشگاه دانش‌های بنیادی
پردیس فرمانیه، تهران

۱۴ و ۱۵ دی‌ماه ۱۴۰۱

برگزار کنندگان

IPM
پژوهشگاه دانش‌های بنیادی

انجمن انفورماتیک ایران

حامیان

ترمس مکانیک هفشجان
شماره ثبت : ۵۹۸۸

ILMAN
ILMAN TEJARAT BORNA CO.

شرکت ذوق صنعت هفشجان
(سهامی خاص)
ثبت ۳۵۸۵

فهرست مطالب

عنوان **صفحه**

پیام دبیران کنفرانس .... ت
برنامه کنفرانس در یک نگاه .... ج
برگزارکنندگان کنفرانس .... چ
محورهای کنفرانس .... ذ
سخنرانی‌های علمی .... ز
برنامه زمانی کنفرانس .... س
میزگرد .... ض
مقالات کنفرانس ۱

پیام دبیران کنفرانس

با کمال خوشوقتی برگزاری چهارمین کنفرانس ملی انفورماتیک ایران را طی روزهای ۱۴و۱۵ دی ماه ۱۴۰۱ به اطلاع میرسانیم. این کنفرانس با همت و همکاری پژوهشگاه دانش‌های بنیادی و انجمن انفورماتیک ایران برگزار میگردد. هدف اصلی از برگزاری این رویداد، گردهم آوری متخصصان، پژوهشگران و صنعتگران مرتبط با این حوزه مهم از علوم و مهندسی می‌باشد تا در محیطی علمی و تخصصی به تبادل مسائل و یافته‌های مهم خود بپردازند.

در این کنفرانس، بر اساس دسته بندی مرسوم در انجمن‌ها و محافل تخصصی انفورماتیک، مقالات در چهار شاخه ۱) سیستم، ۲) تئوری، ۳) هوش مصنوعی و ۴) زمینه‌های بین رشته‌ای، ارسال و داوری شدند. برای هر یک از مقالات ارسالی در هر شاخه، داوری حرفه‌ای با حداقل سه داور انجام شد و نهایتا ۲۷ مقاله که از کیفیت بسیار بالایی برخوردار بودند برای ارائه شفاهی در کنفرانس پذیرفته شدند.

در کنار ارائه مقالات عادی کنفرانس، کمیته علمی کنفرانس اقدام به دعوت از دانشمندان و پژوهشگران فعال و شناخته شده داخل و خارج کشور در این حوزه برای ارائه سخنرانی‌های علمی کرد. خوشوقتیم به اطلاع برسانیم که در نتیجه این اقدام، ۵ سخنرانی علمی بسیار جذاب توسط پژوهشگران فعال در حوزه انفورماتیک در برنامه کنفرانس گنجانده شد. در روز دوم کنفرانس علاوه بر سخنرانیهای در نظر گرفته شده، یک میزگرد در خصوص علوم و فناوری‌های همگرا با حضور افراد شناخته شده در این حوزه خواهیم داشت.

بلافاصله پس از برگزاری جلسات کنفرانس در روزهای ۱۴ و ۱۵ دی ماه، در روز جمعه ۱۶ دی ماه، ۳ کارگاه آموزشی با موضوعهای متنوع و جذاب، توسط دانشمندان و متخصصان متبحر حوزه انفورماتیک در دو نوبت صبح و بعدازظهر برگزار خواهد شد.

مراتب سپاس خود را از سخنران افتتاحیه جناب آقای دکتر بهروز پرهامی از دانشگاه سانتا باربارا که دعوت کنفرانس را پذیرفتند تا در افتتاحیه کنفرانس به ایراد سخنرانی بپردازند ابراز میکنیم. همچنین از دکتر مهدی جلیلی از دانشگاهRMIT ، دکترسید مجید زاهدی از دانشگاه واترلو ، دکتر پرهام مرادی از دانشگاهRMIT و دکتر حمید بیگی از دانشگاه صنعتی شریف که به دعوت کنفرانس پاسخ مثبت دادند و با ایراد سخنرانی، جدیدترین یافته‌های پژوهشی خود را ارائه دادند کمال تشکر و قدردانی را می‌نماییم.

همچنین لازم میدانیم از دکتر بابک حسین خلج، دکتر علیرضا ولی‌زاده و دکتر میثم فرج‌اللهی که با شرکت در میزگرد به هرچه پربارتر شدن این رویداد کمک کردند، تشکر و قدردانی نماییم.

مایلیم از اعضای کمیته‌های مختلف کنفرانس و بخصوص کمیته علمی، مسئولین شاخه‌های آن و داوران مقالات که با دقت بسیار به بررسی مقالات و تشکیل برنامه کنفرانس پرداختند تشکر کنیم و قدردان نویسندگان مقالات ارسال شده، چه آنان که پذیرفته شدند و چه آنان که پذیرفته نشدند، می‌باشیم.

همچنین از حامیان کنفرانس، که با حمایت مالی خود در برگزاری مناسب کنفرانس ما را یاری کردند تشکر می‌کنیم.

در انتها از تمامی عزیزانی که به هر نحو ممکن ما را در برگزاری کنفرانس یاری کردند و بخصوص شما شرکت کنندگان در جلسات مختلف کنفرانس و کارگاه‌ها تشکر میکنیم و امیدواریم برنامه‌های این کنفرانس مورد استفاده و توجه شما قرار گیرد.

دکتر پژمان لطفی کامران، دبیر کنفرانس

دکتر حمید سربازی آزاد، دبیر کمیته علمی

برنامه‌ی کنفرانس در یک نگاه

روز اول: چهارشنبه ۱۴ دی‌ماه ۱۴۰۱	
۹:۰۰-۱۰:۰۰	افتتاحیه کنفرانس و سخنرانی افتتاحیه
۱۰:۰۰-۱۱:۳۰	ارایه مقالات گروه هوش مصنوعی
۱۱:۳۰-۱۲:۰۰	استراحت
۱۲:۰۰-۱۲:۴۵	سخنرانی کلیدی ۱
۱۲:۴۵-۱۴:۰۰	استراحت
۱۴:۰۰-۱۵:۱۵	ارایه مقالات گروه سیستم
۱۵:۱۵-۱۵:۴۵	استراحت
۱۵:۴۵-۱۶:۴۵	ارایه مقالات گروه زمینه های بین رشته ای
۱۶:۴۵-۱۷:۳۰	سخنرانی کلیدی ۲
روز دوم: پنج‌شنبه ۱۵ دی‌ماه ۱۴۰۱	
۹:۰۰-۱۰:۱۵	ارایه مقالات گروه هوش مصنوعی
۱۰:۱۵-۱۰:۴۵	استراحت
۱۰:۴۵-۱۱:۳۰	سخنرانی کلیدی ۳
۱۱:۳۰-۱۲:۳۰	ارایه مقاالت علمی گروه تئوری و پژوهش دانشجویان دکترا
۱۲:۳۰-۱۴:۰۰	استراحت
۱۴:۳۰-۱۶:۰۰	میزگرد
۱۶:۰۰-۱۶:۴۵	سخنرانی کلیدی ۴
۱۶:۴۵-۱۷:۱۵	استراحت
۱۷:۱۵-۱۸:۱۵	ارایه مقالات گروه هوش مصنوعی
۱۸:۱۵-۱۸:۳۰	اختتامیه کنفرانس

برگزارکنندگان کنفرانس

روسای کنفرانس: محمدجواد اردشیر لاریجانی، رئیس پژوهشگاه دانش‌های بنیادی
اسلام ناظمی، عضو هیات‌مدیره انجمن انفورماتیک ایران

دبیر کنفرانس: پژمان لطفی‌کامران، پژوهشگاه دانش‌های بنیادی

دبیر کمیته علمی: حمید سربازی‌آزاد، دانشگاه صنعتی شریف و پژوهشگاه دانش‌های بنیادی

دبیران شاخه‌های علمی:
دکتر مهدی مدرسی، دانشگاه تهران (شاخه سیستم)
دکتر محسن ابراهیمی مقدم، دانشگاه شهید بهشتی (شاخه هوش مصنوعی)
دکتر حمید بیگی، دانشگاه صنعتی شریف (شاخه تئوری)
دکتر محمد عبدالهی‌ازگمی، دانشگاه علم و صنعت (شاخه زمینه‌های بین‌رشته‌ای)

کمیته اجرایی:
سید محمد حسینی و محسن باقری پژوهشگاه دانش‌های بنیادی (تبلیغات)
نظام رهبانی، پژوهشگاه دانش‌های بنیادی (اینترنت)
ابراهیم اردشیر لاریجانی، دانشگاه علم و صنعت ایران (میزگردها)
سمیرا حسین‌قربان و فاطمه بهاری‌فرد ، پژوهشگاه دانش‌های بنیادی (ارتباط با صنعت و امور حامیان)
فاطمه آقایی‌پور، پژوهشگاه دانش‌های بنیادی (انتشارات)
نعیمه امیدوار، پژوهشگاه دانش‌های بنیادی (پژوهش دانشجویان دکترا)
سپیده صفری، پژوهشگاه دانش‌های بنیادی (ارتباط با پایگاه استنادی جهان اسلام)
امیرحسین هادیان، پژوهشگاه دانش‌های بنیادی (کارگاه‌ها)
سلما سادات مهدوی، پژوهشگاه دانش‌های بنیادی (مسئول برنامه‌ریزی و ارتباطات)
زهرا رضوانی و حامد رستگار، پژوهشگاه دانش‌های بنیادی و محمد خالویی، دانشگاه صنعتی امیرکبیر (سامانه مجازی)
حمیدرضا شهرابی‌فراهانی، پژوهشگاه دانش‌های بنیادی (امور اجرایی)

اعضای کمیته علمی

اعضای کمیته علمی شاخه‌ی سیستم:

- مصطفی ارسالی صالحی نسب (دانشگاه تهران)
- محسن انصاری (دانشگاه صنعتی شریف)
- امیر امینی‌فر (دانشگاه Lund-سوئد)
- سیاوش بیات سرمدی (دانشگاه صنعتی شریف)
- حاکم بیت‌اللهی (دانشگاه علم و صنعت)
- فرهاد پاکدامن (دانشگاه تهران)
- سید محمد حسینی (پژوهشگاه دانش‌های بنیادی)
- مسعود دانشطلب (دانشگاه mdh-سوئد)
- مهدی دولتی (پژوهشگاه دانش‌های بنیادی)
- مسعود دهیادگاری (دانشگاه صنعتی خواجه نصیر)
- دارا رحمتی (دانشگاه شهید بهشتی)
- مهدی رضایی (پژوهشگاه دانش‌های بنیادی)
- نظام رهبانی (پژوهشگاه دانش‌های بنیادی)
- حمیدرضا زرندی (دانشگاه صنعتی امیرکبیر)
- سیما سینایی (دانشگاه mdh-سوئد)
- حسین شفیعی (دانشگاه صنعتی خواجه‌نصیرالدین طوسی)
- سپیده صفری (پژوهشگاه دانش‌های بنیادی)
- حامد فربه (دانشگاه صنعتی امیرکبیر)
- هاجر فلاحتی (پژوهشگاه دانش‌های بنیادی)
- مهدی کمال (دانشگاه تهران)
- سعید گرگین (سازمان پژوهش‌های علمی و صنعتی ایران)
- پرهام مرادی (دانشگاه کردستان)
- محمود نادران طهان (دانشگاه شهید چمران اهواز)

اعضای کمیته علمی شاخه‌ی هوش مصنوعی:

- فاطمه آقایی‌پور (پژوهشگاه دانش‌های بنیادی)
- نعیمه امیدوار (پژوهشگاه دانش‌های بنیادی)
- سعید بیگدلی (شرکت ایزایران)
- زهرا رضوانی (پژوهشگاه دانش‌های بنیادی)
- علیرضا رضوانیان (دانشگاه علم و فرهنگ)
- زهرا ریاحی (دانشگاه پنسیلوانیا)
- محمد سبکرو (پژوهشگاه دانش‌های بنیادی)
- یاسر شکفته (دانشگاه شهید بهشتی)
- مهرنوش شمس فرد (دانشگاه شهید بهشتی)
- احمد علی آبین (دانشگاه شهید بهشتی)
- محمد مهدی فقیه (دانشگاه صنعتی کرمان)
- هشام فیلی (دانشگاه تهران)
- حامد ملک (دانشگاه شهید بهشتی)
- آرمین سلیمی بدر (دانشگاه شهید بهشتی)
- رضا خسروآبادی (دانشگاه شهید بهشتی)
- علیرضا طالب‌پور (دانشگاه شهید بهشتی)
- محسن سریانی (دانشگاه علم و صنعت ایران)
- منصور جم‌زاد (دانشگاه صنعتی شریف)
- حسام عمران‌پور (دانشگاه نوشیروانی)

ا عضای کمیته علمی شاخه‌ی تئوری:

- محمدعلی آبام (دانشگاه صنعتی شریف) آرش احدی (دانشگاه خوارزمی)
- ابراهیم اردشیر لاریجانی (دانشگاه علم و صنعت ایران)
- محمد ایزدی (دانشگاه صنعتی شریف)
- علیرضا باقری (دانشگاه صنعتی امیرکبیر)
- فاطمه بهاری‌فرد (پژوهشگاه دانش‌های بنیادی)
- سمیرا حسین‌قربان (پژوهشگاه دانش‌های بنیادی)
- منصور داودی منفرد (دانشگاه تحصیلات تکمیلی علوم پایه زنجان)
- زاهد رحمتی (دانشگاه صنعتی امیرکبیر)مسعود صدیقین (دانشگاه خاتم)

- حمید ضرابی‌زاده (دانشگاه صنعتی شریف)
- مریم طهماسبی آبدر (دانشگاه شهید بهشتی)
- شراره علیپور (دانشگاه خاتم)
- علی غلامی رودی (دانشگاه صنعتی نوشیروانی بابل)
- امین غیبی (دانشگاه صنعتی امیرکبیر) محمد فرشی (دانشگاه یزد)
- محمد هادی فروغمند اعرابی (دانشگاه صنعتی شریف)
- محمدمهدی کیخا (دانشگاه سیستان و بلوچستان)
- احمد مرادی (دانشگاه مازندران)
- معصومه مرادیان (پژوهشگاه دانش‌های بنیادی)
- سلما سادات مهدوی (پژوهشگاه دانش‌های بنیادی)
- عباس نوذری دالینی (دانشگاه تهران)
- مصطفی نوری بایگی (دانشگاه فردوسی مشهد)
- زهرا نیلفروشان (دانشگاه خوارزمی)
- مهدیه هاشمی نژاد (دانشگاه یزد)

ا عضای کمیته علمی شاخه‌ی زمینه‌های بین رشته‌ای:

- مهرداد آشتیانی (دانشگاه علم و صنعت ایران)
- محمدرضا ابراهیمی دیشابی (دانشگاه آزاد اسلامی)
- جعفر الماسی‌زاده (دانشگاه اصفهان)
- رضا انتظاری ملکی (دانشگاه علم و صنعت ایران)
- علیرضا باقری (دانشگاه صنعتی امیرکبیر)
- سوده حسینی (دانشگاه شهید باهنر کرمان)
- سید حامد رستگار (پژوهشگاه دانش‌های بنیادی)
- بهمن زمانی (دانشگاه اصفهان)
- حامد سپهرزاده (دانشکده فنی و حرفه‌ای شهید شمسی‌پور)
- سعید صدیقیان کاشی (دانشگاه خواجه نصیرالدین طوسی)
- مهدی کارگهی (دانشگاه تهران)
- حسن مطلبی پاقلعه (دانشگاه تحصیلات تکمیلی صنعتی کرمان)
- علی مهجور (دانشگاه صنعتی مالک اشتر)

محورهای کنفرانس

مقالات کنفرانس ملی انفورماتیک ایران امسال در چهار محور اصلی زیر سازماندهی شده‌اند:

۱- سیستم

- معماری کامپیوتر
- سیستم حافظه و ذخیره سازی داده
- شبکه کامپیوتری
- امنیت داده و کامپیوتر
- پایگاه داده
- سیستم‌های تعبیه شده و کم توان
- سیستم‌های بی‌درنگ
- پردازش با کارایی بالا
- ارزیابی و تحلیل کارایی
- سیستم‌های عامل
- زبان‌های برنامه‌سازی
- مهندسی نرم افزار
- دیگر موضوع‌های مرتبط با این شاخه

۲- هوش مصنوعی

- هوش مصنوعی
- یادگیری ماشین
- رایانش نرم
- شناسایی الگو
- داده کاوی
- پردازش زبان طبیعی
- بینایی ماشین و پردازش تصویر
- دیگر موضوع‌های مرتبط با این شاخه

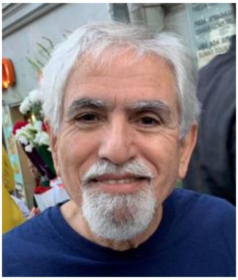
۳- تئوری

- طراحی و تحلیل الگوریتم‌ها
- هندسه محاسباتی
- الگوریتم‌های تقریبی و تصادفی
- پیچیدگی محاسبات
- الگوریتم‌های کوانتومی
- الگوریتم‌های موازی و توزیع‌شده
- منطق و اعتبارسنجی
- روش‌های صوری
- دیگر موضوع‌های مرتبط با این شاخه

۴- زمینه‌های بین‌رشته‌ای

- پردازش داده‌های کلان
- وب و بازیابی اطلاعات
- بیوانفورماتیک
- پردازش موبایل
- گرافیک کامپیوتری
- اقتصاد و محاسبات
- تعامل انسان و کامپیوتر
- رباتیک
- مصورسازی
- رایانش ابری
- اینترنت اشیا
- زنجیره‌های بلوکی
- سایر موضوع‌های میان رشته‌ای

سخنرانیهای علمی چهارمین دوره‌ی کنفرانس ملی انفورماتیک ایران



دکتر بهروز پرهامی

دانشگاه سانتا باربارا

دکتر مهدی جلیلی

دانشگاهRMIT

دکترسید مجید زاهدی

دانشگاه واترلو

دکتر پرهام مرادی

دانشگاهRMIT

دکتر حمید بیگی

دانشگاه صنعتی شریف

برنامه‌ی چهارمین کنفرانس ملی انفورماتیک ایران

روز اول: چهارشنبه ۱۴ دی‌ماه ۱۴۰۱

افتتاحیه کنفرانس	۹:۰۰–۱۰:۰۰	خوش‌آمدگویی مسئولین کنفرانس
		سخنرانی افتتاحیه: A Realistic Assessment of Intelligent Behavior and Machine Learning دکتر بهروز پرهامی، دانشگاه سانتا باربارا - آمریکا
مقالات علمی گروه هوش مصنوعی	۱۰:۰۰–۱۰:۱۵	پیش‌بینی اثرات جانبی داروها به کمک گراف افزوده‌شده‌ی شباهت دارو–دارو و الگوریتم‌های شبکه‌عصبی گرافی مریم صباحی
	۱۰:۱۵–۱۰:۳۰	روشی جدید برای استخراج تداخل دارویی از متن: استفاده از تفکیک دو مرحله ای مبتنی بر جنگل تصادفی در مجموعه های با کلاس های چندگانه و توزیع نامتعادل هانا فتحی
	۱۰:۳۰–۱۰:۴۵	استفاده از یک طبقه‌بندی ترکیبی بر اساس الگوریتم جنگل تصادفی و الگوریتم جایا برای طبقه بندی سرطان سینه سپیده چهره
	۱۰:۴۵–۱۱:۰۰	سیستم توصیه‌گر مبتنی بر نشست بر اساس شبکه توجه ریحانه داورزنی
	۱۱:۰۰–۱۱:۱۵	یک مدل خودتوجه تجزیه‌شده برای سیستم‌های توصیه‌گر دنباله‌ای پگاه جعفریان
	۱۱:۱۵–۱۱:۳۰	احتمال تعمیم‌پذیر بودن روش های ارزیابی کیفیت مبتنی بر بردار پشتیبان پوریا چراغی
	۱۱:۳۰–۱۲:۰۰	استراحت
سخنرانی کلیدی ۱	۱۲:۰۰–۱۲:۴۵	Influence Maximisation in Network Systems: Beyond Heuristic Approaches دکتر مهدی جلیلی، دانشگاه RMIT - استرالیا
	۱۲:۴۵–۱۴:۰۰	استراحت
مقالات علمی گروه سیستم	۱۴:۰۰–۱۴:۱۵	پیش‌بینی عملکرد تحصیلی دانشجویان مبتنی بر شبکه کانولوشن گراف و لایه توجه وجیهه باقری
	۱۴:۱۵–۱۴:۳۰	شتاب دهنده بازپیکربند برای اجرای کارآمد شبکه های عصبی کانولوشن پریا دربانی
	۱۴:۳۰–۱۴:۴۵	بررسی استفاده از برداشتگرهای انرژی برای شبکه‌های حسگر بی‌سیم علی جوادی
	۱۴:۴۵–۱۵:۰۰	مقایسه تطبیقی کارایی زیرساخت ابری ارائه شده در ابرهای عمومی داخلی محمدعلی کشاورز
	۱۵:۰۰–۱۵:۱۵	بهبود عملکرد محاسباتی در تبدیل موجک گسسته با استفاده از سیستم اعداد ماندهای زهراسلطانی
	۱۵:۱۵–۱۵:۴۵	استراحت
مقالات علمی گروه زمینه‌های بین‌رشته‌ای	۱۵:۴۵–۱۶:۰۰	ارائه ی یک روش کارا برای شناسایی گره‌های تأثیرگذار در شبکه‌های اجتماعی بزرگ نوشین شریفی
	۱۶:۰۰–۱۶:۱۵	ارائه روشی برای ارزیابی کیفیت تصاویر محتوای صفحه با تأکید بر ویژگی‌های توصیف‌گر لبه، وزن‌دهی شده توسط نواحی چشمگیر تصویر ریحانه مسافر
	۱۶:۱۵–۱۶:۳۰	تعیین سطح دانشجویان در آموزش مجازی با تحلیل سری زمانی جریان کلیک حدیث عزیزی
	۱۶:۳۰–۱۶:۴۵	ساخت و جاسازی گراف ناهمگن برای خوشه‌بندی مؤثر محتوای متنی شبکه اجتماعی توییتر ملیکا حیدری دستجردی
سخنرانی کلیدی ۲	۱۶:۴۵–۱۷:۳۰	یادگیری چند-مؤلفه‌ای برای مدیریت بار در مرکز داده اشتراکی دکتر سید مجید زاهدی، دانشگاه واترلو - کانادا

روز دوم: پنج‌شنبه ۱۵ دی‌ماه ۱۴۰۱

مقالات علمی گروه هوش مصنوعی	۹:۰۰-۹:۱۵	**دسته‌بندی ابرنقاط سه‌بعدی با استفاده از یک معماری عمیق توسعه‌یافته** ایمان احمدی
	۹:۱۵-۹:۳۰	**قطعه بندی خودکار دو مرحله‌ای اسبک مغز در تصاویر MRI با استفاده از شبکه های عصبی عمیق** علیرضا صادقی
	۹:۳۰-۹:۴۵	**پتانسیل پژوهشی طرح‌های رمزنگاری متقارن در بستر رایانش ابری: عملیات مجموعه‌ای محرمانه و طرح‌های پیوند تطابق‌پذیر چندگانه** مجتبی رفیعی
	۹:۴۵-۱۰:۰۰	**بررسی تأثیر نظرات کاربران در مورد اخبار در قیمت ارز دیجیتال با استفاده از مدل زبانی برت** سهیل راهساز
	۱۰:۰۰-۱۰:۱۵	**طراحی سیستم معاملاتی نیمه خودکار برای اعلام سیگنال‌های معاملاتی بایننس‌کوین** امیرحسین برادران
	۱۰:۱۵-۱۰:۴۵	**استراحت**
سخنرانی کلیدی ۳	۱۰:۴۵-۱۱:۳۰	**Graph Neural Networks for Recommender Systems** دکتر پرهام مرادی، دانشگاه RMIT - استرالیا
مقالات علمی گروه تئوری و پژوهش دانشجویان دکترا	۱۱:۳۰-۱۱:۴۵	**بهبود دقت وارسی مدل آماری برای فرآیندهای تصمیم مارکوف** محمدصادق محققی
	۱۱:۴۵-۱۲:۰۰	**بهبود کیفیت تبدیل سری زمانی به گراف پدیداری نفوذپذیر محدود، با ایجاد نسخه وزن‌دار و پیش‌بینی پیوند** الهام امیری
	۱۲:۰۰-۱۲:۳۰	**کدهای متریک-گرافی (پژوهش دانشجویان دکترا)** مهری عقبایی
	۱۲:۳۰-۱۴:۰۰	**استراحت**
میزگرد	۱۴:۳۰-۱۶:۰۰	**علوم و فناوری های همگرا: از پژوهش تا نوآوری** دکتر ابراهیم اردشیر لاریجانی، استادیار دانشکده ریاضی دانشگاه علم و صنعت ایران و پژوهشگاه دانش های بنیادی دکتر بابک حسین خلج، استاد دانشکده برق دانشگاه صنعتی شریف و پژوهشکده علوم و فناوری‌های همگرای دانشگاه صنعتی شریف دکتر علیرضا ولی زاده، استاد دانشکده فیزیک مرکز تحصیلات تکمیلی علوم پایه زنجان دکتر میثم فرج اللهی، استادیار دانشکده فناوری‌های نوین دانشگاه علم و صنعت ایران و مرکز فناوری‌های راهبردی معاونت علمی ریاست جمهوری
سخنرانی کلیدی ۴	۱۶:۰۰-۱۶:۴۵	**Deep Sub-Modular Networks and Its Applications to Document Summarization** دکتر حمید بیگی، دانشگاه صنعتی شریف
	۱۶:۴۵-۱۷:۱۵	**استراحت**
مقالات علمی گروه هوش مصنوعی	۱۷:۱۵-۱۷:۳۰	**ارائه‌ی بهینه‌ساز بر پایه‌ی گرادیان کاهشی کسری چند متغیره و توزیع شده برای شبکه‌های عصبی** امیرحسین هادیان
	۱۷:۳۰-۱۷:۴۵	**تقریب پارامترهای مدل‌های تصادفی: رویکرد شبکه‌های عصبی پیچشی** امیرحسین ابراهیمی
	۱۷:۴۵-۱۸:۰۰	**نگاشت عمق تصاویر به تصاویر دید از بالا: رویکرد شبکه‌های عصبی عمیق** آبتین ماهیار
	۱۸:۰۰-۱۸:۱۵	**ANN-EAM: یک چارچوب شبکه عصبی مصنوعی – انباشت شواهد برای مدل سازی تصمیم‌گیری واژگان انسانی** آرش دادرس
	۱۸:۱۵-۱۸:۳۰	**اختتام کنفرانس**

روز سوم: جمعه ۱۶ دی‌ماه ۱۴۰۱ (کارگاه‌های کنفرانس)

مقدمه‌ای بر پردازش زبان‌های طبیعی (سخنران: نغمه جمالی)	
1. Introduction to NLP 2. NLP Challenges 3. Different types and formats of Text Data	۸:۰۰-۹:۳۰
استراحت	
1. Word Embedding & Representation (non-Contextualized, Contextualized) a. Monolingual NLP Models b. Multilingual NLP Models 2. Sentence Embedding & Representation	۱۰:۰۰-۱۱:۳۰
استراحت	
آشنایی با کوبرنتیز و استفاده از آن در مقیاس‌های کوچک و بزرگ (سخنران: امیر میرزائی)	
1. Background on Containers a. Containers and OCI standard Overview. b. How Containers Work. c. How OCI images are built. 2. Container Orchestration & Kubernetes Introduction a. Service oriented and micro-service architectures overview. Microservices and use of containers. Container Orchestration b. Kubernetes (Architecture Overview, Major Concepts, Abilities and Limits, Overview of next part)	۱۲:۰۰-۱۳:۳۰
استراحت	
1. Kubernetes at any scale. Using Kubernetes for different workloads a. Challenges for small scale deployments! Challenges for large scale deployments! b. Why to use Kubernetes at small scales. Why to use Kubernetes at large scales. c. Solutions for challenges d. Kubernetes modular approach, Distributions, Addons and Extensions + Operators + Tools e. Different Workloads (e.g., compute intensive workloads, heterogeneous clusters, …)	۱۴:۳۰-۱۶:۰۰
استراحت	
مقدمه‌ای بر داده‌کاوی کاربردی (سخنران: امیر حسین هادیان)	
1. Introduction to machine leaning & machine learning tools 2. machine learning applications Examples 3. Life cycle of a machine learning project a. Data collection b. Data cleaning & data preprocessing c. Model selection d. Model evaluation	۱۶:۳۰-۱۸:۰۰
استراحت	
1. Example of implementing a project using python a. Data collection b. Data cleaning & data preprocessing c. Model selection d. Model evaluation	۱۸:۳۰-۲۰:۰۰

میزگرد در چهارمین کنفرانس ملی انفورماتیک ایران

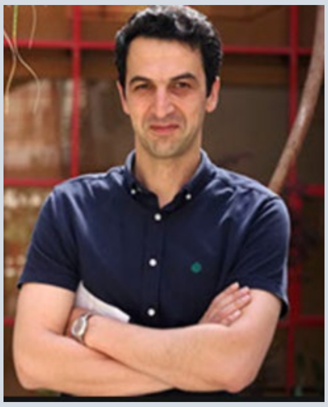
عنوان

علوم و فناوری های همگرا: از پژوهش تا نوآوری

تاریخ برگزاری کنفرانس: چهارشنبه و پنجشنبه، ۱۴ و ۱۵ دی‌ماه ۱۴۰۱
زمان برگزاری میزگرد: ساعت ۱۴:۳۰-۱۶:۰۰ پنجشنبه ۱۵ دی‌ماه ۱۴۰۱

دکتر ابراهیم اردشیر لاریجانی
استادیار دانشکده ریاضی دانشگاه علم و صنعت ایران و پژوهشگاه دانش های بنیادی و موسسه رهیافت های نوآور پاسارگاد

دکتر بابک حسین خلج
استاد دانشکده برق، ، مدیر مرکز سیستم های اطلاعات و علوم داده دانشگاه صنعتی شریف و پژوهشکده علوم و فناوری های همگرای دانشگاه صنعتی شریف



دکتر علیرضا ولی زاده
استاد دانشکده فیزیک دانشگاه تحصیلات تکمیلی علوم پایه زنجان و موسسه رهیافت های نوآور پاسارگاد

دکتر میثم فرج اللهی
استادیار دانشکده فناوری‌های نوین دانشگاه علم و صنعت ایران و مرکز فناوری‌های راهبردی معاونت علمی ریاست جمهوری

موضوع میزگرد:

- آشنایی با علوم و فناوری‌های همگرا . چرا فناوری‌های همگرا در اقتصاد «نوآوری» محور مهم هستند؟
- چگونه فناوری نانو، زیستی، اطلاعات و شناختی (NBIC) در واقعیت همگرا می‌شوند؟
- تاثیر NBIC در اقتصاد و تجارت چگونه است؟ در ایران افق چنین فناوری‌هایی در صنعت و بازار داخلی چیست؟

میهمانان میزگرد:

- دکتر بابک حسین خلج
- دکتر علیرضا ولی زاده
- دکتر میثم فرج اللهی

مجری میزگرد:

- دکتر ابراهیم اردشیر لاریجانی

شرکت برای عموم علاقمندان آزاد است
جهت کسب اطلاعات بیشتر به وبسایت کنفرانس مراجعه فرمایید: https://cs.ipm.ac.ir/nic/1401

پیش‌بینی اثرات جانبی داروها به کمک گراف افزوده‌شده‌ی شباهت دارو-دارو و الگوریتم‌های شبکه‌عصبی گرافی

مریم صباحی[1]، حسین رحمانی[2]، سینا ضیایی[3]

[1] دانشکده کامپیوتر، دانشگاه علم و صنعت ، تهران،
Maryam_sabahi@comp.iust.ac.ir

[2] استادیار، دانشکده کامپیوتر، دانشگاه علم و صنعت ، تهران
H_rahmani@iust.ac.ir

[3] دانشکده کامپیوتر، دانشگاه علم و صنعت ، تهران
Sinaziaee99@gmail.com

چکیده

داروها به هدف درمان بیماری ساخته می‌شوند، با این حال بیشتر آن‌ها هنگام درمان بیماری ممکن است اثرات منفی تحت عنوان عوارض جانبی داشته باشند. تشخیص اثرات جانبی داروها همواره فرآیندی زمان‌بر و پرهزینه بوده‌است. به طور مثال برآورد می‌شود که فرایند تولید دارو، از شناسایی هدف تا استفاده بالینی، می‌تواند تا ۱۴ سال و هزینه ۸۰۰ میلیون دلار طول بکشد. رویکردهای متعارف برای پیش‌بینی اثرات جانبی در طی فرایند تولید دارو سنجش‌های آزمایشگاهی و حیوانی است. با این حال امکان تشخیص تمام عوارض جانبی در محیط آزمایشگاهی وجود ندارد و این اثرات جانبی کشف نشده می‌توانند ضررهای بسیاری برای بیمار و همچنین شرکت تولیدکننده‌ی دارو داشته‌باشند. به‌طور مثال در آمریکا تخمین زده می‌شود که عوارض جانبی چهارمین عامل اصلی مرگ‌ومیر باشند و همچنین عامل یک سوم از عدم موفقیت داروها در طی فرایند تولید، اثرات جانبی دارو شناخته شده‌است. با توجه به نکات گفته‌شده کارامدتر است که در آغاز فرآیند تولید دارو، با بیشترین دقت ممکن اثرات جانبی آن محاسبه و سنجیده شود. در نتیجه مطالعه در رابطه با اثرات جانبی داروها و پیش‌بینی آن‌ها از اهمیت بالایی برخوردار است. در این پژوهش ما به مسئله پیش‌بینی اثرات جانبی از منظر پیش‌بینی برچسب در گراف نگاه کردیم و روشی مبتنی بر شبکه‌های عصبی گرافی ارائه دادیم. به این منظور گراف شباهت دارو-دارو را ساخته و در رابطه با معیارهای شباهت دارویی بحث کردیم. با توجه به معیار AUC روش ما در مقایسه با روش‌های پیشین ۵ درصد بهبود داشته است.

کلمات کلیدی

عوارض جانبی، یادگیری عمیق، شبکه‌های عصبی عمیق، شبکه‌های عصبی گرافی، یادگیری ماشین، پیش‌بینی اثرات جانبی

۱- مقدمه

داروها به هدف درمان بیماری ساخته‌می‌شوند، با این حال بیشتر آن‌ها هنگام درمان بیماری ممکن است اثرات منفی تحت عنوان عوارض جانبی داشته‌باشند. اثر جانبی کشف‌نشده برای یک داروی تأییدشده می‌تواند برای انسان مضر باشد و همچنین خطرات بزرگی برای یک شرکت دارویی داشته‌باشد[1]. رویکردهای متعارف برای پیش‌بینی اثرات جانبی در طی فرایند تولید دارو سنجش‌های آزمایشگاهی و حیوانی است. با این‌حال، چنین پیش‌بینی‌های تجربی دارای هزینه‌ی بالا و زمان‌بر هستند[2]. در سال‌های اخیر، روش‌های محاسباتی بسیاری برای پیش‌بینی عوارض جانبی دارو ایجاد شده‌است. این روش‌ها می‌توانند دامنه اثرات جانبی دارو را کاهش دهند[3]. روش‌های محاسباتی موجود برای پیش‌بینی عوارض جانبی داروها، فرض می‌کنند که داروهای مشابه از نظر خصوصیات شیمیایی و بیولوژیکی مانند ساختارها و اهداف آن‌ها دارای خواص قابل مقایسه‌اند[4]. چندین روش مبتنی بر داده‌کاوی و یادگیری‌ماشین برای پیش‌بینی عوارض جانبی داروها بر اساس ویژگی‌های مختلف از جمله ساختارهای شیمیایی داروها، اثر متقابل دارو–پروتئین، اثر متقابل پروتئین–پروتئین، فعالیت در شبکه‌های متابولیکی، مسیرها، اطلاعات فنوتیپی و تفسیر ژن ارائه شده‌است[5].

روش‌های مختلفی در حوزه‌ی پیش‌بینی اثرات جانبی داروها ارائه شده است. این روش‌ها با توجه به پارامترهایی نظیر ویژگی‌های موجود از داروهای در حال بررسی و الگوریتم‌های داده‌کاوی مختلف به چندین دسته تقسیم می‌شوند. به طور کلی برای پیش‌بینی اثر جانبی داروهای جدید، اطلاعات زیادی برای توصیف ویژگی‌های دارو در دسترس نیست و پژوهش‌گران در این مسائل تلاش می‌کنند با اطلاعات پایه‌ی موجود از دارو مانند ساختار شیمیایی و فیزیکی دارو بهترین نتایج پیش‌بینی را بدست بیاورند [6], [7]. همچنین در مواردی که سعی بر پیش‌بینی اثرات جانبی شناخته‌نشده‌ی داروهای تأیید شده است، اطلاعات جامع‌تر و بیشتر نسبت به دارو در اختیار پژوهش‌گر قرار دارد که با استفاده‌ی مناسب از این ویژگی‌ها می‌تواند نتایج به مراتب بهتری بدست آورد[8]. با در نظر گرفتن این موارد در اکثر روش‌ها تلاش می‌شود مدل بدست‌آمده توانایی پیش‌بینی برای هر دو گروه با بهترین نتیجه را داشته باشد. روش‌های محاسباتی بسیاری نیز در این حوزه ارائه شده‌اند. الگوریتم‌های استفاده شده را می‌توان به چندین دسته‌ی رده‌بندی [3], [4], [9]، تجزیه و تکمیل ماتریس [8], [10]، یادگیری‌عمیق [5], [6], [9]، گراف‌کاوی [10] و سایر روش‌های نوآورانه [3], [4], [7], [9] تقسیم کرد.

در دسته‌ی یادگیری ماشین، یکی از پرتکرارترین و متنوع‌ترین الگوریتم‌ها، الگوریتم‌های رده‌بندی هستند. رده‌بندها نقش زیادی در پیش‌بینی اثرات جانبی داشته‌اند و از تعداد بسیاری از آن‌ها برای پیش‌بینی استفاده شده است. یکی از پرکاربردترین رده‌بندهای استفاده شده، جنگل تصادفی بوده است [3] [4] [9] [11] که بهترین نتیجه را نیز در بین رده‌بندهای موجود داشته است. این مقایسه را می‌توان در مقاله‌ی لیانگ و همکاران [9] مشاهده کرد. در این پژوهش از سه رده‌بند SVM، جنگل تصادفی و ANN استفاده شده و بهترین نتیجه را جنگل تصادفی بدست آورده است. از دیگر رده‌بندهای استفاده شده در پیش‌بینی اثرات جانبی می‌توان به رگرسیون لجستیک [4] [11] ، بیز ساده [4] [11] ، knn [11] و XGboost [4] یاد کرد.

روش محاسباتی دیگر در دسته یادگیری ماشین، تجزیه ماتریس است. هدف تجزیه ماتریس یافتن دو یا چند ماتریس با مرتبه‌ی پایین‌تر (low-rank) است که ضرب آن‌ها، تقریب خوبی از ماتریس اصلی می‌دهد. در پیش‌بینی اثرات جانبی ماتریس اصلی، در واقع ماتریس دارو–اثر جانبی است. ژانگ و همکاران [1] روش محاسباتی جدیدی را پیشنهاد می‌کنند که عوارض جانبی مشاهده نشده را برای داروهای تایید شده بر اساس ارتباط دارو–عوارض جانبی و خصوصیات داروها پیش‌بینی می‌کند. به طور کلی، روش‌های کلاسیک تجزیه‌ی ماتریس از ماتریس روابط شناخته شده برای پیش‌بینی استفاده می‌کنند، اما در مسئله‌ی پیش‌بینی عوارض جانبی به دلیل وجود اطلاعاتی در زمینه‌ی زیست‌پزشکی که می‌تواند مفید باشد، ژانگ و همکاران تلاش کرده‌اند تا این اطلاعات زیست‌پزشکی در مورد داروها را در تجزیه‌ی ماتریس بگنجایند و به عملکردیی با دقت بالا برسند.

امروزه کشف کردن دانش جدید از داده‌های گرافی با استفاده از روش‌های گراف کاوی ممکن شده است. یکی از دامنه‌هایی که گراف‌کاوی در چند سال اخیر در آن بسیار مورد توجه قرار گرفته است، دامنه‌ی زیستی است. مسائل بسیاری مانند پیش‌بینی عملکرد پروتئین، پیش‌بینی پروتئین‌های هدف برای داروها، پیش‌بینی اثرات جانبی داروها و ... وجود دارند که بهره‌گیری از

ساختار گرافی و به کاربردن روش‌های گراف‌کاوی باعث بهبود در نتایج آن‌ها شده است.

Timilsina و همکاران [10] با تشکیل گراف دارو-دارو و دارو-اثر جانبی و سپس ترکیب دو گراف ذکر شده، توانسته‌اند مسئله پیش‌بینی اثرات جانبی را به یک مسئله‌ی گراف‌کاوی تبدیل کرده و از الگوریتم انتشار حرارت به عنوان راه حل مسئله استفاده کنند.

بیشتر روش‌هایی که تاکنون مطرح کردیم، شناسایی اثر جانبی را به عنوان یک مشکل رده‌بندی در نظر می‌گیرند، که ترتیب کشف اثرات جانبی را در نظر نمی‌گیرد. بنابراین، روند پر هزینه بوده و هم چنین منجر به از دست رفتن اطلاعات قابل توجهی در مورد روابط دارو-اثر جانبی در مرحله آموزش مدل می‌شود. علاوه بر این، این رویکردها قادر به پیش‌بینی اثر جانبی داروهای جدید نیستند، بنابراین تشخیص اثر جانبی را دشوارتر می‌کند.

برای رسیدگی به این محدودیت‌ها، وانگ و همکاران[12] از یک مدل شبکه عصبی عمیق (DNN) برای تشخیص اثرات جانبی داروها استفاده کردند. این مدل ۲ هدف دارد: «شناسایی اثرات جانبی»، که مستلزم کشف اثرات جانبی بالقوه یک دارو از سوابق اثرات جانبی شناخته شده است، و «پیش‌بینی اثرات جانبی»، که به پیش‌بینی اثرات جانبی احتمالی برای یک داروی جدید مربوط می‌شود. جهت بهبود نتایج آن‌ها از رویکرد جاسازی کلمه و تابع نگاشت برای پردازش داروهای جدیدی که در مجموعه داده ظاهر نمی‌شوند استفاده کردند. در این راستا با الهام از مدل Word2Vec، مدل Drug2Vec خود را معرفی کردند. در این مدل به جای کلمه به ازای هر نام دارو یک بردار تخصیص داده می‌شود. سپس برای رسیدگی به داروهایی که در مرحله آموزش جاسازی مشاهده نشدند، Drug2Vec را با معرفی یک تابع نگاشت توصیف دارو گسترش دادند.

عوارض جانبی حاصل از چند-دارویی باید به دقت برای توسعه داروی جدید در نظر گرفته شود. با این حال، در نظر گرفتن تمام تداخلات دارو-دارو پیچیده‌ای که باعث عوارض جانبی چند-دارویی می‌شود، چالش برانگیز است. اخیراً، مدل‌های شبکه عصبی گرافی (GNN) این تعاملات پیچیده را با موفقیت مدیریت کرده‌اند و عملکرد بالایی در پیش‌بینی نشان داده‌اند.

با این وجود، مدل‌های GNN در ارائه فاکتورهای قابل درک پیش‌بینی برای کارشناسان حوزه زیست‌پزشکی و دارویی مشکل دارند. بانگ و همکاران[13] یک رویکرد جدید، به نام شبکه توجه ویژگی‌های گراف (GFAN)، برای پیش‌بینی قابل تفسیر عوارض جانبی چند-دارویی با تأکید بر ژن‌های هدف متفاوت ارائه داده‌اند. برای شبیه‌سازی مصنوعی حالت چند-دارویی، که در آن دو داروی مختلف با هم مصرف می‌شوند، یک مسئله رده‌بندی گره را با استفاده از مفهوم گراف خطی در تئوری گراف فرمول‌بندی کرده‌اند. فرآیند کلی این روشبه این صورت است: a) ساخت شبکه دارویی: ابتدا شبکه تعامل دارو-دارو ساخته شده است به گونه‌ای که گره‌ها دارو هستند و اطلاعات عوارض جانبی چند-دارویی یال‌های آن را تشکیل می‌دهند و ویژگی هر گره، ژن‌های هدف آن دارو هستند. b) ساخت شبکه چند-دارویی: در ادامه، با پیاده‌سازی مفهوم گراف خطی در نظریه گراف، شبکه دارویی به شبکه چند-دارویی تبدیل شده است. در واقع هر یال در شبکه دارویی، گرهی در شبکه چند-دارویی در نظر گرفته شد. درنتیجه گره‌های چند-دارویی ایجاد شده و ویژگی هر گره ترکیب ژن‌های هدف دو دارو هستند. در صورت داشتن داروهای مشترک بین دو گره با یال به هم متصل می‌شوند. c) سپس، روش پیشنهادی GFAN چند-برچسب را برای عوارض جانبی چند-دارویی پیش‌بینی می‌کند و ویژگی‌هایی را استخراج می‌کند که به طور قابل‌توجهی به پیش‌بینی کمک می‌کنند.

در دسته‌ی آخر که نام سایر را برای آن انتخاب کردیم، روش‌های محاسباتی قرار می‌گیرد که فقط در یکی از دسته‌بندی‌های بالا جای نمی‌گیرد و یا از روشی نوین استفاده شده است.

ودوا و همکاران [11] بر اساس ویژگی‌های دو بعدی و سه بعدی دارویی و در نظر گرفتن ماتریس با هم‌آیی دارو-اثر جانبی در تلاش ساخت مدل پیش‌بینی خود شدند. آن‌ها قصد داشتند از تعدادی از رده‌بندهای پرکاربرد برای مقایسه‌ی دقت پیش‌بینی عوارض جانبی در سطوح مختلف آناتومی (عوارض جانبی، اندام‌ها، زیر-دستگاه‌ها و دستگاه‌ها) استفاده کنند. رده‌بندهای استفاده شده شامل رگرسیون لجستیک، SVM، kNN، GaussianNB و جنگل تصادفی بوده‌اند. علاوه بر آن برای نوآوری در عملکرد، به کمک سه‌تا از رده‌بندهای پیشین که بهترین عملکرد را داشتند، رده‌بندی را برای ساخت یک مدل ترکیبی مبتنی بر رأی‌گیری نرم بهینه کردند.

سئو و همکاران [4] ، ویژگی‌های دارویی مرتبط با تغییر کاربری دارو را نیز در روش خود در نظر گرفته‌اند. در این روش در ابتدا برای مشخص کردن مجموعه‌ای از ویژگی‌ها برای زوج دارو-اثر جانبی، داروی موجود در زوج دارو-اثر جانبی با داروهایی که با اثر جانبی موجود در زوج گفته شده، مرتبط بوده‌اند، با یکدیگر بر اساس ویژگی‌های انتخابی مقایسه می‌شوند و بیشترین مقدار شباهت انتخاب می‌شود. همین روند به طور برعکس و برای یافتن بیش‌ترین شباهت بین عوارض جانبی انجام می‌شود. در آخر برای هر زوج دارو-اثر جانبی برداری ایجاد می‌شود. مدل ترکیبی یادگیری مبتنی بر پشته شامل چهار رده‌بند جنگل تصادفی، رگرسیون لجستیک، بیز ساده و XGboost به عنوان رده‌بندهای پایه و الگوریتم شبکه عصبی به عنوان رده‌بند متا آن اعمال شده است.

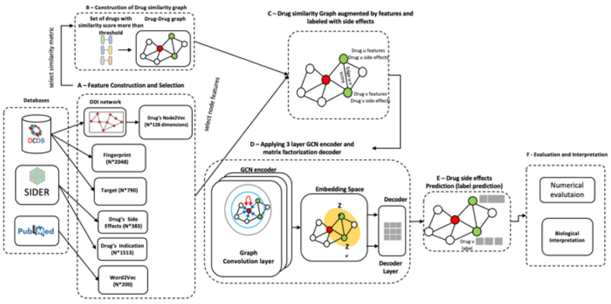
همان‌طور که در بخش‌های پیشین نیز بیان شد ارتباط بین داروها و اثرات جانبی را به روش‌های مختلف می‌توان بدست آورد و مجموعه داده‌های فراوانی برای این منظور فراهم شده است. عده‌ای از محققین برای افزایش دقت مدل‌های پیش‌بینی خود، علاوه بر این مجموعه داده، به روش‌های مختلف مجموعه داده‌ای از نمونه‌های منفی نیز ایجاد کرده‌اند. منظور از نمونه‌ی منفی، زوج‌های دارو-اثر جانبی است که کمترین احتمال وجود رابطه و باهم‌آیی بین آن‌ها باشد. در ادامه نمونه‌ای از این رویکرد را بررسی می‌کنیم.

ژائو و همکاران [3] از رده‌بندی دودوئی برای پیش‌بینی اثرات جانبی داروها استفاده کرده‌اند. در این مدل زوج دارو-اثر جانبی را به عنوان نمونه در نظر گرفته و مسئله‌ی اصلی را به یک مسئله‌ی رده‌بندی دودوئی تبدیل می‌کنند در نتیجه هدف «تعیین توانایی جفت شدن دارو با اثر جانبی داده‌شده» است. برای ایجاد مجموعه داده‌ی منفی (NDS) از تمام زوج دارو-عوارض جانبی‌های ممکن، به غیر از زوج‌های موجود در مجموعه داده‌ی مثبت (PDS) به صورت تصادفی زوج انتخاب شده است. تعداد اعضای NDS با PDS برابر در نظر گرفته شده و برای جلوگیری از تاثیر مقادیر موجود در NDS در نتیجه پژوهش، پنج NDS متفاوت ایجاد شده است. هر جفت، توسط پنج ویژگی که از خصوصیات دارو بدست آمده، نشان داده می‌شوند. الگوریتم یادگیری ماشین جنگل تصادفی به عنوان موتور پیش‌بینی انتخاب شده‌است.

در پژوهشی دیگر توسط لیانگ و همکاران [9] تلاش شده است با تغییر روش انتخاب مقادیر NDS نتایج بدست‌آمده را بهبود دهند. در این روش ابتدا شبکه‌ای بر اساس اثر متقابل شیمیایی (CCI) داروها ساخته شده است. در این شبکه داروها گره‌های شبکه را تشکیل می‌دهند و هر جفت دارو یک امتیاز بر اساس CCI بین آن‌ها دارد که این مقدار (در صورت مخالف صفر بودن) وزن یال بین دو دارو را مشخص می‌کند. حال در هر مرحله به ازای یک عارضه‌ی جانبی موجود در مجموعه داده‌ی SIDER داروهایی که به آن مرتبط هستند به عنوان گره‌های بذر در نظر گرفته شده و الگوریتم راه رفتن تصادفی (RWR) بر روی شبکه اجرا می‌شود. وقتی الگوریتم RWR متوقف می‌شود، به هر گره در شبکه یک احتمال اختصاص داده شده است. واضح است که یک گره (دارو) با احتمال زیاد، ارتباط نزدیکی با گره‌های بذر دارد و در نتیجه استنباط می‌شود که چنین گره‌ای به احتمال زیاد آن اثر جانبی را دارد. برعکس، در گره‌ها (داروها) با احتمال کم، امکان بسیار کمی وجود دارد که دارای آن اثر جانبی باشند. بدین ترتیب با در نظر گرفتن یک آستانه احتمال ε ، داروهایی که احتمالات کمتر از ε را دریافت می‌کنند استخراج شده و زوج آن‌ها با عارضه‌ی جانبی، به عنوان نمونه‌ی منفی کاندید در نظر گرفته می‌شود. پس از انجام این روند برای تمام عوارض جانبی، با جمع آوری تمام نمونه‌های منفی نامزد برای هر عارضه جانبی، NDS ساخته می‌شود. این مجموعه با PDS ترکیب شده تا مجموعه داده آموزش را تشکیل دهد. ساخت بردار ویژگی‌ها برای هر زوج مانند پژوهش قبل است. در آخر سه الگوریتم RF، SVM و MLP بر روی مجموعه داده آموزش ایجاد شده اجرا می‌شود.

ما در این پژوهش قصد داریم به کمک ویژگی‌های دارویی استخراج‌شده و الگوریتم‌های داده‌کاوی به پیش‌بینی اثرات جانبی داروها بپردازیم. در این راستا ما با در نظر گرفتن این مسئله به عنوان یک مسئله پیش‌بینی چند-برچسب در یک گراف و با کمک الگوریتم‌های شبکه‌عصبی گرافی به حل آن پرداختیم. روش پیشنهادی ما از مراحل مختلفی تشکیل شده است که در شکل ۱ سلسه‌مراتب کارهای انجام‌شده نشان داده شده است. در این روش مسئله، پیش‌بینی برچسب در گراف دارو-دارو است. در نتیجه ابتدا گراف شباهت دارو-دارو ساخته می‌شود. این گراف بر اساس ویژگی‌های دارویی به

عنوان معیار شباهت و مقدار آستانه شباهت مشخص، ساخته می‌شود. سپس دیگر ویژگی‌های دارویی به گراف افزوده شده است. در مرحله بعد گراف ساخته‌شده به عنوان ورودی و برای آموزش به یک مدل مبتنی بر شبکه‌عصبی گرافی داده می‌شود. این مدل به پیش‌بینی برچسب هر گره (یعنی هر دارو) می‌پردازد. در پایان به بررسی عملکرد مدل پرداختیم و اثرات جانبی پیش‌بینی شده تعدادی از داروها که در مجموعه داده اولیه بیان نشده بودند، در مقالات پزشکی روز بررسی شدند.



شکل ۱ : مراحل روش پیشنهادی: مراحل اصلی این روش شامل ساخت گراف شباهت دارو–دارو، افزودن ویژگی‌ها به گراف، آموزش مدل، پیش‌بینی برچسب (اثرات جانبی دارو) و ارزیابی نتایج است.

۲- روش

همان‌طور که در شکل ۱ مشاهده شد، ویژگی‌های دارویی از مجموعه‌داده‌های مختلف بدست آمده‌اند. ابتدا به معرفی مجموعه‌داده‌های استفاده شده در این پژوهش می‌پردازیم و چگونگی آماده‌سازی ویژگی‌های دارویی مختلف از این مجموعه‌داده‌ها را توضیح می‌دهیم. سپس روش پیشنهادی به طور کامل توضیح داده می‌شود.

۲-۱- مجموعه‌داده‌ها و استخراج ویژگی‌ها

پایگاه‌داده ترکیبات دارویی (DCDB) [14]، اولین پایگاه‌داده موجود است که با هدف تسهیل کشف داروی جدید، اطلاعات مربوط به ترکیبات دارویی را جمع‌آوری و سازماندهی کرده است. ما در این پژوهش از نسخه ۲ (آخرین نسخه موجود) این پایگاه داده استفاده کردیم. این پایگاه داده با تمرکز بر روی مطالعات ترکیبات دارویی اطلاعات مختلفی از داروها را نظیر اهداف دارویی، ترکیبات دارویی آنزیم‌ها و ... جمع‌آوری کرده است. اطلاعات ۹۰۴ دارو در DCDB وجود دارد.

فرمول شیمیایی داروها را می‌توان توسط رشته‌ی SMILES (Simplified Molecular Input Line Entry System) نشان داد. SMILES یک سیستم علامت‌گذاری خطی است که برای توصیف ساختار گونه‌های شیمیایی با استفاده از رشته‌های کوتاه ASCII استفاده می‌شود[15]. رشته‌ی SMILES هر دارو در پایگاه‌داده DCDB وجود دارد. اثر انگشت ECFP_4 را با استفاده از رشته‌های SMILES موجود و یک ابزار انفورماتیک شیمی منبع‌باز و یادگیری‌ماشین به نام RDKit[16] ایجاد کردیم. این اثر انگشت‌ها به عنوان یک بردار دودویی ۲۰۴۸ بعدی شامل ویژگی‌های شیمیایی دارو هستند.

هدف دارویی، مولکولی در بدن انسان است، معمولا یک پروتئین، که به یک فرآیند بیماری خاص مرتبط است و یک دارو می تواند این مولکول را برای ایجاد اثر درمانی مورد نظر تحریک کند[17]. پروتئین‌های هدف داروها را از DCDB استخراج کرده‌ایم. یک دارو فهرستی از پروتئین‌های هدف دارد که توسط یک بردار دودویی ۷۹۰ بعدی نشان داده می‌شود. این بردار وجود یا عدم وجود هر پروتئین را نشان می‌دهد. یک ماتریس ویژگی با اندازه ۷۶۰ در ۷۹۰ ایجاد شده است. ۷۹۰ تعداد اهداف دارویی موجود است و اگر پروتئینی هدف دارو باشد، ۱ و در غیر این صورت ۰ برای آن ثبت شده است.

به منظور استخراج بردار ویژگی هر دارو از شبکه تعاملات بیولوژیکی داروها ابتدا گراف تاثیرات دارو–دارو ساخته شد. هر گره در این گراف یک دارو است و یال‌ها تاثیرات میان داروها را نمایش می‌دهند. DCDB شامل تداخلات دارویی در ترکیبات دارویی است که می‌توانند به تداخلات فارماکوکینتیک و فارماکودینامیک تقسیم شود. هنگامی که اثرات یک دارو با حضور داروی دیگری در محل اثر دارو تغییر می‌کند، به آن تداخل دارویی فارماکودینامیک گفته می‌شود، در حالی که تداخلات فارماکوکینتیک زمانی رخ می‌دهد که حضور داروی دیگری جذب، توزیع، متابولیسم و فرآیندهای دفع یک دارو را تغییر دهد. همه‌ی این تداخلات برای ایجاد شبکه‌ای از تداخلات دارو–دارو استفاده می‌شود که درآن گره‌ها دارو را نشان می‌دهند و یال‌ها تداخلات را نشان می‌دهند. در مرحله بعد، با اعمال روش Node2vec که از گشت تصادفی برای تولید مجموعه‌ای از «جملات» از یک شبکه خاص استفاده می‌کند، یک ماتریس ویژگی به اندازه ۷۵۲ * ۱۲۸ ایجاد کردیم که ۱۲۸ نشان دهنده تعداد بعد تعبیه بدست آمده برای هر یک از ۷۵۲ دارو است.

یکی از مجموعه‌داده‌های شناخته شده برای اثرات جانبی داروها، مجموعه‌داده‌ی Sider [18] است. در این پژوهش ما از جدیدترین نسخه‌ی آن (Sider 4.1) استفاده کردیم. این مجموعه‌داده دارای ۱۴۳۰ دارو و ۵۸۶۸ اثر جانبی مختلف است. فراوانی هر اثر جانبی دارو نیز در این مجموعه داده ذکر شده است. این فراوانی مشخص می‌کند که اثر جانبی برای دارو در چند درصد مواقع گزارش شده است. با توجه به درصد فراوانی اثر جانبی، سازمان جهانی بهداشت و سایر سازمان‌های بهداشتی احتمال یا شانس تجربه اثر جانبی را به این صورت توصیف می‌کنند[19]:

- بیشتر از ۱۰ درصد: خیلی رایج
- بین ۱ تا ۱۰ درصد: رایج
- بین ۰.۱ تا ۱ درصد: غیر معمول
- بین ۰.۰۱ تا ۰.۱ درصد: نادر
- کمتر از ۰.۰۱ درصد: خیلی نادر

کمیسیون اروپا توصیه می‌کند که فهرست اثرات جانبی یک دارو فقط باید حاوی اثراتی باشد که «حداقل احتمال معقولی» وجود داشته باشد که ناشی از مصرف دارو است و درصد فراوانی باید نشان‌دهنده فقط نرخ‌های بروز باشد[19]. در نتیجه در این مقاله ما فقط اثرات جانبی داروهایی که طبق تعریف جزء موارد خیلی رایج یا رایج بودند را در نظر گرفتیم. پس از مرتبط کردن این داروها با ویژگی‌های دارویی که در قسمت‌های قبل توضیح دادیم، تعداد ۲۳۵ دارو و ۱۰۹۰ اثر جانبی بدست آمد. تعدادی از این اثرات جانبی فقط برای یک یا دو دارو از ۲۳۵ دارو بروز داده بودند. این مسئله آموزش مدل انتخابی ما را با مشکل مواجه می‌کرد، در نتیجه اثرات جانبی که در کمتر از ۵ دارو مختلف ظاهر شده بودند را نیز حذف کردیم. در پایان ماتریسی با ۲۳۵ سطر (داروها) و ۳۸۳ ستون (اثرات جانبی) ایجاد شد. که در صورت بروز اثر جانبی ۱ و در صورت عدم بروز آن ۰ در ماتریس قرار گرفت.

اندیکاسیون‌های داروها را نیز از مجموعه‌داده Sider استخراج کردیم. ماتریسی به اندازه ۱۵۱۳*۳۸۳ ایجاد شد که در آن ۱۵۱۳ تعداد اندیکاسیون‌ها است. مشابه اهداف دارویی، ۱ به این معنی است که دارو برای آن اندیکاسیون استفاده شده است و ۰ به این معنی است که دارو برای آن هدف استفاده نشده است.

سایت PubMed[20] یک پایگاه اطلاعاتی غنی از مقالات زیستی است. با توجه به این‌که مقالات موجود در PubMed به صورت متمرکز حاوی اطلاعات زیستی بوده و اطلاعات غیرمرتبط را شامل نمی‌شوند، از اهمیت بالایی در زمینه تحلیل‌های متن‌کاوی در زمینه زیستی برخوردار هستند. یکی از ویژگی‌های استفاده شده برای هر دارو، بردار استخراج شده توسط مدل Word2Vec از داده‌های متنی مقالات پزشکی موجود در مجموعه‌داده PubMed است. در این پژوهش از مدل Wor2Vec ساخته شده توسط الهقلی و همکاران[21] استفاده شده است. برای ساخت مدل Word2Vec آن‌ها مجموعه‌داده متنی چکیده مقالات PubMed را به عنوان ورودی در نظر گرفتند. قبل از ساخت مدل Word2Vec باید داده‌های متنی پیش‌پردازش شوند. به منظور پیش‌پردازش چکیده مقالات PubMed، مراحل پاک‌سازی، نرمال‌سازی حروف، ریشه‌یابی، حذف ایست‌واژه‌ها و جداسازی انجام شده است . پس از پیش‌پردازش چکیده مقالات، دنباله‌ای از جملات و کلمات آن‌ها بدست می‌آید که به عنوان ورودی الگوریتم Word2Vec درنظر گرفته می‌شود. این مدل بر روی ۳۰ میلیون چکیده مقالات PubMed برای استخراج تعبیه داروها در ادبیات زیست‌پزشکی آموزش داده شده بود و در نهایت هر دارو در متن را با یک بردار ۲۰۰ بعدی نشان می‌دهد.

۲-۲- ساخت گراف شباهت دارو–دارو

هدف این مرحله ایجاد گراف شباهت داروها است که در مراحل بعدی برای پیش‌بینی برچسب‌های گره‌ها (اثرات جانبی) مورد استفاده قرار می‌گیرد. همانطور که در بخش‌های قبل گفتیم، یکی از فرضیه‌های موجود در پیش‌بینی اثرات جانبی دارو، بر شباهت داروها تاکید دارد. یعنی داروهایی که با هم مشابه‌اند اثرات جانبی مشابه دارند. در نتیجه در این مرحله ما ۵ گراف مختلف بر اساس معیار شباهت‌های مختلف بدست آوردیم. در این گراف‌ها گره‌ها داروها در نظر گرفته شدند و در صورت بیش‌تر بودن میزان شباهت دو دارو از مقدار آستانه‌ی مشخص شده بین دو دارو یال در نظر گرفته می‌شود و میزان شباهت به عنوان ویژگی یال در نظر گرفته می‌شود. ۵ معیار شباهت مختلف شامل بردار Node2Vec، کد ATC، اثرانگشت، بردار Word2Vec و اهداف دارویی هستند. ما این ۵ گراف مختلف را در مراحل بعدی استفاده کردیم تا بر اساس ارزیابی‌ها بهترین معیار شباهت مشخص شود.

۲-۳- افزودن ویژگی‌ها به گراف

در این مرحله ویژگی‌های دارویی هر گره (دارو) را به گراف اضافه کردیم. ۳ گراف متفاوت بر اساس ویژگی‌های دارویی مختلف بدست آمد. این ویژگی‌ها شامل Nod2vec، Word2Vec و اثر انگشت دارویی بودند. ما این ۳ گراف مختلف را در مراحل بعدی استفاده کردیم تا بر اساس ارزیابی‌ها بهترین ویژگی دارویی مشخص شود.

۲-۴- آموزش مدل

مدل مورد استفاده در این روش، خودرمزگذار گرافی است که شامل ۲ بخش اصلی رمزگذار و رمزگشا است. هدف اصلی رمزگذار ترسیم ویژگی‌های گراف ورودی در یک نمایش فضای تعبیه جدید است. رمزگشا این نمایش‌های جدید را به عنوان ورودی می‌گیرد و گراف ورودی را در فضای جدید بازسازی می‌کند. برچسب‌های گره گراف جدید می‌توانند اثرات جانبی جدید برای دارو پیشنهاد دهند. تمام پارامترهای قابل آموزش این رمزگذارها و رمزگشاها برای دستیابی به مشابه‌ترین نمودار بازسازی شده ممکن در فضای جدید بهینه شده‌اند.

شبکه کانولوشن گراف یکی از رایج‌ترین معماری‌های مورد استفاده در شبکه‌های عصبی گرافی است. آن‌ها از لایه‌های کانولوشنال گراف مشابه لایه‌های شبکه‌عصبی کانولوشن (CNN) در آرایه‌های اقلیدسی برای یافتن ویژگی‌ها و الگوهای داده‌های محلی استفاده می‌کنند. مهم‌ترین مزیت شبکه کانولوشن گراف یا به عبارتی شبکه‌های عصبی گرافی در مقایسه با شبکه‌عصبی کانولوشن، توانایی آن‌ها برای اعمال در مجموعه‌داده‌های بدون ساختار است. مدل ما از سه لایه GCN تشکیل شده است.

خروجی این مدل یک گراف بازسازی‌شده است به گونه‌ای که برچسب‌های هر گره (دارو) اثرات جانبی آن دارو را پیش‌بینی می‌کند. در واقع هر گره به عنوان برچسب، برداری ۳۸۳ بعدی (تعداد اثرات جانبی در مجموعه‌داده ما) دارد. مدل ما احتمال وقوع هر اثرجانبی را در بردار مشخص می‌کند. در ادامه نتایج بدست‌آمده به کمک این روش از دو منظر ارزیابی عددی و تفسیر بیولوژیکی بررسی می‌شود. کنید.

۳- ارزیابی و پیاده‌سازی

در این فصل به نحوه‌ی پیاده‌سازی و ارزیابی نتایج روش پیشنهادی خواهیم پرداخت. علاوه بر اعتبارسنجی عددی نتایج، به تفسیر بیولوژیکی آن نیز می‌پردازیم. در پایان نتایج بدست آمده را با کارهای پیشین مقایسه می‌کنیم.

۳-۱- نحوه‌ی پیاده‌سازی مدل

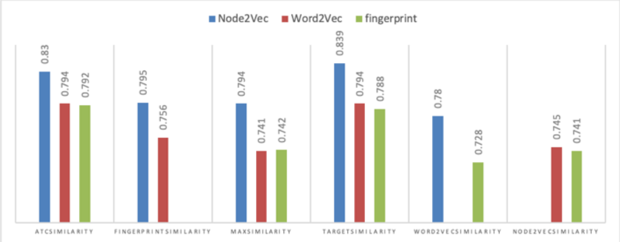
خودرمزگذار گرافی به کمک torch_geometric پیاده‌سازی شده است. مدل ما از سه لایه GCN تشکیل شده است. همه لایه‌ها در مدل رمزگذار یک کانال ورودی و خروجی دارند. کانال‌ها در سه لایه مدل رمزگذار ما به شرح زیر است: (تعداد ویژگی‌ها، ۱۲۸)، (۱۲۸، ۱۲۸)، (۱۲۸، تعداد اثرات جانبی). مدل ما با بهینه‌ساز Adam و نرخ یادگیری ۰.۰۱ آموزش داده می‌شود. برای جلوگیری از بیش‌برازش از حد، لایه‌های dropout با نرخ افت ۰.۵ بین هر دو لایه از لایه‌های GNN وجود دارد. تابع فعالسازی ReLu بین هر لایه GNN و بعد از لایه‌های dropout استفاده می‌شود. مرحله آموزش با ۵۰۰ دوره برای هر یک از مدل‌ها انجام می‌شود. تابع Loss مورد استفاده برای آموزش، BCEWithLogitsLoss است. این تابع در واقع شامل یک لایه‌ی sigmoid و BCELoss با هم است.

همچنین از یک روش اعتبارسنجی متقابل ۵ لایه برای ارزیابی عملکرد استفاده کردیم.

۲-۳- نتایج اعتبار سنجی عددی

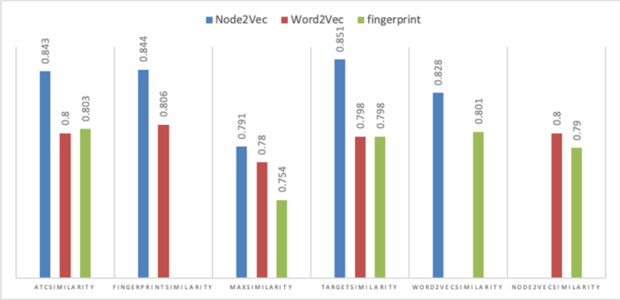
گراف شباهت دارو–دارو از داروهایی تشکیل شده است که در صورت شباهت بین آن‌ها یال قرار می‌گیرد. بر روی هر یال میزان شباهت دارو به عنوان ویژگی یال در نظر گرفته شده است. برای محاسبه‌ی میزان شباهت داروها از شباهت کسینوسی و یا جاکارد بنا بر نوع داده استفاده شد. در این مسئله ما پنج معیار شباهت داشتیم در نتیجه برای مقایسه بهترین نتیجه ۵ گراف با معیار شباهت مختلف ساخته شد. همچنین یک آستانه شباهت تعریف کردیم. اگر میزان شباهت دو دارو از میزان آستانه کمتر باشد بین دو دارو یالی رسم نمی‌شود. برای بررسی نتایج ما سه آستانه‌ی ۰.۲، ۰.۵ و ۰.۷ را برای تمام گراف‌ها در نظر گرفتیم تا بتوانیم آستانه‌ی شباهت مناسب را پیدا کنیم. پس از ساخت گراف ما آن را توسط یکی از سه ویژگی Node2Vec، Word2Vec و اثر انگشت دارویی به گراف افزوده تبدیل می‌کنیم. همه‌ی گراف‌های تهیه شده به مدل شبکه‌عصبی گرافی ما داده می‌شوند تا بهترین نتیجه را بدست بیاوریم.

در نمودار شکل ۲ نتایج AUC بدست آمده با در نظر گرفتن آستانه شباهت ۰.۲ برای همه‌ی گراف‌های ساخته شده قابل مشاهده است. همانطور که مشاهده می‌شود بهترین نتیجه برای گراف با معیار شباهت هدف دارویی و ویژگی افزوده شده Node2Vec است.



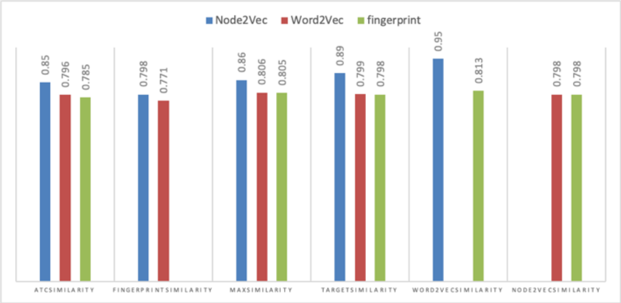
شکل ۲ : مقایسه‌ی نتایج گراف‌ها با آستانه شباهت ۰.۲

در نمودار شکل ۳ نتایج AUC بدست آمده با در نظر گرفتن آستانه شباهت ۰.۵ برای همه‌ی گراف‌های ساخته شده قابل مشاهده است. همانطور که مشاهده می‌شود بهترین نتیجه برای گراف با معیار شباهت هدف دارویی و ویژگی افزوده شده Node2Vec است.



شکل ۳ : مقایسه نتایج گراف‌ها با آستانه شباهت ۰.۵

در نمودار شکل ۴ نتایج AUC بدست آمده با در نظر گرفتن آستانه شباهت ۰.۷ برای همه‌ی گراف‌های ساخته شده قابل مشاهده است. همانطور که مشاهده می‌شود بهترین نتیجه برای گراف با معیار شباهت هدف دارویی و ویژگی افزوده شده Word2Vec است.



شکل ۴ : مقایسه گراف‌ها با آستانه شباهت ۰.۷

بهترین AUC بدست آمده مختص گراف با معیار شباهت هدف دارویی، آستانه ۰.۷ و ویژگی افزوده شده Word2Vec است. این گراف ۱۹۵ گره یا دارو دارد. در ادامه به بررسی اثرات جانبی پیش‌بینی شده توسط این گراف می‌پردازیم.

۳-۳- تفسیر بیولوژیکی

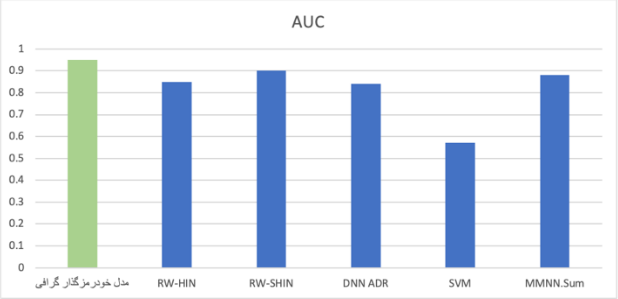
در این بخش، قصد داریم علاوه بر ارزیابی عددی، مدل پیشنهادی را بر اساس تفسیر بیولوژیکی نیز بررسی کنیم. برای این منظور، ما تمام اثرات جانبی داروها را که به عنوان مثبت کاذب وجود داشتند را جمع‌آوری کردیم (در واقع اثرات جانبی که مدل ما پیش‌بینی کرده است اما در مجموعه داده ذکر نشده‌اند) و در مرحله بعد برای اعتبارسنجی این پیش‌بینی‌ها، آن‌ها را در ادبیات پزشکی بررسی کردیم.

- برخی از این اثرات جانبی پیش‌بینی شده در Sider به عنوان اثرات جانبی نادر آن دارو ذکر شده‌اند و مدل ما توسط این اثرات جانبی نادر آموزش ندیده است اما مشاهده می‌شود که توانسته است به خوبی آن‌ها را پیش‌بینی کند. اثرات جانبی نادر در بعضی موارد می‌توانند برای بیمار خطرناک و حتی کشنده باشند در نتیجه دانستن زود هنگام این عوارض می تواند محقق داروساز را یاری دهد.
- برخی دیگر از اثرات جانبی پیش‌بینی شده برای یک دارو در Sider ذکر نشده‌اند. در حالی که در سایت‌ها و مقالات پزشکی بروز این اثرات جانبی برای آن دارو بیان شده است. با توجه به آخرین بروز رسانی Sider که در سال ۲۰۱۵ بوده است. این اثرات جانبی می‌توانند در سال‌های بعد گزارش شده باشند.

در جدول ۱ برای نمونه ۱۰ دارو، اثرات جانبی نادر پیش‌بینی شده توسط مدل ما و اثرات جانبی موجود در مقالات و سایت‌های پزشکی مشاهده می‌شود.

۴-۳- مقایسه با کارهای پیشین

در این بخش به مقایسه نتایج مدل پیشنهادی با کارهای پیشین می‌پردازیم. در این مقایسه مدل مبتنی بر شبکه‌عصبی گرافی (AUC برابر با ۰.۹۵) را با ۵ مدل مرتبط (DNN ADR[12]، RW-HIN[22]، RW-SHIN[22] و MMNN.sum[5]) مقایسه کردیم که در نمودار شکل ۵ نشان داده شده است. مشاهده می‌شود با توجه به معیار AUC روش ما در مقایسه با روش‌های پیشین ۵ درصد بهبود داشته است.



شکل ۵ : مقایسه مدل پیشنهادی و کارهای پیشین: با توجه به معیار AUC روش ما در مقایسه با روش‌های پیشین ۵ درصد بهبود داشته است

۴- نتیجه گیری

داروها در آزمایشگاه‌ها با هدف درمان بیماری تولید می‌شوند اما می‌توانند اثرات منفی نیز داشته باشند. کشف زود هنگام این موارد می‌تواند در زمان و هزینه صرفه جویی کند. رویکردهای سنتی برای تشخیص اثرات جانبی داروها در طی فرایند تولید دارو سنجش‌های آزمایشگاهی و حیوانی است. با این حال امکان کشف تمام اثرات جانبی در محیط آزمایشگاهی وجود ندارد و این اثرات جانبی کشف نشده می‌توانند ضررهای بسیاری برای بیمار و همچنین شرکت تولیدکننده‌ی دارو داشته‌باشند. امروزه روش‌های محاسباتی مختلفی در حوزه‌ی پیش‌بینی اثرات جانبی داروها ارائه شده است.

در این پژوهش نیز با انتخاب ویژگی‌های دارویی مختلف و با استفاده از شبکه‌های عصبی گرافی به پیش‌بینی اثرات جانبی داروها پرداختیم. نتایج بدست‌آمده در این مقاله در کنار تلاش‌های انجام شده در آزمایشگاه‌های پزشکی، می‌تواند به محققان علوم زیستی کمک کند تا در کارهای خود با صرف هزینه‌ی کمتر، نتایج دقیق‌تری را کسب نمایند.

تمرکز این پژوهش بر روی پیش‌بینی اثرات جانبی داروها به کمک روش‌های مبتنی بر شبکه‌های عصبی گرافی بوده است. در این راستا ما این مسئله را از رویکرد شبکه‌های عصبی گرافی مورد بررسی قرار دادیم. در ابتدا ویژگی‌های دارویی را از مجموعه‌داده‌های DCDB، PubMed و Sider استخراج کردیم. این ویژگی‌ها تاثیرات متفاوتی در پیش‌بینی اثرات جانبی داشتند. درنتیجه به مقایسه‌ی تاثیر این ویژگی‌ها در عملکرد پیش‌بینی پرداختیم. در ادامه یک راه حل نوآورانه به کمک شبکه‌های عصبی گرافی و GCNها معرفی کردیم. در این روش ابتدا یک گراف شباهت دارو–دارو ساخته می‌شود. این گراف با معیارهای شباهت و آستانه‌ی شباهت مختلف بارها ساخته شد و نتایج بدست‌آمده از گراف‌های مختلف جهت پیدا کردن بهترین آستانه و معیار شباهت با هم مقایسه شد. گراف بدست آمده به عنوان ورودی به یک خودرمزگذار گرافی داده می‌شود و در نهایت با استفاده از اعتبار سنجی متقابل پنج لایه، ما موفق به AUC برابر ۹۵ درصد شدیم. در پایان نتایج و پیش‌بینی‌های انجام شده توسط مدل معرفی شده در منابع به روز پزشکی بررسی شد.

لازم به ذکر است نتایج بدست آمده و اثرات پیش‌بینی شده توسط مدل پیشنهادی، قطعی نیستند و در کنار عوامل دیگر می‌توانند متخصصان را در ساخت دارو و انجام آزمایشات مربوطه یاری دهند.

نوآوری‌های این مقاله عبارتند از:

- ارائه گراف شباهت دارو–دارو جدید که می‌تواند در زمینه‌های مختلف پزشکی مورد استفاده قرار گیرد.
- ارائه روشی مبتنی بر پیش‌بینی برچسب در گراف برای پیش‌بینی اثرات جانبی
- استفاده از الگوریتم‌های GCN برای پیش‌بینی اثرات جانبی داروها، این الگوریتم‌ها سابقا فقط جهت پیش‌بینی اثرات جانبی چند–دارویی استفاده شده است.

به عنوان کارهای آتی برای اصلاح و گسترش تحقیقات در این زمینه چندین نوآوری جدید می‌توان در نظر گرفت:

- می‌توان روش‌های پیشنهادی را روی مجموعه‌داده‌های بزرگ‌تر با تعداد دارو و تنوع ویژگی‌های دارویی بیشتر اجرا کرد.
- همچنین می‌توان با تمرکز بر داروهای با اهداف درمانی یکسان (مثلا اهداف شیمی درمانی و ضد سرطان)، روش پیشنهادی را محدود به داروهای مشابه کرد.
- به‌دلیل میزان مصرف دارو در شدت و یا بروز اثر جانبی، در آینده می‌توان دوز داروها را نیز در طراحی مدل خود لحاظ کنیم.
- در این پژوهش اثرات جانبی مصرف تک‌دارو بررسی شد، عوارض جانبی داروها می‌توانند به دلیل تداخلات دارویی حاصل شوند. در آینده می‌توان اثرات جانبی حاصل از مصرف چند–دارو و تداخلات آن‌ها را پیش‌بینی کرد.

جدول ۱ : نمونه‌ای از صحت‌سنجس پیش‌بینی‌های مدل در منابع پزشکی

	Drug name	Rare side effects Exists in sider	New side effects	New side effects source
1	**Clobazam**	anxiety, constipation, cough, decreased appetite, depression, dermatitis, diarrhoea, dizziness, dry mouth, hyperhidrosis, pruritus, rash, somnolence, tremor, vomiting	Abdominal pain, chest pain, dyspepsia, gastrointestinal pain, paraesthesia, pharyngitis	[23]
2	**Phenytoin**	agitation, arthralgia, asthenia, body temperature increased, chestpain, constipation, cough, decreased appetite, dizziness, drymouth, dyspepsia, dyspnoea, fatigue, feeling abnormal, hyperhidrosis, hypertension, hypotension, infection, insomnia, mental disability, musculoskeletal discomfort, myalgia, nervousness, paraesthesia, pharyngitis, pruritus, somnolence, tension, tremor, vertigo	Abdominal pain, anxiety, depression, gastrointestinal pain, increased appetite, pain, weight increased	[25] ,[24]
3	**Repaglinide**	-	asthenia, dizziness	[26]
4	**ranitidine**	headache, nausea, vomiting	asthenia	[27]
5	**Lapatinib**	abdominal pain, asthenia, constipation, dermatitis, dyspepsia, fatigue, gastrointestinal pain, headache, insomnia, nausea, rash, vomiting	Body temperature increased, dizziness	[29] ,[28]
6	**Azathioprine**	diarrhoea, dizziness	cough, dyspnoea	[30]
7	**Fluconazole**	dizziness, somnolence	cough	[31]
8	**Tenofovir**	diarrhoea, dizziness, headache, nausea	cough	[32]
9	**Oxcarbazepine**	anxiety', chest pain, dry mouth, fatigue, musculoskeletal discomfort, pharyngitis, somnolence, tremor	depression, paraesthesia	[33]
10	**Boceprevir**	anxiety, depression, dyspepsia, fatigue, insomnia, pain, pruritus	somnolence, tremor	[34]

مراجع

[1] W. Zhang, X. Liu, Y. Chen, W. Wu, W. Wang, and X. Li, "Feature-derived graph regularized matrix factorization for predicting drug side effects," *Neurocomputing*, vol. 287, pp. 154–162, 2018,

[2] Y. Zheng, H. Peng, S. Ghosh, C. Lan, and J. Li, "Inverse similarity and reliable negative samples for drug side-effect prediction," *BMC Bioinformatics*, vol. 19, no. Suppl 13, p. 554, Feb. 2019

[3] X. Zhao, L. Chen, and J. Lu, "A similarity-based method for prediction of drug side effects with heterogeneous information," *Math Biosci*, vol. 306, pp. 136–144, 2018

[4] S. Seo, T. Lee, M.-H. Kim, and Y. Yoon, "Prediction of Side Effects Using Comprehensive Similarity Measures," *Biomed Res Int*, vol. 2020, p. 1357630, Feb. 2020

[5] O. Uner, R. Cinbis, O. Tastan, A. C.- Biorxiv, and undefined 2019, "DeepSide: a deep learning framework for drug side effect prediction," *biorxiv.org*, doi: 10.1101/843029.

[6] S. Dey, H. Luo, A. Fokoue, J. Hu, and P. Zhang, "Predicting adverse drug reactions through interpretable deep learning framework," *BMC Bioinformatics*, vol. 19, no. Suppl 21, p. 476, Dec. 2018, doi: 10.1186/s12859-018-2544-0.

[7] S. Shabani-Mashcool, S.-A. Marashi, and S. Gharaghani, "NDDSA: A network- and domain-based method for predicting drug-side effect associations," *Information Processing & Management*, vol. 57, no. 6, p. 102357, 2020

[8] A. Poleksic and L. Xie, "Predicting serious rare adverse reactions of novel chemicals," *Bioinformatics*, vol. 34, no. 16, pp. 2835–2842, Aug. 2018

[9] H. Liang, L. Chen, X. Zhao, and X. Zhang, "Prediction of Drug Side Effects with a Refined Negative Sample Selection Strategy," *Comput Math Methods Med*, vol. 2020, p. 1573543, May 2020

[10] M. Timilsina, M. Tandan, M. d'Aquin, and H. Yang, "Discovering Links Between Side Effects and Drugs Using a Diffusion Based Method," *Sci Rep*, vol. 9, no. 1, p. 10436, Jul. 2019

[11] S. Wadhwa, A. Gupta, S. Dokania, R. Kanji, and G. Bagler, "A hierarchical anatomical classification schema for prediction of phenotypic side effects," *PLoS One*, vol. 13, no. 3, pp. e0193959–e0193959, Mar. 2018

[12] C. S. Wang, P. J. Lin, C. L. Cheng, S. H. Tai, Y. H. K. Yang, and J. H. Chiang, "Detecting Potential Adverse Drug Reactions Using a Deep Neural Network Model," *J Med Internet Res*, vol. 21, no. 2, Feb. 2019

[13] S. Bang, J. H. Jhee, and H. Shin, "Polypharmacy side-effect prediction with enhanced interpretability based on graph feature attention network," *Bioinformatics*, vol. 37, no. 18, pp. 2955–2962, Sep. 2021

[14] Y. Liu, Q. Wei, G. Yu, W. Gai, Y. Li, and X. Chen, "Database update DCDB 2.0: a major update of the drug combination database," *Database*, vol. 2014, p. 124, 2014

[15] D. Weininger, "SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules," *J Chem Inf Comput Sci*, vol. 28, no. 1, pp. 31–6, Feb. 1988

[16] "RDKit." https://www.rdkit.org/ (accessed Aug. 19, 2022).

[17] D. Amaratunga, H. Göhlmann, and P. J. Peeters, "Microarrays," *Comprehensive Medicinal Chemistry II*, pp. 87–106, 2007

[18] M. Kuhn, I. Letunic, L. J. Jensen, and P. Bork, "The SIDER database of drugs and side effects," *Nucleic Acids Res*, vol. 44, no. D1, pp. D1075–D1079, 2016

[19] Council for International Organizations of Medical Sciences., "Guidelines for preparing core clinical-safety information on drugs : report of CIOMS Working Groups III and V : including new proposals for investigator's brochures.," 1999.

[20] "PubMed." https://pubmed.ncbi.nlm.nih.gov/ (accessed Aug. 19, 2022).

[21] M. Allahgholi, H. Rahmani, D. Javdani, G. Weiss, and D. Módos, "ADDI: Recommending alternatives for drug–drug interactions with negative health effects," *Comput Biol Med*, vol. 125, p. 103969, Oct. 2020

[22] B. Hu, H. Wang, and Z. Yu, "Drug Side-Effect Prediction Via Random Walk on the Signed Heterogeneous Drug Network," *Molecules*, vol. 24, no. 20, Oct. 2019.

[23] "Clobazam (Oral Route) Side Effects - Mayo Clinic." https://www.mayoclinic.org/drugs-supplements/clobazam-oral-route/side-effects/drg-20075333?p=1 (accessed Aug. 25, 2022).

[24] J. Antel and J. Hebebrand, "Weight-reducing side effects of the antiepileptic agents topiramate and zonisamide," *Handb Exp Pharmacol*, vol. 209, no. 209, pp. 433–466, 2012, doi: 10.1007/978-3-642-24716-3_20.

[25] "Phenytoin: MedlinePlus Drug Information." https://medlineplus.gov/druginfo/meds/a682022.html (accessed Aug. 25, 2022).

[26] "Repaglinide (Oral Route) Side Effects - Mayo Clinic." https://www.mayoclinic.org/drugs-supplements/repaglinide-oral-route/side-effects/drg-20067881 (accessed Aug. 25, 2022).

[27] "Ranitidine Hydrochloride Capsules and Tablets (ranitidine hydrochloride) dose, indications, adverse effects, interactions... from PDR.net." https://www.pdr.net/drug-summary/Ranitidine-Hydrochloride-Capsules-and-Tablets-ranitidine-hydrochloride-2965 (accessed Aug. 25, 2022).

[28] "Lapatinib: MedlinePlus Drug Information." https://medlineplus.gov/druginfo/meds/a607055.html (accessed Aug. 25, 2022).

[29] "lapatinib oral: Uses, Side Effects, Interactions, Pictures, Warnings & Dosing - WebMD." https://www.webmd.com/drugs/2/drug-147806/lapatinib-oral/details (accessed Aug. 25, 2022).

[30] "Azathioprine (Oral Route) Side Effects - Mayo Clinic." https://www.mayoclinic.org/drugs-supplements/azathioprine-oral-route/side-effects/drg-20067180 (accessed Aug. 25, 2022).

[31] "Fluconazole (Oral Route) Side Effects - Mayo Clinic." https://www.mayoclinic.org/drugs-supplements/fluconazole-oral-route/side-effects/drg-20071428?p=1 (accessed Aug. 25, 2022).

[32] "Tenofovir (Oral Route) Side Effects - Mayo Clinic." https://www.mayoclinic.org/drugs-supplements/tenofovir-oral-route/side-effects/drg-20066221 (accessed Aug. 25, 2022).

[33] "Oxcarbazepine (By mouth) | Allina Health." https://account.allinahealth.org/library/content/45/1023 (accessed Aug. 25, 2022).

[34] "Boceprevir Side Effects: Common, Severe, Long Term - Drugs.com." https://www.drugs.com/sfx/boceprevir-side-effects.html (accessed Aug. 25, 2022).

چهارمین کنفرانس ملی انفورماتیک ایران
پژوهشگاه دانش‌های بنیادی، پردیس فرمانیه، تهران
۱۴ و ۱۵ دی ماه ۱۴۰۱

پیش‌بینی عملکرد تحصیلی دانشجویان مبتنی بر شبکه کانولوشن گراف و لایه توجه

وجیهه باقری بزآباد[1]، آزاده سلطانی[2]، رابعه روانی‌فرد[3]

[1] گروه کامپیوتر ، دانشکده فنی و مهندسی، دانشگاه بجنورد، بجنورد،
vajihe.bagheri.۷٦@gmail.com

[2] گروه کامپیوتر ، دانشکده فنی و مهندسی، دانشگاه بجنورد، بجنورد،
a.soltani@ub.ac.ir

[3] گروه کامپیوتر ، دانشکده فنی و مهندسی، دانشگاه بجنورد، بجنورد،
ravanifard@ub.ac.ir

چکیده

پیش‌بینی عملکرد دانشجویان، یکی از موضوع‌های اساسی در آموزش مجازی[1] است. عدم تعامل بین استاد و دانشجو در آموزش مجازی سبب می‌شود، شرایط تحصیلی دانشجویان به طور صحیح و کامل تشخیص داده نشود. در سیستم‌های آموزش مجازی، داده‌های مربوط به فعالیت‌های آموزشی دانشجویان جمع‌آوری می‌شود. ما با تحلیل این داده‌ها با استفاده از تکنیک‌های داده‌کاوی[2] می‌توانیم الگوهای آموزشی دانشجویان را تعیین کنیم. این الگوها به اساتید در شناسایی دانشجویان در معرض خطر کمک می‌کند.

در تجزیه و تحلیل داده های آموزشی، از مدل‌های مختلفی از جمله شبکه‌های عصبی[3]، ماشین‌های بردار پشتیبان[4]، بیزین[5]، درخت تصمیم‌گیری[6]، نزدیک‌ترین هم‌سایه و ... استفاده شده است. نتایج نشان می‌دهد که شبکه عصبی دقت بهتری نسبت به سایر روش‌ها دارد. در این تحقیق، روش جدیدی برای پیش‌بینی عملکرد تحصیلی دانشجویان ارائه کرده‌ایم که اولا برای تحلیل اطلاعات مربوط به پیشینه دانشجو از گراف دانش استفاده می‌کند و همچنین به منظور استفاده از داده‌های فعالیتی دانشجو در طول ترم، از مکانیزم توجه[7] بهره می‌برد. نتایج به دست آمده نشان می‌دهد روش پیشنهادی توانسته است دقت بهتری نسبت به مدل مبتنی بر LSTM کسب کند.

کلمات کلیدی

یادگیری عمیق[8]، آموزش مجازی، شبکه عصبی، پیش‌بینی عملکرد دانشجویان.

۱- مقدمه

با پیشرفت روزافزون تکنولوژی، در سیستم‌های آموزشی مبتنی بر فناوری‌های اطلاعات و ارتباطات، همچنین همه‌گیری بیماری کرونا در سراسر جهان، استفاده از محیط و آموزش مجازی در دانشگاه‌ها، مدارس و مراکز علمی و پژوهشی رشد چشم‌گیری پیدا کرده است. آموزش مجازی به صورت دوره‌هایی، از طریق برنامه‌های آنلاین به دانشجویان ارائه می‌شود؛ این دوره‌ها آموزش از راه دور را با شرکت در جلسات آنلاین و آفلاین در سامانه پشتیبانی می‌کنند. آموزش مجازی علی‌رغم نیاز فعلی، بنا به شرایط موجود کمک‌حال دانشجویانی که توانایی شرکت در کلاس‌های حضوری را ندارند، می‌باشد. علاوه بر این موارد، آموزش مجازی به‌عنوان یک ابزار حمایت‌کننده در کلاس‌های حضوری نیز شناخته شده و کاربرد دارد. در سیستم‌های یادگیری آنلاین، به دلیل عدم تعامل کافی اساتید با دانشجو بر خلاف سیستم آموزشی سنتی، در صورت وجود هرگونه کم‌کاری یا هرگونه افت تحصیلی دانشجو، اساتید قادر به ارزیابی جامع هر دانشجو نیستند [۱]. طبق تحقیقات انجام شده در مقالات اخیر، بیش از ۸۰ درصد از دانشجویان آموزش مجازی، در معرض خطر شکست می‌باشند [۲]. در نتیجه، محققین و دانشمندان به دنبال روش‌هایی هستند که بتوانند عملکرد دانشجویان را پیش‌بینی کرده و با مداخله‌ی به موقع، در موفقیت دانشجویان در معرض خطر، نقش بیشتری داشته باشند. پس از تحقیقات و بررسی‌های انجام شده و نتایج حاصل، اثبات شده است که رویکرد عمیق درصد موفقیت بیشتری نسبت به الگوریتم‌های پایه داشته است [۳]. یکی از مهم‌ترین قابلیت‌های آموزش آنلاین، توانایی ذخیره‌ی اطلاعات و جزئیات فعالیت‌های آموزشی دانشجویان به صورت سیستماتیک است. با پردازش و تحلیل این داده‌ها می‌توان، دیدگاهی در رابطه با وضعیت فعلی و آینده‌ی دانشجویان پیدا کرد. این داده‌ها، عموماً بر دو نوع پیشینه و داده‌ی فعالیتی دانشجویان است. نوع پیشینه، شامل اطلاعات بیوگرافی دانشجویان (موقعیت جغرافیایی، شرایط اقتصادی، میزان تحصیلات، جنسیت، سن، تجربه‌ی‌کاری، مدرک تحصیلی و ...) و اطلاعات مربوط به دروس اخذ شده (درس مورد نظر وابسته به چه درس‌هایی در دوره‌های قبل می‌باشد، تخصصی یا عمومی بودن درس، واحد درس و ...) است. در طول هر درس، دانشجویان و اساتید تنها مجاز به استفاده از سامانه الکترونیکی و اسناد الکترونیکی می‌باشند؛ لذا، داده‌ی فعالیتی دانشجویان شامل

اطلاعات فعالیتی دانشجو در جلسات آنلاین (ساعات شروع و اتمام کلاس، حضور و غیاب، میزان تاخیر دانشجو، فعالیت دانشجو در کلاس و ...) و اطلاعات فعالیتی دانشجو در سامانه (دانلود ویدئو، تمرین‌ها، امتحانات، ارسال و زمان ارسال تمرین، میزان بررسی سامانه، تعداد کلیک‌ها، تعداد فایل‌های ارسالی و ...) است. معمولا در مقالات از یکی از دو نوع داده‌ی فعالیتی یا پیشینه و یا هر دو، جهت پیش‌بینی عملکرد دانشجویان استفاده شده است.

تاکنون مقالات مختلف در حوزه‌ی پیش‌بینی عملکرد دانشجویان کار شده است؛ در این مقالات از روش‌های گوناگون پردازش داده استفاده شده است: روش‌های مبتنی بر بردار ویژگی (یادگیری بدون نظارت، خودرمزنگار، شبکه عصبی چند لایه، پیش پردازش داده‌ها و ...)، روش‌های مبتنی بر دنباله زمانی (RNN، LSTM، GRU، CNN، مکانیزم توجه و ...)، روش‌های مبتنی بر گراف (GCN، MLP و ...) و روش‌های مبتنی بر مدل‌های ترکیبی که کامل‌ترین نوع از لحاظ پردازش داده بوده و هر دو نوع داده‌ی فعالیتی و پیشینه را پوشش می‌دهد؛ مدل DOPE که مدلی مبتنی بر مدل ترکیبی است، برای پردازش داده‌های فعالیتی از الگوریتم LSTM استفاده می‌کند. در LSTM برای پیش‌بینی حالت نهایی خروجی مراحل قبل به صورت یک بردار به مدل وارد می‌شود که این موضوع، یک نقطه ضعف برای این مدل محسوب می‌شود. در اصل اطلاعات حالت‌های قبل باید در یک بردار خلاصه شوند که ممکن است بخشی از داده‌ها را از دست بدهیم. همچنین یک مدل متوالی خواهیم داشت که نتیجه یک حالت به حالت‌های قبلش وابسته است و امکان اجرای موازی در آن وجود ندارد. روش مکانیزم توجه می‌تواند این مشکل‌ها را برطرف کند [٤]. مکانیزم توجه داده‌های ورودی را بر مبنای اهمیت هر یک از داده‌ها نسبت به سایر داده‌ها وزن‌دهی کرده و در تصمیم‌گیری از داده‌های مربوط به تمام حالت‌های قبلی استفاده می‌کند. هدف از این پژوهش، ارائه‌ی یک مدل برای پیش‌بینی عملکرد تحصیلی دانشجویان است که مبتنی بر مدل DOPE بوده و از مکانیزم توجه استفاده کند.

۲- کارهای مرتبط

الگوریتم‌های مختلفی جهت پیش‌بینی عملکرد دانشجویان ارائه شده است. این روش‌ها به چهار دسته‌ی مختلف تقسیم بندی می‌شوند: ۱- شبکه عصبی مبتنی بر بردار ویژگی، ۲- شبکه عصبی مبتنی بر دنباله زمانی، ۳- شبکه عصبی مبتنی بر گراف و ۴- شبکه عصبی مبتنی بر مدل ترکیبی. در ادامه، مطالعات انجام شده در هر دسته را بیان خواهیم کرد.

الف) شبکه عصبی مبتنی بر بردار ویژگی:

در این دسته از روش‌ها، اطلاعات موجود در قالب بردارهای ویژگی به شبکه وارد شده و با استفاده از شبکه‌های عصبی چند لایه پیش‌بینی انجام می‌شود. الگوریتم SPPN [٥]، یک الگوریتم مبتنی بر یادگیری عمیق است که از داده‌ی ورودی از نوع پیشینه و شبکه عصبی چند لایه برای گروه‌بندی دانشجویان استفاده می‌کند. در این الگوریتم از یادگیری بدون نظارت و خودرمزنگار پراکنده استفاده می‌شود. با استفاده از خودرمزنگار، مدل خود آموزش دیده و ورودی خود را بازسازی می‌کند، در نتیجه ویژگی‌های مناسب به طور خودکار استخراج می‌شوند. آموزش نودهای لایه‌های پنهان نیز بر مبنای تابع خطای بین داده‌ی اصلی و داده‌ی باز سازی شده صورت می‌گیرد. در نهایت در لایه‌ی خروجی مدل، گروه‌بندی صورت می‌گیرد. در الگوریتم مبتنی بر یادگیری عمیق [٦]، پیش‌بینی به طور هر هفته و بر مبنای داده‌ی فعالیتی دانشجو، از اول دوره تا هفته جاری بر روی شبکه عصبی عمیق می‌باشد. همچنین دانشجویان براساس احتمال ترک‌تحصیل رتبه‌بندی می‌شوند. این رتبه‌بندی می‌تواند به معلمان کمک کند تا مداخله خود را برای دانشجویانی که به کمک فوری نیاز دارند، شخصی کنند؛ سپس به تدریج به سمت دانشجویان ریسک پایین‌تر بروند. این مدل داده ورودی را از طریق تعداد زیادی از لایه‌های پنهان پردازش می‌کند. در نهایت، بر مبنای اهمیت هر ویژگی، ارزیابی مدل به صورت احتمال شکست دانشجو صورت می‌گیرد. الگوریتم NN [۷]، یک الگوریتم بر مبنای شبکه عصبی، با داده‌ی ورودی از نوع فعالیتی است که از شبکه عصبی چند لایه جهت پیش‌بینی موفقیت یا شکست دانشجویان استفاده می‌کند. در این الگوریتم براساس داده‌های جمع‌آوری‌شده، پروفایل دانشجویان ساخته می‌شود. پروفایل هر دانشجو شامل اطلاعات مربوط به فعالیت‌های او در سامانه و مشارکت در جلسات آنلاین است. پیش از استفاده از پروفایل دانشجویان، جهت حذف داده‌های بدون ربط تحت یک پیش پردازش (قوانین فرد خبره) قرار می‌گیرند. از طرفی، اگر بخش کوچکی از داده‌ی پروفایل دانشجویان از دست رفته باشد، با میانگین یا مقدار میانه جایگزین می شود. در نهایت با استفاده از داده‌های پردازش شده فرایند طبقه‌بندی، روی یک شبکه عصبی پیشخور اعمال می‌شود. الگوریتم مبتنی بر ANN [۸] یکی از الگوریتم های حوزه‌ی یادگیری عمیق است. مبنای الگوریتم ANN، سه هدف شناسایی دانشجویان در معرض خطر، شناسایی دانشجویان مایل به کناره‌گیری از درس و شناسایی ویژگی‌های موثر در پیش‌بینی است. دو روش جهت پیش‌بینی عملکرد دانشجویان استفاده می‌شود. روش اول: پیش‌بینی بر اساس داده‌های فعالیتی ناشی از تعامل دانشجو با سامانه و داده‌ی پیشینه می‌باشد؛ در این روش جهت انتخاب ویژگی‌ها و رسیدن به ویژگی‌های کاربردی‌تر از تکنیک کاهش پراکندگی استفاده شده است و در نهایت، داده‌های مهندسی شده به مدل داده می شود. روش دوم، پیش‌بینی بر اساس اطلاعات ناشی از داده‌ی فعالیتی دانشجویان در چارک‌ها می‌باشد. در این روش، احتیاجی به اطلاعات پایان دوره جهت پیش‌بینی نبوده و دوره‌ی آموزشی یک درس چارک‌بندی شده و در نهایت بر مبنای خروجی آخرین لایه‌ی پنهان، پیش‌بینی صورت می‌گیرد. ارزیابی در ANN، تشخیص دانشجو جزو طبقه‌بندی چهارگانه "ممتاز"، "پاس"، "شکست"و "حذف" است. همچنین در این الگوریتم، تعداد کلیک برای هر نوع فعالیت، قبل از شروع دوره، دامنه دیگری از ویژگی‌ها در نظر گرفته می‌شود. استفاده از فعالیت‌های قبل از شروع دوره، به جهت شناسایی دانشجویان مایل به حذف از دانشجویان ممتاز است.

ب) شبکه عصبی مبتنی بر دنباله زمانی:

در این روش‌ها فعالیت‌های هفتگی دانشجویان به صورت یک سری زمانی در نظر گرفته می شود و از روش‌هایی استفاده می شود که این قابلیت را دارند که با داده‌های دو بعدی کار کنند به عبارت دیگر، بعد زمان هم اضافه خواهد شد. الگوریتم GRU_reg [۹] بر مبنای شبکه‌ی عصبی بازگشتی (RNN) می‌باشد. روال کلی الگوریتم براساس یک برنامه هفتگی ساخته شده است؛ به صورتی که داده‌های لاگ مشاهده شده در هفته اول، به طور مستقیم به مجموعه داده‌ها در هفته دوم و نیز برای هفته‌های دیگر اضافه می‌شوند. در نهایت، از داده‌های جمع‌آوری‌شده تا هفته جاری برای پیش‌بینی نتیجه دانشجو در انتهای دوره‌ی تحصیلی استفاده می شود. در گام بعد، ویژگی‌های مهندسی از داده‌های خام استخراج و مدل بر روی آنها آموزش داده می‌شود. معماری

مدل شامل سلول‌های GRU است. در نهایت، عملکرد نهایی دانشجویان با استفاده از یک مسئله‌ی رگرسیون با خروجی بین بازه‌ی (۰- ۱۰۰ ٪) پیش‌بینی می شود؛ از نتیجه‌ی بدست آمده جهت دسته‌بندی چند کلاسه استفاده شده است. الگوریتم SMWT یا مدل دانش‌آموز آموزش‌دیده از مدل معلم [۱۰] یکی از الگوریتم‌های مبتنی بر یادگیری عمیق است. مدل دانش‌آموز به جای آموزش مستقیم روی تمام ویژگی‌ها، خروجی شبکه‌ی بزرگ‌تر (معلم) را تقلید می‌کند. پیش‌بینی در این الگوریتم، بر مبنای داده های فعالیتی و پیشینه دانشجویان است. در مرحله اول، از همه ویژگی‌ها برای آموزش مدل معلم استفاده می‌شود. در مرحله دوم، با توجه به عمکرد مدل معلم، تنها از چند ویژگی اساسی برای آموزش مدل دانش‌آموز استفاده می‌شود و در عین حال از خروجی مدل معلم به عنوان یک سیگنال اضافی برای هدایت آموزش مدل دانش‌آموز استفاده می‌شود. معماری هر دو مدل معلم و دانش‌آموز، ترکیبی از یک LSTM و یک شبکه CNN است، تفاوت بین آن‌ها اندازه لایه پنهان و تعداد ویژگی‌های استفاده شده در آنها است. از نظر تابع خطا، خروجی مدل معلم با خروجی حقیقی، مورد ارزیابی قرار گرفته و آموزش می‌بیند. اما مدل دانش‌آموز بایستی خروجی مشابه با خروجی مدل معلم و خروجی حقیقی داشته باشد. در نهایت، مدل دانش‌آموز دانشجویان را، تعیین گروه می‌کند. در [۱۱]، الگوریتمی مبتنی بر RNN معرفی شده است که از داده فعالیتی به عنوان ورودی استفاده کرده و هدفش تولید احتمال برای طبقه‌بندی و پیش‌بینی رفتار دانشجویان، در یک طبقه‌بندی دوتایی است. در این الگوریتم نشان داده شده است که RNN ساده، از سه معماری مختلف (RNN ساده، GRU و LSTM) بهتر عمل می‌کند. علت این موضوع تغییر در رفتار یادگیری دانشجویان در طول دوره‌ی آموزشی است. الگوریتم CLSA [۳]، از سری الگوریتم‌های یادگیری عمیق است که شبکه‌ی عصبی پیچشی (CNN)، شبکه‌ی حافظه کوتاه‌مدت (LSTM) و مکانیسم توجه را ادغام می‌کند. اول، ورودی یک ماتریس دو بعدی، به ازای هر هفته و تمام ویژگی‌ها، وارد مدل می‌شوند. سپس، به طور اتوماتیک، ویژگی‌های مهم از رفتار دانشجو، با استفاده از لایه‌ی کانولوشن CNN استخراج می شود؛ در ادامه، یک لایه کاملا متصل خروجی متراکم شده‌ی لایه‌ی کانولوشن را به حالت مسطح به مدل LSTM می‌دهد. سپس، LSTM برای استخراج ویژگی‌های سری زمانی استفاده می‌شود. پس از LSTM از مکانیزم توجه جهت وزن‌دهی به ویژگی‌ها استفاده شده است. در نهایت، خروجی مدل پیش‌بینی دو کلاسه می‌باشد. در این مدل، با استفاده‌ی کامل از اطلاعات سری زمانی، امکان پیش‌بینی هفتگی عملکرد دانشجویان نیز فراهم است.

ج) شبکه عصبی مبتنی بر گراف:

این مدل‌ها سعی می‌کنند، ارتباط‌های موجود بین دروس و دانشجویان را در قالب یک گراف دانش مدل کرده و از آن‌ها در فرایند یادگیری بهره ببرند. الگوریتم ACGN [۲] برای پیش‌بینی عملکرد دانشجویان در یک درس، به همه درس‌هایی که در ترم‌های قبل توسط دانشجو اخذ می‌شود توجه دارد. ورودی مدل، دو ماتریس مجاورت و ماتریس ویژگی است. برای استخراج اطلاعات مربوط به همسایگان، از چندین لایه GCN استفاده شده است. خروجی نهایی GCN گراف تعبیه حاوی اطلاعات از تمام درس‌هایی است که یک دانشجو اتخاذ کرده است. پس از لایه‌ی GCN لایه‌ی توجه است؛ با محاسبه‌ی امتیاز توجه، ارزش اطلاعات ارزیابی می‌شود. در نهایت، در مرحله‌ی Flattened خروجی ماتریس تعبیه گراف وزن‌دار مسطح شده و داده‌ها به صورت پشت سر هم و در غالب برداری وارد لایه‌ی کاملا متصل (تابع MLP) می‌شود و خروجی نهایی که همان پیش‌بینی گروه‌بندی هردانشجو در هر درس است تولید می‌شود.

د) شبکه عصبی مبتنی بر مدل ترکیبی:

در این دسته از الگوریتم‌ها هدف ترکیب مدل‌های مختلف شبکه عصبی و گراف دانش است. الگوریتم DOPE یکی از الگوریتم‌های مبتنی بر یادگیری عمیق و گراف دانش است [۱]. در این الگوریتم ابتدا روابط درس ها و دانشجویان به عنوان یک گراف دانش مدل سازی می شود و به جای استفاده مستقیم از ویژگی‌ها، از چند لایه GCN استفاده شده است. به کمک این گراف، بر اساس ویژگی‌های دروس و دانشجویان و همچنین ارتباط‌های بین دانشجویان و دروس (چه دانشجویی چه درسی را اخذ کرده است) بردارهای تعبیه برای دانشجویان و دروس به دست می‌آید. همچنین از یک شبکه عصبی بازگشتی (LSTM)، جهت استخراج داده‌های رفتاری دانشجو استفاده می‌شود. در نهایت، خروجی تعبیه درس و دانشجو و شبکه LSTM ترکیب شده و برای پیش‌بینی عملکرد دانشجو، به یک طبقه‌بند فرستاده می شوند. داده‌ی ورودی این الگوریتم داده‌ی فعالیتی و پیشینه بوده و هدف گروه‌بندی دانشجویان بر مبنای تعداد هفته‌های گذرانده شده است.

۳- پیش زمینه

در این بخش به توصیف الگوریتم‌های استفاده‌شده می‌پردازیم.

۳-۱- الگوریتم GCN

اساس شبکه گراف کانولوشن، بر مبنای CNN می‌باشد. این الگوریتم، بر روی داده‌ها با ساختار گراف اعمال می‌شود. وجود ارتباطات در گراف باعث ایجاد شباهت بین گره‌های گراف شده است. GCN وظیفه‌ی ایجاد بردارهای تعبیه به ازای هر گره را دارد. GCN به کمک ماتریس همجواری بدست آمده از گراف دانش، تعبیه هر گره را تولید می‌کند. لازم به ذکر است، تعبیه هر گره با گره‌های همسایه‌اش شباهت دارد. در حقیقت حین ضرب وزن هر گره در داده‌اش، وزن گره‌های همسایه آن گره نیز در داده‌هایشان ضرب و با هم جمع می‌شوند؛ در نتیجه، گره‌های همسایه هر گره در تولید بردار تعبیه آن گره تاثیرگزار هستند.

۳-۲- مکانیزم توجه

این مدل، از مکانیزم توجه انتخابی توسط چشم انسان الهام می‌گیرد. مکانیزم توجه قادر است به اطلاعات مهم با وزن بالاتر توجه کرده و اطلاعات بی‌ربط با وزن کم‌تر را نادیده بگیرد. هدف اصلی مکانیزم توجه اختصاص وزن بالاتر به اطلاعات مناسب‌تر است. در مکانیزم توجه، یک بردار وزن‌دار داریم که سیستم آن را آموزش می‌بیند و به کمک آن وزن و ارزش هر یک از اطلاعات حاصل می‌شود.

۴- روش پیشنهادی

در این بخش به شرح مسئله و روش پیشنهادی می‌پردازیم.

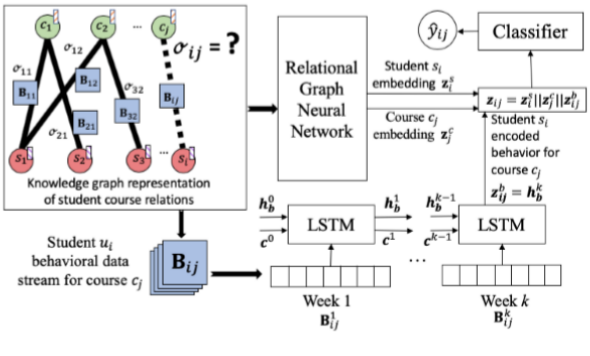
۴-۱- شرح مسئله

در این مسـئله فرض می‌شـود که m درس $C = \{c_۱, c_۲, c_۳, \dots, c_m\}$ و n دانشـجو $S = \{s_۱, s_۲, s_۳, \dots, s_n\}$ وجود دارد که هر دانشـجو حداقل یکی از دروس C را اخذ کرده اسـت. هر درس یکسـری ویژگی‌های بدیهی از قبیل: عنوان، واحد و … دارد که با بردار F_j نشان داده می‌شود. هر دانشجو نیز یک سری ویژگی‌های دموگرافی از قبیل: سن، جنسیت، مدرک تحصیلی و … دارد که با بردار G_i نشـان می‌شـود. در طول دوره‌ی برگذاری درس، داده‌های رفتاری هر دانشجو s_i برای درس c_i، به ازای هر هفته و برای تعداد مشخصی از فعالیتها (از قبیل تعداد کلیک، نمرات آزمون و کلیک) جمع آوری شده و در B_{ij} ذخیره می شود؛ بنابراین B_{ij}، $(B_{ij} = [B_{ij}^۱, B_{ij}^۲, \dots, P_{ij}^k])$ دنباله‌ای از k بردار (k تعداد هفته های جمع‌آوری داده) خواهد بود که هر بردار اطلاعات رفتاری دانشجو s_i برای درس c_i، در یک هفته را نشان می‌دهد. به طور دقیق‌تر $B \in \mathbb{R}^{n \times m \times k \times q}$ که q تعداد فعالیت‌هایی است که اطلاعات آنها به ازای هر هفته ذخیره می‌شـود. در نهایت o_{ij}، کارایی دانشـجوی s_i را برای درس c_j نشـان می‌دهد. در این مسـئله به دنبال مدل پیش‌بینی هسـتیم که با داشتن B_{ij}، F_j و G_j بتواند o_{ij} را پیش‌بینی کند.

۴-۲- روش پیشنهادی

همانطور که در بخش‌های بالاتر مطالعه کردید روش‌های گوناگون پردازش داده برای پیش‌بینی عملکرد دانشـجویان به کار گرفته شـده اسـت: روش‌های مبتنی بر بردار ویژگی (یادگیری بدون نظارت، خودرمزنگار، شبکه عصبی چند لایه، پیش پردازش داده‌ها و …)، مبتنی بر دنباله زمانی (RNN، LSTM، GRU، CNN، مکانیزم توجه و …)، مبتنی بر گراف (GCN، MLP و …) و مبتنی بر مدل‌های ترکیبی که کامل‌ترین نوع از لحاظ پردازش داده بوده و هر دو نوع داده‌ی فعالیتی و پیشـــینه را پوشـش می‌دهد؛ مدل DOPE که مدلی مبتنی بر مدل ترکیبی است، همانطور که در شکل (۱) نشان داده شده اسـت برای پردازش داده‌های پیشـــینه از الگوریتم GCN و برای پردازش داده‌های فعالیتی از الگوریتم LSTM اسـتفاده می‌کند و در نهایت جهت رسیدن به خروجی، بردارهای حاصل از دو مدل با یکدیگر ترکیب شده و به یک طبقه‌بند تزریق می‌شود.

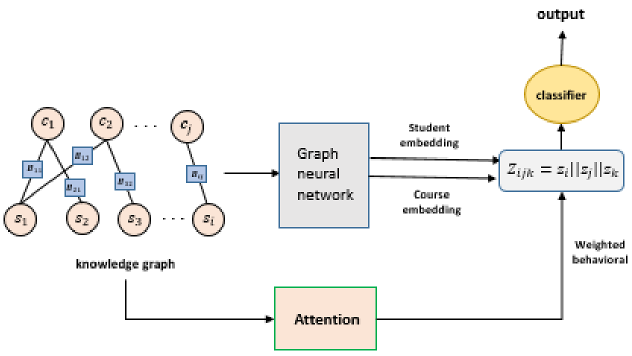
در روش پیشـنهادی به جای اسـتفاده از LSTM از یک لایه توجه اسـتفاده شده است، مکانیزم توجه داده‌های ورودی را بر مبنای اهمیت هر یک از داده‌ها نسبت به سایر داده‌ها وزن‌دهی می‌کند.



شکل (۱): معماری الگوریتم DOPE [۱]

۴-۳- معماری الگوریتم پیشنهادی

مدل GCN_Att مدلی مبتنی بر DOPE اسـت. در این الگوریتم، هدف ترکیب مدل شـبکه عصـبی و گراف دانش اسـت. ابتدا ارتباطهای موجود بین دروس و دانشجویان در قالب یک گراف دانش مدل سازی شده و به الگوریتم GCN جهت کدگذاری داده‌های حاصل از گراف دانش، داده می‌شود. سپس از یک مکانیزم توجه چند واحدی[۹]، جهت اسـتخراج داده‌های رفتاری دانشـجو اسـتفاده می‌شـود. در نهایت، بردارهای تعبیه مربوط به دانشـجو و درس که از نمودار دانش استخراج شده است با بردار نمایش رفتارهای فعالیتی دانشجو که خروجی لایه توجه اسـت، الحاق می‌شـوند. از الحاق این سـه بردار تعبیه در طبقه‌بند، جهت رسـیدن به یک خروجی نهایی اسـتفاده می‌شـود. شـکل (۲) ساختاری از این مدل را نمایش می‌دهد.



شکل (۲): معماری الگوریتم GCN_Att

مدل GCN_Att یک الگوریتم مبتنی بر شبکه کانولوشن گراف و لایه توجه است که از داده‌ی ورودی از نوع پیشینه و فعالیتی برای گروه‌بندی دانشـجویان به دو گروه "قبول شـده" و "مردود شـده" اسـتفاده می‌کند. همچنین، در این الگوریتم برای پیش‌بینی احتیاجی به داده‌های پایان دوره نبوده و پیش‌بینی تنها با استفاده از داده‌های ۲۰ هفته‌ی اول آموزشی صورت می‌گیرد.

۵- مجموعه داده

دیتاست مورد استفاده از دانشگاه بریتیش کلمبیا، کانادا بوده و شامل اطلاعات در مورد دوره‌ها، دانشـجویان و تعاملات آن‌ها با محیط یادگیری مجازی برای هفت درس اسـت که در ماه‌های فوریه و اکتبر ارائه می‌شـود [۱۲]. دیتاسـت شامل هفت جدول داده است که به ترتیب شامل: اطلاعات دروس، اطلاعات ارزیابی دروس، اطلاعات امکانات آموزشی موجود، اطلاعات دموگرافی دانشجو، اطلاعات ثبت نام دانشجو، اطلاعات ارزیابی عملکرد دانشجو و جدول اطلاعات فعالیت دانشجو در سامانه می‌باشد. هر دوره‌ی آموزشی برای هر درس شامل ۴۰ هفته است.

۶- ارزیابی و آزمایش

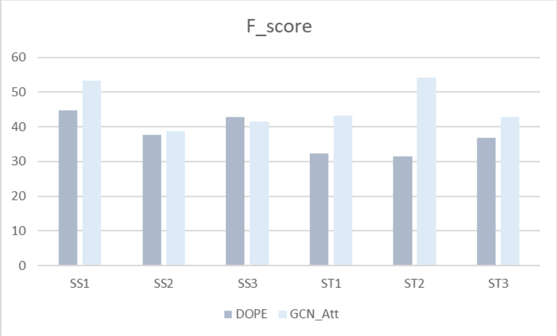
برای ارزیابی مدل GCN_Att، عملکرد آن با مدل DOPE مورد مقایسـه قرار می‌گیرد؛ این آزمایش با مقدار گام عملیاتی[۱۰] ۲۰۰، نرخ یادگیری[۱۱] ۰٫۰۰۱

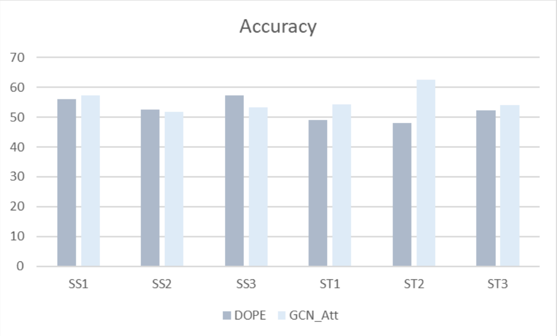
و لایه‌ی توجه با ۴ واحد، بر روی داده‌های ۶ درس مختلف ارائه شده در تاریخ ۲۰۱۳_۲۰۱۴ انجام و نتایج آن در جدول (۱) نمایش داده شده است.

معیارهای ارزیابی مورد استفاده جهت بررسی میزان دقت در این مقایسه، معیارهای: F_score و Accuracy است. همانطور که در جدول مشهود است، میانگین تمامی معیارهای ارزیابی مدل GCN_Att نسبت به الگوریتم DOPE بهتر بوده است که نشان دهنده‌ی عملکرد بهتر الگوریتم GCN_Att نسبت به الگوریتم DOPE است. نتایج حاصل از ارزیابی فوق در نمودار شکل (۳) نیز آورده شده است.

جدول (۱): مقایسه عملکرد روش پیشنهادی با DOPE

			DOPE		GCN_Att	
course	num student	periods	F_score	F_score	Accuracy	Accuracy
SS1	702	2013J,2014J	44/65	53/31	56/08	57/4
SS2	3757	2013J,2014J	37/65	38/68	52/61	51/91
SS3	1590	2013J,2014J	42/77	41/47	57/23	53/4
ST1	3312	2013J,2014J	32/21	43/3	48/97	54/33
ST2	2302	2013J,2014J	31/49	54/05	48	62/49
ST3	2034	2013J,2014J	36/9	42/86	52/32	54/02
Average			37/61167	45/61167	52/535	55/5917





شکل (۳): نمودار مقایسه عملکرد روش پیشنهادی با DOPE

۷- نتیجه گیری

در این مقاله، الگوریتم مبتنی بر شبکه کانولوشن گراف و لایه توجه مورد بررسی قرار گرفته و پیش‌بینی عملکرد دانشجویان به طور هفتگی و در طول دوره‌ی آموزشی انجام شده است. مدل‌های مبتنی بر دنباله زمانی قادر نیستند پیچیدگی و انعطاف‌پذیری داده‌های دانشجویان را به طور کامل ثبت کنند. از طرفی، استفاده از شبکه عصبی مبتنی بر گراف نیز به تنهایی به دلیل عدم پوشش داده‌های سری زمانی توصیه نمی شود. لذا بهترین حالت، استفاده از ترکیبی از دو شبکه عصبی مبتنی بر گراف و مبتنی بر دنباله زمانی است؛ این مقاله بر مبنای همین اصل انجام شده است. در این تحقیق نشان دادیم که جایگزین شدن مکانیزم توجه به جای مدل LSTM، جهت پردازش داده‌ی فعالیتی دقت مدل را به طور چشم‌گیری بهبود می‌بخشد.

مراجع

[۱] H. Karimi, T. Derr, J. Huang, and J. Tang, “*Online Academic Course Performance Prediction Using Relational Graph Convolutional Neural Network.*,” International Educational Data Mining Society, ۲۰۲۰.

[۲] Q. Hu and H. Rangwala, “*Academic performance estimation with attention-based graph convolutional networks*,” arXiv preprint arXiv:۲۰۰۱,۰۰۶۳۲, ۲۰۱۹.

[۳] Q. Fu, Z. Gao, J. Zhou, and Y. Zheng, “*CLSA: A novel deep learning model for MOOC dropout prediction*,” Computers & Electrical Engineering, vol. ۹۴, p. ۱۰۷۳۱۵, ۲۰۲۱.

[۴] A. Vaswani et al., “*Attention is All you Need*,” in Advances in Neural Information Processing Systems, ۲۰۱۷, vol. ۳۰. Accessed: Sep. ۲۰, ۲۰۲۲. [Online]. Available: https://proceedings.neurips.cc/paper/۲۰۱۷/hash/۳f۵ee۲۴۳۵۴۷dee۹۱fbd۰۵۳c۱c۴a۸۴۵aa-Abstract.html

[۵] B. Guo, R. Zhang, G. Xu, C. Shi, and L. Yang, “*Predicting students performance in educational data mining*,” in ۲۰۱۵ International Symposium on Educational Technology (ISET), ۲۰۱۵, pp. ۱۲۵–۱۲۸.

[۶] W. Xing and D. Du, “*Dropout prediction in MOOCs: Using deep learning for personalized intervention*,” Journal of Educational Computing Research, vol. ۵۷, no. ۳, pp. ۵۴۷–۵۷۰, ۲۰۱۹.

[۷] K. Chytas, A. Tsolakidis, N. N. Karanikolas, and C. Skourlas, “*An integrated system for predicting students' academic performance in smart universities*,” in ۲۴th Pan-Hellenic Conference on Informatics, ۲۰۲۰, pp. ۴۱۶–۴۱۹.

[۸] H. Waheed, S.-U. Hassan, N. R. Aljohani, J. Hardman, S. Alelyani, and R. Nawaz, “*Predicting academic performance of students from VLE big data using deep learning models*,” Computers in Human behavior, vol. ۱۰۴, p. ۱۰۶۱۸۹, ۲۰۲۰.

[۹] G. K\Horösi and R. Farkas, “*MOOC performance prediction by deep learning from raw clickstream data*,” in International Conference on Advances in Computing and Data Sciences, ۲۰۲۰, pp. ۴۷۴–۴۸۵.

[۱۰] T. Kang, Z. Wei, J. Huang, and Z. Yao, “*Mooc student success prediction using knowledge distillation*,” in ۲۰۲۰ International Conference on Computer Information and Big Data Applications (CIBDA), ۲۰۲۰, pp. ۳۶۳–۳۶۷.

[۱۱] J. E. M. Fotso, B. Batchakui, R. Nkambou, and G. Okereke, “*Algorithms for the Development of Deep Learning Models for Classification and Prediction of*

Behaviour in MOOCS.," in ۲۰۲۰ IEEE Learning With MOOCS (LWMOOCS), pp. ۱۸۰–۱۸٤.

[۱۲] J. Kuzilek, M. Hlosta, and Z. Zdrahal, "*Open university learning analytics dataset,*" Scientific data, vol. ٤, no. ۱, pp. ۱–۸, ۲۰۱۷.

پانویس ها

۱ Online education
۲ Data mining
۳ Neural Network
۴ Support vector machine
۵ Bayesian
۶ Decision tree
۷ Attention mechanism
۸ Deep learning
۹ Multi head
۱۰ epoch
۱۱ learning rate

چهارمین کنفرانس ملی انفورماتیک ایران
پژوهشگاه دانش‌های بنیادی، پردیس فرمانیه، تهران
۱۴ و ۱۵ دی ماه ۱۴۰۱

شتاب‌دهنده بازپیکربند برای اجرای کارآمد شبکه‌های عصبی کانولوشن

پریا دربانی[1]، حاکم بیت‌الهی[2]، پژمان لطفی‌کامران[3]

[1] دانشکده کامپیوتر، دانشگاه علم و صنعت ایران ، تهران
paria_darbani@cmps2.iust.ac.ir

[2] دانشکده کامپیوتر، دانشگاه علم و صنعت ایران ، تهران
beitollahi@iust.ac.ir

[3] پژوهشکده علوم کامپیوتر، پژوهشگاه دانش‌های بنیادی، تهران
plotfi@ipm.ir

چکیده

لایه‌های شبکه عصبی کانولوشن نسبت به یکدیگر اندازه‌های منحصربه‌فرد و متفاوتی دارند. به دلیل وجود وابستگی داده‌ای، لایه‌ها یکی‌یکی اجرا می‌شوند. شتاب‌دهنده[1]های آرایه‌ای یکی از انواع سکو[2]ها برای اجرای شبکه‌های عصبی هستند که در دستهٔ شتاب‌دهنده‌های سفارشی[3] قرار دارند. ساختار آرایه‌ای آن‌ها امکان بهره‌مندی از پردازش موازی و استفاده چندباره[4] از داده را فراهم می‌کند و به این ترتیب از کارآمدی و مصرف پایین‌تر انرژی بهره می‌برند. اندازه ابعاد آرایه‌ها بر مبنای میانگینی از اندازه تمام لایه‌های شبکه عصبی تعیین می‌شود. هنگام اجرای برخی از لایه‌ها به دلیل عدم همپوشانی کامل آن لایه با ابعاد پردازنده، تعدادی از منابع بی‌استفاده باقی‌می‌مانند. یعنی در کنار کمبود منابع محاسباتی، اختلاف اندازه لایه‌های شبکه عصبی با ابعاد شتاب‌دهنده، مانع از دستیابی به حداکثر کارآیی می‌شود.

معماری بازپیکربند پیشنهادی، مشکل بی‌استفاده ماندن منابع را بهبود داده و در برخی موارد کاملا برطرف کرده‌است. در این معماری اندازهٔ هر عنصر پردازشی، تعداد ردیف‌های عنصر پردازشی و تعداد ستون‌های عنصر پردازشی مطابق با لایهٔ آمادهٔ اجرا تعیین می‌شود. همچنین در این معماری یک میانگیر بازپیکربند مخصوص کم‌هزینه نیز پیشنهاد شده است.

معماری پیشنهادی منجر به افزایش سرعت محاسبات شبکه ResNet تا ۴۵/۹٪ شده است. همچنین دفعات مراجعه به حافظه داخلی ۳۶/۵٪ کاهش پیدا کرده است.

کلمات کلیدی

شبکه عصبی کانولوشن، یادگیری ماشین، شتاب‌دهنده، منابع بی‌استفاده، معماری بازپیکربند.

۱- مقدمه

امروزه هوش مصنوعی در تمام جنبه‌های زندگی کاربرد دارند. از استفاده‌های صنعتی گرفته تا پزشکی و اموزشی. شبکه‌های عصبی کانالوشن با دقت بالایی که فراهم می‌کنند در بخش‌های مختلف از جمله تشخیص تصویر استفاده می‌شوند [1]. تشخیص چهره در تلفن‌های همراه، مانع در ماشین‌های خودران، شی مشکوک در مکان‌های عمومی و یا تشخیص ناحیه سرطانی در عکس‌های پزشکی نمونه‌هایی از کاربردهای شبکه‌های عصبی کانولوشنی هستند.

شبکه‌های عصبی کانولوشنی با محاسبات بسیار زیاد و ارتباطات فشرده با حافظه، نتایج مورد انتظار را با دقت خوبی فراهم می‌کنند. در بعضی از موارد سرعت تشخیص تصویر حیاتی است. برای مثال اگر اتومبیل‌های خودران مانع را با سرعت و دقت بالا تشخیص ندهند ممکن است جان یک انسان را به خطر بیاندازند. در برخی موارد مانند استفاده در سیستم‌های سیار مصرف بهینهٔ انرژی نقش پررنگی را ایفا می‌کند. در نتیجه با توجه به محدودیت ظرفیت‌های سخت‌افزاری، تکنیک‌هایی که منجر به پردازش کارآمد و بهبود مصرف انرژی می‌شوند برای سیستم‌های هوش مصنوعی ضروری هستند.

شبکه‌های عصبی بر روی سکوهای مختلف، مثل پردازنده‌های معمولی، پردازنده گرافیکی[5]، FPGA و معماری‌های سفارشی قابل اجرا هستند که هر کدام از این سکوها نسبت به دیگری دارای مزایا و معایبی است. توان گذردهی[6] پردازنده‌های معمول ضعیف است و پردازنده گرافیکی توان مصرفی بالایی دارد. FPGA از انعطاف بالایی برخوردار است ولی نسبتا زمان زیادی صرف برنامه‌ریزی و پیکربندی آن می‌شود. با وجود هزینه بالای طراحی و ساخت، شتاب‌دهنده سفارشی به دلیل کارآیی بالا در هر وات[7] و مساحت کمترش، برای پیاده‌سازی انواع شبکه عصبی عمیق[8] مورد استفاده قرار گرفته است. شتاب‌دهنده آرایه‌ای که از دسته معماری سفارشی است، ساختار کارآمد و شناخته‌شده‌ای برای اجرای شبکه‌های عصبی عمیق است [2]. این معماری شامل آرایه‌ای از عناصرپردازشی متصل به هم است که گذردهی بالایی را فراهم می‌کنند[6 ,5 ,4 ,3].

به علت وجود وابستگی داده‌ای بین لایه‌ها در شبکه‌های عصبی، لایه‌ها یکی پس از دیگری اجرا می‌شوند. اندازه ابعاد شتاب‌دهنده‌های پیشرفته[9] مانند Laconic [5]، TPU [7] و Eyeriss [4] بر اساس میانگینی از اندازه‌های دسته‌ای از لایه‌های مختلف تعیین شده‌اند. اما در خیلی از مواقع اندازه ابعاد لایه به طور کامل بر آرایه عناصرپردازشی منطبق نمی‌شود [8 ,3]. برای مثال، کوچکتر بودن یکی از ابعاد لایه شبکه عصبی نسبت به بُعد متناظرش در شتاب‌دهنده، منجر به بی‌استفاده ماندن تعدادی از عناصرپردازشی می‌شود. از طرفی اگر بُعد دیگر لایه بزرگتر از بُعد متناظر شتاب‌دهنده باشد، عملیات خواندن از حافظه و پردازش داده باید بارها تکرار[10] شود تا محاسبه لایه پایان پذیرد. این بدان معنی است که با اینکه در هر تکرار تعدادی از عناصرپردازشی

بی‌استفاده بودند، ولی برای جبران کمبود عناصرپردازشی در بُعدِ دیگر، شتاب‌دهنده مجبور به تکرار چندباره عملیات شد. در نتیجه عدم استفاده بهینه از تمام منابع محاسباتی منجر به چرخه‌های محاسباتی ناکارآمد، افزایش زمان اجرا و مصرف توان بیشتر می‌شود[3, 9, 10]. مشکل منابع بی‌استفاده به طور گسترده در شتاب‌دهنده‌ها وجود دارد [2] و مانع دستیابی آن‌ها به حداکثر کارآیی ممکن می‌شود [4, 5, 7, 10].

DaDianNao [11] ساختار مبنا[11] معماری آرایه‌های تکرارشونده است که نسبت به پردازنده‌های گرافیکی تا ۳۳۰ برابر انرژی کمتری مصرف می‌کند [12]. DaDianNao از ۱۶ کاشی[12] تشکیل شده که هر کدام از آن‌ها حاوی تعدادی عنصرپردازشی هستند. هر عنصرپردازشی ۱۶ وزن و ۱۶ ورودی را از حافظه می‌خواند، آن‌ها را جفت‌جفت در هم ضرب می‌کند و به وسیله درخت تجمیعی[13] یک خروجی جزئی را در میانگیر انباشتگر[14] ذخیره می‌کند. پس اندازه ورودی لایه‌های مدلی که قرار است به وسیله DaDianNao اجرا شود باید مضربی از ۲۵۶ باشد تا تمام عناصرپردازشی همزمان و به طور کامل به کار گرفته شوند. اما در واقعیت مدلی با این ویژگی وجود ندارد [12] و همه وزن‌های لایه‌های یک مدل مضرب صحیحی از ۲۵۶ نیستند. برای مثال ٪۹۰ لایه‌های مدل AlexNet مضرب ۲۵۶ نیستند.

ابعاد ثابت در معماری آرایه‌ای مانع از دست‌یابی به یک ساختار کارآمد برای انواع مدل‌های شبکه‌های عصبی کانولوشن با ابعاد مختلف می‌شود [8, 12, 13, 14]. وجود انعطاف در این ساختار می‌تواند به بهینه شدن استفاده از عناصرپردازشی و افزایش کارآیی کمک کند. از این رو ما یک ساختار آرایه‌ای را معرفی کردیم که در آن شتاب‌دهنده بتواند اندازه ابعاد خود را تا جای ممکن به اندازه ابعاد لایه‌ای که قرار است اجرا شود نزدیک کند. معماری پیشنهادی منجر به استفاده بهتر از منابع شده، همچینین زمان اجرا را کاهش داده است. ساختار پیشنهادی سرعت محاسبات مدل گوگل‌نت را حدود ٪۳۲/۱ نسبت به DaDianNao (ساختار پایه) بهبود داده است. این نتایج مربوط به معماری دارای ۱۶ کاشی، هر کاشی دارای ۱۶ ستون و ۱۶ ردیف از عناصرپردازشی است که هر عنصر پردازشی دارای ۱۶ ورودی (تعداد ضرب‌کننده‌ها) است. مسأله دیگر شتاب‌دهنده‌ها مراجعهٔ مکرر به حافظه است که باعث افزایش تاخیر و مصرف انرژی زیاد می‌شود [1]. معماری پیشنهادی تعداد دفعات خواندن از حافظه داخلی را برای مدل گوگل‌نت حدود ٪۲۲ کاهش داده است.

۱-۱- کارهای پیشین

خانوادهٔ DaDianNao ساختار پایه در این دسته و برای این مقاله محسوب می‌شوند. ShiDianNao [10] با کم کردن ابعاد عناصرپردازشی، بهره‌وری را افزایش داده است اما از طرفی این کار باعث بدتر شدن گذردهی شده‌است. طوری که بالاترین کارآیی گزارش شده از این معماری ۱۹۴ GOP/S است [3]. Laconic [5]، ShapeShifter [6] و Strips [12] از پردازش سریال یا پردازش موازی در سطح بیت برای محاسبات مدل‌های شبکه عصبی استفاده می‌کنند. در کنار آن از واحدهای سخت‌افزاری برای مدیریت محاسبات بهره می‌برند. SNAP [15] و stitch-x [16] برای استفاده بهینه از منابع به نگهداری و مدیریت نمایه[15] ورودی‌ها در یک واحد سخت‌افزاری جداگانه می‌پردازند. تمام این روش‌ها باعث افزایش سربار مساحتی و مصرف توان شده‌اند در حالی که مسأله را به طور موثر حل نکرده‌اند. راه دیگر برای کاهش ضررهای بی‌استفاده ماندن منابع، روش‌های متعادل کردن بارمحاسباتی مانند استفاده مجدد از داده، تُنُک‌کردن[16] یا هَرَس[17] کردن است. اما به دلیل اختلاف اندازه‌های وزن‌ها و ورودی‌ها روش‌هایی مثل هرس کردن خود می‌تواند باعث بی‌استفاده ماندن بیشتر منابع شود [17]. Eyeriss [4] یک ساختار بازپیکربند است که از روش‌های استفاده مجدد از داده[18] با هزینه سخت‌افزاری زیاد استفاده می‌کند. Eyeriss در توزیع وزن‌ها دارای محدودیت‌هایی است و در بدترین حالت حدود ٪۵۰ از زمان اجرایش صرف انتقال داده می‌شود. TPU [7] یک آرایه از عناصرپردازشی با ابعاد ۱۲۸ در ۱۲۸و اتصالاتِ سختِ بینشان است که ویژگی مهم آن استفاده مجدد از داده است. Sigma [2] برای حل مشکل منابع بی‌استفاده در TPU از ساختار مائری [8] الهام گرفته و درخت‌جمع‌کننده‌ای به TPU اضافه کرده است. به این ترتیب انعطاف TPU بیشتر شده ولی ٪۸۲ مصرف توان بیشتر و ٪۳۸ مساحت بیشتری به آن تحمیل کرده است. با وجود تمام تلاش‌ها، مسأله بی‌استفاده ماندن منابع به علت ثابت بودن ابعاد پردازنده، همچنان در شتاب‌دهنده‌های امروزی وجود دارد.

۲- معماری پیشنهادی

مسالهٔ منابع بی‌استفاده می‌تواند منجر به تکرارهای بیشتر برای تکمیل پردازش شود [5, 7, 15, 18]. مثلا معماری کانونت [19] هنگام اجرای مدل VGG با اینکه با کمبود منبع مواجه است ولی فقط از ٪۵۵ درصد از منابعش استفاده می‌کند. FPGAها وقتی در حالت تک‌پردازنده، مدل AlexNet را اجرا می‌کنند حدود ٪۷۶ از منابعشان بی‌کار می‌ماند [18].

برای بهبود این مشکل یک شتاب‌دهنده آرایه‌ای را پیشنهاد داده‌ایم که به وسیلهٔ قابلیت بازپیکربندی، امکان تغییر اندازه عناصرپردازشی (تغییر تعداد ضرب‌کننده‌ها در هر عنصر پردازشی)، تغییر تعداد عناصرپردازشی، تعداد ردیف‌ها و تعداد ستون‌ها را می‌دهد. این تغییرات به این معنی است که معماری پیشنهادی می‌تواند اندازه ابعاد خود را به اندازه ابعاد لایه در حال اجرا نزدیک یا بر آن منطبق کند.

مدت زمان پیاده‌سازی یک لایه به ابعاد ورودی، خروجی و وزن لایه و نسبت آن به ابعاد آرایه بستگی دارد. سیکل‌های مورد نیاز برای اجرای یک لایه از شبکه برابر است با:

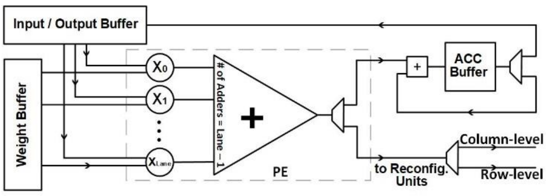
$$C_{cycle} = \lceil \frac{Filters}{Tiles \times Rows} \rceil \times \lceil \frac{Channels}{PE\ size} \rceil \times \lceil \frac{Windows}{Columns} \rceil \times K_x \times K_y \qquad (۱)$$

در معادلهٔ (۱)، Filters، Channels و Windows تعداد فیلترهای وزن، تعداد کانال‌ها و تعداد پنجره‌های ورودی هستند. K_x و K_y ابعاد هر کانال از وزن را نمایش می‌دهد.

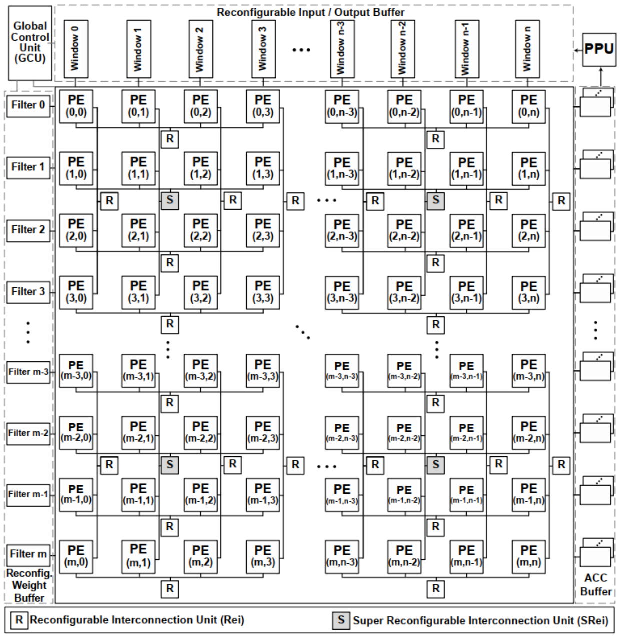
در این معادله صورت کسرها مربوط به مشخصات مدل شبکه عصبی است و مخرج کسرها مربوط به مشخصات آرایه‌ای که آن را اجرا می‌کند. با توجه به این فرمول برای کم کردن چرخه‌های محاسباتی مطلوب است که جواب هر کدام از کسرها تا جای ممکن به ۱ نزدیک شود. در نتیجه صورت و مخرج هر کسر هرچه به هم نزدیک‌تر باشند، بهتر است. این همان ایده‌ایست که معماری پیشنهادی دنبال می‌کند.

این معماری به صورت پیش‌فرض از ۱۶ کاشی تشکیل شده است. عناصر پردازشی در هر کاشی در تعدادی سطر و ستون جایگذاری شده‌اند. تمام این

عناصر به صورت موازی با یکدیگر کار می‌کنند و هر کدام در هر چرخه محاسباتی یک خروجی جزئی تولید می‌کنند. عنصرپردازشی پیشنهادی که در شکل (۱) به وسیله نقطه‌چین محدود شده، ۱۶ ورودی و ۱۶ وزن را از میانگیرهای مربوطه‌شان می‌خوانند؛ آن‌ها را جفت‌جفت به وسیله ۱۶ ضرب‌کننده در هم ضرب می‌کند. سپس خروجی ضرب‌کننده‌ها به وسیله یک درخت‌تجمیعی که شامل ۱۵ جمع‌کننده است با هم جمع می‌شوند. نتیجه این محاسبات می‌تواند در میانگیر انباشتگر ذخیره گردد یا به طور مستقیم به سمت واحدهای بازپیکربند فرستاده شود. اگر به سمت واحدهای بازپیکربند ارسال شود باز دو راه پیش رو است. به سمت واحدهای بازپیکربند مربوط به پیکربندی در سطح ردیف ارسال شود یا به واحدهای بازپیکربند مربوط به سطح ستون. انتخاب یکی از این سه حالت و اجرای آن، توسط واحد مدیریت بازپیکربندی انجام می‌شود. سپس به وسیله برنامه‌ریزی واتسهیم‌گر[۱۹]‌های ۱ به ۲ عملی می‌شود. برنامه‌ریزی واتسهیم‌گر به کمک یک حافظه کوتاه‌مدت یک بیتی با دسترسی تصادفی[۲۰] انجام می‌شود. از زمان ورود داده به عنصرپردازشی تا زمان ذخیره‌شدن در میانگیر انباشتگر در یک چرخه محاسباتی صورت می‌گیرد. شکل (۲) ساختار یک کاشی از معماری پیشنهادی و نحوه قرارگیری واحدهای بازپیکربند را نشان می‌دهد.

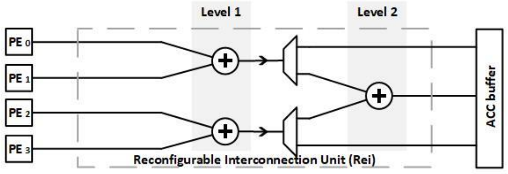


شکل (۱): ساختار عنصرپردازشی پیشنهادی

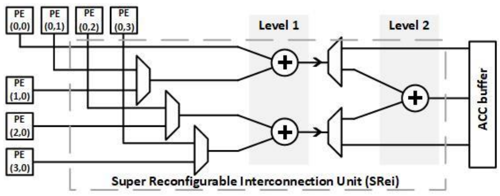


شکل (۲): ساختار کاشی در معماری پیشنهادی

در معماری پیشنهادی دو نوع واحد بازپیکربند برای تامین انعطاف‌پذیری معماری و ادغام عناصر پردازشیِ همسایه در یک ردیف و یا در یک ستون پیشنهاد شده است. بر اساس مطالعاتی که بر روی ابعاد مفروض برای معماری پیشنهادی انجام شد و همچنین با در نظر گرفتن ابعاد لایه‌های چند مدل از شبکه‌های عصبی، واحدهای بازپیکربندی برای ادغام دو تایی و چهار تایی عناصر پردازشی پیشنهاد شد.



شکل (۳): واحد بازپیکربند نوع اول



شکل (٤): واحد بازپیکربند نوع دوم

- واحد بازپیکربند نوع اول

معماری این واحد بازپیکربند در شکل (۳) قابل مشاهده است. ورودی‌های واحدِ بازپیکربندِ نوعِ اول به خروجیِ ۴ عدد عنصر پردازشیِ مجاور یکدیگر متصل است. این چهار عنصر پردازشی در یک ستون و یا همگی در یک ردیف قرار دارند. خروجی‌های جزئیِ دو عنصر پردازشی مجاور، به وسیلۀ جمع‌کننده‌های سطح یک با یکدیگر جمع می‌شوند. اگر هدف معماری ادغام هر دو عنصر پردازشی باشد. تسهیم‌کننده‌ها طوری برنامه‌ریزی می‌شوند که خروجی جمع‌کننده‌های سطح اول به میانگیر انباشتگر منتقل شود. اما اگر هدف معماری ادغام چهار عنصر پردازشی باشد، تسهیم‌کننده‌ها داده را به سمت جمع‌کننده‌ای که در سطح دوم قرار دارد، هدایت می‌کنند.

- واحد بازپیکربند نوع دوم

واحد بازپیکربند نوع دوم که در شکل (۴) قابل مشاهده است، در واقع به هفت عدد عنصر پردازشی تعلق دارد. چهارتا از این عناصر پردازشی در یک ردیف و چهار تا در یک ستون قرار دارند. این واحد پیکربند نیز قابلیت ادغام هر دو عنصر پردازشی مجاور و یا هر چهار عنصر پردازشی مجاور متعلق به یک ستون یا متعلق به یک ردیف را دارد. معماری پیشنهادی قابلیت ادغام عناصر پردازشی در سطح ردیف‌ها و یا در سطح ستون‌ها را دارد. از این رو اگر تصمیم معماری بر ادغام در سطح ردیف باشد، سه ورودیِ بالاییِ واتسهیم‌گر به خروجی منتقل می‌شود و عملیات ادغام بر رویش انجام می‌شود. اگر بازپیکربندی در سطح ستون باشد، سه ورودیِ پایینیِ واتسهیم‌گر به خروجی منتقل می‌شود و عملیات ادغام در سطح ستون انجام می‌گیرد. عملکرد باقی قسمت‌ها همانند واحد بازپیکربندِ نوع اول است.

در شکل (۲) جایگاه واحدهای پیکربندی نوع اول و نوع دوم قابل مشاهده است. هر چهار عنصر پردازشی در یک ردیف یا در یک ستون به یک واحد بازپیکربند نوع اول متصل هستند. در مرکز هر ۱۶ عنصر پردازشی با آرایش ۴ × ۴ یک عدد واحد بازپیکربند نوع دوم، جایگزین نوع اول شده است.

واحدهای بازپیکربند به شتاب‌دهنده این توانایی را می‌دهند که بتواند بر اساس ابعاد لایه اندازۀ عنصر پردازشی، تعداد ردیف‌ها و تعداد ستون‌ها و اندازۀ عناصر پردازشی خود را تغییر دهد. یعنی یک شتاب‌دهنده می‌تواند دارای چند

پیکربندی مختلف برای اجرای لایه‌ها باشد که از میان آن‌ها پیکربندی بهینه را برای اجرای لایه برمی‌گزیند.

- **میانگیر بازپیکربند**

میانگیرهای وزن و ورودی، در معماری‌های پایه به صورت هم‌پخشی[۲۱] عمل می‌کنند. برای مثال یک بانک از میانگیر به تمام عناصر پردازشی یک ستون متصل است. از آنجایی که معماری پیشنهادی، قابلیت تعیین اندازه عناصر پردازشی برای هر لایه را می‌دهد، پس باید میانگیر دارای انعطاف کافی برای به اشتراک‌گذاری داده‌ها به عناصر با ابعاد قابل تغییر را داشته باشد. از آنجایی که نقشهٔ پخش داده با تغییر اندازهٔ عناصر پردازشی متفاوت می‌شود، وجود یک میانگیر با قابلیت بازپیکربندی برای مدیریت جریان داده لازم است.

به جای استفاده از میانگیرهای بازپیکربند متداول که از شبکه‌های توزیعی مانند توری استفاده می‌کنند ما یک میانگیرِ بازپیکربند را پیشنهاد داده‌ایم. ویژگی معماری پیشنهادی این است که همهٔ بانک‌های میانگیر ورودی/وزن لازم نیست که به تمام عناصر پردازشی در یک ستون/ردیف دسترسی داشته باشند، بلکه هر بانک کافی است که قابلیت انتقال داده به تعداد محدودی از عناصر پردازشی را داشته باشد. در این میانگیر از یک شبکه توزیعی انتخابی به وسیله منطق کنترل سوئیچ[۲۲] استفاده شده است. اندازهٔ این میانگیرِ بازپیکربند C برابر اندازه میانگیر در شتاب‌دهندهٔ آرایه‌ای پایه است. که C برابر است با بیشترین تعداد عناصر پردازشی که امکان ادغام با یکدیگر توسط یک واحد بازپیکربند را دارند. این منطق سوئیچ دارای تاخیر کم ($O(1)$)، و سربار و پیچیدگی کمی است.

معماری پیشنهادی دارای یک واحد کنترل‌کننده برای زمان‌بندی عملیات است که سیگنال‌های لازم را برای خواندن یا نوشتن از/ به حافظه، تعیین اندازه بهینه عناصرپردازشی و برنامه‌ریزی تسهیم‌گرها را برعهده دارد. مراحل اجرای معماری پیشنهادی به صورت خلاصه به این ترتیب است: ۱) واحد کنترل‌کننده بر اساس اطلاعات لایه، پیکربندی بهینه را مشخص می‌کند و سیگنال‌های کنترلی را به عناصر پردازشی، واحدهای بازپیکربند و میانگیر بازپیکربند ارسال می‌کند. ۲) اطلاعات ورودی اولیه (عکس، ویدیو، ...) و وزن‌های لایه نخست از حافظه خارجی خوانده می‌شود. ۳) وزن‌ها درون میانگیرهای وزنی مربوط به هر کاشی توزیع می‌شوند. دادهٔ ورودی درون میانگیر مخصوص داده ورودی ثبت می‌شود. ۴) در هر چرخه یک بلوک از داده ورودی به همه واحدهای پردازشی هم‌پخشی[۲۳] می‌شود. هر عنصرپردازشی یک فیلتر (۱۶ کانال وزنی) و ۱۶ کانال ورودی را از میانگیر می‌خواند و نتیجه را در یک بلوک از میانگیر خروجی ذخیره می‌کند. تا پایان محاسبه یک لایه این روند تکرار می‌شود. ۵) اطلاعات از میانگیرخروجی به حافظه خارجی منتقل می‌شود. سپس برای پردازش لایه بعدی به عنوان ورودی به میانگیر مربوطه‌اش انتقال می‌یابد.

۲– ۱- نحوه اجرای لایهٔ کانولوشن

به وسیلهٔ اجزای بازپیکربند گفته شده، امکان تغییر اندازه عناصر پردازشی (تغییر تعداد ضرب‌کننده‌ها در هر عنصر پردازشی) و تغییر تعداد ردیف‌ها و ستون‌ها (تغییر تعداد عناصرپردازشی) را فراهم می‌کند. این تغییرات به این معنی است که معماری پیشنهادی می‌تواند اندازه ابعاد خود را بر اندازه ابعاد لایهٔ در حال اجرا منطبق یا به آن نزدیک کند.

روشی برای پیدا کردن اندازه‌های بهینه پیشنهاد شده است. این روش به این صورت است که تمام حالت‌های ممکن بررسی می‌شود و به وسیله معادله (۱) چرخه محاسباتی آن پیش‌بینی می‌گردد. سپس آن پیکربندی که منجر به تعداد چرخه‌های کمتری شده، برگزیده می‌شود.

برای مثال لایه ششم از مدل ResNet دارای ۱۲۸ فیلتر و ۶۴ کانال است که ابعاد هر کانال وزن آن ۳ در ۳ است. همچنین هر کانال وزن آن ۳۱۳۶ پنجره ورودی دارد. واحد مدیریت پیکربندی مشخصات لایه را در معادله (۱) جایگذاری می‌کند و محاسبات را بر اساس پیکربندی‌های مختلف (اندازه عنصرپردازشی × ردیف × کاشی) واحد بازپیکربند انجام می‌دهد. به این ترتیب تعداد چرخه‌های مورد نیاز برای اجرای لایه ششم به وسیلهٔ ۵ پیکربندی شتاب‌دهنده تخمین زده می‌شود. در این پنج پیکربندی، اندازه عنصر پردازشی، تعداد ردیف‌ها و تعداد ستون‌ها با یکدیگر متفاوت است. برای مثال پیکربندی شماره یک به معنی است که شتاب‌دهنده با ابعاد اولیه خود کار می‌کند. اما در پیکربندی دوم، تعداد ردیف‌ها از ۱۶ به ۸ کاهش پیدا کرده، در عوض اندازهٔ عنصر پردازشی به ۳۲ افزایش یافته است.

1) **Config. 1** $(16\times16\times16)$:
$\lceil 128/(16\times16)\rceil \times \lceil 64/16\rceil \times \lceil 3136/16\rceil \times 3\times 3 = 7056$

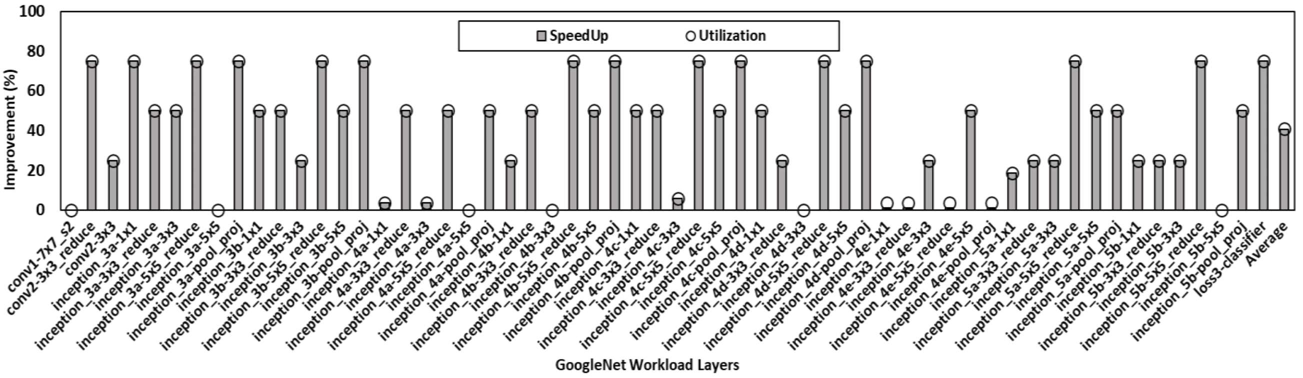
2) **Config. 2** $(8\times16\times32)$:
$\lceil 128/(16\times 8)\rceil \times \lceil 64/32\rceil \times \lceil 3136/16\rceil \times 3\times 3 = 3528$

3) **Config. 3** $(16\times8\times32)$:
$\lceil 128/(16\times 16)\rceil \times \lceil 64/32\rceil \times \lceil 3136/8\rceil \times 3\times 3 = 7056$

4) **Config. 4** $(4\times16\times64)$:
$\lceil 128/(16\times 4)\rceil \times \lceil (64/64\rceil \times \lceil 3136/16\rceil \times 3\times 3 = 3528$

5) **Config. 5** $(16\times4\times64)$:
$\lceil 128/(16\times16)\rceil \times \lceil 64/64\rceil \times \lceil 3136/4\rceil \times 3\times 3 = 7056$

واحد مدیریت بازپیکربندی نتایج محاسبات را با هم مقایسه می‌کند و نتیجه می‌گیرد که دو تا از پیکربندی‌ها (پیکربندی شماره دو و شماره چهار) نسبت به باقیِ پیکربندی‌ها نتیجهٔ بهتری دارد. در این شرایط واحد مدیریت



شکل (٦): بهبود سرعت محاسبات در لایه‌های مدل گوگل‌نت توسط معماری پیشنهادی نسبت به معماری DaDianNao

بازپیکربندی پیکربندی با عنصر پردازشی کوچکتر را انتخاب می‌کند. یعنی پیکربندی شمارهٔ دو به عنوان پیکربندی بهینه برگزیده می‌شود. در این پیکربندی عناصر پردازشی با اندازهٔ ۳۲ در ۸ ردیف و ۱۶ ستون قرار دارند. یعنی بازپیکربندی باید در سطح ستون‌ها اتفاق بیافتد. واحد کنترل کننده سیگنال‌های کنترلی را در جهت ادغام هر دو عنصر پردازشی در یک ستون برای ساختن عنصر پردازشی بزرگتر (اندازهٔ ۳۲) ارسال می‌کند. این به این معناست که عنصر پردازشی اولیه، طوری بازپیکربندی می‌شود که خروجی‌های جزئی خود را به خروجی مربوط به بازپیکربندی در سطح ستون ارسال می‌کند. این خروجی‌های جزئی وارد واحدهای بازپیکربندی می‌شوند. واحدهای بازپیکربندی طوری برنامه‌ریزی شده‌اند که ورودی‌های خود را بعد از جمع‌کننده‌های سطح اول مستقیما به میانگیر انباشتگر بفرستند (جمع‌کنندهٔ سطح دوم بی‌کار است). به این ترتیب رفتارِ عناصرِ پردازشی اولیه و پیکربندیِ واحدهای بازپیکربند، رفتار عناصر پردازشی با تعداد ورودی‌های ۳۲تایی را شبیه‌سازی می‌کند. گویی که واحد بازپیکربند در امتداد درخت جمع‌کننده در عنصر پردازشی قرار گرفته باشد.

به این ترتیب برای اجرای بهینهٔ لایهٔ ششم از ResNet عناصر پردازشی در هر ستون دو تا دو تا با هم ادغام می‌شوند. این پیکربندی تا پایان اجرای لایه ثابت می‌ماند. همین روال برای تمام لایه‌ها انجام می‌شود. بنابراین در نهایت بهره‌وری نسبت به شتاب‌دهنده‌های آرایه‌ای که دارای ابعاد ثابت و از پیش تعیین شده هستند، افزایش می‌یابد.

۳- نتایج شبیه‌سازی‌ها

برای ارزیابی معماری پیشنهادی از یک شبیه‌ساز DNNsim استفاده شد. از معیارهای گوگل‌نت و ResNet برای ارزیابی معماری‌ها استفاده شده است. همچنین ImageNet به عنوان مجموعه داده[۲۴] در نظر گرفته شده‌است.

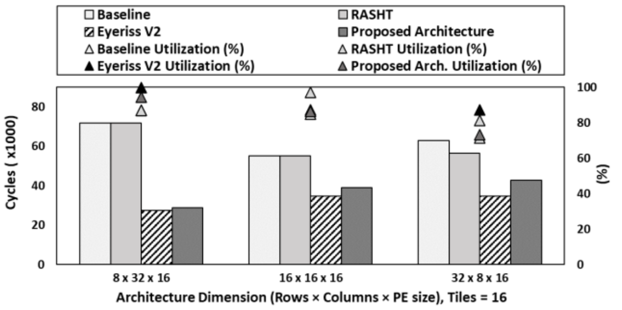
ما معماری پیشنهادی را با معماری مبنای آرایه‌ای، نسخه دوم Eyeriss و RASHT [20] مقایسه کردیم. جریان دادهٔ انتخابی برای ارزیابی‌ها ورودی ثابت[۲۵] است. ما کد HDL معماری پیشنهادی را با استفاده از مجموعه ابزار Design Compiler برای یک کتابخانه سلولی باز ۴۵ نانومتری NanGate، با در نظر گرفتن وزن‌های تصادفی و سایر سیگنال‌های داخلی برای تخمین توان، سنتز کردیم.

شکل (۶) نشان می‌دهد که معماری پیشنهادی سرعت و میزان استفادهٔ بهینه از منابع را در هر لایه از گوگل‌نت به چه اندازه بهبود داده است. برخی از لایه ها بهبود قابل توجهی را نشان می‌دهند. به عنوان مثال، لایه دوم (conv2-3×3_reduce) گوگل‌نت دارای ۶۴ فیلتر، ۶۴ کانال و ۳۱۳۶ پنجره است. بر اساس فرمول (۱)، تعداد چرخه‌های محاسباتی مورد نیاز به وسیلهٔ این پیکربندی ۷۸۴ چرخه است. چهار تا از ۱۶ کاشی (جمعا ۶۴ ردیف) برای ۶۴ فیلتر استفاده می‌شود، در حالی که ۱۲ کاشی باقی‌مانده بیکار هستند. درمعماری پیشنهادی استفاده از تمام منابع موجود، عملکرد این لایه را تا ۴ برابر بهبود داده است. علاوه بر این، چون از تعداد چرخه‌ها کاسته شده، میانگین تعداد مراجعه به حافظه روی تراشه ٪۷۰/۸ کاهش یافته‌است.

ولی، معماری پیشنهادی نتوانسته لایه ۸ (inception_3a-5×5) گوگل‌نت را بهبود دهد. این لایه دارای ۱۶ کانال و ۳۲ فیلتر است که پیکربندی اولیه کاملا بر آن منطبق است. یا به زبان دیگر، پیکربندی بهینه برای این لایه

پیکربندی شماره ۱ تشخیص داده شده. بنابراین، معماری پیشنهادی ترجیح می‌دهد که عناصر پردازشی را ادغام نکند.

در شکل (۷) عملکرد سه پیکربندی مختلف از معماری پیشنهادی با تعداد ثابت منابع با نسخهٔ دوم Eyeriss، RASHT و DaDianNao مقایسه شده‌است. مانند برخی دیگر از شتاب‌دهنده‌ها مثل Sigma، معماری پیشنهادی کمی کندتر از نسخه دوم Eyeriss است. زیرا نسخهٔ دوم Eyeriss وزن‌ها و ورودی‌ها را در حافظه با دسترسی تصادفی ایستا که برای هر عنصر پردازشی وجود دارد، ذخیره می‌کند. در حالی که در معماری پیشنهادی هر دسته از عناصر پردازشی از یک میانگیر استفاده می‌کنند و تنها یک عملوند را ثابت نگه می‌دارند. نسخهٔ دوم Eyeriss به دلیل مسیریاب‌ها، شبکه‌های ارتباطی و حافظه، سربار سخت افزار و مصرف توان قابل توجهی را تحمیل می‌کند. علاوه بر این، نسخهٔ دوم Eyeriss برای شبکه‌های عصبی کانولوشنی کوچک موبایل بهینه شده است، که استفاده از آن را برای مدل‌های بزرگ‌تر محدود می‌کند. این در حالی است که با گذشت زمان شبکه‌های عصبی کانولوشنی بزرگ‌تر می‌شوند. اما معماری پیشنهادی مقیاس‌پذیر است و قابلیت پیاده‌سازی مدل‌های بزرگ را نیز دارد.



شکل (۷): مقایسه تعداد چرخه‌های محاسباتی چند معماری با منابع محاسباتی ثابت

٤- نتیجه‌گیری

اگرچه منابع محاسباتی بیشتری برای محاسبات کارآمد شبکهٔ عصبی کانولوشنی مورد نیاز است، شکل‌ها و اندازه‌های لایه‌های مختلف منجر به بی‌کار ماندن تعدادی از منابع محاسباتی می‌شود. مشکل عدم استفاده بهینه از منابع، که به طور گسترده در معماری‌های آرایه‌ای وجود دارد، مانع از رسیدن آنها به اوج عملکرد می‌شود و باعث افزایش زمان اجرا و مصرف توان بیشتر می‌گردد.

معماری پیشنهادی با استفاده از قابلیت بازپیکربندی، اندازه ابعاد شتاب‌دهنده را برای هر لایه طوری تغییر می‌دهد که تا جای ممکن نزدیک یا مطابق با اندازه ابعاد هر لایه شود. مزیت کلیدی معماری پیشنهادی وجود دو واحد ساده اما موثر و قابل بازپیکربندی است. به این ترتیب با مصرف بهینهٔ منابع، سرعت انجام محاسبات مدل ResNet، نسبت به معماری پایه ٪۴۵/۳ افزایش یافته است. در کنار تمام این مزایا از دقت محاسبات کاسته نشده است.

مراجع

[1] Chen, Kun-Chih, et al. "NoC-based DNN accelerator: A future design paradigm." Proceedings of the 13th IEEE/ACM International Symposium on Networks-on-Chip. 2019.

[2] Qin, Eric, et al. "Sigma: A sparse and irregular gemm accelerator with flexible interconnects for dnn training." 2020 IEEE International Symposium on High Performance Computer Architecture (HPCA). IEEE, 2020.

[3] Liu, Bosheng, et al. "Addressing the issue of processing element under-utilization in general-purpose systolic deep learning accelerators." Proceedings of the 24th Asia and South Pacific Design Automation Conference. 2019.

[4] Chen, Yu-Hsin, et al. "Eyeriss: An energy-efficient reconfigurable accelerator for deep convolutional neural networks." IEEE journal of solid-state circuits 52.1 (2016): 127-138.

[5] Sharify, Sayeh, et al. "Laconic deep learning inference acceleration." 2019 ACM/IEEE 46th Annual International Symposium on Computer Architecture (ISCA). IEEE, 2019.

[6] Chen, Shang-Tse, et al. "Shapeshifter: Robust physical adversarial attack on faster r-cnn object detector." Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Springer, Cham, 2018.

[7] Jouppi, Norman P., et al. "In-datacenter performance analysis of a tensor processing unit." Proceedings of the 44th Annual International Symposium on Computer Architecture. 2017.

[8] Kwon, Hyoukjun, Ananda Samajdar, and Tushar Krishna. "Maeri: Enabling flexible dataflow mapping over dnn accelerators via reconfigurable interconnects." ACM SIGPLAN Notices 53.2 (2018): 461-475.

[9] Dave, Shail, et al. "Hardware Acceleration of Sparse and Irregular Tensor Computations of ML Models: A Survey and Insights." arXiv preprint arXiv:2007.00864 (2020).

[10] Du, Zidong, et al. "ShiDianNao: Shifting vision processing closer to the sensor." Proceedings of the 42nd Annual International Symposium on Computer Architecture. 2015.

[11] Kwon, Hyoukjun, Ananda Samajdar, and Tushar Krishna. "A Communication-centric approach for designing flexible DNN accelerators." IEEE Micro 38.6 (2018): 25-35.

[12] Judd, Patrick, et al. "Stripes: Bit-serial deep neural network computing." 2016 49th Annual IEEE/ACM International Symposium on Microarchitecture (MICRO). IEEE, 2016.

[13] Alwani, Manoj, et al. "Fused-layer CNN accelerators." 2016 49th Annual IEEE/ACM International Symposium on Microarchitecture (MICRO). IEEE, 2016.

[14] Parashar, Angshuman, et al. "Scnn: An accelerator for compressed-sparse convolutional neural networks." ACM SIGARCH Computer Architecture News 45.2 (2017): 27-40.

[15] Zhang, Jie-Fang, et al. "SNAP: A 1.67—21.55 TOPS/W sparse neural acceleration processor for unstructured sparse deep neural network inference in 16nm CMOS." 2019 Symposium on VLSI Circuits. IEEE, 2019.

[16] Lee, Ching-En, et al. "Stitch-x: An accelerator architecture for exploiting unstructured sparsity in deep neural networks." SysML Conference. 2018.

[17] Liu, Zhi-Gang, Paul N. Whatmough, and Matthew Mattina. "Systolic Tensor Array: An Efficient Structured-Sparse GEMM Accelerator for Mobile CNN Inference." IEEE Computer Architecture Letters 19.1 (2020): 34-37.

[18] Shen, Yongming, Michael Ferdman, and Peter Milder. "Maximizing CNN accelerator efficiency through resource partitioning." 2017 ACM/IEEE 44th Annual International Symposium on Computer Architecture (ISCA). IEEE, 2017.

[19] Moons, Bert, and Marian Verhelst. "An energy-efficient precision-scalable ConvNet processor in 40-nm CMOS." IEEE Journal of solid-state Circuits 52.4 (2016): 903-914.

[20] P. Darbani, N. Rohbani, H. Beitollahi and P. Lotfi-Kamran, "RASHT: A Partially Reconfigurable Architecture for Efficient Implementation of CNNs," in IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 30, no. 7, pp. 860-868, July 2022, doi: 10.1109/TVLSI.2022.3167449.

[۲۱] پریا دربانی، نظام رهبانی، پژمان لطفی کامران و حاکم بیت الهی، "معماری یک شتاب‌دهنده بازپیکربند برای شبکه‌های عصبی عمیق جهت افزایش بهره‌وری"، دومین کنفرانس ملی انفورماتیک ایران، NIC 99.

[۲۲] پریا دربانی، نظام رهبانی، پژمان لطفی کامران و حاکم بیت الهی، "شتاب‌دهنده دارای قابلیت بازپیکربندی برای اجرای کارآمد شبکه‌های عصبی"، نشریه علوم رایانشی، پیاپی ۲۰، صفحه ۱۲ – ۲۱، بهار ۱۴۰۰.

پانویس ها

---

1 Accelerator
2 Platform
3 ASIC
4 Data reusing
5 GPU
6 Throughput
7 Performance per watt
8 Deep Neural Network
9 State-of-the-art
10 Iteration
11 Baseline architecture
12 Tile
13 Reduction-tree
14 Accumulator buffer
15 Index
16 Sparsity
17 Pruning
18 Data reusability
19 De-multiplexer
20 1-bit SRAM
21 Multi-Casting
22 Switching control logic
23 Broadcast
24 Data-set
25 Input stationary

چهارمین کنفرانس ملی انفورماتیک ایران
پژوهشگاه دانش‌های بنیادی، پردیس فرمانیه، تهران
۱۴ و ۱۵ دی ماه ۱۴۰۱

بهبود دقت وارسی مدل آماری برای فرایندهای تصمیم مارکوف

برای چهارمین کنفرانس ملی انفورماتیک ایران

محمدصادق محققی[1]، زینب قاسمی[2]

[1] استادیار، گروه علوم کامپیوتر، دانشگاه ولیعصر رفسنجان،
mohagheghi@vru.ac.ir

[2] فارغ التحصیل کارشناسی ارشد، گروه علوم کامپیوتر، دانشگاه ولیعصر رفسنجان
Zeinab.ghasemi۹۶۹۶@gmail.com

چکیده

انفجار حالات یکی از مشکلات اصلی وارسی مدل در همه انواع آن است. وارسی مدل آماری به عنوان یکی از راه حل‌های پیشنهادی برای این مشکل ارائه شده است. این روش به طور ویژه برای راستی‌آزمایی سیستمهایی که دارای برخی جنبه‌های تصادفی هستند مورد توجه قرار گرفته است. تعمیم وارسی مدل آماری به تحلیل فرآیندهای تصمیم مارکوف مستلزم حل عدم قطعیت برای حالت‌های هر مدل است. هر چند تا کنون راه حل‌هایی برای این مسئله ارائه شده‌اند هر یک از این راه حل‌ها دارای محدودیت‌هایی هستند. در این مقاله راه حل جدیدی مبتنی بر یادگیری ماشین و با استفاده از درخت تصمیم برای حل مسئله عدم قطعیت در فرآیندهای تصمیم مارکوف ارائه می‌شود. استفاده از این روش دو راهکار مهم را برای غلبه بر مشکل انفجار حالات به وجود می‌آورد. راهکار نخست استفاده از درخت تصمیم برای حل عدم قطعیت در وارسی مدل آماری است با هدف تقریب احتمال دسترس‌پذیری است. راهکار دوم به شناسایی حالت‌های دسترس‌پذیر از حالت شروع با هدف کاهش اندازه مدل مورد نیاز می‌پردازد. نتایج تجربی نشان می‌دهد دقت محاسبات وارسی مدل آماری با استفاده از این روش بین ۹۰ تا ۹۹ درصد است و اندازه مدل بین ۳۰ تا ۹۸ درصد کاهش می یابد.

کلمات کلیدی

وارسی مدل احتمالاتی، فرایندهای تصمیم مارکوف، کاهش مدل، روش‌های تقریبی آماری، سیاست بهینه، یادگیری ماشین، درخت تصمیم

۱- مقدمه

وارسی مدل یکی از روش‌های راستی‌آزمایی صوری برای تحلیل ویژگی‌های کیفی و کمی سیستم‌های کامپیوتری است. در این روش از یک سیستم تغییروضعیت با ساختاری مبتنی بر گراف برای مدل‌سازی انتزاعی سیستم و از یک منطق زمانی برای توصیف ویژگی‌ها و نیازهای مورد نظر استفاده می‌شود. یک ابزار نرم‌افزاری (که وارسی‌گر مدل نامیده می‌شود) بررسی می کند که آیا مدل ارائه شده ویژگی‌های مورد نظر را براورده می‌کند یا خیر. بسته به ماهیت سیستم و محیط مورد نظر انواع رده‌های وارسی مدل مانند وارسی مدل سیستم‌های بلادرنگ و وارسی مدل احتمالاتی تعریف و توسعه داده شده‌اند [۱،۲،۳]. هدف از وارس مدل احتمالاتی تحلیل ویژگی‌های کمی و کیفی سیستم‌هایی با ویژگی‌های رفتاری تصادفی و غیرقطعی است. در این رده، از زنجیره‌های مارکوف با زمان گسسته و یا فرایندهای تصمیم مارکوف برای مدلسازی استفاده می‌شود [۳،۵].

با اینکه نمونه‌ها و مثال‌های متعدد نشان از موفقیت به‌کارگیری وارسی مدل در راستی‌آزمایی سیستم‌های کامپیوتری دارند [۲،۳] چالش‌هایی برای توسعه و استفاده همه‌گیر از آن وجود دارد. یکی از این چالش‌ها مسئله انفجار حالات است. این مسئله به دلیل رشد انفجارگونه فضای حالت‌های سیستم که معمولا از مرتبه نمایی نسبت به تعداد مولفه‌های آن است اتفاق می‌افتد. در نتیجه به دلیل کمبود حافظه در عمل امکان انجام کامل فرایند وارسی مدل وجود نخواهد داشت. برای مقابله با این چالش روش‌های گوناگونی ارائه شده‌اند که از جمله می‌توان به روش‌های کاهشی [۶،۷،۸،۱۲،۱۷] و استفاده از ساختار داده فشرده [۱۰] اشاره کرد. یکی از روش‌های موفق برای مقابله با انفجار حالات وارسی مدل آماری است که به طور خاص در مورد سیستم‌های احتمالاتی قادر به تحلیل مدل‌هایی با تعداد 10^{200} حالت و بیشتر هستند [۴،۱۸]. در این روش به جای ایجاد مجموعه فضای حالات و انجام روش‌های تکراری عددی یا مبتنی بر گراف از شبیه‌سازی مجموعه‌ای از اجراها استفاده می‌شود. سپس با تحلیل نتایج هر شبیه سازی احتمال رسیدن به حالت پایانی یا امید ریاضی پاداش انباشت شده تخمین زده می‌شود. مزیت اساسی این روش پیچیدگی حافظه مصرفی پایین آن و زمان اجرای منطقی آن (معمولا چند دقیقه حتی بر روی کامپیوترهای قدیمی) است [۱۵].

هرچند وارسی مدل آماری برای تحلیل سیستم‌هایی که با زنجیره‌های مارکوف مدلسازی شده‌اند نتایج راضی کننده‌ای را نشان می‌دهد، تعمیم آن برای فرایندهای تصممم مارکوف (MDPs) نیازمند حل عدم قطعیت برای حالت‌های یک مدل است. روش استاندارد حل عدم قطعیت استفاده از مفهوم سیاست است که در آن به هر حالت یکی از کنش‌های آن با هدف بهینه‌سازی

جواب مورد نظر نگاشت می‌شود. به دلیل پیچیدگی خطی حافظه مصرفی سیاست‌ها نسبت به اندازه مدل، امکان ذخیره کامل اطلاعات سیاست‌ها برای کاربرد در وارسی مدل آماری تنها برای مدل‌های کوچک امکان‌پذیر است [۸]. برای مقابله با این مشکل تا کنون چندین راه حل ارائه شده‌اند. روش نمونه-برداری هوشمند سبک [۱۵] با تولید تصادفی مجموعه‌ای از سیاست‌ها سعی می‌کند تقریب مناسبی از مقدار واقعی ارائه دهد. عملکرد این روش به ساختار مسئله و فراوانی سیاست‌های نزدیک به بهینه است و در بدترین حالت تعداد سیاست‌های مورد نیاز آن نسبت به اندازه مدل از مرتبه نمایی است. استفاده از برون‌نمایی برای تقریب سیاست بهینه در [۱۳] ارائه و بررسی شده است. استفاده از یادگیری عمیق برای تقریب سیاست بهینه در یادگیری تقویتی در [۱۴,۲۱] مورد بررسی قرار گرفته است. با توجه به محدودیت‌های روش‌های موجود در تخمین سیاست بهینه برای استفاده در وارسی مدل آماری در این مقاله روشی جدید برای تخمین سیاست بهینه با استفاده از یادگیری ماشین ارائه می‌شود. در این روش از تعدادی مدل کوچکتر برای مسئله داده شده استفاده می‌شود. ابتدا سیاست بهینه برای هر یک از این مدل‌ها محاسبه شده و مجموعه حالت‌های این مدل‌ها و کنش‌های بهینه برای مرحله آموزش استفاده می‌شوند. برای این هدف از درخت تصمیم به عنوان یک دسته بند استفاده می‌شود. در مرحله بعد از درخت تصمیم ساخته شده برای حل عدم قطعیت‌های یک مدل بزرگ از همان نوع مسئله استفاده می‌شود تا فرایند وارسی مدل آماری بتواند انجام شود. وجود درخت تصمیم برای پیش‌بینی سیاست بهینه همچنین می‌تواند در وارسی مدل احتمالاتی استاندارد برای کاهش فضای حالات مدل استفاده شود. در این روش می‌توان مجموعه حالت‌های دسترس‌پذیر از حالت شروع را بدون نیاز به محاسبه کل فضای حالات به دست آورد. مزیت اصلی روش ارائه شده در این مقاله نسبت به روش‌های پیشین مبتنی بر یادگیری ماشین برای تخمین سیاست بهینه عمومیت آن و عدم وابستگی به اطلاعات خاص مساله است. در حالی که روش ارائه شده در [۱۴،۲۱] بر مساله پاداش مورد انتظار تمرکز دارد و از فاکتور تخفیف[1] برای تخمین کنش‌های بهینه استفاده می‌کند روش ارائه شده ما بر احتمال دسترس‌پذیری بهینه (که در وارسی مدل اهمیت ویژه‌ای دارد) متمرکز است. برای این دسته از ویژگی‌ها به دلیل عدم امکان استفاده از فاکتور تخفیف، امکان تعمیم روش‌های [۱۴،۲۱] وجود ندارد. ضمنا روش ارائه شده در [۱۳] مبتنی بر داشتن اطلاعات اضافه از تعریف مساله است و قابلیت تعمیم آن به سایر مسائل مشخص نیست. در مقابل، روش ارائه شده در مقاله ما قابل استفاده در تمام مواردی است که در تعریف مدل یک یا چند پارامتر برای تغییر اندازه مدل ساخته شده وجود دارد.

هرچند استفاده از درخت تصمیم در [۲۰] نیز مورد استفاده بوده است اما استفاده از آن برای بهبود عملکرد وارسی مدل آماری و نیز کاهش فضای حالات تا کنون مورد بررسی قرار نگرفته است. به طور خلاصه نوآوری‌های اصلی مقاله جاری به شرح زیرند:

- استفاده از درخت تصمیم برای پیش بینی سیاست بهینه در حل مسائل دسترس‌پذیری احتمالاتی برای فرایندهای تصمیم مارکوف با هدف بهبود دقت محاسبات
- استفاده از سیاست پیشنهادی درخت تصمیم آموزش دیده برای کاهش فضای حالات یک فرایند تصمیم مارکوف
- افزودن حداقل یک ویژگی اضافه به مجموعه ویژگی‌های داده-های مورد استفاده در آموزش با هدف افزایش دقت درخت‌های تصمیم در پیش‌بینی سیاست بهینه

در ادامه این مقاله ابتدا مفاهیم و تعاریف اولیه ارائه می‌شوند. سپس در بخش ۳ روش‌های پیشنهادی ارائه و توضیح داده می‌شوند. در بخش ۴ نتایج تجربی ارائه می‌شوند و سرانجام در بخش ۵ نتیجه‌گیری و موارد مربوط به کارهای آینده ارائه می‌شوند.

۲- تعاریف و مفاهیم اولیه

یک توزیع احتمال بر روی مجموعه S تابعی مانند $\mu : S \to [0,1]$ است که در آن $\sum_{s \in S} \mu(s) = 1$

تعریف ۱. یک فرایند تصمیم مارکوف (MDP) یک چندتایی به صورت $M = (S, s_0, Act, \delta, G)$ است که در آن S یک مجموعه متناهی به نام فضای حالات، $s_0 \in S$ یک حالت آغازین، Act مجموعه‌ای متناهی از کنش‌ها است به شکلی که برای هر حالت $s \in S$ تعدادی کنش تعریف می‌شود و با Act(s) نشان داده می‌شود، $\delta : S \times Act \to \mu$ یک تابع تغییروضعیت احتمالاتی است که به هر حالت و کنش یک توزیع احتمال را نسبت می‌دهد.

هر حالت $s \in S$ یکی از پیکربندی‌های ممکن برای سیستم مدل شده را نشان می‌دهد. رفتار پویای سیستم با دنباله‌ای از حالت‌ها و کنش‌ها با شروع از حالت s_0 دنبال می‌شود: با قرار داشتن در یک حالت ممکن از $s \in S$ یکی از کنش‌های $\alpha \in Act(s)$ به طور غیرقطعی انتخاب می‌شود و حالت بعدی s' به طور تصادفی با احتمال $\delta(s,\alpha)(s')$ انتخاب می‌شود. به طور رسمی یک مسیر به صورت $\pi = s_0\alpha_0 s_1\alpha_1 ...$ نشان داده می‌شود که در آن $\alpha_i \in Act(s_i)$ و $\delta(s_i, \alpha_i)(s_{i+1}) > 0$ برای هر $i > 0$ برقرار است. یک زنجیره مارکوف زمان-گسسته (DTMC) یک MDP است که در آن $|Act(s)| = 1$ برای هر حالت $s \in S$ برقرار باشد یعنی به ازای هر حالت آن فقط یک کنش وجود داشته باشد. روش استاندارد برای حل عدم قطعیت در MDP ها استفاده از مفهوم سیاست به عنوان تابعی از یک مسیر متناهی به Act است. برای هر سیاست σ یک معیار احتمالاتی قابل تعریف است. جزئیات بیشتر در مورد مفهوم سیاست و استفاده از آن در حل عدم قطعیت در [۱،۲،۸،۹] در دسترس است.

۲-۱- ویژگی‌های دسترس‌پذیری

رده مهمی از ویژگی‌های مورد بررسی در وارسی مدل احتمالاتی به محاسبه ویژگی‌های دسترس‌پذیری (محدود یا نامحدود) تقلیل می‌یابند. این ویژگی‌ها به صورت احتمال رسیدن به یکی از حالت‌های هدف G تعریف می‌شوند. در مورد MDP ها، ویژگی‌های دسترس‌پذیری به صورت احتمال کمینه یا بیشینه رسیدن به یک حالت هدف (بر روی مجموعه سیاست‌های ممکن) قابل تعریفند. با در نظر گرفتن x_s به عنوان احتمال بهینه (بیشینه یا کمینه) رسیدن به یکی از حالت‌های هدف می‌توان این مقدار را با حل معادله زیر (معروف به معادله *بلمن*) [۵] به دست آورد:

$$x_s = \begin{cases} 1 & \text{if } s \in G \\ \max_{\alpha \in Act(s)} \sum_{s' \in S} \delta(s,\alpha)(s') \times x_{s'} & \text{if } s \notin G \end{cases} \quad (۱)$$

همچنین سیاست بهینه سیاستی است که با توجه به بهترین کنش‌های انتخاب شده به دست می‌آید:

$$\sigma(s) = \arg\max_{\alpha \in Act(s)} \sum_{s' \in S} \delta(s,\alpha)(s') \times x_{s'} \quad (۲)$$

برای حل معادله بلمن می‌توان از روش‌های مستقیم مانند برنامه ریزی خطی و یا روش‌های تکراری استفاده کرد [۱،۵]. بـه دلیـل پیچیـدگی زمـانی روش برنامه‌ریزی خطی، محاسبه ویژگی‌های دسترس‌پذیری معمـولا یـا اسـتفاده از روش‌های تکرار شونده و در دو مرحله انجام می‌شـود: در مرحلـه نخسـت بـا استفاده از تحلیل‌های مبتنی بر گراف مجموعه حالت‌هایی که احتمال رسـیدن به G برای آنها صفر یا یک است محاسبه می‌شوند. سپس بـرای حالـت‌هـای باقی‌مانده از روش‌های تکراری عددی (مانند روش تکرار مقدار[۲]) برای محاسبه احتمال‌های دسترس‌پذیری استفاده می‌شود. شرط پایان تکرار رسیدن حـداکثر اختلاف مقادیر بین دو تکرار متوالی به کمتر از یک حـد آسـتانه اسـت. بـرای تضمین دقت محاسبات انجام شده در روش‌هـای تکـراری، روش تکـرار بـازه پیشنهاد شده است [۹].

۲-۲- زبان مدلسازی PRISM

در وارسی مدل از زبان‌های مدل‌سازی سطح بالا بـرای توصـیف مولفـه‌هـای برنامه استفاده می‌شود. در این مقاله از PRISM به عنوان یک زبان مدلسازی سطح بالا برای وارسی مـدل احتمـالاتی [۴] اسـتفاده مـی‌شـود. هـر برنامـه PRISM شامل تعدادی مولفه است که بر روی یک یا چند متغیر بـا دامنـه اعداد صحیح در یک بازه مشخص تعریف می‌شوند. حد بالای مقادیر هر متغیر یا به طور صریح و یا بر اساس یک پـارامتر مشـخص مـی‌شـود. یـک متغیـر *پارامتری* نامیده می‌شود اگر حد بالای آن توسط یک پـارامتر مشـخص شـود. دستورات گاردار[۳] چگونگی تغییر مقدار متغیرها را مشخص می‌کنند. هر برنامه PRISM به یک سیسـتم تغییروضـعیت سـطح پـایین (ماننـد MDP طبـق تعریف ۱) کامپایل می‌شود. سپس محاسبات مبتنی بر گراف و تکراری عـددی بر روی مدل کامپایل شده انجام می‌شوند.

مثال ۱. بخشی از یک برنامه PRISM مربـوط بـه مسـئله consensus از مثال‌های استاندارد ابزار PRISM [۴] در شکل ۱ نشان داده شده است. ایـن برنامه دارای تعدادی ثابت است که مقدار ثابت دوم بـه عنـوان یـک پـارامتر توسط کاربر تعیین می‌شود. این برنامـه دارای یـک متغیـر سراسـری بـه نـام counter است که حد بالای مقادیر آن به عنوان تابعی از پارامتر K محاسبه می‌شود. همچنین بخشی از تعریف مولفه process۱ در شکل (۱) نشان داده شده است. این مولفه شامل دو متغیر است کـه دامنـه مقـادیر آن مشـخص و ثابت است. دو دستور تغییر مقادیر متغیرهای این برنامه نیز در شکل نشان داده شده است.

۳-۲- یادگیری ماشین برای وارسی مدل

یادگیری ماشین یکی از شاخه‌های هوش مصنوعی اسـت کـه شـامل توسـعه الگوریتم‌های آماری برای استخراج دانش جدید از داده‌های موجود است. هدف

```
mdp
// constants
const int N=4;
const int K;
const int range = 2*(K+1)*N;
const int counter_init = (K+1)*N;
const int left = N;
const int right = 2*(K+1)*N - N;
// shared coin
global counter : [0..range] init counter_init;
module process1
        // program counter
        pc1 : [0..3];
        coin1 : [0..1];
        // flip coin
        [] (pc1=0)  -> 0.5 : (coin1'=0) & (pc1'=1) +
           0.5 : (coin1'=1) & (pc1'=1);
        // write tails -1  (reset coin to add regularity)
        [] (pc1=1) & (coin1=0) & (counter>0) ->
           (counter'=counter-1) & (pc1'=2) & (coin1'=0);
```

شکل (۱) بخشی از یک برنامه PRSIM

اصلی آن فراهم آوردن سازو کار خودکار برای پیش‌بینی نتایج برای داده‌هـای جدید است. در مسائل یادگیری ماشین داده‌ها بر روی مجموعه‌ای از ویژگی‌ها تعریف می‌شوند و هر نمونه داده نشان‌دهنده یکی از مقدارگـذاری‌هـای معتبـر ویژگی‌های فضای داده است. در یادگیری با نظارت[۴] مجموعـه‌ای از داده‌هـای برچسب‌دار برای مرحله آموزش استفاده می‌شوند. دسـته‌بنـدی یکـی از انـواع یادگیری ماشین با نظارت است که در آن هـر نمونـه داده بـه یکـی از رده‌هـا نسبت داده می‌شود. در این مقاله از درخت تصمیم به عنوان یکی از روش‌های دسته‌بندی استاندارد به منظور یادگیری سیاست بهینه استفاده می‌کنیم.

۳- روش ارائه شده

در این بخش روش‌ها و نوآوری‌های اصلی این مقاله ارائه می‌شوند.

۳-۱- یادگیری سیاست بهینه با استفاده از درخت تصمیم

یکی از مشکلات مهم در وارسی مدل احتمالاتی دشـواری محاسـبه سیاسـت بهینه برای یک MDP داده شده است. در صورتی که فضای حالات مدل به اندازه‌ای باشد که بتوان همه اطلاعات آن را در حافظه اصلی قـرار داد امکـان استفاده از روش‌های استاندارد برای حل معادله بلمن و محاسبه سیاست بهینه وجود دارد. اما برای مدل‌های بـزرگ بـه دلیـل محـدودیت حافظـه (حتـی در صورت استفاده از ساختار داده‌ای فشرده) این امکان وجود ندارد. در این صورت می‌توان از یادگیری ماشین برای تخمین سیاسـت بهینـه اسـتفاده کـرد. ایـده استفاده از درخت تصمیم برای نمایش فشـرده سیاسـت بهینـه قـبلا در [۲۰] بررسی شده است. در این مقاله از این ایده برای تخمین سیاست بهینـه بـرای MDPهای بزرگ استفاده می‌کنیم. با در نظر گـرفتن Prg بـه عنـوان یـک برنامـه PRISM شـامل n متغیـر $var_1, var_2, ..., var_n$ و var_i بـه عنـوان یـک متغیـر پـارامتری بـا حـداکثر مقـدار param، از $Prg(param = val)$ برای برنامه‌ای که در آن مقدار حـداکثر متغیر پارامتری برابر val است استفاده می‌کنیم. همچنین برای MDP حاصـل از کامپایل یک برنامه از $MDP(Prg(param = val))$ استفاده می‌کنیم. ایده روش مورد استفاده در شکل ۲ ارائه شده است.

ورودی: برنامه Prg به زبان PRISM و پارامتر param

خروجی: درخت تصمیم DT برای سیاست بهینه مدل‌های ممکن برای Prg

۱- با در نظر گرفتن ۳ یا ۴ مقدار کوچک برای param مدل‌های MDP معادل را محاسبه کن.

۲- برای هر مدل ایجاد شده سیاست بهینه را با توجه به ویژگی داده شده محاسبه کن

۳- با درنظر گرفتن حالت‌های مدل‌ها و برچسب معادل هر حالت (با توجه به سیاست بهینه) درخت تصمیم DT را برای دسته بندی حالت‌ها محاسبه کن.

شکل ۲. ایده مبتنی بر یادگیری درخت تصمیم برای سیاست بهینه

برای مرحله آموزش و ساخت درخت تصمیم با در نظر گرفتن مقادیر کوچک برای param و کامپایل برنامه‌های متناظر تعدادی MDP به دست می‌آید. با حل مساله دسترس‌پذیری احتمالاتی و محاسبه سیاست بهینه برای هر یک از این MDP ها تعدادی حالت و کنش بهینه (به عنوان برچسب) به دست می‌آید. پس از ساخت یک درخت تصمیم با استفاده از داده‌های فراهم شده می‌توان از این درخت برای حل عدم قطعیت در مدل‌های بزرگتر، که از اعمال مقادیر بزرگ برای پارامتر برنامه به دست می‌آیند استفاده کرد. در واقع درخت تصمیم ساخته شده قادر است تقریبی از سیاست بهینه را ارائه دهد که با در نظر گرفتن این سیاست و تعیین یک کنش برای هر حالت می‌توان به یک DTMC رسید. معمولا به DTMC حاصل از اعمال یک سیاست به یک MDP مشخص DTMC خارج قسمت (یا القایی) گفته می‌شود. باید دقت داشت که با داشتن درخت تصمیم (که معمولا حافظه مصرفی پایینی دارد) می‌توان بدون ساختن صریح MDP اصلی و DTMC خارج قسمت از روش وارسی مدل آماری برای تخمین احتمال رسیدن به حالت هدف استفاده کرد [۸،۱۳]. در این صورت درخت تصمیم ساخته شده به عنوان یک راه حل برای تخمین سیاست بهینه و در نتیجه اعمال وارسی مدل آماری به عنوان روش موثر مقابله با انفجار حالات استفاده خواهد شد. برای این هدف می‌توان هر شبیه سازی را با شروع از S_0 و استفاده از درخت تصمیم برای رفتن به حالت بعدی انجام داد.

مثال ۲. نگاشت کنش بهینه به هر یک از حالت‌های MDP معادل برنامه مثال شکل ۱ به ازای K=۴ در شکل (۳) نشان داده شده است. در این شکل ابتدا لیست متغیرهای برنامه مشخص شده است. سپس لیست حالت‌های MDP و مقدارگذاری هر متغیر برای حالت‌ها نشان داده شده است. در مقابل هر حالت شماره مربوط به کنش بهینه در انتهای سمت راست مشخص شده است. درخت تصمیم معادل این مساله دارای ۹ ویژگی خواهد بود و بسته به مقدار متغیرهای مربوط به هر حالت و کنش منتسب به آن به گونه‌ای ساخته خواهد شد که بهترین تخمین را بر روی داده های آموزشی ارائه دهد. باید دقت داشت برای استفاده از درخت تصمیم برای حل عدم قطعیت یک مدل لازم است مجموعه MDP های آموزشی و MDP هدف از یک برنامه PRISM به دست آمده باشند. در نتیجه در صورتی که برنامه داده شده فاقد پارامتر باشد امکان استفاده مستقیم از این روش وجود ندارد. اما در بیشتر مسائل و نمونه‌های موجود این ویژگی برقرار است [۴].

```
(counter,pc1,coin1,pc2,coin2,pc3,coin3,pc4,coin4)
0:(1,2,0,2,0,2,0,2,0)  > 0
1:(1,2,0,2,0,2,0,3,0)  > 0
2:(1,2,0,2,0,3,0,2,0)  > 1
3:(1,2,0,2,0,3,0,3,0)  > 0
4:(1,2,0,3,0,2,0,2,0)  > 2
5:(1,2,0,3,0,2,0,3,0)  > 0
6:(1,2,0,3,0,3,0,2,0)  > 1
7:(1,2,0,3,0,3,0,3,0)  > 1
8:(1,3,0,2,0,2,0,2,0)  > 0
9:(1,3,0,2,0,2,0,3,0)  > 0
10:(1,3,0,2,0,3,0,2,0) > 1
```

شکل (۳). بخشی از فضای حالات یک MDP و کنش‌های بهینه آن

۳-۲- استفاده از ویژگی اضافه در درخت تصمیم

هرچند روش ارائه شده در بخش قبل می‌تواند سیاست بهینه را با دقت خوبی تخمین بزند اما دارای چندین ایراد است. نخست اینکه امکان تمایز بین برخی حالت‌های مدل‌های مختلف وجود ندارد. به عنوان مثال حالت‌های نمایش داده شده در شکل (۲) برای مقادیر K=۳ و K=۴ و K=۵ یکسان هستند در صورتی که ممکن است کنش بهینه برای این مدل‌ها متفاوت باشد. دوم اینکه بخش قابل توجهی از حالت‌های یک مدل بزرگ مشابه هیچ یک از حالت‌های مدل‌های آموزشی نیستند. به عنوان مثال اگر از مقادیر K<=۵ برای ساخت مدل‌های آموزشی برای برنامه شکل (۱) استفاده شود مقدار حد بالای متغیر counter برابر ۴۸ خواهد بود اما اگر از K=۱۰۰ برای ساخت یک مدل بزرگ (مدلی با بیش از یک میلیون حالت) استفاده شود حد بالای counter برابر ۸۰۸ خواهد شد که نشان می‌دهد بخش قابل توجهی از حالت‌ها مشابه هیچ یک از حالت‌های مدل‌های آموزشی نیستند. مشکل سوم روش بخش قبل به ماهیت درخت تصمیم به عنوان یک دسته بند برمیگردد. یکی از مشکلات عمومی این دسته‌بند گرایش به داده‌های آموزشی و عدم توانایی آن در عمومیت بخشیدن به داده‌های غیر مشابه است. این مشکل که به overfitting معروف است مورد مطالعه فراوان قرار گرفته و راه حل‌های گوناگونی برای غلبه بر آن، بسته به نوع دسته‌بند ارائه شده است [۱۹]. یکی از راه حل‌های ارائه شده استفاده از ویژگی یا ویژگی‌های اضافه به مساله است. این ویژگی‌ها می‌توانند با انجام محاسبه بر روی ویژگی‌های جاری تعریف و به مساله اضافه شوند. ایده ما برای غلبه بر این مشکل در نظر گرفتن تفاضل مقدار متغیر پارامتری یک مدل با مقدار حداکثر آن به عنوان ویژگی اضافه مدل است. با افزودن این ویژگی درخت تصمیم بهتر می‌تواند در مورد حالت-هایی که مقدار متغیر پارامتری آن‌ها به مقدار حداکثر نزدیک است تصمیم بگیرد. از نقطه نظر یادگیری ماشین افزودن یک ویژگی به فضای داده‌های مساله تاثیر چندانی بر حافظه و زمان مصرفی فرایند یادگیری ندارد اما می‌تواند به افزایش دقت پیش‌بینی برای داده‌ها جدید منجر شود.

مثال ۳. با در نظر گرفتن counter به عنوان متغیر پارامتری برای برنامه شکل (۱) و range به عنوان حد بالای مقادیر آن مقدار range − counter را می‌توان به عنوان ویژگی جدید برای داده‌های مربوط به مدل‌های مساله consensus در نظر گرفت. در این صورت برای هر حالت مدل‌های مربوطه ۱۰ ویژگی خواهیم داشت.

۳-۳- کاهش فضای حالات MDP ها با استفاده از سیاست تخمین زده شده

یکی از راهکارهای کاهش فضای حالات یک MDP با هدف مقابله با انفجار حالات کنار گذاشتن حالت‌های بی اهمیت یا کم اهمیت از محاسبات است. در روش استاندارد برای کامپایل یک برنامه PRISM و محاسبه MDP معادل آن ابتدا حالت شروع s_0 در نظر گرفته شده و مجموعه حالت‌هایی که بلافاصله بعد از آن قرار می‌گیرند محاسبه می‌شوند. سپس با اعمال جستجوی اول سطح هر یک از حالت‌های به دست آمده در نظر گرفته شده و حالت‌های جدید پس از آن به مجموعه حالت‌های مدل اضافه می‌شوند. از طرف دیگر در صورت داشتن سیاست بهینه، برای هر حالت تنها یکی از کنش‌های آن برای محاسبه حالت‌های بعدی استفاده خواهند شد. در نتیجه ممکن است برخی از حالت‌های MDP ساخته شده در روش استاندارد تحت سیاست بهینه دسترس‌پذیر نباشند و در نتیجه با کنار گذاشتن آن از مدل می‌توان به یک مدل کاهش یافته رسید. ایده این روش در الگوریتم شکل (۴) توضیح داده شده است.

Algorithm 1 Computing Useful and Useless States.

input: an MDP $M = (S, s_0, Act, \delta, G)$ and the optimal policy σ^*
output: The Set $S_{uf} \subset S$ of useful states.
Set of states $S_{uf} := \{s_0\}$;
Queue of states $Q := \emptyset$;
Add(s_0,Q);
while Q is not empty **do**
  remove s from Q;
  for all $s' \in S \backslash S_{uf}$ where $\delta(s, \sigma(s))(s') > 0$ **do**
    Add(s', Q);
    $S_{uf} := S_{uf} \cup s'$;
  end for
end while
return S_{uf};

شکل (۴) الگوریتم محاسبه مدل کاهش یافته با داشتن سیاست بهینه

در این الگوریتم از S_{uf} به منظور نگه داشتن لیست حالت‌های مفید و نیز از Q به عنوان صفی از حالت‌های جدید استفاده می‌شود. با توجه به اینکه محاسبه سیاست بهینه σ^* در عمل ساده نیست از سیاست پیشنهاد شده توسط درخت تصمیم برای محاسبه استفاده می‌شود. الگوریتم در هر تکرار یکی از حالتها را از صف Q خارج و با استفاده از سیاست σ^* داده شده حالت‌های پس از آن را در صورتی که در لیست S_{uf} قرار نداشته باشد به لیست و صف اضافه می‌شوند. محاسبات تا زمانی ادامه می‌یابد که صف Q خارج شود.

۴- نتایج تجربی

برای بررسی نقاط قوت و ضعف روش ارائه شده تعدادی نمونه موردی از مسائل استاندارد PRISM را انتخاب می‌کنیم. این نمونه‌های موردی شامل consensus, zeroconf, firewire, wlan, mer هستند که دارای برنامه‌های PRISM پارامتری هستند. این نمونه‌های در بسیاری از کارهای قبلی مورد استفاده قرار گرفته‌اند [۲۲]. برای هر نمونه مساله ابتدا تعدادی مدل با در نظر گرفتن مقادیر کوچک برای متغیرهای پارامتری ایجاد شده و سپس با اعمال روش تکرار مقدار استاندارد سیاست بهینه محاسبه شده است. سپس مجموعه حالت‌ها و کنش بهینه هر حالت به عنوان داده‌های آموزشی در نظر گرفته درخت تصمیم متناظر ساخته می‌شود. برای این کار از زبان برنامه نویسی پایتون و توزیع Anaconda استفاده شده است. برای ساخت درخت-های تصمیم از کلاس decisiontreeclassifier با پارامترهای پیش‌فرض استفاده شده است. تمامی اجراها بر روی یک کامپیوتر شخصی با ۸ گیگابایت حافظه اصلی و پردازنده core-i7 انجام شده‌اند. خلاصه نتایج مربوط به مرحله آموزش در جدول (۱) ارائه شده است.

جدول (۱) نتایج اجراهای مرحله آموزش

مدل	مقادیر پارامتر	اندازه مدل-ها	زمان وارسی مدل	زمان آموزش
Consensus	K=۴،۵،۶	۱۶۰،۱۲۸	۵،۶	.۱
Zeroconf	K=۲،۳،۴	۵۷۷،۱۲۸	۶،۱	۱۷،۷
Firewire	ddl=۸۰۰،۱۰۰۰،۱۲۰۰	۱،۲۸۵،۶۴۴	۱۰،۳	۱۵
Wlan	n=۲۰،۳۵،۵۰	۶۲۴،۹۲۷	.۶	۱۴،۲
Mer	Ttm=۲۰،۳۵،۶۰	۴،۰۱۹،۱۰۴	۳،۹	۸۴،۳

در این جدول پارامترهای به کار رفته برای ساخت مدل‌های مرحله آموزش و مجموع اندازه مدل‌ها و زمان وارسی مدل برای MDP های ساخته شده و نیز زمان آموزش ارائه شده است. برای بررسی میزان کارایی روش‌های ارائه شده در بخش قبل، تعدادی مدل با در نظر گرفتن مقادیر بزرگ برای پارامترها ساخته شده و زمان اجرای وارسی مدل احتمالاتی برای هر مدل با زمان اجرای مدل کاهش یافته مقایسه شده است. همچنین از روش مبتنی بر برنامه‌ریزی پویای بلادرنگ که در [۸] ارائه شده به عنوان یکی از روش‌های مشابه پیشین نیز استفاده شده است. نتایج حاصل در جدول (۲) ارائه شده‌اند.

جدول (۲) نتایج اجرای روش‌های استاندارد و روش‌های ارائه شده

مدل	مقدار پارامتر	اندازه مدل	VI روی مدل اصلی	VI روی مدل کاهش یافته	RTDP	دقت
Consensus	۱۰	۱۰۴K	۵۹	۱۲	T.O	۹۹،۹
	۵۰	۵۱۴K	۵۱۵۲	۸۶۴	T.O	۹۸
	۱۰۰	۱۰۲۶K	۳۰۹۴۱	۴۹۴۴	T.O	۹۳
Zeroconf	۱۰	۳۰۰۲K	۲۴	۸،۵	۱۳،۳	۹۹،۹
	۱۴	۴۴۲۷K	۴۱	۱۳،۴	۱۴،۱	۹۹،۹
	۱۸	۵۴۷۷K	۵۷	۱۶،۶	۱۷،۷	۹۹،۹
Firewire	۱۵۰۰	۷۳۷K	۱۱	۴،۷	۱۰۲۸۸	۹۹،۹
	۵۰۰۰	۲۸۹۳K	۹۰	۳۹	T.O	۹۹،۹
	۱۰۰۰۰	۵۹۷۳K	۳۸۴	۲۰۰	T.O	۹۹،۹
Wlan	۲۰۰	۱۵۹۴K	۱۱	۴،۸	۷۲۹	۹۹،۹
	۱۰۰۰	۲۸۵۰K	۲۴	۲۹	T.O	۹۹،۹
	۵۰۰۰	۹۲۳۰K	۹۷	۶۴	T.O	۹۹،۹
Mer	۲۰۰	۱۱۸۳K	۱،۲	.۶	T.O	۹۹،۹
	۱۰۰۰	۵۹۰۹K	۶	۲،۸	T.O	۹۹،۹
	۵۰۰۰	۲۹۵۳۷K	۲۷،۸	۱۳	T.O	۹۹،۹

در این جدول اندازه هر مدل با توجه به مقدار پارامتر آن، زمان اجرای روش تکرار مقدار بر روی مدل اصلی و بر روی مدل کاهش یافته و نیز زمان

اجرای روش RTDP ارائه شده‌اند. اجراهایی که بیش از ۱۲۰۰۰ ثانیه به طول انجامیده‌اند به صورت time out گزارش شده‌اند. نتایج تجربی نشان می‌دهد درخت تصمیم ساخته شده در بیشتر موارد قادر است با خطای کمتر از ۰,۱ درصد احتمالات دسترس‌پذیری را درست پیش‌بینی کند. همچنین زمان مصرفی حاصل از اجرای روش‌های تکراری بر روی مدل‌های کاهش یافته کاهش قابل توجهی را نشان می‌دهد.

۵- نتیجه‌گیری

در این مقاله یک روش جدید مبتنی بر یادگیری ماشین برای پیش‌بینی سیاست بهینه برای مدلهای پارامتری ارائه شده است. با استفاده از این روش می‌توان وارسی مدل آماری را بر روی MDP های بزرگ با دقت خوبی انجام داد و همچنین اندازه مدل‌ها را کمتر کرد بدون اینکه دقت محاسبات کاهش یابد. برای کارهای آینده می‌توان از تعمیم روش پیشنهادی برای سایر ویژگی-های مورد نظر در وارسی مدل احتمالاتی استفاده کرد.

٦- مراجع

[١] Baier, C., Katoen, J.P.: *Principles of Model Checking.* MIT Press, Cambridge, ٢٠٠٨

[٢] Baier C., Hermanns H., Katoen JP. *The ١٠,٠٠٠ Facets of MDP Model Checking.* In: Steffen B., Woeginger G. (eds) Computing and Software Science. Lecture Notes in Computer Science, vol ١٠٠٠٠. Springer, Cham, ٢٠١٩. https://doi.org/١٠,١٠٠٧/٩٧٨-٣-٣١٩-٩١٩٠٨-٩_٢١

[٣] Klein, Joachim, Christel Baier, Philipp Chrszon, Marcus Daum, Clemens Dubslaff, Sascha Klüppelholz, Steffen Märcker, and David Müller. "*Advances in probabilistic model checking with PRISM: variable reordering, quantiles and weak deterministic Büchi automata.*" International Journal on Software Tools for Technology Transfer ٢٠, no. ٢: pp. ١٧٩-١٩٤. ٢٠١٨

[٤] Kwiatkowska, M., Norman, G., Parker, D.: *The PRISM benchmark suite.* In: ٩th International Conference on Quantitative Evaluation of SysTems, pp. ٢٠٣-٢٠٤. IEEE CS Press, ٢٠١٢

[٥] Puterman, M.L.: *Markov Decision Processes: Discrete Stochastic Dynamic Programming.* Wiley, Hoboken, ٢٠١٤

[٦] Kwiatkowska, M., Norman, G., Parker, D.: *Symmetry reduction for probabilistic model checking.* In: Ball, T., Jones, R.B. (eds.) CAV ٢٠٠٦. LNCS, vol. ٤١٤٤, pp. ٢٣٤-٢٤٨. Springer, Heidelberg, ٢٠٠٦

[٧] Ciesinski, F., Baier, C., Groesser, M., Klein, J.: *Reduction techniques for model checking Markov decision processes.* In: Fifth International Conference on Quantitative Evaluation of Systems, pp. ٤٥-٥٤. IEEE, ٢٠٠٨

[٨] Br´azdil, T., Chatterjee, K., Chmel´ık, M., Forejt, V., Kˇret´ınsk´y, J., Kwiatkowska, M., Parker, D., and Ujma, M. Verification of Markov decision processes using learning algorithms. In ١٢th International Symposium on Automated Technology for Verification and Analysis. LNCS, vol. ٨٨٣٧, pp. ٩٨-١١٤. Springer, Cham , ٢٠١٤.

[٩] B aier, C., Klein, J., Leuschner, L., Parker, D., Wunderlich, S.: *Ensuring the reliability of your model checker: interval iteration for Markov decision processes.* In: Majumdar, R., Kuncak, V. (eds.) CAV ٢٠١٧. LNCS, vol. ١٠٤٢٦, pp. ١٦٠-١٨٠. Springer, Cham. https://doi.org/١٠,١٠٠٧/٩٧٨-٣-٣١٩-٦٣٣٨٧-٩ ٨CrossRef

[١٠] Maisonneuve, Vivien. "*Automatic heuristic-based generation of MTBDD variable orderings for PRISM models.* Internship report.", ٢٠٠٩.

[١١] Hensel, C., Junges, S., Katoen, JP. et al. *The probabilistic model checker Storm.* Int J Softw Tools Technol Transfer, ٢٠٢١. https://doi.org/١٠,١٠٠٧/s١٠٠٠٩-٠٢١-٠٠٦٣٣-z

[١٢] Mohagheghi, M., Karimpour, J., Isazadeh, A. *Prioritizing methods to accelerate probabilistic model checking of discrete-time Markov models.* The Computer Journal, Vol. ٦٣ No.١ , pp. ١٠٥-١٢٢. ٢٠٢٠

[١٣] Rataj, A., Wozna-Szczesniak, B. *Extrapolation of an Optimal Policy using Statistical Probabilistic Model Checking.* Fundamenta Informaticae, Vol. ١٥٧, No.٤, pp. ٤٤٣-٤٦١. ٢٠١٨

[١٤] Gros T.P., Hermanns H., Hoffmann J., Klauck M., Steinmetz M. *Deep Statistical Model Checking.* In: Gotsman A., Sokolova A. (eds) Formal Techniques for Distributed Objects, Components, and Systems. FORTE ٢٠٢٠. Lecture Notes in Computer Science, vol ١٢١٣٦. Springer, Cham, ٢٠٢٠

[١٥] De Argenio, P., Legay, A., Sedwards, S., Traonouez, L. M. (٢٠١٥). *Smart sampling for lightweight verification of Markov decision processes.* International Journal on Software Tools for Technology Transfer, ١٧(٤), ٤٦٩-٤٨٤.

[١٦] Hartmanns, A., Hermanns, H. *Explicit model checking of very large MDP using partitioning and secondary storage.* In International Symposium on Automated Technology for Verification and Analysis (pp. ١٣١-١٤٧). Springer, Cham. ٢٠١٥

[١٧] Groote, J. F., Rivera Verduzco, J., De Vink, E. P. *An efficient algorithm to determine probabilistic bisimulation.* Algorithms, ١١(٩), ١٣١, ٢٠١٨

[١٨] Daca, P., Henzinger, T. A., Kretnsk, J., Petrov, T. *Faster statistical model checking for unbounded temporal properties.* ACM Transactions on Computational Logic (TOCL), Vol. ١٨ No. ٢, pp. ١-٢٥. ٢٠١٧

[١٩] Mohri, Mehryar, Afshin Rostamizadeh, and Ameet Talwalkar. *Foundations of machine learning.* MIT press, ٢٠١٨.

[٢٠] Ashok, P., Jackermeier, M., Kˇret´ınsk´y, J., Weinhuber, C., Weininger, M., Yadav, M. dtControl ٢,٠: *Explainable Strategy Representation via Decision Tree Learning Steered by Experts.* In International Conference on Tools and Algorithms for the Construction and Analysis of Systems, pp. ٣٢٦-٣٤٥. Springer, Cham, ٢٠٢١

[٢١] Bacci, Edoardo, and David Parker. "*Verified probabilistic policies for deep reinforcement learning.*" In *NASA Formal Methods Symposium*, pp. ۱۹۳-۲۱۲. Springer, Cham, ۲۰۲۲.

[٢٢] Budde, C. E., Hartmanns, A. Klauck, M. Křetínský, J. Parker, D. Quatmann, T. Turrini, A and Zhang, Z. *On correctness, precision, and performance in quantitative verification*. In *International Symposium on Leveraging Applications of Formal Methods*, pp. ۲۱۶-۲۴۱. Springer, Cham, ۲۰۲۱.

---

١ Discount factor
٢ Value iteration
٣ Guarded commands
٤ Supervised learning

چهارمین کنفرانس ملی انفورماتیک ایران
پژوهشگاه دانش‌های بنیادی، پردیس فرمانیه، تهران
۱۴ و ۱۵ دی ماه ۱۴۰۱

ارائه‌ی یک روش کارا برای شناسایی گره‌های تأثیرگذار در شبکه‌های اجتماعی بزرگ

نوشین شریفی، مهدی باطنی

کارشناس ارشد مهندسی نرم‌افزار، گروه کامپیوتر، دانشگاه شیخ‌بهائی، اصفهان
nooshin.sharifi@shbu.ac.ir
گروه علوم و مهندسی کامپیوتر، پردیس خوانسار، دانشگاه اصفهان
m.bateni@khc.ui.ac.ir

چکیده

به دلیل اهمیت مبحث انتشار و استحکام معماری در شبکه، روش‌های زیادی برای کشف گره‌ی تأثیرگذار ارائه شده است. با این وجود هر کدام از معیارهای ارائه شده دارای محدودیت‌های خاص خود هستند. الگوریتم GLS یکی از متدهای کشف و رتبه‌بندی گره‌های تأثیرگذار است. GLS با ادغام دو بعد اطلاعات سراسری و محلی گره‌ها، از دقت بیشتری نسبت به متدهای کلاسیک و نوین مشابه برخوردار است اما، به اندازه کافی سریع و کارآمد نیست.

در این مقاله الگوریتم NGLS برای بهبود کارایی و مقیاس پذیری GLS توسعه یافته است. این تغییر با در نظرگرفتن همسایه‌های مشترک هر گره با همسایه‌هایش تنها تا شعاع دو در بخش محاسبه‌ی تأثیر سراسری انجام شده است. *زمان اجرای NGLS به طور متوسط در شبکه‌های دنیای واقعی ۸۵ درصد و در شبکه‌های شبیه سازی شده ۹۷ درصد کاهش یافته است، در حالی که NGLS با GLS از نظر دقت یکسان هستند. در نتیجه، از NGLS درشبکه‌های دنیای واقعی بزرگتر می‌توان استفاده کرد.*

کلمات کلیدی

گره‌های تأثیرگذار، اطلاعات سراسری و محلی، شبکه‌های بزرگ، معیارهای مرکزیت.

۱- مقدمه

با ظهور شبکه های اجتماعی آنلاین (OSN) و نیاز به تجزیه و تحلیل شبکه‌های اجتماعی (SNA) ، دانش در مورد نقش گره‌های شبکه برای ایجاد روش‌هایی جهت جلوگیری یا تسریع انتشار اهمیت بسیاری یافته است[۱]. طبق بسیاری از مطالعات، این روند به شدت با توپولوژی شبکه‌ها ارتباط دارد [۲]. گره‌های تأثیرگذار اطلاعات سرأسری یا موقعیتی بیشتری را در مقایسه با سایر گره‌ها در اختیار دارند. [۳] با توجه به اهمیت کشف گره‌های تأثیرگذار تاکنون متدهای بسیاری با دیدگاه‌های مختلف ارائه شده است که در ادامه به طور خلاصه به برخی از آن‌ها اشاره می‌نماییم.

۲- بیان مسئله

روش‌های ارائه شده برای کشف گره‌های تاثیرگذار اغلب درای محدودیت هایی مانند عدم دقت کافی، پیچیدگی زمانی بالا یا محدودیت‌هایی نسبت به توپولوژی هستند. بنابراین، کشف گره‌های تاثیرگذار همچنان یک موضوع باز است. با توجه به این موضوع، الگوریتم GLS در سال ۲۰۱۹ معرفی شد. GLS از ترکیبی از تأثیر جهانی و محلی برای محاسبه تأثیر نهایی گره‌ها استفاده می کند. این الگوریتم نیازی به دانش قبلی و تنظیم پارامتر ندارد و قادر است به طور خودکار گره‌های تأثیرگذار را شناسایی کند. با توجه به گزارش روش GLS دقیق تر از بسیاری از روش‌های کلاسیک و نوین است اما پیچیدگی آن $O(n^2)$ است. بنابراین با توجه به رشد دائمی شبکه‌های اجتماعی، استفاده از این الگوریتم در شبکه‌های بزرگ پرهزینه و در برخی موارد غیرممکن است. هدف ما در این پژوهش کاهش پیچیدگی زمانی این الگوریتم است به گونه‌ای که بتوان از مزایای آن در شبکه‌های دنیای واقعی بزرگ نیز استفاده نمود.

۳- پژوهش‌های پیشین

روش‌های مبتنی بر آنتروپی، اجماع و تصمیم‌گیری چند معیاره: در بخشی از پژوهش‌ها، از مفهوم آنتروپی استفاده شده است، در [۴] ، متد (CbM) آنتروپی یک پیاده‌روی تصادفی از یک گره به هر جامعه را در نظر می‌گیرد. نتایج حاکی از عملکرد برتر آن نسبت به متدهای کلاسیک است. برای حل کمبود معیار واحد در ارزیابی اهمیّت گره، چندین روش رتبه‌بندی براساس تصمیم‌گیری چند معیاره (MCDM) پیشنهاد شده است. در [۵] از VIKOR برای ترکیب مرکزیت درجه، مرکزیت نزدیکی و مرکزیت میانی استفاده شده و توزین آنتروپی (EW) برای محاسبه‌ی وزن هر معیار بکار رفته است. VIKOR باوجود دقت بالا گره‌هایی با رتبه‌ی یکسان دارد. [۶] یکی دیگر از تکنیک‌های MCDM، AHP است که با تقلیل تصمیمات پیچیده به یک سری از مقایسه‌های دو به دو، و سپس ترکیب آن‌ها عمل می‌کند و نسبت به متدهای کلاسیک و W-TOPSIS عملکرد بهتری داشته است.

روش‌های مبتنی بر مرکزیت: مهم‌ترین معیارها در زمینه‌ی شناسایی تأثیرگذارترین گره، مرکزیت‌ها هستند. با این حال روش‌های ساده‌ی مرکزیت از دقّت کافی برخوردار نمی‌باشند، و روش‌های دقیق محاسباتی بالایی را به

همراه دارند [7]. در مرکزیت‌های طیفی نیز، مانند مرکزیت بردار ویژه(EC) ، رتبه‌بندی صفحه (PR) و ... محدودیت‌های نسبت به توپولوژی شبکه وجود دارد[8] [۹]. در برخی از پژوهش‌ها از مرکزیت‌های کلاسیک به طور مستقیم برای کشف گره‌های تأثیر گذار استفاده شده است. به عنوان مثال در [۱۰] با استفاده از مرکزیت میانی، مرکزیت نزدیکی، مرکزیت درجه و مرکزیت بردارویژه، گره‌های یک شبکه تروریستی در وب تاریک رتبه‌بندی شده است. همچنین در [۱۱] نیز پنج مرکزیت مذکور برای فعالیت‌های بازاریابی در زمینه‌ی زیبایی و سلامت بکاررفته‌اند. در [۱۲] نیز معیارهای مذکور از طریق وزندهی، جهت کشف گره‌های تأثیرگذار فریبکار استفاده شده‌اند. دستیابی به گره‌های تأثیرگذار در OppNet به دلیل پویایی توپولوژی و عدم دید سراسری نسبت به شبکه، آسان نیست. بنابراین [۱۳] از روش‌های آماری توقف بهینه برای انتخاب بین گره‌های مختلف کاندیدای تحویل استفاده نموده است.

به علت محدودیت‌های روش‌های کلاسیک برخی از پژوهشگران از ایده‌ی اولیه‌ی آن‌ها استفائه نموده‌اند. در [۱۴] به منظور استفاده از مفهوم مرکزیت نزدیکی، به جای محاسبه‌ی نزدیکی یک گره با کل گره‌های شبکه، کوتاه‌ترین مسیرها فقط تا گره‌های مرکزی جوامع اندازه‌گیری شود. بنابراین پیچیدگی زمانی مرکزیت نزدیکی از $O(n^3)$ به $O(n^2)$ کاهش یافته است.

در برخی از پژوهش‌ها محققین به ارائه‌ی روش‌های فرا ابتکاری پرداخته‌اند. به عنوان مثال: H-index میزان تأثیرگذاری گره‌ها را بر اساس توانایی انتشار آن‌ها در نظر می‌گیرد اما بخشی از اطلاعات مربوط به گره‌های همسایه در این روش نادیده گرفته می‌شود، که از دقت آن کاسته است[۱۵]. در [۱۶] متدی ارائه شده است که براساس همه‌ی اطلاعات همسایگان، قابلیت گسترش گره را تعیین می‌نماید و توانایی بالایی در تمایز بین گره‌ها و رتبه‌بندی آن‌ها، در مقایسه با سایر روش‌ها دارد. از دیگر متدهای فرابتکاری مشابه، k-shell است. این متد براساس درجه‌ی گره‌ها آن‌ها را در لایه‌های مختلف قرار داده و بر این اساس به آن‌ها یک شاخص نسبت می‌دهد. ایراد اصلی این متد قرار دادن تعداد زیادی از گره‌ها را در یک لایه است، [۱۷] به و سیله‌ی یک پارامتر آزاد ضعف ذکر شده را رفع کرده است. تنظیم این پارامتر نیازمند آزمون و خطا می‌باشد و بار پردازشی زیادی را به همراه دارد، بنابراین، در [۳] یک مدل ریاضی بر اساس شعاع همسایگی ارائه شده است.

در گروهی از پژوهش‌ها ترکیب معیارهای مرکزیت مورد برر سی قرار گرفته‌اند. [۱۸] با ایده از فرمول گرانش نیوتون، مرکزیت درجه‌ی گره‌ها را به عنوان جرم، و برای تأثیر خود گره شاخص k-shell آن را به کار برده است.

در برخی دیگر از پژوهش‌ها از معیارهای مرکزیت با توجه به سراسری بودن یا محلی بودن معیارها استفاده شده است. در [7] میزان مشارکت همسایه‌ها تا شعاع ۲ در اثرگذاری گره مورد نظر اندازه‌گیری شده است. اما در در برخی از شبکه‌ها به علت عدم توجه به ویژگی‌های سراسری، ضعیف عمل نموده است. در مقابل در [۱۹] مدل کمیGIN از بعد سراسری به کشف گره‌های تأثیرگذار پرداخته است، که در شبکه‌های غیرمتصل بسیار توانمند است.

در نهایت در [۸] الگوریتمی ارائه شده است که هم بعد سراسری هم بعد محلی گره را مدنظر قرار داده است، و هر دوی آن‌ها را با هم ترکیب کرده است. نمودار بلوکی این الگوریتم در شکل (1) آورده شده است.

قبل از توضیح روش پیشنهادی، به برخی تعاریف اساسی مورد استفاده در بخش‌های بعدی پرداخته شده است. با فرض اینکه G=(V, E) یک شبکه‌ی بدون وزن و جهت، دارای |V| گره و |E| لبه باشد:

تعریف ۱ (مرکزیت درجه): مرکزیت درجه ساده‌ترین شاخص برای اندازه‌گیری تأثیر یک گره در شبکه است، که با $DC(v_i)$ معرفی شده است و که به شرح زیر تعریف می شود:

$$DC(v_i) = \frac{d(v_i)}{n - 1} \qquad (۱)$$

که v_i یک گره در شبکه G است، n تعداد کل گره‌ها است و $d(v_i)$ نشان دهنده‌ی درجه‌ی گره v_i است.

تعریف ۲ (درجه متوسط): $averD(v_i)$ متوسط درجه تمام گره‌های همسایه گره v_i است:

$$averD(v_j) = \frac{\sum_{vj=\eta 1(vi)} d(v_j)}{d(v_i)} \qquad (۲)$$

که $d(v_i)$ و $d(v_j)$ به ترتیب درجه گره v_i و v_j را نشان می‌دهند و $\eta 1(v_i)$ مجموعه‌ی نزدیکترین همسایه‌های v_i است.

تعریف ۳ (همسایگان مشترک): گره های v_i و v_j نشانگر دو گره متفاوت در شبکه G هستند. گره‌های همسایه که بین این دو گره مشترک هستند توسط $Com(v_i, v_j)$، نشان داده می‌شوند، که به شرح زیر تعریف می شود:

$$Com(v_i, v_j) = Nei(v_i) \cap Nei(v_j) \qquad (۳)$$

$Nei(v_i)$ و $Nei(v_j)$ نشانگر نزدیکترین همسایه‌های گره v_i و v_j هستند.

تعریف ۴ (مجموع توان): برای اندازه‌گیری توانایی به اشتراک گذاشتن اطلاعات توسط گره v_i، مجموع ارتباطات آن در سه فاصله شعاع ۱ و ۲ و فراتر از آن، با عنوان $SIGMA$ محاسبه می‌گردد که به صورت زیر تعریف شده است:

$$SIGMA = \sum_{vj=\eta 1(vi), \eta 2(vi), \eta x>2(vi)} pow(A, \#Com(v_i, v_j)) \qquad (۴)$$

در رابطه‌ی بالا $vj=\eta_1(vi), \eta_2(vi), \eta_{x>2}(v_i)$ نشانگر همسایه‌های گره v_i در شعاع یک و دو و بیش از آن است. $\#Com(v_i, v_j)$ تعداد همسایه‌های مشترک دو گره v_i و v_j را مشخص می‌کند و A یک ثابت است که برای سهولت محاسبات ۱/۱ در نظر گرفته شده است.

تعریف ۵ (تأثیر سراسری): توانایی گره v_i برای به اشتراک گذاشتن منابع با گره‌های دیگر یکی از مؤلفه‌های محاسبه‌ی تأثیر گره است. تأثیر سراسری گره v_i با $glbInflu(v_i)$ مشخص می‌شود و به صورت زیر تعریف شده است:

$$glbInflu(v_i) = d(v_i) * SIGMA \qquad (۵)$$

تعریف ۶ (احتمال مشارکت): معکوس متوسط درجه به عنوان احتمال مشارکت گره‌های همسایه در تأثیر گره v_i در نظر گرفته شده است و با $p(v_i)$ مشخص می‌شود.

$$p(v_i) = \frac{1}{averD(v_i)} \qquad (۶)$$

تعریف ۷ (تأثیر محلی): تأثیر محلی به تأثیر گره v_i در ساختار محلی (تا شعاع یک) اشاره دارد و با $locInf(v_i)$ مشخص می‌شود، که به شرح زیر تعریف می‌شود:

$$locInflu(v_i)= \sum_{v_j=\eta1(v_i)} DC(v_j)\, p(v_i) \qquad (۷)$$

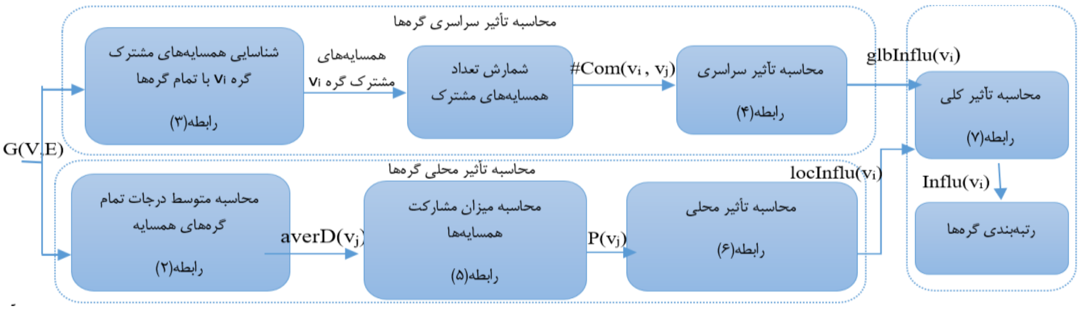
در اینجا مرکزیت درجه‌ی هر گره‌ی همسایه به عنوان نفوذ آن، و میزان مشارکت هر گره در گره v_i برابر با احتمال p در نظر گرفته شده است.

تعریف 8 (تأثیر نهایی): تأثیر گره v_i نه تنها به نفوذ خود گره مربوط است، بلکه به محیط اطراف آن نیز بستگی دارد. بنابراین از ترکیب این تأثیر سراسری و محلی به عنوان شاخصی برای اندازه‌گیری تأثیر گره v_i استفاده می‌شود که با $Influ(v_i)$ مشخص می‌شود.

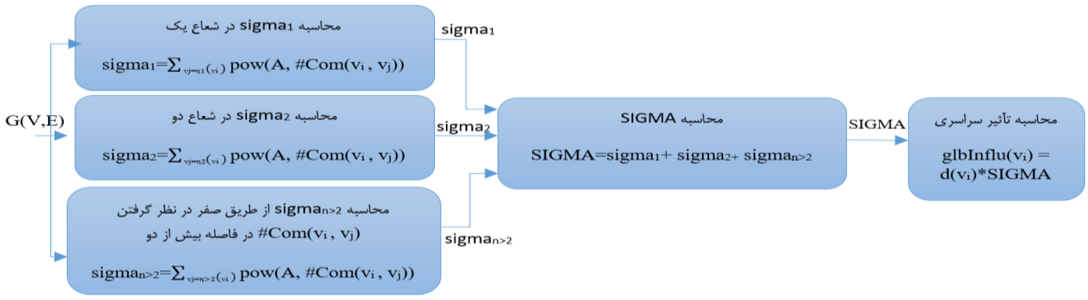
$$Influ(v_i) =glbInflu(v_i)\, locInflu(v_i) \qquad (۸)$$

۴- روش پیشنهادی

چنانچه اشاره شد GLS به دلیل پیچیدگی زمانی بالا در شبکه‌های بزرگ قابل استفاده نیست. در روش پیشنهادی با در نظرگرفتن این فرض که هر گره فقط با همسایگان تا شعاع دو قادر است همسایه‌ی مشترک داشته باشد، و تعداد همسایگان مشترک آن با سایر گره‌ها برابر صفر است، می‌توان محدودیت زمانی الگوریتم GLS را رفع نمود. علت این فرض آن است که احتمال داشتن همسایه‌ی مشترک بین هر گره با نزدیک‌ترین همسایگان آن وجود دارد، در حالی که با همسایگانش تا شعاع دو، قطعاً دارای حداقل یک همسایه‌ی مشترک است؛ اما با همسایگان در مسیرهای طولانی‌تر امکان وجود هیچ همسایه‌ی مشترکی نیست. بنا بر این فرض، متد GLS که دارای دو حلقه‌ی تو در تو به سایز n در محاسبه‌ی تأثیر سراسری است، به سه حلقه‌ی تو در تو با اندازه‌ی m,m, n تبدیل شده است که پیچیدگی زمانی الگوریتم را به $O(nm^2)$ کاهش می‌دهد. نمودار بلوکی بخش محاسبه‌ی تأثیر سراسری در NGLS در شکل ۲ نمایش داده شده است.



شکل (۱) : نمودار بلوکی الگوریتم GLS



شکل (۲) : نمودار بلوکی تأثیر سراسری الگوریتمNGLS

به طور کلی الگوریتم **NGLS** شامل بخش‌های زیر است:

محاسبه‌ی تأثیر سراسری: جریان اطلاعات بین گره‌های موجود در شبکه تأثیر گره‌ها را نشان می‌دهد. بنابراین، باید تأثیر قابلیت اشتراک اطلاعات گره با گره‌های دیگر در کل شبکه لحاظ شود. محاسبه‌ی تأثیر سراسری گره در روش پیشنهادی بدین شرح است: ۱- شناسایی تعداد همسایه‌های مشترک در شعاع یک و محاسبه‌ی $sigma_1$ بر اساس رابطه‌ی ۸.

$$sigma_i=\sum_{v_j=\eta} pow(A, \#Com(v_i , v_j)) \qquad (۸)$$

۲- شناسایی تعداد همسایه‌های مشترک در شعاع دو و محاسبه‌ی $sigma_2$ بر اساس رابطه‌ی ۸. ۳- صفر در نظر گرفتن تعداد همسایه‌های مشترک با سایر گره‌ها(در شعاع بیش از دو) و محاسبه‌ی $sigma_3$ بر اساس رابطه‌ی ۸. ۵- محاسبه مجموع $sigma_i$ها ۴- محاسبه‌ی تأثیر سراسری گره طبق رابطه‌ی ۹.

$$glbInflu(v_i)= d(v_i) *SIGMA \qquad (۹)$$

محاسبه‌ی تأثیر محلی: محاسبه‌ی تأثیر محلی شامل مراحل زیر است:۱- محاسبه میانگین درجات تمام گره‌های همسایه طبق رابطه‌ی ۲. ۲- محاسبه‌ی میزان مشارکت گره‌های همسایه در تأثیر گره توسط رابطه‌ی ۵. ۳- محاسبه‌ی مقدار نهایی تأثیرمحلی بر اساس رابطه‌ی ۶.

رتبه‌بندی گره‌ها: ۱-در این مرحله تأثیر کلی هر گره که ترکیبی از تأثیر سراسری و محلی گره است براساس رابطه‌ی ۷ محاسبه می‌گردد. ۲- گره‌های موجود در شبکه بر اساس تأثیر بدست آمده برای آن‌ها مرتب‌سازی می‌شوند.

۵- ارزیابی نتایج

ارزیابی NGLS در این مقاله در دو بخش ارزیابی دقت و زمان اجرا انجام شده است. این الگوریتم‌ها با زبان پایتون نسخه‌ی3.8.5 با استفاده از Visual StudioCode روی سیستم عامل windows10 پیاده‌سازی شده‌اند، همچنین سیستم مورد استفاده برای ارزیابی دارای 8 گیگابایت رم و پردازنده‌ی اینتل Core i7 6700HQ است.

۱–۵– دقت الگوریتم

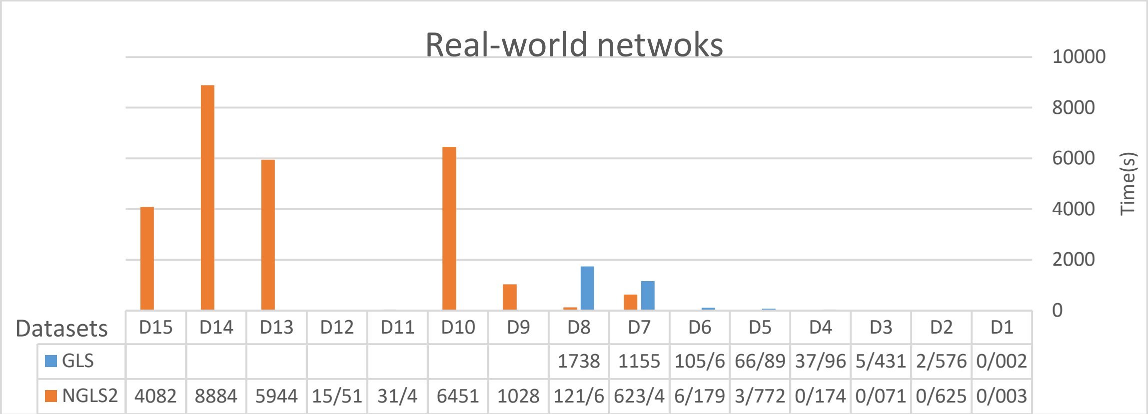
تغییر الگوریتم GLS جهت رفع محدودیت زمانی آن، هیچ تغییری در خروجی آن ایجاد نکرده، و دقت بالای الگوریتم حفظ شده است. این مسئله از طریق دریافت کندال τ برابر با یک، از خروجیNGLS وGLS روی داده‌ی دنیای واقعی و شبکه‌های شبیه سازی شده با پارامترهای متفاوت اثبات شده اســت. همچنین کندال τ روی خروجی هر دو الگوریتم و مدل انتشــار SIR با $t=500$ و $\lambda=0.1$ محا سبه شده ا ست. نتایج آزمایشات در جدول ۱ نشان داده شده است.

جدول (۱) : نتایج آزمایشات سنجش دقت NGLS

τ (SIR ,GLS)	τ (SIR ,NGLS)	τ (NGLS, GLS)	یال‌ها	گره‌ها	شبکه‌ها
۰/۸۲۱۷۴	۰/۸۲۱۷۴۶	۱	۷۸	۳۴	karate
۰/۸۵۱۳۳	۰/۸۵۱۳۳	۱	۵۴۵۱	۱۱۳۳	EMAIL
۰/۸۴۳۱۳	۰/۸۴۳۱۳	۱	۱۷۴۰	۱۷۸۲	Powergrid
۰/۷۹۳۲۲	۰/۷۹۳۲۲	۱	۳۷	۸۰	er-80-0.3
۰/۹۶۵۳۲	۰/۹۶۵۳۲	۱	۱۰۰	۱۰۰	wat-100-3-0.2

۲–۵– زمان اجرای الگوریتم

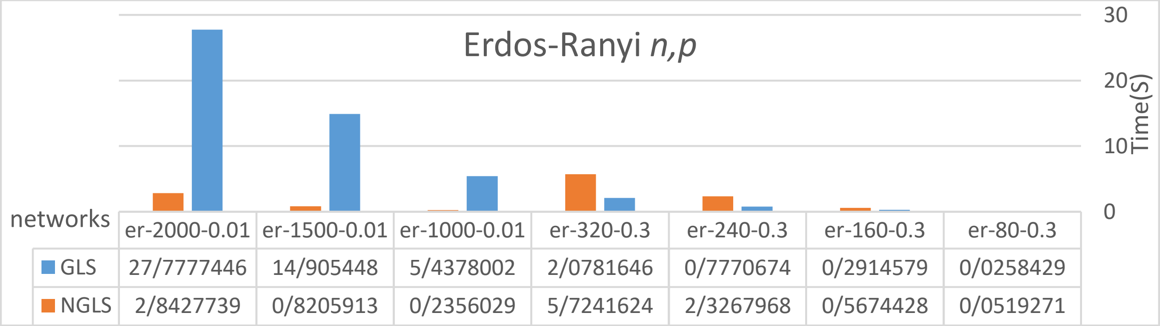
برای مقایسه‌ی زمان اجرای الگوریتم‌های GLS وNGLS در نمودار شکل ۳ زمان اجرای الگوریتم‌ها بر روی شبکه‌های دنیای واقعی بر حسب ثانیه، نمایش داده شده است. در این نمودار محور عمودی نشانگر زمان بر حسب ثانیه، و محور افقی نشانگر تعداد گره‌های شبکه‌ها است. طبق نتایج زمان اجرای الگوریتم GLS به صورت نمایی با افزایش اندازه‌ی شبکه در حال افزایش است. به همین علت با ویژگی‌های سیستم مذکور، زمان اجرای GLS در شکل ۳ تنها روی شبکه‌هایی تا اندازه ۳۲۴۳۰ قابل اندازه‌گیری است. NGLS روی شبکه‌های بزرگ‌تر تا ۳ میلیون گره نیز اعمال شده است. طبق نتایج زمان اجرای NGLS روی هشت مجموعه داده‌ی اولیه نسبت به الگوریتم GLS حداقل ۵۰% کاهش یافته است. زمان اجرای NGLS دارای الگوی رشد ثابت نیست، علت این رخداد این است که پیچیدگی زمانی تنها تابعی از اندازه‌ی شبکه نیست؛ پیچیدگی زمانی الگوریتم NGLS از مرتبه‌ی $O(nm^2)$ است که m متوسط درجات شبکه است. بنابراین این مسئله به ویژگی‌های توپولوژیکی شبکه نظیر متوسط درجات، مدل توزیع درجات و تعداد مثلثات در شبکه وابسته است.



شکل (۳) : نمودار مقایسه‌ی زمان اجرای الگوریتم GLS وNGLS روی شبکه‌های دنیای واقعی

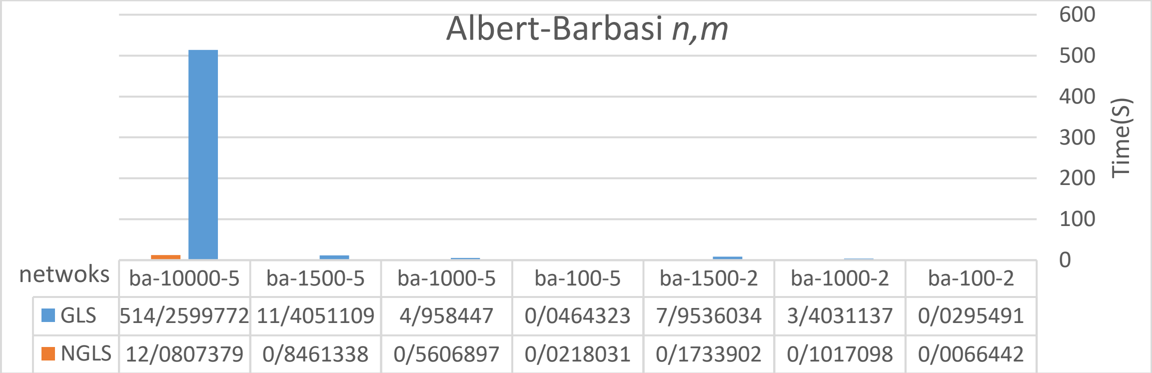
جهت بررسی ویژگی‌های شاخص شبکه‌های شبیه‌سازی شده در زمان اجرای الگوریتم‌ها، زمان اجرای دو الگوریتم بر اساس سه مدل شناخته شده‌ی تصادفی، جهان کوچک و مستقل از مقیاس در شکل‌های (۴)و (۵) و (۶) نیز بررسی شده است. در تمام نمودارهای مذکور محور عمودی نشانگر زمان بر حسب ثانیه و محور افقی نشانگر شبکه‌های شبیه‌سازی شده است.

شبکه‌ها ت صادفی بر ا ساس مدل Erdös-ranyi : طبق نتایج برای p کوچک‌تر از ۰،۰۵ زمان الگوریتم NGLS کمتر از GLS اســت و در مقابل رشــد نمایی آن رشــد خطی دارد. اما با نزدیک شــدن p به یک متوســط طول مســیر کاهش می‌یابد و تجمع گره‌ها در یک خوشـــه‌ی بزرگ زیاد می‌گردد. همچنین توزیع درجه‌ی این شـبکه‌ها براسـاس توزیع پوآسـون اسـت. بنابراین ارتباطات گره‌ها در شعاع یک و دو زیاد شده بار محا سباتی الگوریتم NGLS را زیاد نموده و باعث می‌شـــود در این نمونه‌ی خاص زمان اجرای NGLS بیشتر از قبل شود.



شکل (۴) : نمودار مقایسه‌ی زمان اجرای الگوریتم GLS وNGLS روی شبکه‌های Erdos-Ranyi

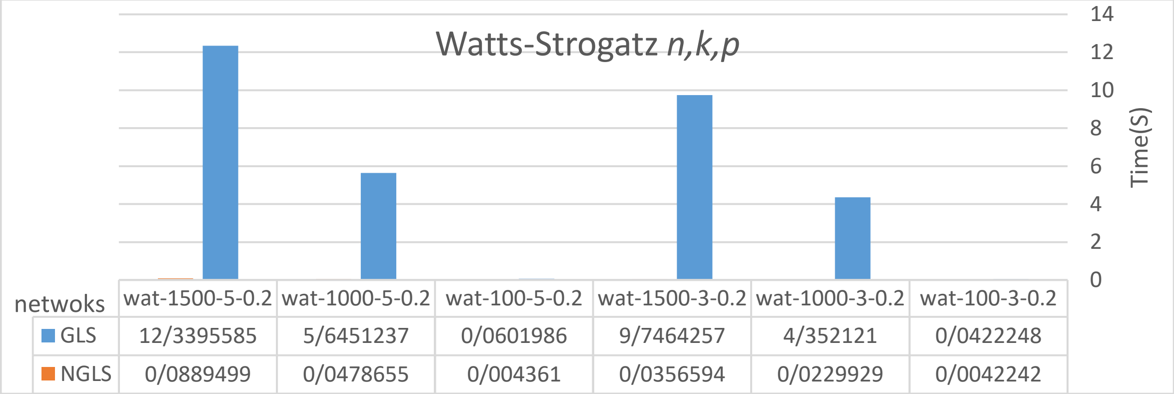
شبکه‌های مستقل از مقیاس بر اساس مدل Albert-Barbasi: آزمایشات در این شبکه‌ها با اندازه‌های متفاوت و دو متوسط درجه‌ی ۲ و ۵ انجام شده است. همانطور که از شکل ۵ مشخص است با افزایش مقدار m به عنوان متوسط درجات، این زمان در هر دو الگوریتم افزایش یافته است. اما همچنان زمان الگوریتم NGLS نسبت به GLS بسیار ناچیز است و حداقل ۸۰ درصد کاهش داشته است. توزیع درجه‌ی توان–قانون و کوچک بودن متوسط طول مسیر از ویژگی‌های مشترک شبکه‌های دنیای واقعی و شبکه‌های Albert-Barbasi است که در بخش‌های قبلی این کاهش زمان در شبکه‌های دنیای واقعی اثبات شده است. علت این کاهش زمان ساختار این شبکه‌ها است که تعداد گره با روابط بالا در آن ها کم است و رفع کردن شمارش تعداد همسایه‌های مشترک تمام گره‌ها با یکدیگر منجر به کاهش بسیار زیاده زمان اجرا شده است.



netwoks	ba-10000-5	ba-1500-5	ba-1000-5	ba-100-5	ba-1500-2	ba-1000-2	ba-100-2
GLS	514/2599772	11/4051109	4/958447	0/0464323	7/9536034	3/4031137	0/0295491
NGLS	12/0807379	0/8461338	0/5606897	0/0218031	0/1733902	0/1017098	0/0066442

شکل (۵) : نمودار مقایسه‌ی زمان اجرای الگوریتم GLS و NGLS روی شبکه‌های Albert-Barbasi

شبکه‌های جهان کوچک بر اساس مدل Watts-Strogatz: در این ارزیابی زمان اجرا روی شبکه‌ها با اندازه‌های یکسان و متوسط درجه‌ی۳ برای p=0.8 و p=0.2 اندازه‌گیری شده است. این شبکه‌ها دارای ضریب خوشگی بالا و متوسط طول مسیر کوتاه است همچنین توزیع درجه‌ی آن‌ها از نوع پوآسون است. متوسط طول مسیر و ضریب خوشگی به مقدار p وابسته است و هرچقدر به یک نزدیکتر باشد به شبکه‌های Erdös-renyi شباهت می‌یابد. طبق نمودار شکل (۶) زمان اجرایNGLS نسبت به نسخه‌ی GLS بسیار ناچیز است.



netwoks	wat-1500-5-0.2	wat-1000-5-0.2	wat-100-5-0.2	wat-1500-3-0.2	wat-1000-3-0.2	wat-100-3-0.2
GLS	12/3395585	5/6451237	0/0601986	9/7464257	4/352121	0/0422248
NGLS	0/0889499	0/0478655	0/004361	0/0356594	0/0229929	0/0042242

شکل (۶) : نمودار مقایسه‌ی زمان اجرای الگوریتم GLS و NGLS روی شبکه‌های Wattst-Strogatzi

۶– جمع‌بندی

در این پژوهش ساختار الگوریتم GLS به گونه‌ای تغییر کرده است که ضمن حفظ دقت، ضعف پیچیدگی زمانی آن رفع شده و امکان اعمال آن برای شبکه‌های بزرگ فراهم شده است. زمان اجرای NGLS در شبکه‌های دنیای واقعی حداقل ۵۰٪ نسبت به GLS کاهش داشته است و در برخی از شبکه‌های بزرگ زمان اجرای الگوریتم بسیار ناچیز است. علت این رویداد این است که زمان اجرا تنها تابع اندازه‌ی شبکه نیست و عوامل توپولوژیک از جمله مدل توزیع درجات، متوسط درجات، در آن دخیل هستند. طبق نتایج آزمایشات روی شبکه‌های شبیه‌سازی شده با پارامترهای متفاوت زمان NGLS بیشترین کاهش را ابتدا روی شبکه‌های جهان کوچک و بعد از آن شبکه‌های مستقل از مقیاس و تصادفی داشته است.

NGLS الگوریتمی با دقت بالا و بدون محدودیت زمانی است که در انواع شبکه‌ها در زمینه‌های مختلف اجتماعی، سیا سی، پز شکی و... قابل ا ستفاده است. ایجاد امکان اعمال NGLS روی شبکه‌های وزن‌دار جزء برنامه‌های آینده‌ی این پژوهش است.

مراجع

[1] S. Rosenthal, "Detecting influencers in social media discussions," 21, XRDS: Crossroads, The ACM Magazine for Students, 1, 2014.

[2] S. Gao, J. Ma, Z. Chen, G. Wang, and C. Xing, "Ranking the spreading ability of nodes in complex networks based on local structure," *Physica A: Statistical Mechanics and its Applications*, vol. 403, pp. 130-147, 2014.

[3] G. Maji, A. Namtirtha, A. Dutta, and M. C. Malta, "Influential spreaders identification in complex networks with improved k-shell hybrid method," *Expert Systems with Applications*, vol. 144, p. 113092, 2020.

[4] M. M. Tulu, R. Hou, and T. Younas, "Identifying influential nodes based on community structure to speed

up the dissemination of information in complex network," *IEEE Access*, vol. 6, pp. 7390-7401, 2018.

[5] Y. Yang, L. Yu, X. Wang, Z. Zhou, Y. Chen, and T. Kou, "A novel method to evaluate node importance in complex networks," *Physica A: Statistical Mechanics and its Applications*, vol. 526, p. 121118, 2019.

[6] T. Bian, J. Hu, and Y. Deng, "Identifying influential nodes in complex networks based on AHP," *Physica A: Statistical Mechanics and its Applications,* vol. 479, pp. 422-436, 2017.

[7] J. Dai, B. Wang, J. Sheng, and G. Duan, "Identifying Influential Nodes in Complex Networks Based on Local Neighbor Contribution," *IEEE Access,* vol. 7, pp. 131719-131731, 2019.

[8] J. Sheng *et al.*, "Identifying influential nodes in complex networks based on global and local structure," *Physica A: Statistical Mechanics and its Applications,* vol. 541, p. 123262, 2020.

[9] T. Bian and Y. Deng, "Identifying influential nodes in complex networks: A node information dimension approach," *Chaos: An Interdisciplinary Journal of Nonlinear Science,* vol. 28, no. 4, p. 043109, 2018.

[10] A. Farooq, G. J. Joyia, M. Uzair, and U. Akram, "Detection of influential nodes using social networks analysis based on network metrics," in *2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*, 2018, pp. 1-6: IEEE.

[11] R. Y. Siti Nurulain Mohd Rum, Lilly Suriani Affendey, "Detecting Influencers in Social Media Using Social Network Analysis (SNA)," *International Journal of Engineering & Technology,* vol. 10, 2018.

[12] A. M. Pudjajana, D. Manongga, A. Iriani, and H. D. Purnomo, "Identification of Influencers in Social Media using Social Network Analysis (SNA)," in *2018 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI),* 2018, pp. 400-404: IEEE.

[13] C. Borrego, J. Borrell, and S. Robles, "Hey, influencer! Message delivery to social central nodes in social opportunistic networks," *Computer Communications,* vol. 137, pp. 81-91, 2019.

[14] C. Salavati, A. Abdollahpouri, and Z. Manbari, "Ranking nodes in complex networks based on local structure and improving closeness centrality," *Neurocomputing,* vol. 336, pp. 36-45, 2019.

[15] L. Lü, T. Zhou, Q.-M. Zhang, and H. E. Stanley, "The H-index of a network node and its relation to degree and coreness," *Nature communications,* vol. 7, no. 1, pp. 1-7, 2016.

[16] A. Zareie and A. Sheikhahmadi, "EHC: Extended H-index Centrality measure for identification of users’ spreading influence in complex networks," *Physica A: Statistical Mechanics and Its Applications,* vol. 514, pp. 141-155, 2019.

[17] A. Namtirtha, A. Dutta, and B. Dutta, "Identifying influential spreaders in complex networks based on kshell hybrid method," *Physica A: Statistical Mechanics and its Applications,* vol. 499, pp. 310-324, 2018.

[18] J. Wang, C. Li, and C. Xia, "Improved centrality indicators to characterize the nodal spreading capability in complex networks," *Applied Mathematics and Computation,* vol. 334, pp. 388-400, 2018.

[19] J. Zhao, Y. Wang, and Y. Deng, "Identifying influential nodes in complex networks from global perspective," *Chaos, Solitons & Fractals,* vol. 133, p. 109637, 2020.

روشی جدید برای استخراج تداخل دارویی از متن: استفاده از تفکیک دو مرحله ای مبتنی بر جنگل تصادفی در مجموعه های با کلاس های چندگانه و توزیع نامتعادل

هانا فتحی[۱]، محمدرضا کیوان پور[۲]، سیدوهاب شجاع الدینی[۳]

[۱] گروه مهندسی کامپیوتر، دانشکده فنی، دانشگاه الزهرا ، تهران، ایران
H.Fathi@student.alzahra.ac.ir

[۲] گروه مهندسی کامپیوتر، دانشکده فنی، دانشگاه الزهرا، تهران، ایران
keyvanpour@Alzahra.ac.ir

[۳] گروه مهندسی پزشکی، پژوهشکده برق و فناوری اطلاعات، سازمان پژوهش های علمی و صنعتی ایران، تهران، ایران
shojadini@irost.ir

چکیده

تشخیص تداخل دارویی یکی از مسایل مطرح در حوزه‌ی سلامت است که با توجه به گستردگی حیطه‌ی داروها و تنوع و پیچیدگی اثرات متقابل آن‌ها به موضوعی برای تکنیک‌های متن‌کاوی بدل شده است. با این مواردی از جمله وجود مجموعه داده‌گان محدود و با کلاس‌های چندگانه و نامتعادل از چالش-های مهم این حوزه می‌باشند. در این مقاله، یک روش استخراج تداخل دارویی مبتنی بر یادگیری ماشین جهت غلبه بر چالش‌های فوق ارایه می‌شود که در آن از راهبردی دو مرحله ای مشتمل بر تشخیص تداخل دارویی و طبقه‌بندی این تداخل‌ها با استفاده از روش درخت تصادفی استفاده می‌شود. نتایج حاصل شده از آزمون این روش بر روی بانک داده‌گان کوچک و نامتعادل حاکی از آن است که ایده‌ی تفکیکی پیشنهادی قادر به تشخیص بهتر تداخل‌های در چهار کلاس عملکردی و به میزان‌های به ترتیب ۱۰ ، ۸، ۱۳ و ۲۸ درصد برای سنجه‌ی امتیاز F1 نسبت به بهترین روش رقیب خود از روش‌های هم خانواده‌ی موجود می باشد و بر این اساس می‌تواند به عنوان روشی با پتانسیل توسعه جهت کاربردهای واقعی تر در نظر گرفته شود.

کلمات کلیدی

تداخل دارویی، متن کاوی، یادگیری ماشین، روش تفکیک دو مرحله ای، طبقه‌بندی جنگل تصادفی.

۱- مقدمه

تداخل دارویی (DDI[1]) به طور کلی به عنوان تغییر در اثر یک دارو به دلیل مصرف داروی دیگر تعریف می‌شود. تداخل‌های دارویی می‌توانند دارویی، فارماکوکینتیک و فارماکودینامیک باشند. تداخل‌های دارویی به دلیل ناسازگاری فیزیکی یا شیمیایی رخ می دهد. تداخل‌های دارویی فارماکوکینتیک زمانی رخ می دهد که یک دارو با فرآیند دارویی دیگر که شامل جذب، توزیع، متابولیسم و دفع است، تداخل ایجاد کند. و تداخل‌های دارویی فارماکودینامیک زمانی ایجاد می‌شود که یک دارو پاسخ فارماکودینامیک را به همان غلظت داروی دیگر تغییر دهد. تداخل دارویی ممکن است زمانی رخ دهد که چندین دارو به طور همزمان تجویز شوند[۱]. اگرچه تداخل‌های دارویی ممکن است اثرات مفیدی داشته باشند، اما گاهی اوقات تداخل دارویی ممکن است منجر به واکنش‌های نامطلوب دارویی مختلف (ADRs[2]) شود که باعث عواقب زیان‌بار اجتناب‌ناپذیر و هزینه‌های زیادی برای ارائه‌دهندگان بهداشت و بیمارستان‌ها می‌شود [۲]. بدین ترتیب، تداخل دارویی یک نگرانی عمده در زمینه نظارت دارویی است که هدف آن شناسایی و تعیین کمیت خطرات مربوط به مصرف دارو برای درک بهتر واکنش‌های نامطلوب دارویی و مکانیسم‌های بیماری‌زا است[۳]. نظارت دارویی شامل دو مرحله است: نظارت قبل از بازاریابی و نظارت پس از بازاریابی. نظارت قبل از بازاریابی اطلاعاتی را که در مورد چگونگی رخ‌دادن واکنش‌های نامطلوب دارویی می‌باشد از مطالعات پیش‌بالینی و آزمایشات

[1] **D**rug-**D**rug **I**nteraction
[2] **A**dverse **D**rug **R**eactions

بالینی جمع‌آوری می‌کند[۴]. از سوی دیگر، نظارت پس از بازاریابی، ایمنی یک دارو را پس از عرضه به بازار بررسی می‌کند و داده‌های موجود در مورد آن دارو را در طول مدت زمانی که وارد بازار شده است را تجزیه و تحلیل می‌کند[۵]. در حال حاضر، اکثر تداخل‌های دارویی به طور تصادفی هنگامی که یک دارو در حال حاضر در بازار وجود دارد کشف می‌شوند. تشخیص زودهنگام تداخل‌های دارویی بر اساس ویژگی‌های دارو به متخصصان ایمنی دارو کمک می‌کند تا منابع تحقیقاتی را تخصیص دهند و اقدامات نظارتی را انجام دهند. محدودیت‌هایی در این حوزه وجود دارد که چالش‌های مهم و فرصت‌های تحقیقاتی قابل توجهی را به همراه دارد. این چالش‌ها به دو دسته اصلی چالش‌های مربوط با نوع داده و چالش‌های مربوط به ساختار شبکه تقسیم می‌شوند. هنگام استخراج و تشخیص تداخل‌های دارویی، چالش‌های متعددی در مجموعه داده‌گان وجود دارد که ممکن است بر کیفیت خروجی یک مدل تأثیر منفی بگذارند. این چالش‌ها شامل عدم وجود مجموعه داده‌گان با کیفیت، حجم زیاد داده و داده‌های نامتعادل می‌باشند. مجموعه‌ای از چالش-ها مربوط به نحوه طراحی یک مدل برای استخراج تداخل دارویی وجود دارند که از جمله این موارد می‌توان به عدم حصول یک مدل پایدار و قابل تفسیر خصوصا برای مجموعه داده گان کوچک اشاره کرد [۶]. علاوه بر این، دلایل زیادی می‌تواند باعث تاخیر در تشخیص تداخل دارویی شوند مانند دشواری تشخیص بالینی تداخل‌های دارویی و ماهیت فرآیند تایید دارو. بنابراین، تمایل زیادی برای توسعه روش‌های محاسباتی برای تسریع در استخراج تداخل دارویی وجود دارد[۷]. اخیراً بسیاری از محققان مدل‌های مبتنی بر یادگیری ماشین را برای غلبه بر این مسئله طراحی کرده‌اند[۲]، که به تجویز ایمن‌تر و مؤثرتر برای بیماران در طول توسعه دارو و همچنین به کاهش زمان و هزینه آزمایشات بالینی کمک می‌کنند. با این دیدگاه استخراج روابط تداخل دارویی یک صورت مساله ی طبقه‌بندی است که هدف آن کشف خودکار روابط دارویی از ادبیات زیست پزشکی برای نسبت دادن آنها به یک نوع تداخل دارویی از پیش تعریف شده است[۸]. روش‌های استخراج تداخل دارویی موجود را می‌توان به دو دسته تقسیم کرد: روش‌های یک مرحله‌ای و دو مرحله‌ای. روش یک مرحله‌ای تشخیص و طبقه‌بندی تداخل دارویی را با آموزش یک الگوریتم هوشمند چند کلاسه به طور همزمان انجام می‌دهد و هر نمونه کاندید را به یکی از پنج نوع تداخل دارویی طبقه‌بندی می‌کند (یعنی توصیه[3]، اثر[4]، مکانیسم[5]، داخلی[6]، منفی[7]). روش دو مرحله‌ای، مسئله یادگیری را به دو مرحله تقسیم می‌کند: در مرحله اول، همه تداخل‌های دارویی شناسایی می‌شوند و در مرحله دوم به یکی از چهار نوع خاص تداخل دارویی (یعنی توصیه، اثر، مکانیسم، داخلی) طبقه‌بندی می‌شوند. روش‌های دو مرحله-ای معمولاً از روش‌های یک مرحله‌ای بهتر عمل می‌کنند[۹]. از دیدگاه مبتنی بر نوع طبقه‌بندی کننده‌ی مورد استفاده، روش‌های استخراج و تشخیص تداخل دارویی را می‌توان در دو گروه مبتنی بر طبقه‌بندی سنتی و مبتنی بر شبکه عصبی دسته‌بندی کرد. در میان روش های کلاسیک معمولا الگوریتم ماشین بردار پشتیبان جایگاه اصلی را دارد. برخی از روش‌های معروف مبتنی بر ماشین بردار پشتیبان از مجموعه داده‌گان DDIExtraction 2011 و DDIExtraction 2013 استفاده کرده‌اند. در بین این روش‌ها، رتبه اول متعلق به روش مبتنی بر ماشین بردار پشتیبان است که توسط چاودوری و همکاران [۱۰] ارائه شده است. روش پیشنهادی شامل دو مرحله است که به طور خاص، یک طبقه‌بندی کننده ماشین بردار پشتیبان باینری برای یافتن جملات کم‌آموزنده آموزش داده می‌شود. در این روش، نمونه‌های منفی از مجموعه داده‌گان آزمایشی حذف شده و سپس یک هسته ترکیبی برای شناسایی تداخل‌های دارویی استفاده شده است. سرانجام با آموزش چهار مدل جداگانه آنها را به یکی از چهار نوع تداخل دارویی اختصاص می دهد. بعدها کیم و همکاران [۱۱] یک روش برای استخراج تداخل دارویی پیشنهاد کردند که در آن پنج نوع ویژگی برای گرفتن داده‌های پیچیده ایجاد کردند. در این روش، ابتدا یک طبقه‌بندی‌کننده باینری برای استخراج تداخل دارویی آموزش داده می‌شود و جفت‌های متقابل را در دسته‌های روابط از پیش تعریف‌شده طبقه‌بندی می‌کنند. سپس، یک طبقه‌بندی‌کننده ماشین بردار پشتیبان خطی با یک تابع ضرر هوبر اصلاح‌شده ساخته می‌شود. این روش عملکرد بهتری را با استفاده از مجموعه داده‌گان پیشین به دست می‌آورد و از ویژگی کلمه، ویژگی جفت کلمه، ویژگی درخت تجزیه و ویژگی هماهنگی محدود به عبارت اسمی استفاده می‌کند. نمونه‌ای از روش تک مرحله‌ای، روشی است که توسط تیم UTurku [۱۲] برای مجموعه داده‌گان DDIExtraction 2013 ارائه شده است. این روش وظایف تشخیص و طبقه‌بندی تداخل دارویی را با آموزش الگوریتم ماشین بردار پشتیبان چند کلاسه به طور همزمان انجام می‌دهد و هر نمونه کاندید را به یکی از پنج نوع تداخل دارویی (توصیه، اثر، مکانیسم، داخلی، منفی)طبقه‌بندی می‌کند. این روش به سومین عملکرد برتر در چالش DDIExtraction 2013 دست یافته است.

رستگار مجرد و همکاران[۱۳] رویکردی را برای طبقه‌بندی تداخل‌های دارویی پیشنهاد کرد. این رویکرد از یک طبقه‌بندی‌کننده ماشین بردار پشتیبان وزن‌دار دو مرحله‌ای برای مدیریت توزیع کلاس بسیار نامتعادل استفاده می‌کند. بدین ترتیب ابتدا، تداخل‌های دارویی با استفاده از طبقه‌بندی‌کننده ماشین بردار پشتیبان با وزن دودویی شناسایی می‌شوند تا نمونه‌های مثبت (یعنی نمونه هایی که تداخل دارو دارند) از نمونه‌های منفی تشخیص داده شوند. در مرحله دوم، نمونه مثبت مرحله اول در یکی از چهار دسته طبقه‌بندی می شود. در این ماشین بردار پشتیبان از مجموعه‌ای از ویژگی‌های واژگانی و معنایی برای تشخیص نمونه‌های مثبت از نمونه‌های منفی استفاده شده است. برای پرداختن بیشتر به مشکل نامتعادل بودن داده، نویسندگان رویکردها و الگوریتم‌های مختلفی از جمله SMOTE [۱۴] و دیگر الگوریتم‌های نمونه‌برداری را بررسی کردند. آنها بهترین نتایج را برای داده‌های آموزش با ماشین بردار پشتیبان با وزن‌های کلاس‌های مختلف به دست آوردند. یکی از اشکالات مهم مدل‌های مبتنی بر ماشین بردار پشتیبان این است که معمولاً از ویژگی‌های پیچیده واژگانی یا نحوی انتخاب شده توسط انسان برای ایجاد نمایش‌هایی از جملات یا نشانه‌ها استفاده می‌کنند. انتخاب ویژگی‌ها به صورت دستی در مقایسه با ویژگی‌های پنهان یادگیری عمیق، پرزحمت و زمان‌بر است. از منظری دیگر، ویژگی‌های انتخاب شده توسط انسان عموماً عملکرد ضعیفی دارند[۱۵]. روش‌های مبتنی بر شبکه‌های عصبی و در آن میان به طور خاص ساختارهای مبتنی بر یادگیری عمیق در استخراج خواص دارویی کارایی بالایی دارند، و از روش‌های مبتنی بر ماشین بردار پشتیبان بهتر عمل می‌کنند[۱۶]. این روش‌ها تأثیر قابل توجهی بر درک زبان طبیعی،

[3] **A**dvice
[4] **E**ffect
[5] **M**echanism
[6] **Int**
[7] **N**egative

تشخیص گفتار و تجزیه و تحلیل احساسات داشته‌اند[۷]، و بدین قرینه برای استخراج تداخل دارویی نیز به کار برده شده اند. به طور مشخص معماری‌هایی از این نوع شبکه‌ها شامل روش‌های مبتنی بر شبکه‌های عصبی بازگشتی (RNN[8])، روش‌های مبتنی بر شبکه‌های عصبی کانولوشنال (CNN[9]) و روش‌های مبتنی بر حافظه بلند کوتاه مدت (LSTM[10]) در این حوزه عمومیت بیشتری دارند. ژانگ و همکاران [۱۷] یک روش شبکه عصبی بازگشتی سلسله مراتبی[۱۱] را برای ادغام کوتاه‌ترین مسیر وابستگی(SDP[12]) و دنباله جمله برای کار استخراج تداخل دارویی ارائه کردند. برای ارزیابی این روش از مجموعه داده‌گان DDIExtraction 2013 استفاده شد که نتایج حاکی از مکمل بودن ترتیب جمله و کوتاه‌ترین مسیر وابستگی با یکدیگر بود. ادغام دنباله جمله با کوتاه‌ترین مسیر وابستگی می‌تواند به طور موثر عملکرد استخراج تداخل دارویی را بهبود بخشد. این روش از Bi-LSTMهای سلسله مراتبی برای ادغام جمله و کوتاه‌ترین مسیر وابستگی استفاده می‌کند، و می‌تواند توانایی مدل RNN را برای مقابله با جملات طولانی و پیچیده بیشتر بهبود بخشد. با این وجود، این روش برای دسته داخلی عملکرد خوبی ندارد که به دلیل عدم وجود داده‌های آموزشی کافی است، چرا که عملکرد این خانواده از روش‌ها به داده‌های آموزشی با کیفیت بالا بستگی دارد. دیوی و همکاران[۸] یک روش به نام DeepCNN را برای استخراج رابطه تداخل دارویی پیشنهاد کردند. این روش از معماری ۱۰ لایه‌ای برای یادگیری بازنمایی ویژگی‌های قوی در سطح بالا استفاده می‌کند. شبکه‌ها با استفاده از Dropout و تنظیم L2 برای دستیابی به بهترین عملکرد تنظیم شدند. روش‌های مبتنی بر شبکه‌های عصبی کانولوشنال می توانند نتایج خوبی را بدست آورند، اما محدودیت‌هایی نیز دارند. از جمله این محدودیت‌ها می توان به عدم در نظر گرفتن فاصله زیاد بین کلمات در نمونه های کاندید DDI و پیچیدگی بالای آن‌ها اشاره کرد[۱۸]. دیوی و همکاران از جاسازی کلمات چند کاناله برای بزرگتر کردن دایره لغات استفاده کردند، و تعداد کلمات ناشناخته را کاهش دادند. ساهو و همکاران [۱۹] سه مدل مبتنی بر LSTM، (یعنی B-LSTM، AB-LSTM، و Joint AB-LSTM) را برای کار استخراج تداخل دارویی ارائه کردند. هر سه مدل از کلمه ساده و تعبیه فاصله به عنوان ویژگی استفاده می‌کنند و همه آنها روش‌های یک مرحله‌ای هستند. همچنین، آنها با استفاده از یک شبکه Bi-LSTM، نمایش ویژگی‌های سطح بالاتر را یاد می‌گیرند. با توجه به نتایج مدل Joint AB-LSTM حتی با معماری‌های ساده بدون استفاده از ویژگی‌های پیچیده زبان مبتنی بر ابزار، عملکرد مناسبی دارد. اما نامتعادل بودن و نویز بر همه مدل‌ها تأثیر منفی دارد. ژو و همکاران [۲۰] یک شبکه عصبی چند وظیفه‌ای آگاه از موقعیت[۱۳] را پیشنهاد کردند. این معماری برای بهبود کارایی مکانیسم توجه، از جاسازی برای تولید وزنه‌های توجه استفاده می‌کند. نویسندگان از مجموعه داده‌گان DDIExtraction 2013 برای ارزیابی مدل استفاده کردند. با توجه به نتایج، این روش مقادیر مناسبی از یادآوری[۱۴] و دقت[۱۵] قابل مقایسه را در بین تمام رویکردهای قبل از خود، به‌دست‌آورد. علیرغم توانایی روشهای خانواده یادگیری عمیق در تشخیص تداخل‌های دارویی، متاسفانه عملکرد این روش‌ها به شدت به نوع و حجم بانک داده گان آموزشی وابسته است که در خصوص تداخل‌های دارویی که بانک داده‌های با حجم بسیار بزرگ، کمتر موجود هستند. عملا استفاده از این روش ها را در بسیاری از کاربردهای عملی محدود می سازد. بنابراین روش‌های کلاسیک که فاقد محدودیت و وابستگی مزبور هستند، همچنان برای تشخیص تداخل‌های دارویی به عنوان راهکاری موثر و با پتانسیل عملی استفاده می‌شوند. در این مقاله روش جدیدی در حوزه‌ی کلاسیک را برای استخراج روابط تداخل دارویی ارائه می شود که شامل دو مرحله تشخیص تداخل دارویی و طبقه‌بندی تداخل دارویی است. در مرحله اول، همه تداخل‌های دارویی شناسایی می‌شوند و در مرحله دوم با استفاده از از مجموعه‌داده‌گان نامتعادل DDIExtractions 2013 و الگوریتم طبقه‌بندی جنگل تصادفی[۱۶] به یکی از چهار نوع خاص تداخل دارویی طبقه‌بندی می‌شوند. مهم‌ترین فایده استفاده از طبقه بندی جنگل تصادفی در این پژوهش وجود هایپرپارامترهایی است که برای افزایش قدرت پیش‌بینی مدل و سریع‌تر کردن آن استفاده می‌شود. ازجمله این هایپر پارامترها می‌توان به n_estimators، n_jobs، oob_score اشاره کرد. ساختار مقاله به صورت زیر است. در بخش دوم، جزییات بانک داده و روش پیشنهادی تشریح می شود. بخش سوم به بستر تست، آزمایش ها، نتایج و مقایسه با رقبا اختصاص دارد و سرانجام نتیجه گیری به عمل می‌آید.

۲- روش پیشنهادی

الگوریتم(I) یک شبه کد از روش پیشنهادی را نشان می‌دهد. با توجه به این که هدف اصلی این کار استخراج تداخل دارویی از متون است، برای آموزش و ارزیابی روش، مجموعه‌ای از متن‌ها که حاوی تداخل‌های دارویی می‌باشند را استفاده کرده‌ایم. در این بخش ابتدا این مجموعه داده‌گان و سپس جزییات روش اعمال شده شامل پیش پردازش، استخراج ویژگی و طبقه‌بندی تشریح می شود.

۲-۱- مجموعه داده

مدل پیشنهادی بر روی مجموعه داده‌گان تداخل دارویی از چالش DDIExtraction 2013 آموزش و ارزیابی شده است. هدف اصلی این کار، طبقه‌بندی هر تداخل دارویی به یکی از انواع تداخل زیر است:

توصیه (نظریه یا توصیه در مورد یک تداخل دارویی می‌باشد)، داخلی (تعاملی که به سادگی در یک جمله بیان شده یا توصیف شده‌است)، مکانیسم (تداخل‌های فارماکودینامیک بین جفت داروها را توصیف می‌کند)، اثر (تداخل‌های فارماکوکینتیک بین جفت داروها را توصیف می‌کند) و منفی (عدم وجود تداخل دارویی بین دو دارو) می‌باشد. که ما نمونه‌هایی مجموعه داده‌گان DDIExtraction 2013 از دو بخش بانک دارو[۱۷] و مدلاین[۱۸] تشکیل شده است. مدلاین شامل مقالات تحقیقاتی زیست پزشکی است، در حالی که بانک دارو شامل متن‌هایی است که به صورت دستی از منابع مختلف و تایید شده

---

[8] **R**ecurrent **N**eural **N**etwork
[9] **C**onvolutional **N**eural **N**etwork
[10] **L**ong **S**hort-**T**erm **M**emory
[11] **H**ierarchical **R**ecurrent **N**eural **N**etwork
[12] **S**hortest **D**ependency **P**ath
[13] **P**osition-**A**ware **M**ulti-**T**ask **N**eural **N**etwork
[14] **R**ecall
[15] **P**recision
[16] **R**andom **F**orest
[17] **D**rug**B**ank
[18] **M**edLine

Algorithm I

```
Input: get sentence_text and relation_type from dataset
        for sentence_dom in sentences:
            entity_dict = get_entity_dict(sentence_dom)
            pairs =
    sentence_dom.getElementsByTagName('pair')
            sentence_text = sentence_dom.getAttribute('text')
            for pair in pairs:
                ddi_flag = pair.getAttribute('ddi')
                if not os.path.isfile('types'):
                    types.add(pair.getAttribute('type'))
                if ddi_flag == 'true':
                    e1 = pair.getAttribute('e1')
                    e2 = pair.getAttribute('e2')
                    relation_type = pair.getAttribute('type')
                    lol.append([sentence_text, entity_dict[e1],
                    entity_dict[e2], relation_type])
Vector_Feuture, lable = get data from tarined dataset
X_train, X_test, y_train, y_test = split data tarin and test
(Vector_feuture,Lable)
model = RandomForestClassifier (set some Configuration for
process Model)
model.fit(X_train, y_train)
Output: Classifire DDI
```

توسط کارشناسان معتبر جمع‌آوری‌شده است. همان طور که قبلا گفتیم ما مواردی را که منفی هستند را حذف می‌کنیم. آمار اولیه مجموعه داده‌گان تست و آموزش برای بانک دارو و مدلاین در جدول(۱) نشان داده شده است.

جدول (۱)- آمار اولیه مجموعه داده‌گان آموزش و تست.

آموزش	منفی	اثر	مکانیسم	داخلی	توصیه
بانک دارو	۲۲۲۱۷	۱۵۳۵	۱۲۵۷	۱۷۸	۸۱۸
مدلاین	۱۵۵۵	۱۵۲	۶۲	۱۰	۸
تست	منفی	اثر	مکانیسم	داخلی	توصیه
بانک دارو	۴۳۸۱	۲۹۸	۱۷۸	۹۴	۲۱۴
مدلاین	۲	۶۲	۲۴	۲	۷

۲-۲- پیش‌پردازش

شکل (۱) کل فرآیند پردازش مورد استفاده در این مقاله را نشان می‌دهد. در اولین گام، پیش‌پردازش مناسب می‌تواند عملکرد را به میزان قابل توجهی افزایش دهد. در روش پیشنهادی، دو عملیات پیش‌پردازش انجام می‌شود، یعنی نرمال‌سازی[۱۹] و توکن‌سازی[۲۰]. ما از نرمال‌سازی نشانه‌ها برای جلوگیری از تغییرات املایی استفاده می‌کنیم. نرمال‌سازی به ما کمک می‌کند تا اطلاعات n-gram را بدست آوریم. فرآیند توکن‌سازی یکی از مراحل پیش-پردازش استاندارد است. به عنوان مثال اگر نام دارو حاوی یک کاراکتر فضای خالی بلافاصله قبل و بعد از مرزهای آن نباشد، کاراکترهای فاصله را در آن موقعیت‌ها در داخل جمله مربوطه وارد می‌کنیم. ما کلمات تکراری را حذف می‌کنیم.

۲-۳- استخراج ویژگی

در این روش ما بعد از انجام پیش‌پردازش‌های لازم، ویژگی‌های لازم را استخراج می‌کنیم که شامل ویژگی‌های کلمه مانند کلمات فردی در یک جمله و دنباله‌ای از کلمات نشان داده. از این رو، در روش پیشنهادی از ویژگی‌های n-gram تا اندازه ۲ استفاده می‌کنیم، به عبارتی ما ویژگی‌های unigrams و bigrams را استخراج می‌کنیم. این ویژگی‌ها قادر به کشف الگوهای مربوط به کلمات دور در یک جمله نیستند. یک راه حل ساده برای گرفتن الگوهای کلمات دور این است که تمام ترکیب‌های ممکن کلمات را از یک مجموعه داده‌گان آموزشی استخراج کنیم. با این حال، این رویکرد تعداد ویژگی‌ها را به میزان قابل توجهی افزایش می دهد.

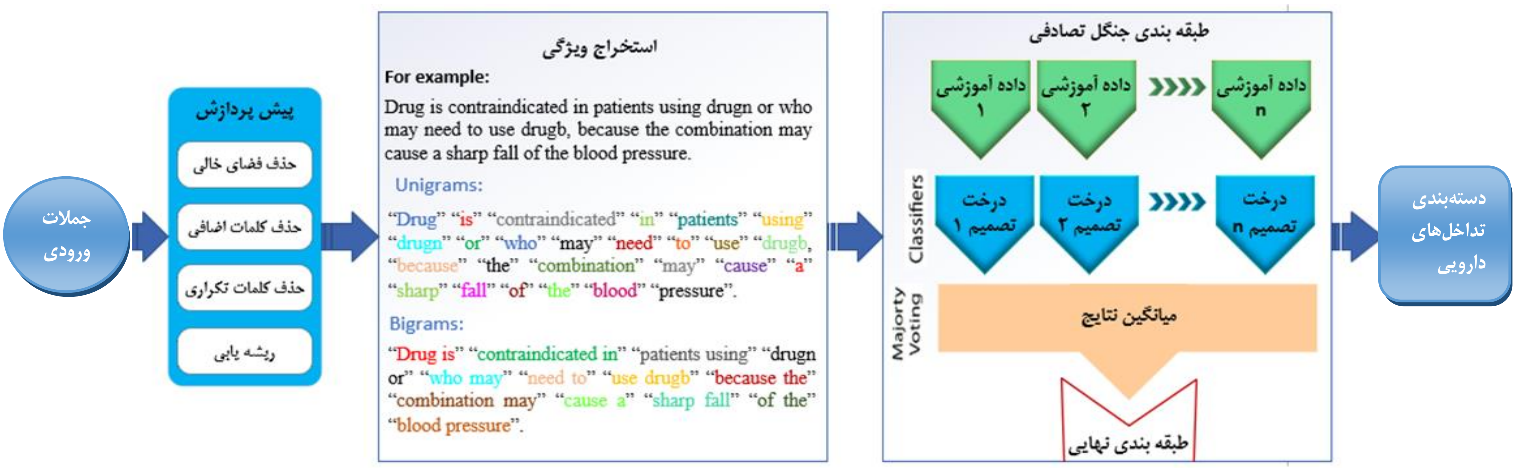
۲-۴- طبقه‌بندی کننده

روش پیشنهادی ما وظیفه استخراج تداخل دارویی را به مرحله تشخیص تداخل دارویی و مرحله طبقه‌بندی تداخل دارویی تقسیم می‌کند. سه مرحله برای تشخیص تداخل دارویی وجود دارد: جملات کم‌آموزنده را کنار می-گذاریم، نمونه‌های اطلاعاتی کمتر را کنار می‌گذاریم، و سیستم (یک مدل واحد بدون در نظر گرفتن انواع تداخل دارویی) بر روی نمونه‌های آموزشی باقی‌مانده و تداخل‌های دارویی احتمالی شناسایی شده از نمونه‌های آزمایشی باقی‌مانده آموزش داده می‌شود. در مرحله طبقه‌بندی تداخل دارویی ما از طبقه بندی کننده جنگل تصادفی با ۴ کلاس بر روی تداخل دارویی‌های استخراج شده مثبت برای آموزش و استخراج نوع تداخل در مجموعه داده‌گان آزمایشی استفاده کرده‌ایم. جنگل تصادفی، یک روش یادگیری گروهی است که درختان طبقه‌بندی زیادی را رشد می‌دهد، جایی که هر درخت یک طبقه‌بندی را ارائه می‌دهد. طبقه‌بندی که حداکثر آرا را دارد از جنگل انتخاب می‌شود. یکی از ویژگی‌های جالب این طبقه‌بندی کننده این است که یک اندازه طبیعی از اطمینان طبقه‌بندی را برای هر نمونه آموزشی ارائه می‌دهد. معیار استاندارد عملکرد برای جنگل تصادفی، تخمین خطا (OOB [21]) است که شبیه اعتبارسنجی متقاطع یک خروجی است. هر درخت در جنگل تصادفی تنها با استفاده از حدود ۲/۳ از داده‌های آموزشی موجود ساخته شده است. بقیه نقاط داده به عنوان داده‌های "خارج از کیسه[۲۲]" برای آن درخت شناخته می‌شوند. بنابراین، می‌توان کل جنگل را ساخت، سپس هر نمونه آموزشی را تنها با استفاده از درختانی که برای آنها تخمین خطا بود، طبقه‌بندی کرد. یکی از ویژگی‌های مهم جنگل تصادفی این است که معیاری طبیعی از قطعیت طبقه‌بندی آن برای هر نمونه آموزشی ارائه می‌کند. به عنوان مثال، کسری از درختانی که رای "تعامل" داده اند. با رتبه‌بندی مسیرها برای یک جفت داروی خاص بر اساس تعداد رای‌های «بله» که هر کدام از جنگل تصادفی دریافت کرده‌اند، می‌توانیم تعیین کنیم که کدام مسیر(ها) محتمل‌ترین مکانیسم (های) تعامل را برای آن جفت نشان می‌دهد [۲۱].

این رویکرد راهی جدید برای اتصال حقایق مختلف از سراسر متن‌های زیست پزشکی برای ارائه توضیحات مکانیکی از تداخل‌ها ارائه می‌دهد. در مورد جفت‌های دارویی که قبلاً برای تداخل شناخته شده‌اند، استفاده از این رویکرد فهرستی از مکانیسم‌های تداخل‌های مهم را ارائه می‌کند که ممکن است به کشف مکانیسم‌های جدیدی که هنوز بخشی از دانش پزشکی فعلی نیستند کمک کند. این رویکرد بینش‌های ارزشمندی را در مورد مکانیسم‌های احتمالی تداخل دارویی اراده می‌دهد..هر درخت در جنگل تصادفی تنها از یک زیر مجموعه از ویژگی‌ها برای طبقه‌بندی استفاده می‌کند، که تضمین می‌کند

[19] Normalization
[20] Tokenization

[21] out-of-bag
[22] out-of-bag



شکل(۱) – فرآیند روش پیشنهادی.

که آرای درخت‌های مختلف همبستگی ندارند. جنگل تصادفی انعطاف‌پذیری خوبی دارد (یعنی تعاملات را بدون شرایط تعامل مدیریت می‌کند)، و قادر است متغیرهای زیادی را در مقایسه با اندازه مجموعه داده مدیریت کند. این الگوریتم معیاری از سهم هر متغیر در دقت پیش‌بینی (به نام اهمیت متغیر) ارائه می‌کند. این معیار مشخص نمی‌کند که یک ارتباط مثبت یا منفی است، اما برای رتبه‌بندی متغیرهای مستقل استفاده می شود[۲۲].

۳- بحث و نتایج

آزمایش ها بر بستری شامل سیستم ۶۴ بیتی با ۴ گیگا بایت رم و CPU، Intel corei3 انجام شد و برای اجرای کد از پایتون ۳.۹ و کتابخانه‌های pandas ، spacy، tqdm استفاده شد. مقایسه عملکرد بین روش پیشنهادی و روش‌های مطرح دیگر این حوزه در جدول (۲) نشان داده شده است.
معیارهای ارزیابی که استفاده شده است شامل دقت، فراخوانی و امتیاز F1 می‌باشد. روش‌های مبتنی بر یادگیری ماشین که برای استخراج تداخل دارویی ارائه شده‌اند اغلب از طبقه‌بند کننده ماشین بردار پشتیبان استفاده کرده‌اند. ما روش خود را با دو روش ارائه شده توسط تیم‌های WBI-DDI[۲۳] و NIL-UCM[۲۴] مقایسه کردیم. این دو روش برای طبقه‌بندی تداخل‌های دارویی شناسایی شده از الگوریتم ماشین بردار پشتیبان استفاده کرده‌اند. همان طور که در جدول (۲) مشاهده می‌کنید، روش ما نسبت به دو روش مقایسه‌ای برتری دارد. روش ما در تمام کلاس‌ها بخصوص در کلاس داخلی، ۰.۲۳ دقت بالاتری بدست آورده است، دلیل این امر این است که الگوریتم جنگل تصادفی برای کار طبقه‌بندی تداخل‌های دارویی کارایی بالایی دارد. علت بهبود نتایج در روش پیشنهادی نسبت به سایرین آن است که الگوریتم طبقه–بندی جنگل تصادفی این قابلیت را دارد که مشکل طبقه‌بندی چند کلاسه را بهتر از روش‌های ترکیب چند ماشین بردار پشتیبان حل کند، زیرا ماشین بردار پشتیبان در اصل برای مسئله طبقه‌بندی باینری طراحی شده است. از سویی طبقه‌بندی‌کننده جنگل تصادفی این قابلیت را دارد که برای توضیح تداخل‌های شناخته‌شده و برای کشف تداخل‌های جدیدی که هنوز گزارش نشده‌اند استفاده شود. هر درخت در جنگل تصادفی تنها از یک زیرمجموعه از ویژگی‌ها برای طبقه‌بندی استفاده می‌کند، که تضمین می‌کند که آرای درختان مختلف همبستگی ندارند. هنگامی که جنگل تصادفی بر روی مجموعه‌ای از جفت‌های داروی متقابل شناخته شده آموزش داده شد، می‌توان از آن برای پیش‌بینی اینکه آیا هر جفت داروی دیگری تداخل خواهد کرد یا خیر، استفاده کرد. لذا به طور خاص، می‌توان آن را برای یک مجموعه آزمایش جدید که شامل داروهای مجموعه آموزشی اصلی نیست، اعمال کرد. بعلاوه بر این موارد، میانگین گیری یا ترکیب نتایج درخت‌های تصمیم گیری مختلف، می تواند باعث شود که روش پیشنهادی تا حد بیشتری از روش‌های رقیب بر مشکل بیش‌برازش[۲۳] غلبه کند که همه‌ی این عوامل می‌توانند توجیه بهبود حاصل شده در جدول (۲) باشند. علیرغم این موارد، روش پیشنهادی نسبت به رقبای برشمرده شده تا حدی پیچیده‌تر و اجرای آن تا میزان ۲ برابر زمان برتر می‌باشد.

جدول(۲)– نتایج استخراج رابطه در مجموعه آزمایشی.

روش‌ها	کلاس	امتیاز F1 %	Recall ٪	دقت ٪
NIL-UCM [۲۱]	مکانیسم	۵۳	۴۸	۵۸
	توصیه	۶۱	۶۲	۵۹
	اثر	۵۵	۵۸	۵۳
	داخلی	۳۹	۳۳	۴۷
WBI-DDI [۲۰]	مکانیسم	۶۱	۵۸	۶۵
	توصیه	۶۳	۵۸	۶۹
	اثر	۶۱	۶۱	۶۰
	داخلی	۵۱	۴۱	۶۷
روش پیشنهادی این مقاله	مکانیسم	۷۱	۷۱	۷۰
	توصیه	۷۱	۶۸	۷۵
	اثر	۷۴	۷۶	۷۲
	داخلی	۷۹	۷۰	۹۰

[23] **o**verfitting

۴- نتیجه گیری

در این مقاله، یک روش جدید برای کار استخراج تداخل دارویی در حوزه روش‌های کلاسیک ارائه شد که در دو مرحله تشخیص تداخل دارویی و طبقه‌بندی تداخل دارویی را تفکیک می‌نماید. برای طبقه‌بندی از روش طبقه‌بندی جنگل تصادفی استفاده شد که به دلیل پوشش دادن بهتر مسائل چند کلاسه و مساله‌ی بیش‌برازش توانست تداخل‌های شناخته شده دارو را در مجموعه داده‌گان نامتعادل DDI Extractions 2013 با دقت ۹۰ درصد به دست‌آورد. با توجه به کوچک بودن بسیاری از مجموعه داده‌گان موجود و عدم وجود تعادل در آنها به عنوان دو چالش اصلی در این حوزه به نظر می‌رسد که استفاده از روش های چند مرحله‌ای و به کارگیری طبقه‌بندهای با قابلیت کنترل عدم تعادل دسته‌ها می تواند یک راهکار برای انجام پژوهش‌های آتی و سیستم‌های کاربردی تر در این حوزه باشد.

مراجع

[1] S. Liu *et al.*, "Enhancing drug-drug interaction prediction using deep attention neural networks," *IEEE/ACM Trans. Comput. Biol. Bioinforma.*, 2022.

[2] P. K. Shukla *et al.*, "Efficient prediction of drug–drug interaction using deep learning models," *IET Syst. Biol.*, 2020, doi: 10.1049/iet-syb.2019.0116.

[3] R. Ferdousi, R. Safdari, and Y. Omidi, "Computational prediction of drug-drug interactions based on drugs functional similarities," *J. Biomed. Inform.*, 2017, doi: 10.1016/j.jbi.2017.04.021.

[4] C. Y. Lee and Y. P. P. Chen, "Prediction of drug adverse events using deep learning in pharmaceutical discovery," *Briefings in Bioinformatics*. 2021. doi: 10.1093/bib/bbaa040.

[5] I. Heba, A. Amany, S. E. Ahmed, and S. Amr, "Novel data-mining methodologies for detecting drug-drug interactions: a review of pharmacovigilance literature," *Adv Env. Sci Dev Chem*, vol. 20, pp. 301–314, 2014.

[6] R. Safdari, R. Ferdousi, K. Aziziheris, S. R. Niakan-Kalhori, and Y. Omidi, "Computerized techniques pave the way for drug-drug interaction prediction and interpretation," *BioImpacts*. 2016. doi: 10.15171/bi.2016.10.

[7] Y. Qiu, Y. Zhang, Y. Deng, S. Liu, and W. Zhang, "A Comprehensive Review of Computational Methods for Drug-drug Interaction Detection," *IEEE/ACM Trans. Comput. Biol. Bioinforma.*, 2021, doi: 10.1109/TCBB.2021.3081268.

[8] I. N. Dewi, S. Dong, and J. Hu, "Drug-drug interaction relation extraction with deep convolutional neural networks," in *Proceedings - 2017 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2017*, 2017. doi: 10.1109/BIBM.2017.8217933.

[9] Z. Zhao, Z. Yang, L. Luo, H. Lin, and J. Wang, "Drug drug interaction extraction from biomedical literature using syntax convolutional neural network," *Bioinformatics*, 2016, doi: 10.1093/bioinformatics/btw486.

[10] M. Faisal Mahbub Chowdhury and A. Lavelli, "FBK-irst: A multi-phase kernel based approach for drug-drug interaction detection and classification that exploits linguistic information," in **SEM 2013 - 2nd Joint Conference on Lexical and Computational Semantics*, 2013.

[11] S. Kim, H. Liu, L. Yeganova, and W. J. Wilbur, "Extracting drug-drug interactions from literature using a rich feature-based linear kernel approach," *J. Biomed. Inform.*, 2015, doi: 10.1016/j.jbi.2015.03.002.

[12] J. Björne, S. Kaewphan, and T. Salakoski, "UTurku: Drug Named Entity Recognition and Drug-Drug Interaction Extraction Using SVM Classification and Domain Knowledge," in **SEM 2013 - 2nd Joint Conference on Lexical and Computational Semantics*, 2013.

[13] M. Rastegar-Mojarad, R. D. Boyce, and R. Prasad, "UWM-TRIADS: Classifying Drug-Drug Interactions with Two-Stage SVM and Post-Processing," in **SEM 2013 - 2nd Joint Conference on Lexical and Computational Semantics*, 2013.

[14] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, 2002, doi: 10.1613/jair.953.

[15] B. Xu *et al.*, "Full-attention Based Drug Drug Interaction Extraction Exploiting User-generated Content," in *Proceedings - 2018 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2018*, 2019. doi: 10.1109/BIBM.2018.8621281.

[16] T. Zhang, J. Leng, and Y. Liu, "Deep learning for drug-drug interaction extraction from the literature: A review," *Briefings in Bioinformatics*. 2020. doi: 10.1093/bib/bbz087.

[17] Y. Zhang, W. Zheng, H. Lin, J. Wang, Z. Yang, and M. Dumontier, "Drug-drug interaction extraction via hierarchical RNNs on sequence and shortest dependency paths," *Bioinformatics*, 2018, doi: 10.1093/bioinformatics/btx659.

[18] S. Liu, K. Chen, Q. Chen, and B. Tang, "Dependency-based convolutional neural network for drug-drug interaction extraction," in *Proceedings - 2016 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2016*, 2017. doi: 10.1109/BIBM.2016.7822671.

[19] S. K. Sahu and A. Anand, "Drug-drug interaction extraction from biomedical texts using long short-term memory network," *J. Biomed. Inform.*, 2018, doi: 10.1016/j.jbi.2018.08.005.

[20] D. Zhou, L. Miao, and Y. He, "Position-aware deep multi-task learning for drug–drug interaction extraction," *Artif. Intell. Med.*, 2018, doi: 10.1016/j.artmed.2018.03.001.

[21] Percha, B., Garten, Y., & Altman, R. B. (2012). Discovery and explanation of drug-drug interactions via text mining. *Pacific Symposium on Biocomputing*. https://doi.org/10.1142/9789814366496_0040

[22] Cai, R., Liu, M., Hu, Y., Melton, B. L., Matheny, M. E., Xu, H., Duan, L., & Waitman, L. R. (2017). Identification of adverse drug-drug interactions through causal association rule discovery from spontaneous adverse event reports. *Artificial Intelligence in Medicine*.

[23] P. Thomas, M. Neves, T. Rocktäschel, and U. Leser, "WBI-DDI: Drug-Drug Interaction Extraction using Majority Voting," in **SEM 2013 - 2nd Joint Conference on Lexical and Computational Semantics*, 2013.

[24] B. Bokharaeian and A. Díaz, "NIL UCM: Extracting Drug-Drug interactions from text through combination of sequence and tree kernels," in **SEM 2013 - 2nd Joint Conference on Lexical and Computational Semantics*, 2013.

استفاده از یک طبقه‌بندی ترکیبی بر اساس الگوریتم جنگل تصادفی و الگوریتم جایا برای طبقه بندی سرطان سینه

برای چهارمین کنفرانس ملی انفورماتیک ایران

سپیده چهره[۱]

[۱] دانشجوی دکتری مهندسی فناوری اطلاعات(سیستم‌های چندرسانه‌ای)،دانشکده فنی و مهندسی، دانشگاه قم، شهر قم
s.chehreh@stu.qom.ac.ir

چکیده

امروزه تصمیمات متخصصان و ارزیابی داده‌های بیماران، بر تحلیل و درمان سرطان سینه تأثیر می‌گذارد. برای تشخیص زودهنگام سرطان سینه، تکنیک‌های متعدد یادگیری ماشینی نه تنها می‌تواند به بررسی و تشخیص سریع داده‌های پزشکی در درمان سرطان سینه کمک کند، بلکه خطاهایی که ممکن است به دلیل تصمیم‌گیرندگان غیرمتخصص رخ دهد را نیز کاهش می‌دهد. الگوریتم جنگل تصادفی یکی از طبقه بندی کننده‌های معروفی است که در طبقه بندی سرطان تاثیر بسزایی داشته و تنظیمات پارامترهای آن به طور قابل توجهی بر عملکرد آن تأثیر می‌گذارد. بنابراین برای بهبود عملکرد دقت طبقه‌بندی کننده، یک روش طبقه‌بندی هوشمند در حوزه سرطان، مبتنی بر انتخاب یک زیر مجموعه ویژگی و بهینه‌سازی پارامترهای مربوطه در این مقاله پیشنهاد شده است و سبب شده است که الگوریتم جنگل تصادفی از طریق یک الگوریتم هوشمند با عنوان الگوریتم جایا بهبود داده شود. با ترکیب الگوریتم جنگل تصادفی و الگوریتم جایا برای انجام پژوهش از ۶۹۹ نمونه، که به ترتیب شامل ۴۵۸ و ۲۴۱ نمونه برای داده‌های خوش خیم و بدخیم هستند، استفاده شد. علاوه بر این، برای ارزیابی اثربخشی طبقه‌بندی‌کننده پیشنهادی از نظر پیچیدگی محاسباتی و دقت طبقه‌بندی، مقایسه‌ای با الگوریتم جنگل تصادفی به عنوان روش پایه روی پایگاه داده UCI انجام گرفته است.

کلمات کلیدی

یادگیری ماشین، الگوریتم جنگل تصادفی، الگوریتم جایا، طبقه‌بندی، انتخاب ویژگی،

۱- مقدمه

یکی از شایع‌ترین سرطان‌های تشخیص داده شده که علت اصلی مرگ در زنان در سراسر جهان محسوب می شود، سرطان سینه است[1]. سیستم اطلاعات سرطان اروپا ارزیابی می‌کند که در مجموع ۴۰۴،۹۲۰ مورد جدید سرطان سینه در زنان کشورهای اتحادیه اروپا تشخیص داده شده است و حدود ۱۶،۲٪ از زنان در سال ۲۰۱۸ جان خود را از دست داده اند[2]. علیرغم شرایط بحرانی، خوشبختانه مطالعه‌ها نشان داده است که با تشخیص زودهنگام (۴۰٪یا بیشتر) می‌توان میزان مرگ و میر ناشی از سرطان سینه را به طور قابل توجهی کاهش داد[3]. با این حال، تشخیص زودهنگام سرطان اغلب به دلیل عدم وجود علائم اولیه سرطان سینه به عنوان یک موضوع چالش برانگیز است.

رویکردهای مختلفی برای تشخیص، پیشگیری و درمان سرطان سینه استفاده شده است[4]. ماموگرافی، تصویربرداری اولتراسوند یا بیوپسی سینه از موثرترین روش‌ها برای تشخیص سرطان سینه هستند[5]. با این وجود، گزارش شده است که تنها ۵ تا ۱۰ درصد از نتایج ماموگرافی می‌تواند یک پاتولوژی بدخیم را نشان دهد[6]. با وجود ارزشمند بودن غربالگری ماموگرافی[۲] میزان خطاهای مثبت و منفی کاذب قابل توجهی در تفسیرهای برخی از پزشکان و رادیولوژیست‌ها دیده می‌شود، که می‌تواند برای بیماران خطرناک تلقی شود[7]. برای رفع این محدودیت‌ها، در سال‌های گذشته محققان با استفاده از روش‌های شناسایی از طریق داده‌های بالینی علاقمند شده‌اند که در این تلاش‌ها با دو چالش اساسی رو به رو هستند که در اولین مورد می‌توان مقادیر زیادی از داده‌های آموزش را به دست آورد و در دومین مورد می‌توان به تعمیم در بین جامعه آماری برای بهبود تجهیزات و روش‌های تحلیل داده‌ها رسید.

برای جلوگیری از ناراحتی روحی و جسمی بیماران و کاهش هزینه‌های غیر ضروری که برای معاینات صرف می شود، ابزار یا روش‌های مناسب برای شناسایی سرطان سینه و درمان زودهنگام آن ضروری است[8]. رویکردهای یادگیری ماشینی یکی از تکنیک‌های پرکاربرد برای کمک به تشخیص سرطان سینه است[9]. این تکنیک‌ها شناسایی دقیق سرطان سینه را در زمان کوتاه‌تری فراهم می‌کنند.

تا به امروز، رویکردهای مختلف یادگیری ماشین مانند روش یادگیری عمیق[10]، روش منطق فازی[11]، تکنیک داده کاوی[12]، شبکه عصبی مصنوعی[13] و ماشین بردار پشتیبان[۳][14]، استفاده شده است. در میان تمام این تکنیک‌ها در تشخیص سرطان سینه، الگوریتم جنگل تصادفی به دلیل عملکرد پیشرفته برای مجموعه داده‌ها با ابعاد بالا، دارای مزایایی از نظر کارایی محاسباتی و انعطاف پذیری است[15]. هدف اصلی استفاده از الگوریتم جنگل تصادفی برای تشخیص سرطان سینه، ساخت مدلی از طریق این الگوریتم و به صورت حرفه‌ای است تا با دستیابی به خطای کم طبقه بندی،

بتواند وضعیت مبهم بیماران جدید را شناسایی کند. برای استفاده از الگوریتم جنگل تصادفی باید بهینه‌سازی پارامتر و انتخاب ویژگی را مد نظر قرار داد[16].

به منظور انجام طبقه بندی دقیق، انتخاب ویژگی برای تشخیص زیرمجموعه‌ای از ضروری‌ترین ویژگی‌های موجود مورد استفاده قرار می‌گیرد. سه دلیل مهم که نقش انتخاب ویژگی را برجسته کرده است عبارتند از: عملکرد تعمیم، کارایی محاسباتی و همگرایی یادگیری. بنابراین انتخاب ویژگی مناسب نه تنها می‌تواند در زمان محاسباتی الگوریتم به طور قابل توجهی صرفه جویی کند، بلکه مدل‌هایی را نیز می‌سازد که توانایی تعمیم بهتری دارند[17].

بهینه‌سازی پارامترها عامل مهم دیگری است که بر توانایی تعمیم الگوریتم جنگل تصادفی تأثیر می‌گذارد. به عنوان مثال، Xu و همکاران[18] از الگوریتم جنگل تصادفی با الگوریتم ژنتیک موازی برای رسمی کردن مدل هوشمند تشخیص سرطان سینه استفاده کرد. در[19] از الگوریتم ژنتیک برای انتخاب ویژگی به منظور افزایش دقت طبقه بندی استفاده شد. در[20]، از روش الگوریتم جستجوی فاخته برای افزایش عملکرد در طبقه‌بندی سرطان استفاده شد. در[21] ترکیبی از شبکه کانولوشن گراف و شبکه عصبی کانولوشن برای افزایش طبقه‌بندی سرطان سینه ارائه شده است.

در سال‌های اخیر و با پیشرفت یادگیری ماشینی در تحلیل داده‌های پزشکی، روش‌های مناسبی برای بررسی چالش‌های موجود در تفسیر داده پزشکی پیشنهاد شده است. استفاده از مدل‌های یادگیری ماشینی و شناسایی الگوها در تشخیص و تعیین خودکار سرطان باعث کم شدن خطاهای انسانی و افزایش سرعت تشخیص می‌شود. همچنین پردازش داده‌های پزشکی پزشک را در تشخیص راحت‌تر بیماری یاری می‌دهد، بنابراین بیمار در برابرخطرات جبران ناپذیری که پیش روی خواهد داشت، حفظ خواهد شد. چالش اصلی در زمینه تحلیل داده‌های پزشکی آن است که بیشتر روش‌های پیشین در این زمینه فقط وجود یا عدم وجود تومور را تشخیص می‌دهند و در نتیجه، تحقیق‌های کمی در مورد یک رویکرد خودکار برای شناسایی ضایعات خوش‌خیم و بدخیم انجام شده است[22].

علیرغم تحقیق‌های قبلی، هنوز نیاز به تکنیک‌های دقیق‌تری در تشخیص سرطان سینه وجود دارد. هدف کلی این مطالعه پیشنهاد رویکردی است که بتواند سرطان سینه را با دقت بالا و زمان محاسباتی کم، طبقه‌بندی و تشخیص دهد. بنابراین یک روش ترکیبی مبتنی بر الگوریتم جایا و الگوریتم جنگل تصادفی برای تشخیص و طبقه‌بندی سرطان سینه پیشنهاد شده است. الگوریتم جنگل تصادفی یکی از طبقه‌بندی‌کننده‌های معروفی است که قبلاً سهم مهمی در طبقه‌بندی سرطان داشته است. برای بهبود عملکرد دقت طبقه‌بندی کننده الگوریتم جنگل تصادفی برای تشخیص سرطان سینه، یک روش طبقه‌بندی هوشمند در حوزه سرطان مبتنی بر انتخاب یک زیر مجموعه ویژگی و بهینه‌سازی پارامترهای مربوطه در این مقاله پیشنهاد شده است. این طبقه بندی کننده، از طریق یک الگوریتم هوشمند با عنوان الگوریتم جایا[5] بهبود داده شده است. در روش ترکیبی الگوریتم جایا با الگوریتم جنگل تصادفی برای مشخص کردن دقیق سرطان سینه از ۶۹۹ نمونه، که به ترتیب شامل ۴۵۸ و ۲۴۱ نمونه برای داده‌های خوش خیم و بدخیم هستند، استفاده شد.

پایگاه داده مورد استفاده در این روش برگرفته از اطلاعات UCI است و با مقایسه روش مورد مطالعه با روش الگوریتم جنگل تصادفی، امکان سنجش عملکرد آن فراهم آمده است. نتایج تجربی نشان می‌دهد که رویکرد مورد بررسی از روش الگوریتم جنگل تصادفی مانند زمان محاسباتی، حساسیت و دقت طبقه‌بندی، پیشی می‌گیرد. علاوه بر این، در مقایسه با رویکردهای یادگیری ماشینی هوشمند، رویکرد ارائه شده به طور مشابه عملکرد دلگرم کننده‌ای را ارائه می‌دهد.

۲- مطالب اصلی

در بخش ۲-۱ ادبیات موضوع و مفاهیم اولیه و در بخش ۲-۲ طرح مسئله و راه حل پیشنهادی مطرح شده است.

۲-۱- ادبیات موضوع و مفاهیم اولیه

در بخش ۲-۱-۱ یادگیری ماشین ، در بخش ۲-۱-۲ الگوریتم جنگل تصادفی و در بخش ۲-۱-۳ الگوریتم جایا مطرح شده است.

۲-۱-۱- یادگیری ماشین

یادگیری ماشین زمینه نسبتا جدیدی از هوش مصنوعی است که در حال حاضر دوران رشد و تکامل خود را می‌گذراند و یک زمینه تحقیقاتی بسیار فعال در علوم کامپیوتر است. علوم مختلفی از قبیل هوش مصنوعی، روانشناسی، فلسفه، تئوری اطلاعات، آمار و اطلاعات، تئوری کنترل با یادگیری ماشین در ارتباط هستند. یادگیری ماشین عبارت است از اینکه چگونه می‌توان برنامه‌ای نوشت که از طریق تجربه یادگیری کرده و عملکرد خود را بهتر کند[23].

در یادگیری ماشین با استفاده از تئوری اطلاعات مدل‌های ریاضی ساخته می‌شود که می‌تواند برای استنتاج استفاده شوند. تکنیک‌های یادگیری ماشین در شرایطی مناسب است که هیچ گونه دانش اولیه در مورد الگوهای داده‌ها وجود ندارد، به همین دلیل گاهی به این روش‌ها پایین به بالا می‌گویند[24].اهمیت طبقه بندی بیماران روزبه‌روز بیش‌تر می‌گردد. پس با اعمال فرآیند یادگیری ماشین، می‌توان طی کاوش در تعامل‌های ثبت شده، به استخراج الگوها پرداخت[25].

همچنین روش‌های یادگیری ماشین بر اساس ساختارهایشان به دو گروه تقسیم می‌شوند: موازی و ترتیبی. گروه موازی مجموعه‌ای از رویکردهای مختلف یادگیری ماشین است که هر یک به طور موازی یک مدل مستقل ایجاد می‌کنند. اما در ترتیبی به این صورت است که الگوریتم اول به یادگیری و تولید یک مدل می‌پردازد و سپس الگوریتم دوم می‌آموزد که مدل قبلی را اصلاح کند و بنابراین، تا عمق خاص خود ادامه می‌یابد.

۲-۱-۲- الگوریتم جنگل تصادفی

جنگل تصادفی یک روش یادگیری ترکیبی برای طبقه بندی داده‌ها می‌باشد، که از تعداد زیادی درخت تصمیم در زمان آموزش تشکیل شده است و خروجی این الگوریتم برای طبقه‌بندی مجموعه‌ای از داده‌ها مورد استفاده قرار می‌گیرد. جنگل‌های تصادفی برای درختان تصمیمی که در فرآیند آموزش دچار بیش برازش می‌شوند، مناسب هستند. همچنین عملکرد جنگل تصادفی معمولا بهتر از درخت تصمیم است، اما این بهبود عملکرد تا حدی به نوع داده هم بستگی دارد.

نحوه کارکردن الگوریتم جنگل تصادفی به شرح زیر است:

۱- ابتدا با انتخاب نمونه‌های تصادفی از یک مجموعه داده الگوریتم شروع می‌شود.

۲-این الگوریتم برای هر نمونه، یک درخت تصمیم گیری ایجاد می‌کند، سپس نتیجه پیش‌بینی از هر درخت تصمیم به دست می‌آید.

۳- رای گیری برای هر نتیجه پیش بینی شده انجام می‌گیرد.

۴- بیشترین نتیجه پیش‌بینی، به عنوان پیش‌بینی نهایی انتخاب می‌گردد.

در شکل(۱)، کد الگوریتم جنگل تصادفی نیز قابل مشاهده است.

```
Algorithm 1 Random Forest
Precondition: A training set S := (x1, y1), . . . , (xn, yn), features F, and number
    of trees in forest B.
function RandomForest(S , F)
    H ← ∅
    for i ∈ 1, . . . , B do
        S^(i) ← A bootstrap sample from S
        h_i ← RandomizedTreeLearn(S^(i), F)
        H ← H ∪ {h_i}
    end for
    return H
end function
function RandomizedTreeLearn(S , F)
    At each node:
        f ← very small subset of F
        Split on best feature in f
    return The learned tree
end function
```

شکل (۱): کد الگوریتم جنگل تصادفی[26]

۲-۱-۳- الگوریتم جایا

یکی از الگوریتم‌های فرا ابتکاری مناسبی که برای حل مسئله‌های بهینه سازی محدود و نامحدود استفاده می شود، الگوریتم بهینه سازی جایا است. الگوریتم جایا، در فرآیند جستجو سعی می‌کند با فرار از بدترین راه‌حل، با اجتناب از شکست، برای یافتن راه‌حل‌های بهینه به بهینه‌ترین راه حل ممکن نزدیک‌تر شود. الگوریتم جایا روشی ساده و جدید به منظور حل مسائل بهینه سازی می‌باشد. این الگوریتم بر این اصل استوار است که برای رسیدن به بهترین جواب باید راه‌حل‌های موجود در هر تکرار به سمت بهترین راه‌حل ارائه شده حرکت کرده و از بدترین راه حل در همان تکرار نیز دور شوند. از مزایای الگوریتم جایا می‌توان به سادگی، عدم نیاز به هیچ پارامتر کنترلی و سرعت آن اشاره کرد. به بیان دیگر الگوریتم جایا، الگوریتم فرا ابتکاری است که با استفاده از مناسب‌ترین مفاهیم تحریک می‌شود و با اجتناب از راه‌حل غیربهینه، به سمت راه‌حل‌های بهینه می‌رود.

در الگوریتم جایا تابع F(x) تابع هدف مسئله بوده و الگوریتم جایا قصد یافتن مینیمم(ماکزیمم) آن را دارد. در هر تکرار k، الگوریتم جایا دارای n متغیر تصمیم گیری و N_param راه‌حل پیشنهادی است.

مرحله‌های الگوریتم جایا به شریح زیر است:

۱- پارامترهای الگوریتم جایا مانند اندازه جمعیت N و حداکثر تعداد تکرارkmax باید در تمام مسائل بهینه سازی مقداردهی اولیه شوند. داده‌های تعریف شده باید از مجموعه داده‌های مسئله استخراج شوند و همچنین تابع هدف و راه حل پیش بینی شده معمولاً در این مرحله بایستی مشخص گردد. تعریف داده‌های ورودی و مشخص شدن حد بالا و حد پایین مجموعه داده‌ها باید در این مرحله صورت پذیرد.

۲- در این مرحله تشکیل جمعیت اولیه صورت می‌پذیرد. بر اساس داده‌های موجود در مرحله اول، یک جمعیت اولیه به صورت تصادفی به گونه‌ای که محدودیت‌ها رعایت شود، تشکیل می‌گردد. که جامعه تشکیل شده شامل N راه d بعدی است.

۳- محدودیت‌های مسئله در این مرحله مورد بررسی قرار می‌گیرد و بهترین و بدترین راه‌حل‌ها در این مرحله مشخص می‌شود.

۴- الگوریتم جایا به صورت تصادفی هر متغیر تصمیم موجود در هر راه حل را در هر تکرار تغییر می‌دهد. به عبارت دیگر تابع هدف مسئله مشخص می‌شود.

۵- برای استفاده از الگوریتم مورد نظر می‌بایست بدترین و بهترین جواب مشخص شود. بنابراین هدف الگوریتم جایا با استفاده از شرایط فوق تحقق می‌یابد.

۶- الگوریتم جایا روی مجموعه جواب‌های موجود باید اعمال گردد تا در مجموعه جواب‌های محاسبه شده بهبود حاصل گردد.

۷- پس از به‌روز رسانی مجدد جواب‌ها، شماره تکرار یک واحد افزایش می‌یابد و شرط همگرایی بررسی می شود. اگر شرط همگرایی در الگوریتم برآورده شده بود این قاعده پایان می‌یابد و بهترین جواب چاپ می‌شود، در غیر این صورت مجدد مرحله ۵ مورد بررسی قرار می‌گیرد.

در شکل(۲)، کد الگوریتم جایا نیز مشاهده است.

```
Algorithm 1. The Jaya algorithm pseudocode
Initialize input Jaya parameters: r1, r2, X_j^UB, X_j^LB, N, d, k_max
---------- Initializing the initial population----------
For  i = 1:1:N  Do /* Each solution */
     For j = 1:1:d   Do /* Each variable*/
Calculate: X_i,j = X_j^LB + (X_j^UB − X_j^LB) × U(0.1)
Calculate:  f(X_i)
-------------- Define the best and worst solutions in the population-------------
itr (k)=1
while (k < k_max) do
     Determine the best solution in the population (X_best)
     Determine the best solution in the population (X_worst)
For  i = 1:1:N  Do
     For j = 1:1:d   Do
          Set r1 ∈ (0,1)
          Set r1 ∈ (0,1)
Calculate: X'_i,j = X_i,j + r1 × (X_best,j − |X_i,j|) − r2 × (X_worst,j − |X_i,j|)
     End For
      If f(X'_i) ≤ f(X_i)   then
          X'_i = X_i   (update process)
     End If
   End For
   k = k + 1
End while
```

شکل(۲): کد الگوریتم جایا [27]

۲-۲- طرح مساله و راه‌حل پشنهادی

زبان پایتون با مجموعه‌ای از کتابخانه‌های علم داده و یادگیری ماشین عرضه می‌شود که می‌توان از آن‌ها برای پیش‌بینی بر اساس ویژگی یا ویژگی‌های مختلف یک مجموعه داده استفاده کرد. این زبان یکی از پر کاربردترین زبان‌های برنامه‌نویسی برای تجزیه و تحلیل داده‌های مختلف است که دارای کتابخانه‌های مفید و عملکرد داخلی فراوانی است. این مقاله به طبقه بندی داده‌ها برای تشخیص سرطان سینه پپرداخته شده است. کتابخانه SCIKIT-LEARN پایتون یکی از این ابزارها می‌باشد و در این مقاله از این کتابخانه برای طبقه بندی سرطان سینه استفاده شده است.

۲-۲-۱- فاز اول (وارد کردن کتابخانه‌ها)

در این فاز کتابخانه‌های Numpy، pandas، Matplotlib موردنیاز است.

۲-۲-۲- فاز دوم (انتخاب ویژگی)

بعد از جمع آوری مجموعه داده‌های مورد نیاز بر اساس ویژگی‌های کلیدی و نوع صـنعت، در این فاز به انجام پیش پردازش‌های اولیه از جمله حذف نویزها و داده‌های فراموش شـده و تمیز کردن مجموعه داده‌ها پرداخته می‌شود.

۲-۲-۳- فاز سوم (تبدیل داده‌ها)

الگوریتم‌های یادگیری ماشـین با داده‌های عددی کارایی بسـیار بهتری را ارائه می‌دهند. درمجموع داده‌های مورد مطالعه، داده‌هایی که در قالب متنی هسـتند باید به سـتون‌های عددی تبدیل شـوند. یکی از راه‌های تبدیل سـتون‌های دسته‌بندی به ستون‌های عددی این است که می‌توان هر دسته را با یک عدد جایگزین کرد. به‌عنوان‌مثال، در ستون جنسیت، زن را می‌توان با مقدار ۰ و مرد را با مقدار ۱ جایگزین کرد، یا بالعکس. این برای سـتون‌هایی با دودسـته کار می‌کند. یکی از راه‌های تبدیل سـتون‌های دسـته‌بندی به سـتون‌های عددی، استفاده از one-hot encoding است. در این فرآیند، هر کدام از دسته‌ها به عنوان یک ستون مجزا در نظر گرفته می‌شوند و در هر ستون از ۱برای تعیین اینکه دسـته برای ردیف فعلی وجود دارد اسـتفاده شـده اسـت و در غیر این صورت از ۰ به عنوان جایگزین استفاده شده است.

۲-۲-۴- فاز چهارم (پیش پردازش داده‌ها)

در این مرحله از پیش‌بینی، داده‌ها آماده است و می‌توان مدل یادگیری ماشینی را برای آموزش آن در نظر بگیریم. در ابتدا باید متغیری که پیش‌بینی می‌شـود از مجموعه داده‌ها جدا گردد. این فرآیند شامل تمام ستون‌ها به‌جز ستونی است که باید پیش‌بینی شـود. برای پیاده سـازی مدل، داده‌ها به یک مجموعه آموزشی و آزمایشی تقسیم شده است. مجموعه آموزشی حاوی داده‌هایی است که برای آموزش مدل یادگیری ماشـین اسـتفاده می‌شـود و مجموعه آزمایش برای ارزیابی اینکه مدل معرفی شده چقدر خوب است، استفاده خواهد شد. که ۲۰٪ از داده‌ها را برای مجموعه آزمایش و از ۸۰٪ باقیمانده برای مجموعه آموزشی استفاده شده است.

۲-۲-۵- فاز پنجم (آموزش الگوریتم یادگیری ماشین)

در این مرحله از یک الگوریتم یادگیری ماشینی برای پیش بینی استفاده شده است، که الگوها یا روندها را در داده‌های آموزشی شناسایی می‌کند. این مرحله به آموزش الگوریتم معروف است. ویژگی‌ها و خروجی صحیح به الگوریتم داده می‌شـود و بر اسـاس آن داده‌ها، الگوریتم یاد خواهد گرفت که ارتباط بین ویژگی‌ها و خروجی‌ها را بیابد. پس از آموزش الگوریتم، از آن برای پیش‌بینی داده‌های جدید استفاده شده است.

چندین الگوریتم یادگیری ماشـین وجود دارد که می‌توان از آن‌ها برای انجام چنین پیش‌بینی هایی اسـتفاده کرد. در این مقاله از الگوریتم جنگل تصادفی استفاده شده است، زیرا این الگوریتم به عنوان یک از ساده‌ترین و قدرتمندترین الگوریتم‌ها برای مسـائل طبقه‌بندی اسـت که اغلب اوقات نتایج بسـیار خوبی را حتی بدون تنظیم فرا پارامترهای[6] آن، فراهم می‌کند. این الگوریتم هم برای دسته‌بندی[7] و هم رگرسیون[8] به عنوان یکی از پرکاربردترین الگوریتم‌های داده کاوی محسوب می‌شود. جنگل تصادفی، به عنوان یک مدل داده کاوی نظارت شـده[9] یاد می‌گیرد که در فاز آموزش یا برازش مدل، داده‌ها را به خروجی‌ها نگاشت کند. در طول آموزش، داده‌های آموزشی به مدل داده می‌شوند که مرتبط با دامنه مسئله هستند و خروجی مقدار صحیحی است که مدل باید بیاموزد تا بتواند پیش‌بینی کند. بنابراین مدل، رابطه‌های میان داده‌ها و مقادیری که کاربر می‌خواهد پیش‌بینی کند را می‌آموزد. که در پیشـینه پژوهش در رابطه با الگوریتم جنگل تصـادفی و الگوریتم جایا که برای بهینه سازی طبقه بندی استفاده می‌شود، توضیحاتی ارائه شده است.

با توجه به الگوریتم جنگل تصـادفی نتیجه‌های حاصـل از ارزیابی رضـایت‌بخش نبود، بنابراین روش پیشـنهادی ترکیبی از الگوریتم جنگل تصادفی و الگوریتم جایا می‌باشد، که مرحله‌های آن به شرح زیر است:

مرحله اول مقداردهی اولیه: در این مرحله یک تابع توزیع به صـورت تصـادفی برای تعیین پارامترهای الگوریتم Jaya مانند اندازه جمعیت N که بیانگر تعداد راه‌حل‌های ممکن اسـت و حداکثر تعداد تکرار الگوریتم که با kmax نمایش داده می‌شـود، تعریف می‌گردد. همچنین مجموعه پارامترهای الگوریتم جنگل تصادفی تعریف می‌گردد.

مرحله دوم مشخص سازی تابع هدف مناسب: بهترین مقدار برای تابع هدف در مرحله i^{th} به ازای داده ورودی x_i به صـورت f(i) تعریف می‌گردد. به دلیل اینکه الگوریتم جایا هر کدام از راه‌حل‌های انتخابی x_i را برای طراحی درست طبقه‌بندی الگوریتم جنگل تصادفی با جایا بررسی می‌کند، انتخاب صحیح تابع هدف مناسب جز ضروریات است، که m تعداد نمونه‌های مشاهده شده در فرآیند تست است و Error(p) برای بدست آوردن میزان خطا مقدار واقعی از مقدار براورد شده مورد استفاده قرار می‌گیرد. در این مقاله به دقت طبقه بندی و تعداد ویژگی‌های انتخاب شـده، به عنوان دو پارامتر اساسی در تعریف تابع هدف مناسب می‌توان اشاره کرد. بنابراین انتخاب راه‌حل با بالاترین دقت تصـمیم‌گیری و کم‌ترین تعداد ابعاد، به عنوان دو پارامتر برای برآورد بهترین تابع هدف تعریف می‌شـود و میزان مناسـب بودن تابع هدف به کمک فرمول (۱) ارزیابی می‌گردد:

$$finess(i) = \frac{\sum_{i=1}^{m}(Error(p))^2}{m}.p == 1.2 \dots .m \quad (۱)$$

مرحله سوم مشخص سازی بهترین و بدترین کاندیدها: بعد از ارزیابی تابع هدف مناسب، بهترین و بدترین کاندیدها که به ترتیب با X_{best} و X_{worst} تعریف می‌گردد، که این مقدارها از طریق ماکزیمم و مینیمم مقداری که تابع هدف محاسبه کرده است، تعریف شده است.

مرحله چهارم بهبود راه حل: در این مرحلـه بهترین و بـدترین راه‌حل‌های بدست آمده از مرحله قبل ذخیره می‌شوند و سایر راه‌حل‌های بدست آمده به روز رسانی می‌شوند و مجدد تابع هدف مناسب به وسیله مقادیر جدید به دست آمده تخمین زده می‌شود.

مرحله پنجم بررسـی راه حل: اگر مقدار تابع هدف بدسـت آمده در مرحله جدید بهتر از مرحله قبلی باشد، مقدار جدید بدست آمده جایگزین مقدار قبلی می‌شود، در غیر این صورت الگوریتم برای بار دیگر تکرار می‌شود.

مرحله ششم: کلیه مرحله‌های ۲ تا ۵ مادامی که تعداد دفعه‌های تکرار الگوریتم برابر kmax شود، انجام می‌گردد.

۲-۲-۶- فاز ششم (ارزیابی مدل)

بعد از آموزش، ارزیابی صورت می‌گیرد تا بررسی شود که الگوریتم انتخابی در پیش‌بینی سرطان سینه چقدر خوب عمل می‌کند. برای ارزیابی عملکرد یک الگوریتم طبقه‌بندی، متداول‌ترین معیارهای مورداستفاده عبارتند از: دقت؛[1] یادآوری[1] و صحت.[2] در معیار دقت حداکثر مقدار این معیار یک و یا ۱۰۰ در صد است و حداقل مقدار آن صفر است و هرچه مواردی که برنامه به غلط پیش‌بینی کرده است که به آن مثبت کاذب[3] می‌گویند نسبت به پیش بینی‌های درست که به آن مثبت واقعی[4] می‌گویند بیشتر باشد، مقدار دقت کمتر خواهد شد و طبق فرمول (۲) قابل محاسبه است.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \qquad (۲)$$

که در آن TP مثبت واقعی و FP مثبت کاذب است.

در معیار یادآوری حداکثر مقدار این معیار یک و یا ۱۰۰ درصد است و حداقل مقدار آن صفر است و هرچه مواردی که برنامه پیش بینی نکرده است که به آن منفی کاذب[5] می‌گویند نسبت به پیش بینی‌های درست که به آن مثبت واقعی می‌گویند بیشتر باشد، مقدار یادآوری کمتر خواهد شد و طبق فرمول (۳) قابل محاسبه است.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \qquad (۳)$$

که در آن TP مثبت واقعی و FN منفی کاذب است.

در معیار صحت تعداد مواردی که درست پیش‌بینی شده است که به آن مثبت واقعی می‌گویند بر تعداد کل پیش‌بینی‌هایی که انجام شده است، تقسیم می‌گردد و طبق فرمول (۴) قابل محاسبه است.

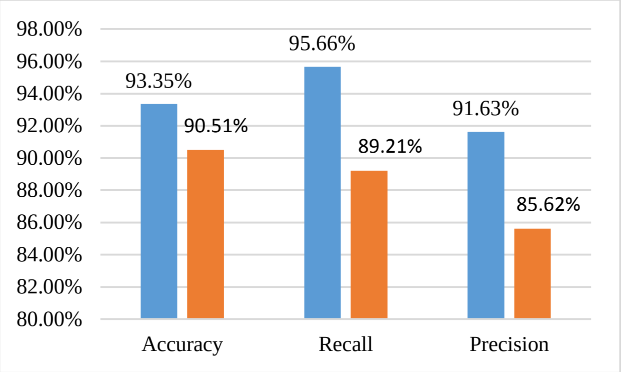
$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \qquad (٤)$$

که در آن TP مثبت واقعی، TN منفی واقعی، FP[6] مثبت کاذب و FN منفی کاذب است.

در کتابخانه scikit-learn پایتون، می‌توانیم از تابع های داخلی برای یافتن همه این مقدارها استفاده کنیم و اگر مدل کارایی لازم را نداشت، دوباره به فاز پنج بر می‌گردیم و تنظیمات پارامترهای اولیه بر روی مدل قبلی برای رسیدن به نتیجه مطلوب تغییر پیدا خواهد کرد. نتایج حاصل از مقاله به ترتیب در جدول(۱) و شکل(۳) آورده شده است.

جدول(۱): نتایج بدست آمده استفاده الگوریتم جنگل تصادفی و روش پیشنهادی

	Accuracy	Recall	Precision
جنگل تصادفی	٪ ۹۰,۵۱	٪ ۸۹,۲۱	٪ ۸۵,۶۲
مدل پیشنهادی	٪ ۹۳,۳۵	٪ ۹۵,۶۶	٪ ۹۱,۴۱



شکل (۳): مقایسه نتایج بدست آمده الگوریتم جنگل تصادفی و روش پیشنهادی

همانطور که قابل ملاحظه است، نتیجه‌های حاصل از مدل پیشنهادی در مقاله به ترتیب در معیارهای Precision، Recall و Accuracy حدودا به میزان ۳ ٪، ۶ ٪و ۷ ٪ از الگوریتم جنگل تصادفی بالاتر می‌باشد.

۳- نتیجه

در این مقاله به منظور تشخیص سرطان سینه از یک مدل دو بخشی استفاده شده است. روش پیشنهادی، ترکیبی از الگوریتم جنگل تصادفی و الگوریتم جایا می‌باشد. در بخش اول از الگوریتم جنگل تصادفی برای جستجو تعداد ویژگی‌های بهینه استفاده شده است. با جستجو بهینه ویژگی‌های موثر انتخاب شده و به عنوان ورودی انتخاب می شود. در بخش دوم نیز از الگوریتم جایا به منظور افزایش دقت طبقه بندی در داده‌های سرطان سینه با تنظیم وزن‌ها استفاده می شود. عملکرد راه‌حل پیشنهادی نقش موثری در بهبود راه‌حل‌های اخیر و ایجاد وزن‌های بهینه داشته است. یکی از دلایلی که وزن‌ها در بخش دوم بهینه سازی می‌شوند، کاهش پیچیدگی زمانی الگوریتم می‌باشد. به طورکلی در روش های انتخاب ویژگی، نیاز به محاسبه به متوالی برازندگی ویژگی‌های انتخاب شده وجود دارد، لذا استفاده از روش‌هایی که کمترین پیچیدگی را داشته باشند، در اولویت است و این امر امکان اجرای الگوریتم را روی پایگاه‌های داده بسیار بزرگ فراهم می سازد و می‌تواند برای کارهای آتی مورد استفاده قرار گیرد.

مراجع

[1] Randi G et al., “*Estimated cancer incidence and mortality in europe for the year 2020*”, Eur J Public Health 30(5): ckaa166.1348,https://doi.org/10.1093/eurpub/ckaa166.1348, 2020.

[2] Ferlay J et al., “*Cancer incidence and mortality patterns in Europe: estimates for 40 countries and 25 major cancers in 2018”*, Eur J Cancer 103:356–387, https://doi.org/10.1016/j.ejca, 2018.

[3] Mehedi IM et al., “*Optimal feature selection using modified cuckoo search for classification of power quality disturbances*”, Appl Soft Comput https://doi.org/10.1016/j.asoc,107897, 2021.

[4] Bruni D, Angell HK, Galon J., “*The immune contexture and immunoscore in cancer prognosis and therapeutic efficacy*”, Nat Rev Cancer 20(11):662–680. https://doi.org/10.1038/s41568-020- 0285-7, 2020.

[5] Yala A, Lehman C, Schuster T, Portnoi T, Barzilay R., “*A deep learning mammography-based model for improved*

[18] Obaid OI, Mohammed MA, Mostafa A, Taha F., "*Evaluating the performance of machine learning techniques in the classification of Wisconsin Breast Cancer Int*", J Eng Technol 7(436):160–166. https://doi.org/10.14419/ijet.v7i4.36.23737, 2018.

[19] AL Fayez F, El-Soud MWA, Gaber T., "*Thermogram breast cancer detection: a comparative study of two machine learning techniques*", Appl Sci 10(2):2020. https://doi.org/10.3390/ app10020551, 2020.

[20] Khan S, Islam N, Jan Z, Ud Din I, Rodrigues JJPC., "*A novel deep learning based framework for the detection and classification of breast cancer using transfer learning. Pattern Recognition*", Lett 125:1–6. https://doi.org/10.1016/j.patrec.2019.03.022 Neural Computing and Applications 123, 2019.

[21] Wang Y, Yang F, Zhang J, Wang H, Yue X, Liu S., "*Application of artificial intelligence based on deep learning in breast cancer screening and imaging diagnosis*", Neural Comput Appl 33(15):9637–9647. https://doi.org/10.1007/s00521-021-05728-x, 2020.

[22] Hamouda S, Hassan A, Wahed ME, Ail M, Farouk O., "*Tuning to optimize SVM approach for breast cancer diagnosis with blood analysis data*", SSRN Electron J. https://doi.org/10. 2139/ssrn.3537067, 2020.

[23] Zhang Y-D, Satapathy SC, Guttery DS, Go'rriz JM, Wang S-H., "*Improved breast cancer classification through combining graph convolutional network and convolutional neural network*", Information Process Manag 58(2):102439. https://doi.org/10. 1016/j.ipm.2021.102439, 2021.

[24] Almasi ON, Khooban MH., "*A parsimonious SVM model selection criterion for classification of real-world data sets via an adaptive population-based algorithm*", Neural Comput Appl 30(11):3421–3429. https://doi.org/10.1007/s00521-017-2930-y, 2018.

[25] Chen Y, Fan L, Bai Y, Feng J, Sareh P., "*Assigning mountain-valley fold lines of flat-foldable origami patterns based on graph theory and mixed-integer linear programming*", Computer Struct 239:106328. https://doi.org/10.1016/j.compstruc. 106328, 2020.

[26] Assiri, A.S.; Nazir, S.; Velastin, S.A. "*Breast Tumor Classification Using an Ensemble Machine Learning Method*", J. Imaging, 6, 39, 2020.

[27] Rao R., "Jaya*: A simple and new optimization algorithm for solving constrained and unconstrained optimization problems*", Int J Ind Eng Comput 7(1):19–34. https://doi.org/10.5267/j.ijiec.8.004, 2015.

breast cancer risk prediction", Radiology 292(1):60–66. https://doi.org/ 10.1148/radiol,2019182716, 2019.

[6] Kim H-E et al., "*Changes in cancer detection and false positive recall in mammography using artificial intelligence: a retrospective, multileader study*", Lancet Digital Health 2(3): e138–e148. https://doi.org/10.1016/S2589-7500(20)30003-0, 2020.

[7] Sloan RJGV, Cohen R, Elder YC., "*Deep learning in ultrasound imaging*", Proceedings of the IEEE 108(1): 11–9 https://doi.org/10.1109/JPROC.2019.2932116, 2020.

[8] Soo MS, Shelby RA, Johnson KS.," *Optimizing the patient experience during breast biopsy*", J Breast Image 1(2):131–138.https://doi.org/10.1093/jbi/wbz001, 2019.

[9] Alrifaey M et al., "*Hybrid deep learning model for fault detection and classification of grid-connected photovoltaic system*", IEEE Access. https://doi.org/10.1109/ACCESS.2022. 3140287, 2022.

[10] Huang M-W, Chen C-W, Lin W-C, Ke S-W, Tsai C-F., "*SVM and SVM ensembles in breast cancer prediction*", PloS ONE 12(1): e0161501. https://doi.org/10.1371/journal.pone. 0161501, 2017.

[11] Abdar M, Makarenkov V.," *Ensemble learning classifier for an accurate diagnosis of breast cancer*", Measurement 146:557–570. https://doi.org/10.1016/j.measure ment.2019.05.022, 2019.

[12] Vijayarajeswari R, Parthasarathy P, Vivekanandan S, Basha AA., "*Classification of mammogram for early detection of breast cancer using SVM classifier and Hough transform*", Measurement 146:800–805. https://doi.org/10.1016/j.measurement,05.083, 2019.

[13] Saturi R, Phani KVS and Chand PPP., "*A frame work to detect breast cancer using KNN and SVM*", (in en), Eur J Mol Clinic Med 8(3): 1432–1438 [Online]. Available: https://ejmcm.com/article_9974.html, 2021.

[14] Kaveh A, Bakhshpoori T., "*Metaheuristics outlines: MATLAB codes and examples*", Springer, New York, 2019.

[15] Huang S, Yang J, Fong S, Zhao Q., "*Artificial intelligence in cancer diagnosis and prognosis*": opportunities and challenges, Cancer Lett 471:61–71. https://doi.org/10.1016/j.canlet.2020.12.007, 2020.

[16] Amrane M, Oukid S, Gagaoua I, EnsarI˙ T ., "*Breast cancer classification using machine learning, in 2018 Electric Electronics*", Computer Science, Biomedical Engineerings Meeting (EBBT), 18–19 pp 1–4, https://doi.org/10.1109/EBBT.2018, 8391453, 2018.

[17] Ganggayah MD, Taib NA, Har YC, Lio P, Dhillon SK., "*Predicting factors for survival of breast cancer patients using machine learning techniques*", BMC Med Informatics Decision Mak 19(1):48. https://doi.org/10.1186/s12911-019-0801-4, 2019.

پانویس‌ها

1 Random forest
2 ECSI
3 Mammography screening
4 Support Vector Machine
5 Jaya
6 Hyper parameter
7 Classifier
8 Regression
9 Supervised
10 Precision
11 Recall
12 Accuracy
13 False Positive
14 True Positive
15 False Negative
16 True Negative

بررسی استفاده از برداشتگرهای انرژی برای شبکه‌های حسگر بی‌سیم

علی جوادی[۱]، امیرمهدی حسینی‌منزّه[۲]

[۱] دانشجوی کارشناسی ارشد، دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت ایران، تهران، ایران
javadi_ali@comp.iust.ac.ir

[۲] استادیار، دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت ایران، تهران، ایران
monazzah@iust.ac.ir

چکیده

حفظ انرژی لازم برای انجام وظایف در شبکه‌های حسگر بی‌سیم در کاربردهای شبکه‌های حسگر بی‌سیم یکی از مهم‌ترین مسائل حوزۀ قابلیت اطمینان در تشکیل شبکه‌ها و صحت عملکرد این شبکه‌ها است. طبق مطالعات انجام شده بخش مسیریابی و ارسال و دریافت بسته‌ها تا حدود ۵۵٪ انرژی مصرفی هر دستگاه را به خود اختصاص می‌دهد. به‌منظور انجام مسیریابی در شبکه‌های کم‌توان و پراتلاف پروتکل RPL به‌عنوان یک استاندارد ارائه شده است. با توجه به نقش مهم توابع هدف در RPL برای مسیریابی، در بسیاری از پژوهش‌های پیشین از معیار انرژی مصرفی در توابع هدف به‌منظور بهینه‌سازی انرژی استفاده شده است. در مقابل، یکی از راه‌های تأمین انرژی و افزایش طول عمر باتری استفاده از برداشتگرهای انرژی در گره‌های شبکه‌های حسگر بی‌سیم است. در این پژوهش ما در گام نخست با بررسی پژوهش‌های پیشین در حوزه‌ی برداشتگرهای انرژی به‌صورت تحلیلی چالش‌ها و امیدهای فناوری‌های مختلف بررسی می‌کنیم. سپس با درنظرگرفتن یک روش مناسب، یک بستر عملی برای بررسی میزان انرژی برداشت شده از محیط و درک چالش‌های موجود در این مسیر معرفی، و نتایج ارزیابی‌ها را مورد تحلیل و بررسی قرار خواهیم داد.

کلمات کلیدی

شبکه حسگر بی‌سیم، انرژی مصرفی، برداشتگر انرژی، مسیریابی، RPL

۱- مقدمه

شبکه حسگر بی‌سیم (WSN[۱]) متشکل از چندین گره حسگر که به‌منظور نظارت بر پارامترهای مختلف طراحی شده‌اند. شبکه‌های حسگر بی‌سیم معمولاً در کاربردهایی مانند تشخیص بلایای طبیعی، سیستم مراقبت‌های بهداشتی، سیستم کنترل ترافیک، سیستم امنیتی ساختمان و غیره استفاده می‌شوند. این دستگاه‌ها به طور عمده برای تأمین انرژی خود از باتری بهره می‌برند.

باتوجه‌به محدودیت در میزان انرژی ذخیره شده در باتری‌ها، لازم است باتری‌ها به‌صورت دوره‌ای تعویض یا شارژ شوند تا از طول عمر گره سنسور اطمینان حاصل شود. جایگزینی باتری‌ها به دلیل استقرار بسیاری از گره‌های حسگر در محل‌هایی با مشکلات دسترسی، ممکن است یک وظیفه سنگین و یا حتی غیرممکن باشد [۱].

اخیراً، برداشت انرژی از محیط‌زیست به یکی از راه‌حل‌های ممکن برای افزایش طول عمر گره‌های حسگر بی‌سیم تبدیل شده است و تحقیقات گسترده‌ای در این حوزه در حال انجام است. در این پژوهش ما به دنبال کاهش انرژی مصرفی گره‌ها در شبکه‌های حسگر بی‌سیم هستیم و برای محقق شدن این مهم از معیار میزان انرژی برداشت شده استفاده می‌کنیم.

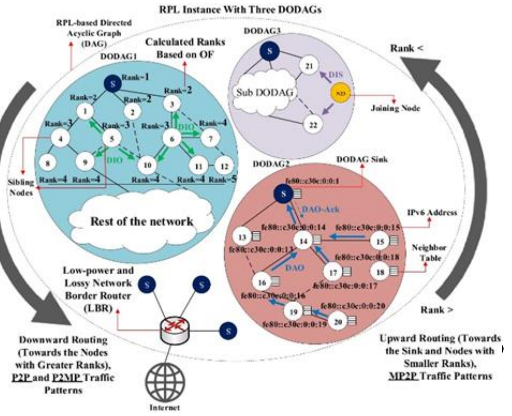
۲- پیش‌نیازها

امروزه گره‌های حسگر بیسیم در کاربردهای مختلفی از جمله برنامه‌های کاربردی سلامت و... کاربرد فراوانی دارند؛ از چالش‌های این نوع شبکه‌ها می‌توان به استفاده دستگاه‌ها از باتری با ظرفیت محدود و برد انتقال کوتاه اشاره کرد بنابراین به علت کوتاه بودن عمر دستگاه‌ها طول عمر شبکه نیز کاهش می‌یابد.

گره‌های شبکه حسگر بی‌سیم باید اطلاعات جمع‌آوری‌شده خود را با همکاری یکدیگر به مقصد اصلی برسانند، ممکن است چندین گره شبکه خارج از پوشش ارتباطی گره دریافت‌کننده اصلی (گره ریشه) باشند برای این موضوع نیاز به استانداردی جهت برقراری ارتباط میان گره‌ها و مسیریابی درست بسته‌ها وجود دارد.

از دیگر چالش‌های شبکه‌های حسگر بی‌سیم می‌توان به محدودیت منابع و قرارگیری گره‌ها در محیط‌هایی با میزان نویز بالا اشاره کرد؛ این محدودیت‌ها منجر به طراحی شبکه کم‌مصرف و پراتلاف (LLN[۲]) می‌شود که به طراحی پروتکل خاصی نیاز دارد که بر کارایی و مسیریابی قوی داده‌ها متمرکز باشد؛ برای پاسخگویی به چنین نیازهایی، IETF[۳] یک پروتکل مسیریابی استاندارد شده برای شبکه‌های LoWPAN۶[۴] به نام RPL[۵] (پروتکل مسیریابی IPv6 برای شبکه‌های کم‌توان و پراتلاف) توسعه داد. RPL برای مدیریت محدودیت‌های دستگاه‌هایی با منابع محدود طراحی شده است و می‌تواند با تغییرات گره و معیارهای ارتباطی در طول زمان سازگار شود. در LLNs، گره‌ها می‌توانند یک یا چند والد داشته باشند که مسیری به سمت ریشه را تشکیل می‌دهند [۲].

در RPL، انتخاب بهترین والدین بر اساس یک تابع هدف است که از مجموعه‌ای از معیارها استفاده می‌کند. این معیارها می‌توانند معیارهای گره محور مانند انرژی، فاصله تا گره مقصد یا معیارهای مبتنی بر پیوند مانند تعداد انتقال مورد انتظار (ETX[۶]) و سطح کیفیت پیوند (LQL[۷]) باشند (شکل ۱). بااین‌حال، استفاده از یک معیار به‌صورت منفرد همان‌طور که در تابع هدف استاندارد MRHOF[۸] و OF0[۹] پیاده‌سازی شده است که به ترتیب با معیارهای ETX و HopCount عمل می‌کنند با محدودیت‌های زیادی همراه است [۲] زیرا زمانی که برای تصمیم‌گیری صرفاً از یک معیار استفاده



شکل ۱- ساختارRPL در شبکه اینترنت اشیاء [4]

شود، شبکه از یک جنبه بهبود پیدا می‌کند و ممکن از سایر جنبه‌ها دچار مشکل شود.

اگرچه یک الگوریتم مبتنی بر معیار ممکن است با درنظرگرفتن کمترین مقدار تعداد انتقال مورد انتظار، بهترین والد را برای یک گره انتخاب کند، این امر باعث مصرف انرژی بیشتر در گره می‌شود که برای گره‌ها با میزان کم‌انرژی مناسب نیست [۳].

برای تشکیل یک شبکه DODAG[۱۰] در شبکه‌های مبتنی بر RPL سه نوع پیام میان گره‌های شبکه جابه‌جا می‌شود که عبارت‌اند از DIO[۱۱]، DIS[۱۲] و DAO[۱۳] که در ادامه به‌صورت کلی به بررسی هر یک از آنها و روند تشکیل یک شبکه DODAG می‌پردازیم (شکل ۱).

برای تشکیل شبکه ابتدا گره ریشه بسته‌های DIO را به گره‌های همسایه خود ارسال می‌کند و گره‌های همسایه نیز در صورت دریافت پیام، ACK را به گره ریشه ارسال می‌کنند، سپس گره‌هایی که به‌تازگی به ریشه متصل شده‌اند نیز پیام DIO را برای همسایگان خویش ارسال می‌کنند تا شبکه به‌مرور تکمیل شود و تمام گره‌ها در محدوده پوشش رادیویی به گره ریشه متصل شوند.

برای افزایش کارایی این سیستم شبکه مسیریابی خود را در بازه‌های مشخص زمانی با ارسال مجدد پیام DIO به‌روزرسانی می‌کند. زمان این به‌روزرسانی توسط زمان‌بندی با نام Trickle Timer مدیریت می‌شود.

دیگر پیام مورداستفاده در شبکه پیام DAO است، این پیام را فرزندان به والدین خود ارسال می‌کنند و در آن اطلاعات مربوط به خود را به والدین منتقل می‌کنند. درصورتی‌که یک گره که به‌تازگی به شبکه اضافه شده بخواهد به شبکه DODAG متصل شود، پیام DIS را به گره‌های اطراف خود ارسال می‌کند و همسایگان پس از دریافت این پیام برای گره جدید پیام DIO را ارسال می‌کنند و گره تازه‌وارد پس از دریافت چند پیام DIO بهترین والد را برای خود انتخاب می‌کند. درصورتی‌که گره جدید پیام DIO را دریافت نکند مجدد در بازه‌های زمانی مشخص پیام DIS را ارسال می‌کند تا بتواند به شبکه متصل گردد.

۳- برداشتگرها انرژی و شبکه‌های حسگر بیسیم

از راهکارهای اصلی افزایش طول عمر باتری دستگاه‌های شبکه‌های حسگر بیسیم استفاده از برداشتگرهای انرژی از محیط است، امروزه فناوری‌های

شکل ۲- برداشتگر انرژی خورشیدی

شکل ۳- برداشتگر انرژی گرمایی شرکت Evergen [6]

برداشت انرژی مختلفی در دسترس است که در ادامه برخی از منابع بالقوه تولید انرژی را بررسی می‌کنیم.

۳-۱- برداشتگر انرژی خورشیدی

فناوری فتوولتائیک نسبت به سایر فناوری‌های برداشت انرژی از محیط دارای مزایای بسیاری از جمله چگالی بالای توان در آن است که استفاده از آن را در محیط‌هایی که دسترسی به نور خورشید وجود دارد توجیه می‌کند [۱] ولی استفاده از پنل خورشیدی در محیط‌های سربسته که دسترسی به نور آفتاب میسر نیست انتخاب بهینه‌ای نیست (شکل ۲).

۳-۲- برداشتگرهای انرژی گرمایی (ترموالکتریک)

این برداشتگرها به دلیل عدم استفاده از فرآورده‌های شیمیایی در آنها دوست‌دار محیط‌زیست محسوب می‌شوند، به دلیل نداشتن ساختار مکانیکی و قطعات متحرک، بی‌صدا عمل می‌کنند و می‌توان آنها را بر روی بسیاری از انواع بسترها مانند سیلیکون، پلیمرها و سرامیک‌ها ساخت. علاوه بر این، دستگاه‌های TEG مستقل از موقعیت هستند، طول عمر طولانی دارند و برای استفاده در دستگاه‌های حجیم و انعطاف‌پذیر مناسب هستند (شکل ۳) [۵].

۳-۳- برداشتگرهای انرژی تشعشعات محیطی

برداشت انرژی از فرستنده‌های رادیویی به نظر شیوه‌ای مناسب و کاربردی در همه نقاط می‌رسد؛ ولی برای دریافت توان در سطحی مفید از این منبع، به یک منطقه بزرگ جمع‌آوری یا نزدیک بودن به منبع انرژی بی‌سیم تابشی نیاز داریم، همان‌طور که در پژوهش [۷] به کمک دستگاه تحلیل‌کننده طیف فرکانسی[۱۴] در چند محیط واقعی بررسی انجام گرفت، هیچ یک از فرکانس‌های موجود در فضا، توان کافی برای فعال‌کردن برداشتگر انرژی رادیویی را نداشتند و در آن تحقیق برای ادامه کار خود از فرستنده توان در نزدیکی گره‌ها استفاده نمودند (شکل ۴).

۳-۴- برداشتگر انرژی جریان سیال

از میان برداشتگرهای انرژی سیال برداشتگر جریان هوا به دلیل دردسترس بودن از اهمیت ویژه‌ای برخوردار است، جریان هوا را می‌توان با فناوری‌های مختلف توربین جمع‌آوری کرد. توربین‌های بادی و سیستم‌های انرژی بادی هوابرد (AWES[۵]) جریان هوا را استخراج می‌کنند. برای مثال، میکرو ژنراتور انحصاری Windbeam شرکت Zephyr Energy، انرژی را از جریان هوا برای شارژ باتری‌ها و تأمین انرژی دستگاه‌های الکترونیکی جذب می‌کند. طراحی جدید Windbeam (شکل ۵) به آن اجازه می‌دهد تا در سرعت‌های باد کمتر از ۲ مایل در ساعت بی‌صدا عمل کند. در ادامه در جدول ۱ انواع روش‌های برداشت انرژی و میزان کارایی هر یک را مشاهده می‌کنید. در میان انواع برداشتگرهای محیطی، برداشت انرژی خورشیدی بیشترین چگالی توان را در مقایسه با سایر منابع فراهم می‌کند. ازآنجایی‌که گره حسگر بی‌سیم به توان الکتریکی در حد میلی‌وات نیاز دارد، انرژی خورشیدی به‌عنوان مؤثرترین انتخاب برای کاربردهایی که به نور خورشید دسترسی دارند در نظر گرفته می‌شود.

۴- پژوهش‌های پیشین

پیش‌ازاین پژوهش‌هایی که در آنها از تابع هدف آگاه به برداشتگر انرژی استفاده شده باشد انجام شده است که در ادامه به برخی از آنها اشاره می‌کنیم، در پژوهش [۳] از تابع هدفی آگاه از انرژی برداشت شده و آگاه از جابه‌جایی گره‌ها استفاده شده است که استفاده از این معیارها در الگوریتم انتخاب والد باعث افزایش قابلیت اطمینان و کاهش مصرف انرژی گره‌ها شده است.

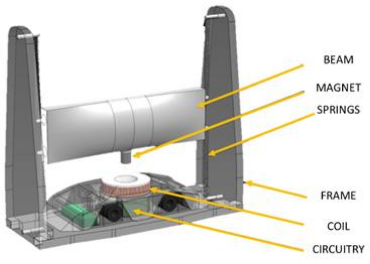
در پژوهش [۹] به افزایش قابلیت پردازشی گره‌ها با بهره‌گیری از تکنیک بارگذاری بر روی لایه مه تأکید شده است که در این پژوهش برای استفاده از قابلیت‌های یادگیری ماشین محققان ناچار به استفاده از برداشتگر انرژی شدند و در پژوهش معیار میزان انرژی برداشت شده را در کنار معیارهای دیگر مورد استفاده قرار دادند که این کار موجب کاهش مصرف انرژی و کاهش تأخیر محاسبات در شبکه شد.

در پژوهش [۱۰] محققان با استفاده از متریک انرژی مصرفی و میزان انرژی باقیمانده و ارسال این پارامتر به نودهایی که قصد انتخاب والد را دارند توانسته‌اند به طول عمر بیشتر شبکه دست‌پیداکنند.

۵- مشاهدات و ایده‌ی پیشنهادی

همان‌طور که در جدول ۱ مشاهده می‌شود در کاربردهایی که دسترسی به نور خوشید وجود ندارد برداشتگر انرژی خورشیدی به کمترین بازدهی در مقایسه با سایر برداشتگرها می‌رسد، در چنین کاربردهایی جهت افزایش طول عمر باتری گره‌ها باید از برداشتگرهای دیگری استفاده نمود؛ از بین سایر تکنولوژی‌های برداشت انرژی، تکنولوژی جذب انرژی از امواج رادیویی به دلیل چگالی پایین این سیگنال‌ها در محیط و همچنین عدم امکان استفاده از آنتن با ابعاد بزرگ، امکان استفاده را ندارند به جز در مواردی که در نزدیکی سنسورها فرستنده توان رادیویی نصب شود که در این صورت نیز قابلیت جابه‌جایی دستگاه به محل کوچکی محدود می‌شود که مناسب کاربردهایی همچون پایش سلامت بیماران نیست.

استفاده از تکنولوژی برداشتگر انرژی سیال هم به دلیل ابعاد بزرگ



شکل ۵- برداشتگر انرژی جریان هوا Wind Beam Zephyr

برداشتگرها در بسیاری از کاربردها قابل‌استفاده نیستند و همچنین شرایط استفاده از آنها در محیطهای داخلی فراهم نیست.

در برخی از کاربردهای شبکه‌های حسگر بی‌سیم گره‌ها در فضای باز قرار داشته و نور خورشید به‌صورت مستقیم به آن‌ها می‌تابد. در این کاربردها می‌توان از پنل خورشیدی برای جذب انرژی تابشی خورشید و تبدیل آن به انرژی الکتریسیته استفاده کرد.

در بسیاری دیگر از کاربردها گره‌ها در دستگاه‌های پوشیدنی استفاده می‌شوند تماس این گره‌ها با بدن انسان و درنظرگرفتن اختلاف دمای سطح پوست با دمای هوا و یا اختلاف دمای سطح مدارات با دمای هوا [۱۱] نیز می‌تواند به‌عنوان منبعی برای تولید انرژی توسط برداشتگرهای ترموالکتریک در نظر گرفته شود. به‌صورت کلی استفاده از برداشتگرهای انرژی در دستگاه‌های به‌کاررفته برای گره‌های حسگر بی‌سیم و درنظرگرفتن میزان انرژی برداشت شده و مصرف شده هر گره در توابع هدف می‌تواند فرصت‌های جدیدی برای افزایش طول عمر شبکه در شرایط محیطی و کاربردهای متفاوت ایجاد نمایند [۱۲].

باتوجه‌به موارد ذکر شده، در محیطهای داخلی نیز به علت عدم کارایی بسیاری از برداشتگرها به سراغ برداشتگرهای ترموالکتریک می‌رویم، این نوع برداشتگرها علی‌رغم بازدهی پایین، در کاربردهای داخلی و دارای میزان جابه‌جایی بالا برای ما بسیار سودمند هستند.

یکی از مهم‌ترین کاربردهای شبکه حسگر بی‌سیم سیستم نظارت بر سلامت بیماران است که وظیفه مشاهده و بررسی علائم حیاتی بیمار را دارا است و بنا به کاربرد این گره‌ها باید بر روی بدن انسان نصب شود تا بتوانند داده‌های موردنیاز از جمله ضربان قلب، اکسیژن خون، دما و... را در تماس با بدن انسان اندازه‌گیری کنند، یکی از مهم‌ترین چالش گره‌های نظارت بر

جدول ۱- منابع برداشت انرژی و چگالی توان مربوط به آنها [۸]

منبع انرژی	انواع	روش برداشت انرژی	تراکم قدرت
تابشی	خورشیدی	سلول‌های خورشیدی (در داخل ساختمان)	<10μW/cm^2
		سلول‌های خورشیدی (در فضای باز، روزهای آفتابی)	15mW/cm^2
	رادیویی	تبدیل الکترومغناطیسی	0.1μW/cm^2 (GSM)
		تبدیل الکترومغناطیسی	0.01μW/cm^2 (WiFi)
مکانیکی	جریان سیال	تبدیل الکترومکانیکی	16.2μW/cm^2
	نویز آکوستیک	پیزوالکتریک	960nW/cm^2
	حرکت جنبش	پیزوالکتریک	330μW/cm^2
حرارتی	گرمای بدن	ترموالکتریک	40μW/cm^2

سلامت بیمار طول عمر باتری آنهاست و این موضوع در طول شب از اهمیت بالاتری برخوردار است زیرا امکان تعویض باتری برای کاربر میسر نیست، باتوجه‌به بررسی‌های انجام شده در بخش سوم می‌توان از برداشتگر انرژی ترموالکتریک در این کاربرد استفاده نمود بدین صورت که با تماس سطح گرم مولد به پوست انسان و سطح سرد آن به گرماگیر[6] (جهت افزایش تبادل حرارت با هوا) توان مناسبی جهت روشن نگاه‌داشتن دستگاه و ذخیره توان مازاد در باتری فراهم آورد.

همان‌طور که پیش‌ازاین گفته شد از دیگر کاربردهای این نوع مولد بر روی سطح مدارات الکترونیکی است، در واقع تمامی مداراتی که در شبکه‌های حسگر بی‌سیم استفاده می‌شوند برای ارتباط از یک ماژول مخابراتی ارسال و دریافت‌کننده استفاده می‌کنند؛ مدارات الکترونیکی در زمان کارکرد خود بخش قابل‌توجهی از انرژی مصرفی خود را به انرژی گرمایی تبدیل می‌کنند و به همین دلیل دمای سطح این قطعات افزایش پیدا می‌کند، این افزایش دما در پردازنده و واحد ارسال‌کننده داده بی‌سیم بیشتر است؛ در صورت اتصال سطح گرم مولد ترموالکتریک با بخش‌های گرم مدار می‌توان بخشی از انرژی اتلاف شده به‌صورت گرما را بازیابی کرد و با این کار طول عمر باتری دستگاه را افزایش داد.

برای افزایش میزان برداشت انرژی در این گره‌ها باید به اختلاف دما بیشتری به دمای هوا رسید؛ درصورتی که بتوان مسیر انتقال بسته‌ها را به نحوی تغییر داد که بسته‌های زیادی از طریق یک والد به گره ریشه منتقل شود باعث افزایش ناگهانی دما آن گره خواهد شد که این افزایش دما در فرایند برداشت انرژی گرمایی برای ما سودمند است؛ برای پیاده‌سازی این مهم می‌توان در الگوریتم انتخاب والد معیار دیگری را جهت میزان انرژی برداشت شده به معیارهای مرسوم اضافه کرد که این معیار باعث به حداکثر رسیدن دمای یک گره در شبکه در هر لحظه از زمان خواهد شد؛ در این عملیات اگر گره‌هایی والد مشترک داشته باشند با یکدیگر برای به حداکثر رساندن دمای آن گره از طریق ارسال بسته‌های خود جهت مسیریابی به آن گره استفاده می‌کنند و پس از مدت زمانی مشخص که کارایی برداشتگر انرژی به حداکثر خود رسید آن گره ویژگی والد بودن را از دست می‌دهد و فرزندان آن گره، بسته‌ها را از مسیری جایگزین برای گره ریشه ارسال می‌کنند و الگوریتم انتخاب والد برای افزایش دمای گره دیگری تلاش می‌کند.

انجام این عملیات تأثیری در توان مصرفی میانگین کل شبکه ندارد و صرفاً در لحظاتی کوتاه توازن شبکه را برای افزایش میزان انرژی برداشت شده تغییر می‌دهد؛ برای بررسی زمان مطلوب جهت افزایش دمای یک گره از طریق مسیریابی بسته‌ها چند آزمایش عملی انجام شده است که در ادامه نحوه آزمایش و نتایج حاصل از آن را می‌بینیم.

برای انجام این آزمایش نیاز به سیستمی جهت اندازه‌گیری دمای سطح گرم و سطح سرد ماژول ترموالکتریک و یک پردازشگر با بار کاری قابل تنظیم و مشخص داریم. در این آزمایش از پردازشگر ESP32 استفاده شده است. این پردازنده برای رسیدن به مصرف انرژی بهینه با فرکانس ۸۰ مگاهرتز کار می‌کند درصورتی‌که قابلیت ارتقا سرعت خود تا فرکانس ۲۴۰ مگاهرتز را دارا است. این نوع میکروکنترلرها دارای فرستنده گیرنده ۲.۴ گیگاهرتز جهت اتصال به WiFi و همچنین دارای قابلیت بلوتوث کم‌مصرف (BLE[7]) در خود است. برای این آزمایش از نوع Wroom32 این پردازشگر استفاده شده است (شکل ۵) که بر روی خود از یک روکش آلومینیومی جهت کاهش تأثیر نویزهای محیطی و انتقال حرارت بهره می‌گیرد. برای این آزمایش ما با حذف روکش آلومینیومی امکان تبادل حرارت را از سطح مدار به ماژول ترموالکتریک بالا بردیم و نیز در محل اتصال از خمیر سیلیکون که هادی گرما بسیار خوبی است استفاده کردیم. همچنین بار کاری در طول آزمایش در جدول ۲ نشان داده شده است.

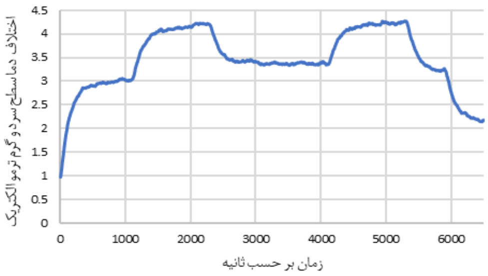
در طول آزمایش پردازشگر به‌صورت مشتری به یک نقطه دسترسی اصلی متصل است، برای این آزمایش نرخ تولید و انتقال دیتا به صورت میانگین ۴۸ بایت در هر ۵ ثانیه در نظر گرفته شده است و تعداد دو گره بسته‌های خود را جهت مسیریابی به گره مورد آزمایش ارسال می‌کنند.

ما برای ارزیابی بازدهی ماژول ترموالکتریک در هر حالت از دمای سطح گرم ماژول ترموالکتریک که با بخش فرستنده و میکروکنترلر در تماس است و دمای سطح سرد ترموالکتریک که از یک گرماگیر در آن به جهت افزایش تبادل دما با هوا در سامانه استفاده شده است بهره بردیم. در محل اتصال ترموالکتریک و گرماگیر از خمیر سیلیکون برای افزایش تبادل حرارت استفاده شده است. به کمک محیط آزمایش تهیه شده (شکل ۷) و اعمال حالت‌های مختلف کاری بر روی ESP32 به نمودار اختلاف دمای سطح سردوگرم رسیدیم که در شکل ۶ مشاهده می‌شود.

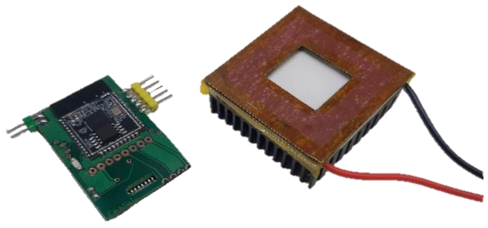
همان‌طور که در شکل۶ مشاهده می‌شود یک گره در حالتی که حداقل توسط دو گره دیگر به عنوان والد انتخاب شود دمای قسمت میکروکنترلر و فرستنده آن افزایش پیدا می‌کند تاحدی‌که اختلاف دمای آن با سطح سرد ماژول ترموالکتریک (با درنظرگرفتن انتقال حرارت از سطح گرم ماژول به سمت سرد آن) به حدود ۴.۵ درجه سانتی‌گراد می‌رسد. این اختلاف دما نسبت مستقیمی با میزان انرژی برداشت شده ترموالکتریکی دارد. این آزمایش در شرایط آزمایشگاهی بدون وجود جریان هوا انجام شده است، از آنجایی که معمولاً در کاربردهای مورد استفاده جریان هوا به دلایل مختلفی وجود دارد، درصورتی‌که بتوان بر روی گرماگیر از جریان هوا بهره‌گیری کرد میزان اختلاف دما و به‌تبع آن میزان انرژی برداشت شده افزایش پیدا می‌کند.

جدول ۲- وضعیت سیستم در مراحل آزمایش

حالت	زمان پایان	زمان شروع
مودم در حالت خواب	۵۰۰	۰
رادیو فعال است	۱۱۰۰	۵۰۰
رادیو فعال + تولید و ارسال بسته‌های تولید شده + مسیریابی بسته‌های فرزندان	۲۳۰۰	۱۱۰۰
رادیو فعال + تولید و ارسال بسته‌های تولید شده	۴۱۰۰	۲۳۰۰
رادیو فعال + تولید و ارسال بسته‌های تولید شده + مسیریابی بسته‌های فرزندان	۵۳۰۰	۴۱۰۰
مودم در حالت خواب	۵۹۰۰	۵۳۰۰
خواب عمیق	۶۵۰۰	۵۹۰۰



شکل ٦- تغییرات اختلاف دمای سطح سرد و گرم ماژول ترمو الکتریک در آزمایش انجام شده بر حسب جدول ۲

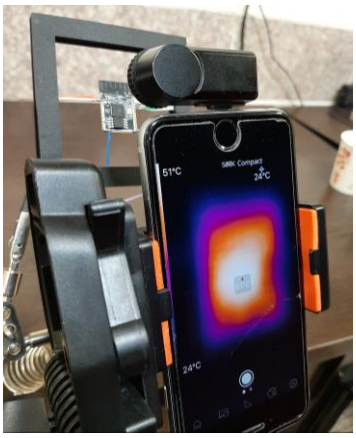


شکل ۷ - محیط آزمایش طراحی شده

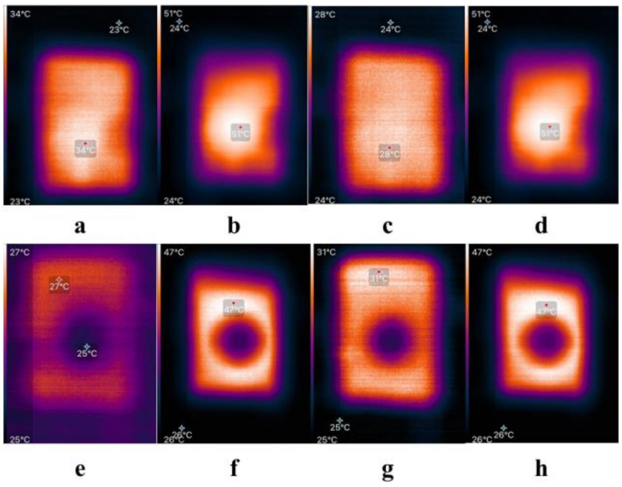
باتوجه‌به اینکه اختلاف دمای ۴.۵ درجه می‌تواند توان الکتریکی برابر با ۰.۰۰۷ وات را تولید کند [13] این میزان جریان برای روشن نگاه داشتن ESP32 در حالت کم‌مصرف کافی است، میکرو در حالت فوق کم‌مصرف مصرفی برابر با ۱۰ میکرو آمپر در ولتاژ ۳.۳ ولت دارد که معادل ۳۳ میکرو وات توان است بنابراین می‌تواند مدت زمان زیادی با انرژی برداشت شده فعال باشد و نیز انرژی مازاد را جهت استفاده در آینده درون باتری خود ذخیره نماید؛ در ادامه جدول توان مصرفی ESP32 را مشاهده می‌کنید. در الگوریتم پیشنهادی برای انتخاب والد پارامترهای HopCount، دمای گره کاندید والد، تعداد فرزندان گره کاندید والد و بار کاری گره در نظر گرفته خواهد شد. این پارامترها، همانند پارامترهای مربوط به الگوریتم‌های مسیریابی مبتنی بر RPL در طول کارکرد شبکه بین گره‌ها به اشتراک گذاشته خواهد شد.

لازم به توضیح است که اگر قیدی برای حداکثر بار کاری (که با حداکثر تعداد فرزندان رابطه دارد) در سامانه در نظر گرفته نشود ممکن است با افزایش تعداد فرزندان یک گره، نرخ ازدست‌دادن بسته‌ها و تأخیر شبکه افزایش یابد. لذا در الگوریتمی که پیشنهاد می‌شود برای هر گره یک مقدار حداکثر بارکاری در نظر گرفته می‌شود که مجاز به تخطی از آن نیستیم.

در انتها لازم به توضیح است که گرم‌کردن دستگاه باعث ایجاد خطاهای گذرا در سیستم نشده و کارایی سیستم را کاهش نمی‌دهد. در واقع باتوجه‌به اینکه مقدار حداکثر گرم‌شدن گره‌ها به‌صورت تجربی به مقدار ۵۱ درجه محدود شده است، دمای گره‌ها همواره در یک مقدار مجاز باقی می‌ماند. به عنوان مثال شرکت Espressif میکروکنترلرهای ESP32 را برای کارکرد بدون خطا تا دمای ۸۵ درجه سانتیگراد تضمین کرده است و حتی در بعضی مدل‌های خاص دمای حداکثر به ۱۲۵ درجه سانتیگراد نیز می‌رسد. لذا دمای در نظر گرفته شده برای کارکرد بدون خطای تراشه توسط سازندگان بسیار بیشتر از حداکثر دمای مشاهده شده در حالت تجربی این الگوریتم است که این موضوع عملکرد بدون خطای سیستم در زمان استفاده از راهکار پیشنهادی را تضمین می‌کند.



شکل ۸- محیط آزمایش طراحی شده بر اساس تصویربرداری گرمایی

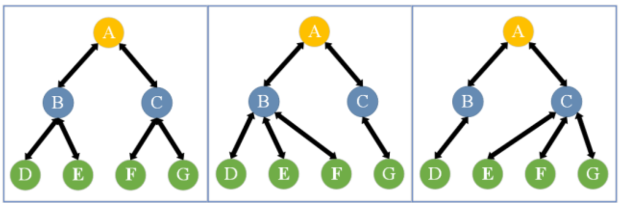


شکل ۹ - نتایج به‌دست‌آمده از تصویربرداری گرمایی

۶- ارزیابی و نتایج

در این فصل به ارزیابی نتایج می‌پردازیم. همان‌طور که در محیط آزمایش شکل ۷ و نتایج به‌دست‌آمده از آن مشاهده شد باتوجه‌به انتقال حرارت ایجاد شده توسط میکروکنترلر به ترموالکتریک و سایر اجزا مدار برای رسیدن به دمای نهایی در هر حالت به زمانی جهت گرم‌شدن همه اجزا نیاز است که ما با طراحی محیط آزمایش شکل ۸ که متشکل از یک پردازنده ESP32 با کمترین میزان اتصال فیزیکی ممکن جهت عملکرد آن است و استفاده از دوربین گرمایی میزان گرمای تولید شده توسط پردازنده را در حالات مختلف اندازه‌گیری کردیم.

همان‌طور که در شکل ۹ مشاهده می‌شود اختلاف دما مابین حالتی که دستگاه دارای یک فرزند و در حالتی که دستگاه دارای سه فرزند باشد در سطح میکروکنترلر به ۱۷ درجه سانتیگراد می‌رسد که این مقدار نشان‌دهنده میزان بالای تأثیر تعداد فرزندان به افزایش دمای والد است. لازم به توضیح است که در شکل ۹ تصاویر a،b،c و d مربوط به سطح بالایی میکروکنترلر و تصاویر e،f،g و h مربوط به سطح زیرین ESP32 هستند. در ستون اول حالت رادیو روشن بدون ارسال یا دریافت اطلاعات را مشاهده می‌کنید و در ستون دوم و چهارم حالتی که گره دارای سه فرزند است و در ستون سوم حالتی که گره دارای یک فرزند است قابل‌مشاهده است.



الف ب پ

شکل ۱۰ – تغییر سناریو شبکه در طول زمان

برای درک بهتر امکان‌پذیری برداشت انرژی توسط برداشتگر ترموالکتریک سناریوهای شکل ۱۰ را در نظر بگیرید. در این سناریوها، درصورتی‌که شبکه متشکل از ۷ گره که یکی از آنها (A) گره ریشه باشد و دو عدد گره میانی (B و C) باشند و چهار عدد از آنها (D،E،F و G) گره برگ باشند اگر شبکه در طول زمان به‌صورت متوازن عمل کند؛ مانند شکل ۱۰-پ میزان گرمای گره‌های میانی به میزان مطلوب برای برداشت انرژی در ترموالکتریک نمی‌رسد؛ ولی اگر در بازه‌های زمانی مشخص بار کاری یک والد با افزایش تعداد فرزندان زیاد شود (همانند شکل ۱۰-الف یا ۱۰-ب) میزان انرژی برداشت شده در طول زمان علی‌رغم ثابت ماندن میزان مصرف شبکه و توازن مصرف باتری در شبکه افزایش می‌یابد. به عنوان مثال با در نظر گرفتن بار کاری سه فرزند برای گره‌ی دارای برداشتگر ترموالکتریک و ایجاد اختلاف دمای ۱۷ درجه‌ی سانتیگراد که در آزمایش قبل مشاهده گردید در هر یک از این دو سناریوی شکل ۱۰-الف و ۱۰-ب می‌توان انرژی قابل توجهی بدست آورد. برای محاسبه میزان انرژی برداشت شده می‌توان از روابط ترمودینامیک و رابطه سیبک استفاده کرد، در صورت که ترموالکتریک با ابعاد میکروکنترلر و با بازدهی بالا در این کاربرد مورداستفاده قرار گیرد می‌توان بازدهی بسیار بالایی را در این کاربرد شاهد بود.

۷- نتیجه‌گیری و کارهای آینده

همان‌طور که در بخش قبل خواندید از مطالعات و آزمایش‌های انجام‌گرفته می‌توان به این نتیجه رسید که استفاده از پارامتر میزان انرژی برداشت شده در معیار انتخاب والد در گره‌هایی که از برداشتگر انرژی ترموالکتریک استفاده می‌کنند می‌تواند برای دستگاه‌هایی که دسترسی به نور خورشید ندارند بسیار سودمند باشد و حتی می‌توان به‌عنوان مکمل در سیستم‌هایی با قابلیت دسترسی به نور خورشید از برداشتگر انرژی ترموالکتریک استفاده نمود.

مراجع

[1] A.Iftikhar, H.Lim Meng, M.Abdelrhman, I.Syed Asad, M.S.Leong, "Scopes, challenges and approaches of energy harvesting for wireless sensor nodes in machine condition monitoring systems: A review," *Measurement*, vol. 183, no. 0263-2241, p. 109856, 2021.

[2] H.Lamaazi, N.Benamar, J.Jara, "RPL-based networks in static and mobile environment: A performance assessment analysis," *Journal of King Saud University - Computer and Information Sciences*, vol. 20, no. 3, pp. 320-333, 2017.

[3] S. R. Lalani, A. A. M. Salehi, H. Taghizadeh, B. Safaei, A. M. H. Monazzah, A. Ejlali, "REFER: A Reliable and Energy-Efficient RPL for Mobile IoT Applications," *2020 CSI/CPSSI International Symposium on Real-Time and Embedded Systems and Technologies (RTEST)*, pp. 1-8, 2020.

[4] B. Safaei, A. M. H. Monazzah and A. Ejlali, "ELITE: An Elaborated Cross-Layer RPL Objective Function to Achieve Energy Efficiency in Internet-of-Things Devices," *IEEE Internet of Things Journal*, vol. 8, pp. 1169-1182, 2021.

[5] NesrineJaziri, AydaBoughamoura, JensMüller, BrahimMezghani, FaresTounsi, MohammedIsmail, "A comprehensive review ofThermoelectric Generators: Technologies and common applications," *Energy Reports*, vol. 6, pp. 264-287, 2020.

[6] I.-V. Marlow, "EHA-PA1AN1-R04 Tecnical Data Sheet," 2020. [Online]. Available: https://ii-vi.com/product/evergen-energy-harvesters/.

[7] Y. Luo, L. Pu, "Practical Issues of RF Energy Harvest and Data Transmission in Renewable Radio Energy Powered IoT," *IEEE Transactions on Sustainable Computing*, 2020.

[8] K.Adu-Manu, N.Adam, C.Tapparello, H.Ayatollahi, "Energy-harvesting wireless sensor networks (EH-WSNs): A review," *ACM Transactions on Sensor Networks*, vol. 14, pp. 1-50, 2019.

[9] M. Min, L. Xiao, Y. Chen, P. Cheng, D. Wu and W. Zhuang, "Learning-Based Computation Offloading for IoT Devices With Energy Harvesting," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 2, pp. 1930-1941, 2019.

پانویس‌ها

---

۱ Wireless Sensor Network
۲ Low-Power and Lossy Network
۳ Internet Engineering Task Force
۴ IPv6 over Low Power Wireless Personal Area Networks
۵ IPv6 Routing Protocol for LLNs
۶ Expected Transmission Count
۷ Link Quality Level
۸ Minimum Rank With Hysteresis Objective Function
۹ Objective Function Zero
۱۰ Destination Oriented Directed Acyclic Graph
۱۱ DODAG Information Object
۱۲ DODAG Information Solicitation
۱۳ Destination Advertisement Object
۱۴ Spectrum Analyzer
۱۵ Airborne Wind Energy System
۱۶ Heat sink
۱۷ Bluetooth Low Energy

چهارمین کنفرانس ملی انفورماتیک ایران
پژوهشگاه دانش‌های بنیادی، پردیس فرمانیه، تهران
۱۴ و ۱۵ دی ماه ۱۴۰۱

مقایسه تطبیقی کارایی زیرساخت ابری ارائه شده در ابرهای عمومی داخلی

محمدعلی کشاورز[۱]، امیر کلباسی[۲]

[۱] دانشجوی کارشناسی ارشد رشته مهندسی کامپیوتر گرایش نرم‌افزار دانشگاه صنعتی امیرکبیر، تهران،
keshavarz.ma13@aut.ac.ir
[۲] استادیار دانشکده مهندسی کامپیوتر گرایش نرم‌افزار دانشگاه صنعتی امیرکبیر، تهران،
kalbasi@aut.ac.ir

چکیده

در سال‌های اخیر رایانش ابری به یکی از موضوعات مورد علاقه‌ی صنعت و دانشگاه تبدیل شده است. انتقال زیرساخت سازمان‌ها به محیط ابری در قالب خدمات ابری در لایه‌ی زیرساخت، به سازمان‌ها کمک می‌کند تا از هزینه‌های نگهداری مراکز داده داخلی صرف نظر کنند و همچنین در هزینه‌ی خرید و تعمیر سخت‌افزار صرفه‌جویی کنند. این شیوه‌ی خدمات ابری به سازمان‌ها انعطاف‌پذیری می‌دهد تا منابع زیرساخت خود را با توجه به تقاضا کم و یا زیاد کنند. استفاده از زیرساخت ابری این امکان را فراهم می‌کند که از هزینه و پیچیدگی خرید و مدیریت سرورهای فیزیکی و زیرساخت مرکز داده صرف نظر شود. هر کدام از منابع می‌تواند به عنوان یک مؤلفه مستقل ارائه شود و مشتری فقط تا زمانی که به آن نیاز دارد هزینه‌ی آن را پرداخت می‌کند.

یکی از چالش‌های اصلی برای سازمان‌هایی که قصد استفاده از خدمات ابری در لایه‌ی زیرساخت را دارند، انتخاب ارائه‌دهنده‌ی زیرساخت ابری است. این پژوهش به بررسی زیرساخت ابری ارائه شده توسط سه شرکت ابر آروان، ابر ستون و ابر ایران پرداخته است و خدمات ارائه شده توسط این سه شرکت، در سه زمینه‌ی شبکه، قدرت پردازش و کیفیت حافظه‌ی مانا بررسی شده است و نقاط قوت و ضعف هر کدام از ارائه‌دهندگان ذکر شده، مورد بررسی قرار گرفته است.

کلمات کلیدی

رایانش ابری، زیرساخت به عنوان خدمت، ابرآروان، ابرایران، ابرستون

۱- مقدمه

اگرچه که استفاده از خدمات ارائه شده توسط یک خدمت‌دهنده ابری مزایای زیادی دارد، اما به کار گرفتن آن‌ها به ماهیت، اندازه و نیازهای خدمت‌گیرنده بستگی دارد. با ظهور خدمت‌دهندگان مختلف ابری، تصمیم گیری برای انتخاب بهترین خدمت‌دهنده که نیازهای یک سازمان را برآورده کند به یک کار پیچیده تبدیل شده است. بنابراین، هر یک از موارد موثر نظیر قیمت‌گذاری، انعطاف‌پذیری، پشتیبانی و سایر پارامترهای مشابه، اثر خود را در انتخاب خدمت‌گیرنده‌ها دارد. بسیاری از شرکت ها مانند آمازون، گوگل، مایکروسافت وارد تجارت رایانش ابری شده‌اند. جامعه متن‌باز[۱] هم در حوزه رایانش ابری در زمینه‌هایی مانند مجازی‌سازی بسیار فعال بوده است. مجازی‌سازی یک فناوری کلیدی برای رایانش ابری است زیرا استفاده کارآمدتر و انعطاف‌پذیرتر از منابع را ممکن می‌سازد.

خدمت‌دهندگان رایانش ابری خدمات متعددی در سطوح مختلف ارائه می‌دهند. معمولا در دسته‌بندی سطوح خدمات در رایانش ابری، سه سطح عنوان می‌شود. سطح اول زیرساخت به عنوان خدمت یا IaaS[۲] است. این سطح که به نسبت دو نوع دیگر خدمات دسترسی بیشتری به تیم توسعه دهنده می‌دهد، نوعی از خدمات رایانش ابری است که منابع محاسباتی، ذخیره‌سازی و شبکه را بر اساس تقاضا و یا بر اساس رزرو قبلی ارائه می‌دهد. سطح دیگر خدمات ابری، بن‌سازه[۳] به عنوان خدمت یا PaaS[۴] است. این سطح، دسته‌ای از خدمات رایانش ابری است که به توسعه‌دهندگان اجازه می‌دهد خدماتی شامل یک بن‌سازه محاسباتی و یا برنامه های کاربردی بیشتر، بدون پیچیدگی ساخت و نگهداری زیرساختی را در اختیار داشته باشند. برای مثال پایگاه‌داده به عنوان خدمت، یکی از انواع خدمات بن‌سازه به عنوان خدمت است؛ به این صورت که استفاده کننده از این خدمت، خود را درگیر استقرار پایگاه‌داده‌ها، نگهداری و به روزرسانی آن‌ها نمی‌کند. نوع دیگر خدمات ارائه شده توسط خدمت‌دهندگان ابری، نرم‌افزار به عنوان خدمت یا همان SaaS[۵] است. این نوع خدمت به این صورت است که خود نرم‌افزار از طریق بستر اینترنت در اختیار کاربر قرار می‌گیرد و دیگر نیازی به نصب و راه‌اندازی به صورت محلی ندارد، پیچیدگی‌های پیاده‌سازی و استقرار نیز به طور کامل از چشم کاربر پوشیده شده است و کاربر از طریق بستر اینترنت از این خدمات استفاده می‌کند.

انتقال زیرساخت‌های سازمان به محیط ابری با استفاده از خدمت زیرساخت به عنوان خدمت، به مشتری کمک می‌کند تا تعمیر و نگهداری مراکز داده داخلی را کاهش دهد، در هزینه‌های سخت‌افزار صرفه‌جویی کند، همچنین با استفاده از خدمات زیرساخت ابری، مشتری میتواند تنها به اندازه‌ی استفاده از خدمات هزینه پرداخت کند. این روش امکان کم یا زیاد کردن منابع بر مبنای میزان تقاضا را فراهم می‌آورد. IaaS به سازمان‌ها این امکان را می‌دهد که از هزینه و پیچیدگی خرید و مدیریت سرورهای فیزیکی و زیرساخت مرکز داده عبور کنند. هر منبع به عنوان یک مؤلفه خدمات جداگانه ارائه می‌شود و فقط تا زمانی که به آن نیاز هست هزینه‌ی آن را پرداخت می‌شود.

علاوه بر تقسیم‌بندی ذکر شده برای خدمات ابری که بر اساس سطح خدمت بود، رایانش ابری از نظر استقرار توسط NIST[۶] نیز به چهار گروه دسته‌بندی شده است [۱].

ابر خصوصی: ابرخصوصی، زیرساخت ابری برای استفاده انحصاری توسط یک سازمان متشکل از چندین مصرف کننده (به عنوان مثال، واحدهای تجاری) است. مدیریت، مالکیت و استقرار این گونه از ابر، ممکن است توسط خود سازمان، شخص ثالث یا ترکیبی از آنها باشد.

ابر انجمنی: این نوع از استقرار ابر، زیرساخت ابری برای استفاده انحصاری توسط جامعه خاصی از مصرف کنندگان شامل سازمان‌هایی که نگرانی‌های مشترکی مانند الزامات امنیتی و کارکردی، دارند را فراهم می‌آورد. مدیریت، مالکیت و استقرار این گونه از ابر، ممکن است توسط خود سازمان، شخص ثالث یا ترکیبی از آنها باشد.

ابر عمومی: در این نوع از سطح دسترسی زیرساخت ابری، امکان استفاده آزاد توسط عموم فراهم شده است. این نوع از خدمات ابری ممکن است تحت مالکیت، مدیریت و اداره یک سازمان تجاری، دانشگاهی، دولتی یا ترکیبی از آنها باشد.

ابر ترکیبی: این نوع از استقرار ابر، ترکیبی از دو یا چند نوع زیرساخت ابری متمایز (خصوصی، انجمنی یا عمومی) است.

در این پژوهش به بررسی خدمت IaaS ارائه شده توسط ابرهای عمومی خواهیم پرداخت. از ارائه دهندگان مشهور خدمات ابری عمومی در جهان می‌توان به ابرگوگل[۷] ، ابرآمازون[۸] (یا همان AWS) یا مایکروسافت آژور[۹] اشاره کرد اما برای این پژوهش سه شرکت ایرانی ابر آروان، ابرستون و ابر ایران مورد بررسی قرار گرفته‌اند و خدمات ارائه شده توسط این سه شرکت، در سه موضوع شبکه، قدرت پردازش و سرعت حافظه‌ی مانا مقایسه شده است.

۲- پژوهش‌های مشابه پیشین

فعالیت‌های زیادی در مورد ارائه‌دهندگان خدمات ابری بزرگ و به طور کلی معیارهای کارایی مرتبط با مجازی‌سازی وجود دارد. در این بخش فعالیت‌های صورت گرفته پیشین مورد بررسی قرار گرفته است و در انتها به تفاوت نگاه این پژوهش نسبت به کارهای مشابه پیشین، پرداخته شده است.

لینتر و همکاران در [۲] با استفاده از ۱۵ آزمایش طراحی شده و بر روی چندین خدمت‌دهنده‌ی ابری، عملکرد کلی خدمات ارائه شده توسط خدمت‌دهندگان را در سطح نمونه مورد بررسی و مقایسه قرار داده‌اند. آزمایش‌های صورت گرفته در [۲] بر روی کیفیت محاسبات، نرخ خواندن و نوشتن بر روی حافظه‌ی مانا و همچنین پهنای باند حافظه مانا متمرکز بوده است. این آزمایش‌ها بر اساس عملکرد یک نمونه با اندازه مشخص صورت گرفته و تلاش شده است که بین ابرهای مختلف از نمونه‌های مشابه استفاده شود. همانطور که از توضیحات ذکر شده مشخص است؛ تحقیقات صورت گرفته روی خدمات در سطح IaaS بوده است.

فلتر و همکاران در [۳] ماشین‌های مجازی میزبانی شده در فراناظر[۱۰] KVM را با Containerهای داکر مقایسه کرده‌اند. از آنجایی که KVM یک فراناظر نوع یک است (نیازی به سیستم عامل میزبان ندارد) و ماشین‌های مجازی میزبانی شده از نوع مجازی سازی کامل هستند، محققین انتظار داشتند که عملکرد به طرز قابل توجهی از Containerها بهتر باشد. این آزمایش‌ها برای اندازه‌گیری محاسبات، پهنای باند حافظه، تأخیر حافظه، پهنای باند شبکه به عنوان IaaS و بررسی خدمات سطح PaaS مانند Redis و MySQL طراحی شده‌اند. نتایج نشان می‌دهد که ماشین‌های مجازی مبتنی بر فراناظر تقریباً برای همه موقعیت‌ها به طور جزئی بهتر از Containerها عمل می‌کنند. اما هزینه اضافی مدیریت ماشین‌های مجازی، مقدار قابل توجهی است. به طور کلی، Containerها به اندازه کافی نزدیک به فراناظر عمل می کنند تا آنها را از نظر سرعت اجرای دستورات به یک جایگزین معتبر برای ماشین‌های مجازی مبتنی بر فراناظر تبدیل کنند. نکته‌ای که در [۳] به آن کمتر توجه شده است پیچیدگی‌های امنیتی حاصل از همسایگی Containerها در هسته‌ی مشترک سیستم‌عامل میزبان است.

اهوجا و همکاران در [۴] خدمت IaaS ارائه شده توسط آمازون و گوگل را با استفاده از ابزارهای Benchmark در دو حوزه‌ی پهنای باند حافظه و پهنای باند حافظه مانا مورد بررسی قرار داده‌اند. برای آزمایش از هر کدام از خدمت دهنده‌ها یک خوشه با مشخصات یکسان تهیه شده است و نتایج بین دو نمونه‌ی تهیه شده مقایسه شده است.

در حوزه‌ی مدیریت و استقرار زیرساخت ابری ابزارهای متن‌بازی نظیر CloudStack، OpenStack و Eucalyptus وجود دارد که در ابرهای خصوصی و گاهی حتی در ابرهای عمومی نیز مورد استفاده قرار می‌گیرند. در [۵، ۶، ۷] در مورد مقایسه تطبیقی عملکرد این ابزارهای متن باز در حوزه‌ی استفاده از منابع، پژوهش‌هایی به عمل آمده است.

علاوه بر مقایسه تطبیقی ذکر شده، در [۸] مدل کیفیتی برای استفاده از نرم‌افزارهای متن‌باز حوزه‌ی مدیریت و استقرار زیرساخت ابری، از دیدگاه پردازش کلان داده‌ها ارائه شده است که به محققین اجازه انتخاب بهتر زیرساخت بر اساس نیازمندی‌های پردازش کلان داده را می‌دهد.

در این پژوهش به طور ویژه خدمات ارائه شده توسط ابرهای عمومی داخلی بررسی شده است و زیرساخت ارائه شده توسط ابرهای ذکر شده، در سه معیار پهنای باند شبکه‌ی اینترنت و شبکه‌ی داخلی، سرعت حافظه‌ی مانا و قدرت محاسباتی، با یکدیگر مقایسه شده است. تفاوت این پژوهش با پژوهش‌های پیشین این است که این پژوهش، در حوزه‌ی خدمت‌دهندگان داخلی صورت گرفته است.

۳- شرح آزمایش صورت گرفته

آزمایش‌ها بر روی خوشه‌های محاسباتی ارائه شده توسط سه خدمت‌دهنده‌ی ابری انجام شده است. این سه خدمت‌دهنده، ابر آروان[۹]، ابر ایران[۱۰] و ابر ستون[۱۱] هستند. ماشین مجازی تهیه شده از ابر آروان، ماشین g2-large بوده است که دارای ۸ پردازنده و ۳۲ گیگ حافظه است. منابع موجود در خوشه‌های ابر ایران دارای ۸ پردازنده و ۱۶ گیگ حافظه و منابع موجود خوشه‌های ابر ستون هم مشابه ابر ایران بود. علت این که میزان حافظه‌ی منابع ابر آروان متفاوت است این است که امکان انتخاب دلخواه منابع در ابر آروان موجود نیست و نسبت تعداد پردازنده به حافظه‌ی اصلی یک به چهار است؛ در حالی که این نسبت برای ابر ستون یک به دو است و برای ابر ایران چنین نسبتی وجود ندارد و به صورت دلخواه می‌توان هر نسبتی را انتخاب کرد. محل قرارگیری مرکز داده‌ی ابر آروان و ابر ستون تهران و محل قرارگیری مرکز داده‌ی ابر ایران شهر مشهد است. به طور کلی مشخصات ماشین‌های مجازی به کار گرفته در خوشه‌های ارائه شده توسط این سه خدمت‌دهنده ابری در جدول(۱) قابل مشاهده است. همه ماشین‌ها از نسخه‌ی اوبونتو ۲۰.۰۴ استفاده می‌کردند. همچنین در جدول(۲) میزان حافظه‌ی نهان به کار گرفته در نمونه‌های تهیه شده ذکر شده است. در رابطه با ساختار شبکه، معماری و شیوه‌ی قرارگیری ماشین‌های فیزیکی و مجازی، در این مراکز داده، اطلاعاتی در وبسایت و مستندهای فنی این سه خدمت‌دهنده وجود نداشت.

یک ماشین معیار با مشخصاتی مشابه ماشین‌های ابر ستون و ابر ایران در دیتاسنتر پارس آنلاین تهران ایجاد شد. علت معیار قرار داده شدن این ماشین، این است که در تست سرعت اینترنت (به خارج مرکز داده) برای ماشین‌های آروان، ایران و ستون، از ماشین معیار استفاده شده است. این آزمون به این صورت انجام شده است که ماشین‌های تهیه شده از ابرهای داخلی عمومی، اقدام به آپلود اطلاعات در اندازه‌های مشخص برروی ماشین معیار کرده‌اند و سرعت بارگذاری آن‌ها مورد ارزیابی قرار گرفته‌است.

جدول(۱) : مشخصات پردازنده و میزان حافظه‌ی ماشین‌های تهیه شده از خدمت‌دهندگان ابری

	ماشین معیار	ابر ایران	ابر ستون	ابر آروان
تعداد هسته	۸	۸	۸	۸
شرکت تولید کننده پردازنده	Intel	Intel	Intel	Intel
نوع پردازنده	Xeon E۵-۲۶۵۰	Fifth Generation Core Processor	Xeon Gold ۶۲۳۸R	Xeon E۳-۱۲xx
فرکانس پردازنده	۲.۳ GHz	۲.۴ GHz	۲.۱۹ GHz	۲.۵ GHz
حافظه	۱۶ GB	۱۶ GB	۱۶ GB	۳۲ GB

جدول(۲) : میزان حافظه‌ی نهان ماشین‌های تهیه شده از خدمت‌دهندگان ابری

	ماشین معیار	ابر ایران	ابر ستون	ابر آروان
حافظه‌نهان۱(دستورات)	۳۲ KB	۳۲ KB	۳۲ KB x ۸	۳۲ KB
حافظه‌نهان۱(داده)	۳۲ KB	۳۲ KB	۳۲ KB x ۸	۳۲ KB
حافظه‌نهان ۲	۲۵۶ KB	۴ MB	۴ MB x ۸	۴ MB
حافظه‌نهان ۳	۶۰ MB	۶۰ MB	۱۶ MB	۱۶ MB

اطلاعات جدول(۲) با استفاده از ابزار geekbench5 به دست آمده است و این اطلاعات به صورت رسمی توسط ارائه‌دهندگان خدمات ابری ارائه نشده است.

۳-۱- ارزیابی کارایی

برای مقایسه‌ی کیفیت خدمات ارائه شده توسط خدمت‌دهندگان ابری سه سنجه‌ی کارایی مورد بررسی قرار گرفته است که در ادامه به معرفی و نحوه‌ی اندازه‌گیری آن‌ها خواهیم پرداخت.

پهنای باند شبکه: اولین معیار پهنای باند شبکه است که به صورت پهنای باند اینترنت (به خارج مرکز داده) و پهنای باند شبکه‌ی داخلی در این پژوهش مورد بررسی قرار گرفته است. برای بررسی پهنای باند شبکه داخلی یک خوشه از دو ماشین مجازی با مشخصات یکسان از یک خدمت‌دهنده گرفته شده است و پهنای باند بین این دو ماشین مورد بررسی و مطالعه قرار گرفته است. برای بررسی وضعیت پهنای باند اینترنت ارتباط یکی از آن دو ماشین ذکر شده با ماشین معیار که در دیتاسنتر پارس آنلاین قرار گرفته، بررسی شده است.

برای بررسی پهنای باند از ابزار iperf3 استفاده شده است [۱۲]. iperf3 ابزاری برای اندازه گیری حداکثر پهنای باند قابل دستیابی در شبکه های IP است. iperf3 با زبان C توسعه داده شده است و به صورت متن باز در اختیار جامعه‌ی توسعه‌دهندگان نرم افزار قرار داده شده است. با توجه به این که استفاده Container درسیستم‌های نرم‌افزاری مستقر در فضای ابری یک بخش جدایی ناپذیر رایانش ابری است؛ محققین در [۱۳] به بررسی عملکرد شبکه‌ی محلی با استفاده از فناوری‌های شبکه‌ی مبتنی بر Container پرداخته‌اند. در این پژوهش، پهنای باند شبکه و تأثیر سیستم شبکه‌های مبتنی بر Container مختلف بر روی ابرهای عمومی محبوب با استفاده از iperf3 مورد مطالعه قرار گرفته است.

سرعت حافظه‌ی مانا: دومین معیار مورد استفاده، حداکثر پهنای باند حافظه‌ی مانا است. این معیار عملکرد خواندن و نوشتن را در میان فایل‌ها به صورت موازی اندازه‌گیری می‌کند. برای اندازه‌گیری این معیار از ابزار fio استفاده شده است. این ابزار نیز با زبان C توسعه داده شده است و به صورت متن باز در اختیار جامعه‌ی توسعه دهندگان نرم افزار قرار داده شده است. به کمک این ابزار میتوان پیکربندی‌های مختلفی برای تست ایجاد کرده از جمله این که می‌توان مشخص کرد که اندازه‌ی هر بخشی که از فایل خوانده می‌شود و نوشته می‌شود چه اندازه باشد[۱۴].

در [۱۵] پژوهشگران یک الگوریتم زمانبندی جدید برای ورودی/خروجی حافظه‌ی مانا در فضای ابری ایجاد کرده‌اند. هدف از این پژوهش ایجاد الگوریتمی بوده است که با اعمال مولفه‌های کیفیت خدمت[۱۱] برای استفاده از حافظه‌ی مانا، بتوانند در نهایت سود ارائه‌دهندگان خدمات ابری و کاربران نهایی را بهبود ببخشند. در این پژوهش با تغییر در هسته لینوکس زمانبند جدیدی برای تعامل با حافظه مانا پیاده شده است. این پژوهش برای بررسی عملکرد زمانبند پیشنهادی، از ابزار fio استفاده کرده است.

قدرت محاسباتی: معیار دیگری که مورد بررسی قرار گرفته است، قدرت محاسباتی ماشین‌های مجازی ارائه شده توسط خدمت‌دهندگان ابری بود. برای بررسی این موضوع از ابزار geekbench5 استفاده شده است. این ابزار پردازشگر را از نظر ابعاد مختلفی مورد ارزیابی قرار می‌دهد و در نهایت کیفیت پردازنده را به صورت یک امتیاز بیان می‌کند. باید به این نکته توجه داشت که ابزار geekbench5 پردازشگر را هم به صورت تک هسته‌ای و هم به صورت چند هسته‌ای مورد بررسی قرار می‌دهد [۱۶].

در [۱۷] برای بررسی عملکرد Containerها به عنوان یک جایگزین سبک‌وزن برای ماشین‌های مجازی، اقدام به مقایسه تطبیقی عملکرد Container و ماشین مجازی در زمینه‌های مختلف شده است. پژوهشگران در این مقاله مسائلی نظیر میزان سربار عملکردی که توسط یک فراناظر در مقایسه با مجازی‌سازی مبتنی بر Container ایجاد می‌شود را بررسی کرده‌اند. علاوه بر این فناوری‌های مبتنی بر Container مختلف را نیز از نظر عملکرد با یکدیگر مقایسه کرده‌اند. در [۱۷] برای بررسی قدرت پردازشی در فناوری‌های مختلف از ابزار Geekbench استفاده شده است.

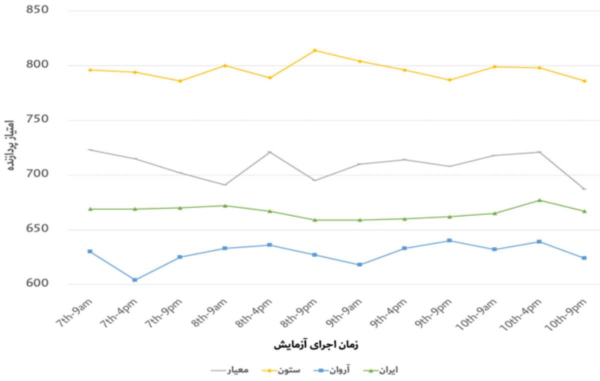
کارایی ترکیبی: علاوه بر استفاده از ابزارهای ذکر شده، با استفاده از یک برنامه‌ی توسعه داده شده با زبان Go با ایجاد ۱۰۰۰ ترد که هر کدام ۱۰۰۰ مدخل بر روی دیتابیس SQLite ایجاد میکند. در این آزمون تلاش شده است که با شبیه‌سازی نیازمندی‌های دنیای واقعی، قدرت محاسباتی و سرعت حافظه مانا را به صورت همزمان مورد بررسی قرار داده شود.

۴- نتایج به دست آمده

برای جمع آوری اطلاعات طی مدت ۴ روز در بازه‌ی سه شنبه ۷ تیرماه ۱۴۰۱ تا روز جمعه ۱۰ تیرماه ۱۴۰۱ هر روز در ساعت‌های ۹، ۱۶ و ۲۱ به اندازه گیری معیارهای ارزیابی پرداخته شده است.

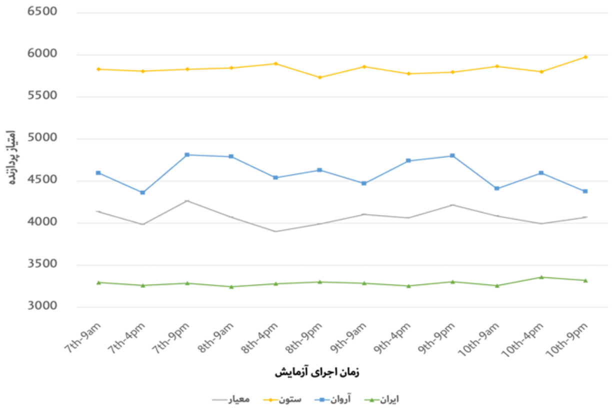
علت انتخاب ۴ روز متفاوت در هفته این است که استفاده‌ی خدمت‌گیرنده‌ها از خدمات ابری در روزهای مختلف هفته ممکن است متفاوت باشد. همچنین ساعات متفاوتی از روز را نیز در نظر گرفتیم که با توجه به استفاده خدمت‌گیرنده‌ها در ساعت‌های مختلف روز از خدمات ابری، بتوانیم عملکرد خدمت‌دهندگان ابری را در همه حالات مورد بررسی قرار دهیم.

ابتدا به بررسی معیارهای قدرت محاسباتی میپردازیم. شکل(۱) امتیاز پردازشگرهای ماشین‌های مجازی ارائه شده توسط خدمت‌دهنده‌های مختلف را به صورت تک هسته‌ای نمایش داده است. امتیاز میانگین برای ابر ایران ۶۶۶.۳ ، ابر آروان ۶۲۸.۴ ، ابرستون ۷۹۵.۷ و برای ماشین معیار ۷۰۸.۷ است. همانطور که به وضوح مشخص است بهترین عملکرد برای ابر ستون و بدترین عملکرد برای ابر آروان بوده است. کمترین ضریب تغییرات بین خدمت‌دهندگان نیز برای ابر ایران با ضریب تغییرات ۰.۰۰۸ بوده است که نشان دهنده‌ی ثبات در کیفیت خدمت‌دهی است.



شکل (۱): امتیاز پردازشگر به صورت تک هسته‌ای.

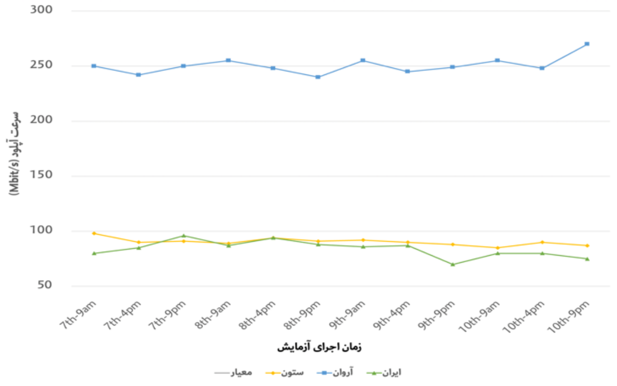
شکل(۲) امتیاز پردازشگرهای ماشین‌های مجازی ارائه شده توسط خدمت‌دهنده‌های مختلف را به صورت چند هسته‌ای نمایش داده است.



شکل (۲): امتیاز پردازشگر به صورت چند هسته‌ای.

امتیاز میانگین برای ابر ایران ۳۲۸۶.۴ ، ابر آروان ۴۵۹۳، ابرستون ۵۸۳۴.۲ و برای ماشین معیار ۴۰۷۲.۳ است. این مقایسه نیز نشان‌دهنده‌ی عملکرد بهتر ابر ستون در زمینه قدرت پردازشی است. در این میان بدترین عملکرد متعلق به ابر ایران بوده است. این درحالیست که امتیاز تک هسته‌ای ابر ایران از ابر آروان بهتر بود ولی در حالت چند هسته‌ای نتایج به نحو دیگری رقم خورده است. این موضوع می‌تواند به این علت باشد که ماشین ابر آروان میزان حافظه‌ی بیشتری در اختیار دارد. نکته‌ی بسیار مهمی که علت عملکرد بهتر ابر ستون را توجیه می‌کند این است که همانطور که در جدول(۲) ذکر شده است؛ در ابر ستون برای هر پردازنده در لایه‌ی اول و دوم حافظه‌ی نهان، از حافظه‌ی جداگانه استفاده شده است که این موضوع سبب میشود در هنگام تغییر Context، حافظه‌ی نهان سرد نشود. در این آزمایش نیز همانطور که از نمودار نیز مشخص است، ابر ایران کمترین نوسانات عملکردی را داشته است و با ضریب تغییرات ۰.۰۰۸ بیشترین ثبات در ارائه‌ی عملکرد را به نمایش گذاشته است.

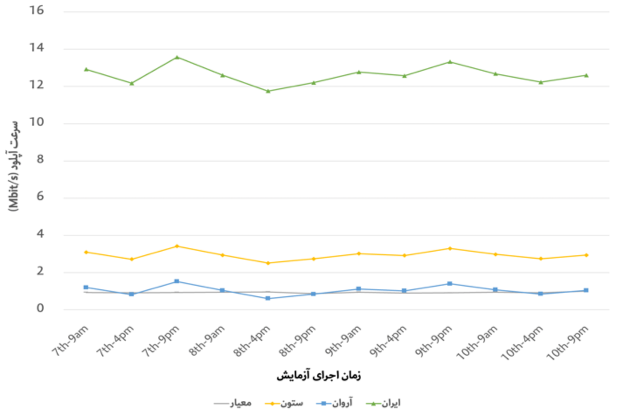
همانطور که گفته شده بود وضعیت شبکه ماشین‌های مجازی، هم به صورت شبکه اینترنت و هم به صورت شبکه‌ی داخل خوشه‌ها مورد بررسی قرار گرفته شده است. شکل(۳) وضعیت شبکه‌ی اینترنت ماشین‌های مجازی را نمایش می‌دهد.



شکل (۳): سرعت آپلود در شبکه‌ی اینترنت.

برای بررسی وضعیت اینترنت ماشین‌های مجازی ما از ماشین معیار استفاده کرده‌ایم. یعنی سرعت آپلود اطلاعات بر روی ماشین معیار که در دیتاسنتر پارس آنلاین قرار داشت توسط ماشین های مجازی ارائه شده خدمت‌دهندگان ابری مورد بررسی قرار گرفته است. سرعت میانگین برای ابر ایران ۸۴.۳۲ Mbit/s، ابر آروان ۲۵۰.۶ Mbit/s و برای ابرستون ۹۰.۴ Mbit/s است. همانطور که در تصویر هم قابل مشاهده است سرعت آپلود در بستر اینترنت ارائه شده توسط دو خدمت دهنده ابر ایران و ابر ستون تفاوت چندانی با یکدیگر ندارند، در نتایج آزمون تی[12] مقدار p برابر با ۰.۰۵۳ به دست آمده است؛ با توجه به این که مقدار آلفا را ۰.۰۵ در نظر گرفته‌ایم، نشان میدهد که تفاوت معناداری بین میانگین نتایج ابر ایران و ستون وجود ندارد. اما ابر آروان با اختلاف معناداری از هر دوی خدمت‌دهندگان دیگر در این زمینه عملکرد بهتری به نمایش گذاشته است.

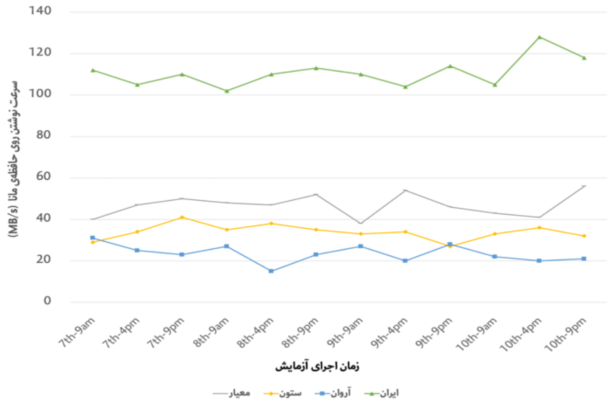
شکل (۴) سرعت شبکه داخلی خوشه‌های ارائه شده توسط هر کدام از خدمت دهندگان ابری را مشاهده میکنیم.



شکل (٤): سرعت آپلود در شبکه‌ی داخلی خوشه‌ها.

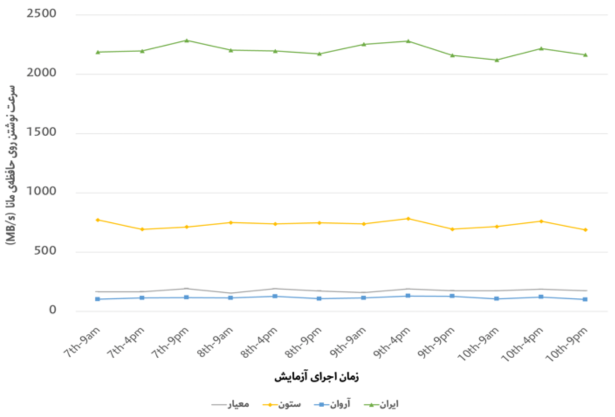
سرعت میانگین در شبکه‌ی داخلی، برای ابر ایران ۱۲.۶ Gbit/s، ابر آروان ۱.۰۴ Gbit/s، ابر ستون ۲.۹۴ Gbit/s و برای ماشین معیار ۰.۹۳ Gbit/s است. همانطور که در شکل(۴) نیز دیده میشود با درنظر نگرفتن ماشین معیار،کمترین سرعت انتقال اطلاعات در شبکه داخلی را ابر آروان به ثبت رسانده است و ابر ایران بهترین عملکرد را نشان داده است. این موضوع را می‌توان به این صورت توجیه کرد که چون زیرساخت ابر ایران به تازگی تهیه شده است، از تجهیزات سخت افزاری به‌روزتری استفاده شده است و همچنین ابر ایران هنوز به طور جدی مورد اقبال کسب و کارهای داخلی قرار نگرفته است، بنابراین احتمالا شبکه‌ی داخلی آن ترافیک کمتری را تحمل می‌کند. همچنین مشخص است که ابر ایران از کارت‌ها و سوییچ‌های شبکه‌ی مناسب برای پهنای‌باند بیش از ۱۰ گیگابیت بر ثانیه استفاده می‌کند. اما با توجه به این که از معماری شبکه‌ی ماشین‌های فیزیکی هیچ کدام از مراکز داده اطلاعاتی در دست نیست، نمی‌توان توجیه قطعی‌ای برای نتایج یافت.

برای بررسی سرعت حافظه مانا، دو حالت را مورد بررسی قرار دادیم. در حالت اول نوشتن بر روی حافظه مانا به صورت قطعات ۴ کیلوبیتی انجام شده است. در این حالت ۲ گیگابایت اطلاعات توسط ۱۶ پردازه[13] بر روی حافظه مانا نوشته شده است. در حالت دوم نوشتن بر حافظه مانا به صورت قطعات ۴۰۰ کیلوبیتی، صورت گرفته است. در این حالت ۱۲ گیگابایت اطلاعات توسط ۱۶ پردازه سیستم عامل بر روی حافظه مانا نوشته شده است. شکل(۵) نتایج حاصل از آزمایش در حالت اول را نشان می‌دهد و شکل(۶) نشان‌دهنده نتایج در حالت دوم است.



شکل (٥): سرعت نوشتن روی حافظه مانا (اندازه هر بار نوشتن ٤ کیلوبیت).

سرعت میانگین در حالت اول برای ابر ایران ۱۱۰.۹ MB/s، ابر آروان ۲۳.۵ MB/s، ابر ستون ۳۳.۹ MB/s و برای ماشین معیار ۴۶.۷ MB/s است. همانطور که دیده می‌شود در حالت اول ابر ایران با اختلاف نتایج بهتری از ابر آروان و ابر ستون ثبت کرده است و عملکرد دو ابر دیگر مشابه همدیگر است. سرعت میانگین در حالت دوم برای ابر ایران ۲۲۰۲.۹ MB/s، ابر آروان ۱۱۵.۰۸ MB/s، ابر ستون ۷۳۲.۴۲ MB/s و برای ماشین معیار ۱۷۵.۱۷ MB/s است. نتیجه‌ی بهتر ابر ایران در حالت دوم نیز برقرار باقی مانده است، اما عملکرد ابر ستون از ابر آروان بهبود قابل توجهی گرفته است. میزان ضریب تغییرات برای حالت اول، برای ماشین ابر ایران ۰.۰۶، ابر آروان ۰.۱۷ و ابر ستون ۰.۱۰ ثبت شده است و در حال دوم، برای ماشین ابر ایران ۰.۰۲، ابر آروان ۰.۰۸ و ابر ستون ۰.۰۴ محاسبه گردیده است. در هر دو حالت آزمایش ابر ایران با کمترین ضریب تغییرات بیشترین ثبات در عملکرد را به نمایش گذاشته است.

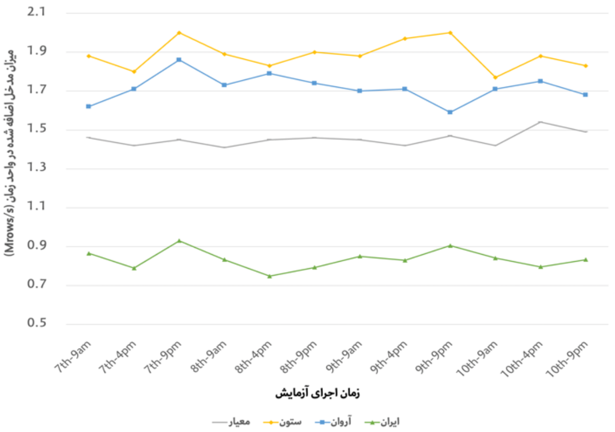


شکل (٦): سرعت نوشتن روی حافظه مانا (اندازه هر بار نوشتن ٤٠٠ کیلوبیت).

شکل(۷) نشان دهنده‌ی سرعت ایجاد مدخل در پایگاه داده SQLite است. در این آزمایش تلاش شده است که یک نیازمندی نزدیک به دنیای واقعی مورد آزمایش قرار بگیرد. هر مدخل اضافه شده به پایگاه داده دارای یک شناسه عددی است و یک ستون از نوع nvarchar بدون محدودیت نیز دارد که در آن نوشته‌هایی با طول تصادفی بین ۳۰ تا ۲۰۰ کاراکتر ایجاد شده است.

نتایج این آزمایش به این صورت است که سرعت میانگین برای ابر ایران ۰.۸۳ Mrows/s، ابر آروان ۱.۷ Mrows/s، ابر ستون ۱.۸ Mrows/s و برای ماشین معیار ۱.۴ Mrows/s ثبت شده است. همانطور که از نتایج مشخص است در این میان بهترین عملکرد متعلق به ابر ستون بوده است و ضعیف‌ترین عملکرد متعلق به ابر ایران است.

از نظر ضریب تغییرات ابر ایران ۰.۰۵۷ ابر آروان ۰.۰۳۹ و ابر ستون ۰.۰۳۷ را ثبت کرده‌اند که نشان میدهد میزان ثبات عملکرد در ابر آروان و ابر ستون به نسبت ابر ایران بهتر است.



شکل (۷): سرعت ایجاد مدخل در دیتابیس SQLite.

۵- نتیجه گیری

در این پژوهش به بررسی خدمات خدمت‌دهندگان ابری مطرح کشور در سطح IaaS پرداختیم. به این صورت که از هر کدام از خدمت‌دهندگان ابر ایران، ابر آروان و ابر ستون خوشه‌های محاسباتی تهیه کردیم و هر کدام را در زمینه‌های قدرت پردازشی، سرعت حافظه مانا و پهنای باند بررسی کردیم.

نتایج به این صورت بود که ابر ستون در زمینه‌ی پردازش بهترین عملکرد را از خود به نمایش گذاشت. بنابراین اگر کارکرد دارای حجم پردازش قابل توجهی است ابر ستون می‌تواند برای این کارکرد گزینه‌ی مناسبی باشد. در زمینه شبکه اینترنت خارج از مرکز داده کیفیت ابرآروان از دیگر رقبا بهتر است و در زمینه‌ی شبکه‌ی داخلی ابر ایران با اختلاف از دیگر رقبا پیشی گرفته است. در نتیجه در صورتی که خدمت گیرنده از خدمات خارج از مرکز داده ابر با پنهای باند بالا بهره می‌برد و یا خدماتی را برای کاربران در خارج از مرکز داده ارائه می‌کند از دیدگاه سرعت شبکه ابر آروان گزینه بهتری می‌تواند باشد. در مقابل اگر تمرکز بر روی استفاده و ارائه خدمات داخلی مرکز داده باشد ابر ایران گزینه مناسب‌تری می‌تواند باشد. اما دیگر سنجه‌ای که مورد بررسی قرار گرفته است، سرعت حافظه مانا است که در این زمینه نیز ابر ایران عملکرد بسیار قابل قبولی از خود ارائه داده است بنابراین چنانچه کارکرد متکی بر استفاده سریع از حافظه مانا است، ابر ایران برای این کارکرد گزینه‌ی مناسبی است. در جدول(۳) خلاصه‌ای از مقایسه نتایج به دست آمده از آزمایش‌ها ذکر شده است.

جدول (۳) : مقایسه‌ی مقادیر میانگین هر کدام از ابرها در معیارهای اندازه گرفته شده.

	سنجه‌های ارزیابی	ماشین معیار	ابر ایران	ابر ستون	ابر آروان
وضعیت پردازش	امتیاز تک هسته‌ای	۷۰۸.۷	۶۶۶.۳	۷۹۵.۷	۶۲۸.۴
	امتیاز چندهسته‌ای	۴۰۷۲.۳	۳۲۸۶.۴	۵۸۳۴.۳	۴۵۹۳
	آزمون SQLite	۱.۴۵ Mrows/sec	۰.۸۳ Mrows/se	۱.۸۸ Mrows/sec	۱.۷۱ Mrows/sec
سرعت حافظه مانا	نرخ نوشتن (۴ کیلوبیت)	MB/s ۴۶.۵	MB/s ۱۱۰	MB/s ۳۳.۹	MB/s ۲۳.۶
	نرخ نوشتن (۴۰۰ کیلوبیت)	MB/s ۱۷۵.۱	۲۲۰۲.۹ MB/s	MB/s ۷۳۲.۴	MB/s ۱۱۵.۱
وضعیت شبکه	سرعت اینترنت	–	۸۴ Mbits/sec	۹۰.۴ Mbits/sec	۲۵۰.۵ Mbits/sec
	سرعت شبکه داخلی	۹۳۴ Mbits/sec	۱۲.۶۱ Gbits/sec	۲.۹۴ Gbits/sec	۱.۰۴ Gbits/sec

از نظر قیمتی نیز می‌بایست این خدمات مورد بررسی و مقایسه قرار بگیرد. هزینه ماشین ابر آروان ماهانه ۲۶،۶۸۰،۰۰۰ ریال، ابر ایران ماهانه ۹،۸۵۵،۳۴۰ ریال بوده است و هزینه ماهانه ابر ستون در بازه‌ی ۲۰،۵۷۱،۰۰۰ تا ۲۶،۵۷۱،۰۰۰ ریال بوده است. نتایج نرمال شده بر اساس قیمت هر کدام از خدمت‌دهندگان در جدول(۴) آورده شده است. برای محاسبه اطلاعات ذکر شده در جدول(۴)، ابتدای ضریب قیمت‌ها محاسبه شده. برای محاسبه‌ی ضرایب به دست آمده از قیمت‌ها، از نرمال سازی بر اساس روش min-max استفاده شده و بعد از اثر دادن قیمت روی نتایج، مقادیر به دست آمده به یک عدد بین ۰ تا ۱ نرمال شده است؛ به این صورت که بهترین نتیجه مقدار یک را گرفته است و دیگر نتایج به صورت خطی مقادیر خود را اتخاذ کرده‌اند.

جدول (٤) : مقایسه‌ی مقادیر نرمال شده بر أساس قیمت ماشین‌های مجازی.

	سنجه‌های ارزیابی	ابر ایران	ابر ستون	ابر آروان
وضعیت پردازش	امتیاز تک هسته‌ای	۱	۰.۴۸	۰.۳۵
	امتیاز چندهسته‌ای	۱	۰.۷۲	۰.۵۱
	آزمون SQLite	۱	۰.۹۲	۰.۷۵
سرعت حافظه مانا	نرخ نوشتن (۴ کیلوبیت)	۱	۰.۱۲	۰.۰۷
	نرخ نوشتن (۴۰۰ کیلوبیت)	۱	۰.۱۳	۰.۰۱
وضعیت شبکه	سرعت اینترنت	۰.۸۹	۰.۴۱	۱
	سرعت شبکه داخلی	۱	۰.۰۹	۰.۰۳

همانطور که از جدول(۴) مشخص است اگر قیمت را در نتایج اثر بدهیم به جز در زمینه‌ی سرعت اینترنت، که ابر آروان بیشترین سرعت را ارائه کرده است، در دیگر زمینه‌های ابر ایران بهترین عملکرد را به نسبت قیمت به نمایش گذاشته است.

مراجع

[1] P. Mell, T. Grance, and others, "The NIST definition of cloud computing," 2011.

[2] P. Leitner and J. Cito, "Patterns in the chaos—a study of performance variation and predictability in public iaas clouds," ACM Transactions on Internet Technology (TOIT), vol. 16, no. 3, pp. 1–23, 2016.

[3] W. Felter, A. Ferreira, R. Rajamony, and J. Rubio, "An updated performance comparison of virtual machines and linux containers," in 2015 IEEE international symposium on performance analysis of systems and software (ISPASS), 2015, pp. 171–172.

[4] Ahuja, S. P., Czarnecki, E., & Willison, S. (2020). Multi-factor performance comparison of amazon web services elastic compute cluster and google cloud platform compute engine. International Journal of Cloud Applications and Computing (IJCAC), 10(3), 1–16.

[5] T. Brummett, P. Sheinidashtegol, D. Sarkar, and M. Galloway, "Performance metrics of local cloud computing architectures," in 2015 IEEE 2nd International Conference on Cyber Security and Cloud Computing, 2015, pp. 25–30. doi: 10.1109/CSCloud.2015.61.

[6] J. P. Mullerikkal and Y. Sastri, "A comparative study of OpenStack and CloudStack," in 2015 Fifth International Conference on Advances in Computing and Communications (ICACC), 2015, pp. 81–84.

[7] M. M. A. AL-Mukhtar and A. A. A. Mardan, "Performance evaluation of private clouds Eucalyptus versus CloudStack," International Journal of Advanced Computer Science and Applications, vol. 5, no. 5, 2014.

[۸] جعفری، مهدی و کلباسی، امیر، **" مطالعه تطبیقی نرم‌افزارهای منبع‌باز استقرار و مدیریت رایانش ابری با استفاده از مدل کیفیت از دیدگاه پردازش کلان داده"**، فصلنامه سیستم های پردازشی و ارتباطی چندرسانه‌ای هوشمند، دوره ۲ شماره ۴، ایران، پاییز ۱۴۰۰.

[9] "ابرآروان." arvancloud. https://www.arvancloud.com/fa/ (Accessed 10 August 2022).

[10] "ابرایران." Abriran. https://www.abriran.com/ (Accessed 11 August 2022).

[11] "ستون." Sotoon. https://sotoon.ir / (Accessed 11 August 2022).

[12] "GitHub - esnet/iperf: iperf3: A TCP, UDP, and SCTP network bandwidth measurement tool" GitHub. https://github.com/esnet/iperf/ (Accessed 11 August 2022).

[13] R. Bankston and J. Guo, "Performance of container network technologies in cloud environments," in 2018 IEEE International Conference on Electro/Information Technology (EIT), 2018, pp. 0277–0283. doi: 10.1109/EIT.2018.8500285.

[14] "GitHub - axboe/fio: Flexible I/O Tester." GitHub. https://github.com/axboe/fio/ (Accessed 11 August 2022).

[15] Y. Wu, B. Jia, and Z. Qi, "IO QoS: A new disk I/O scheduler module with QoS guarantee for cloud platform," in 2012 Fourth International Symposium on Information Science and Engineering, 2012, pp. 441–444. doi: 10.1109/ISISE.2012.105.

[16] "GitHub - cloudharmony/geekbench5: Geekbench 5 Benchmark Harness." GitHub. https://github.com/cloudharmony/geekbench5/ (Accessed 11 August 2022).

[17] A. Bhardwaj and C. R. Krishna, "Virtualization in cloud computing: Moving from hypervisor to containerization—a survey," Arabian Journal for Science and Engineering, vol. 46, no. 9, pp. 8585–8601, 2021.

پانویس ها

[1] Open source community
[2] Infrastructure as a Service
[3] Platform
[4] Platform as a Service
[5] Software as a Service
[6] National Institute of Standards and Technology
[7] Google Cloud
[8] Amazon Web Services
[9] Microsoft Azure
[10] Hypervisor
[11] QoS
[12] T-test
[13] Process

چهارمین کنفرانس ملی انفورماتیک ایران
پژوهشگاه دانش‌های بنیادی، پردیس فرمانیه، تهران
۱۴ و ۱۵ دی ماه ۱۴۰۱

سیستم توصیه‌گر مبتنی بر نشست بر اساس شبکه توجه

ریحانه داورزنی[۱]، سعیدرضا خردپیشه[۲]، هادی فراهانی[۳]

[۱] کارشناسی ارشد، گروه علوم کامپیوتر و داده‌ها، دانشکده علوم ریاضی، دانشگاه شهید بهشتی ، تهران، rdavarzani7@gmail.com

[۲] استادیار، گروه علوم کامپیوتر و داده‌ها، دانشکده علوم ریاضی، دانشگاه شهید بهشتی ، تهران، s_kheradpisheh@sbu.ac.ir

[۳] استادیار، گروه علوم کامپیوتر و داده‌ها، دانشکده علوم ریاضی، دانشگاه شهید بهشتی ، تهران، h_farahani@sbu.ac.ir

چکیده

در سال‌های اخیر، با توجه به افزایش حجم داده‌ها در اینترنت، سیستم‌های توصیه‌گر (RS) به ابزاری مهم برای یافتن اطلاعات مورد نیاز افراد تبدیل شده‌است. روش‌های مختلفی برای مدل‌سازی سیستم‌های توصیه‌گر در طول سالیان ارائه شده است که هر کدام مزایا و معایب متفاوتی دارند. به‌عنوان مثال، سیستم‌های توصیه‌گر سنتی، تعاملات کاربر-آیتم را به‌صورت ایستا مدل‌سازی می‌کنند و تنها می‌توانند ترجیحات عمومی کاربران را به تصویر بکشند. ترجیحات عمومی نشان دهنده رفتارهای درازمدت کاربر است که در طول زمان عموما ثابت است و تنها تکیه بر مدل‌های بلندمدت برای ارائه توصیه‌های خوب کافی نیست، زیرا مدل‌های بلندمدت نمی‌توانند رفتارهای کوتاه‌مدت کاربر را در بسیاری از موقعیت‌ها پوشش دهند. به همین جهت از مفهومی به نام نشست استفاده می‌کنیم. یک نشست متشکل از چندین تعامل کاربر-آیتم است که با هم در یک دوره زمانی پیوسته اتفاق می‌افتند. ما در این مقاله قصد داریم یک سیستم توصیه‌گر مبتنی بر نشست ارائه کنیم که از شبکه توجه استفاده می‌کند، به این منظور که برای هر آیتم، یک وزن با توجه به اهمیت آن آیتم برای کاربر اختصاص می‌دهیم. برای ارزیابی مدل از سه مجموعه داده استفاده کردیم و با روش‌های قبلی به مقایسه پرداختیم که نتایج نسبت به قبل بهبود داشته است.

کلمات کلیدی

سیستم‌های توصیه‌گر، سیستم‌های توصیه‌گر مبتنی بر نشست، سیستم‌های توصیه‌گر متوالی، مدل‌های شبکه عصبی عمیق، مدل توجه، سیستم توصیه‌گر آگاه از زمینه.

۱- مقدمه

سیستم‌های توصیه‌گر به ابزاری اساسی برای اتخاذ انتخاب‌ها و تصمیم‌های آموزنده‌تر، کارآمدتر و مؤثرتر در تقریباً هر جنبه روزمره زندگی، کار، عملیات تجاری، مطالعه و سرگرمی تبدیل شده‌اند. نقش آن‌ها در عصر پربار اقتصاد دیجیتال بیش از پیش مهم شده است زیرا کاربران مجبورند از میان محتویات، محصولات و خدمات معمولاً عظیم و به سرعت در حال افزایش (که به‌طور یکسان آیتم نامیده می‌شوند) انتخاب‌های خود را انجام دهند.

انواع حوزه‌های تحقیقاتی سیستم‌های توصیه‌گر با موفقیت بزرگ پدید آمده‌اند، مانند سیستم‌های توصیه‌گر مبتنی بر محتوا [11]، سیستم‌های توصیه‌گر مبتنی بر فیلترینگ مشارکتی [13] و سیستم‌های توصیه‌گر ترکیبی[1] که دو مورد اول را ترکیب می‌کنند. با این حال، این سیستم‌های توصیه‌گر تمایل دارند از تمام تعاملات تاریخی کاربر-آیتم (تعاملات به اختصار، اشاره به اقدامات مستقیم یا غیرمستقیم کاربر روی آیتم‌ها دارند) برای یادگیری ترجیحات طولانی مدت و ثابت هر کاربر روی آیتم‌ها استفاده کنند. چنین رویه‌ای اغلب با این فرض اساسی مرتبط است که همه تعاملات تاریخی یک کاربر برای ترجیح فعلی او به یک اندازه مهم هستند. این ممکن است در موارد دنیای واقعی واقعیت نداشته باشد و دو دلیل عمده وجود دارد. اولاً، انتخاب کاربر در مورد آیتم‌ها نه تنها به ترجیحات تاریخی بلندمدت او بستگی دارد، بلکه به اولویت‌های اخیر کوتاه‌مدت و زمینه حساس به زمان (به‌عنوان مثال، موارد اخیراً مشاهده یا خریداری شده) بستگی دارد. این اولویت کوتاه مدت در آخرین تعاملات کاربر تعبیه شده است، که اغلب بخش کوچکی از تعاملات تاریخی او را تشکیل می‌دهد. دوم، ترجیح کاربر نسبت به آیتم‌ها به جای ثابت، پویا است، که در طول زمان در حال تغییر است.

برای پرکردن این شکاف‌ها در سیستم‌های توصیه‌گر، سیستم‌های توصیه‌گر مبتنی بر نشست[۱] (SBRS) با توجه فزاینده‌ای در سال‌های اخیر پدیدار شده‌اند. برخلاف سیستم‌های توصیه‌گر فوق الذکر، SBRS ترجیحات کاربران را از نشست‌ها مرتبط و ایجاد شده در طول فرآیند مصرف یاد می‌گیرند. با در نظر گرفتن هر نشست به‌عنوان واحد ورودی اصلی، یک SBRS می‌تواند هم ترجیحات کوتاه‌مدت کاربر را از نشست‌های اخیرش و هم دینامیک اولویت‌ها را که منعکس‌کننده تغییر اولویت‌ها از یک نشست به نشست دیگر است، برای توصیه‌های دقیق‌تر و به موقع‌تر ثبت کند.

مطالعات اخیر اهمیت استفاده از شبکه‌های عصبی بازگشتی[۲] (RNN) را در طیف گسترده‌ای از سیستم‌های توصیه‌گر برجسته کرده‌اند، که در میان آن‌ها استفاده از RNN در سیستم‌های توصیه‌گر مبتنی بر نشست منجر به پیشرفت قابل توجهی در چند سال گذشته شده است [4]. اگرچه ثابت شده‌است که مدل‌های RNN در به‌دست آوردن علایق عمومی کاربران از طریق توالی‌ای از اقدامات مفید هستند [16]، اما یادگیری پیش‌بینی با استفاده از نشست‌ها همچنان یک مشکل چالش برانگیز است که عمدتاً به دلیل عدم قطعیت ذاتی در رفتار کاربر و اطلاعات محدود ارائه شده توسط نشست‌های مرورگر است. در مسائل معمولی SBRS، نشست شامل دنباله‌ای از آیتم‌های نامگذاری شده است و علایق کاربر در این بازخوردهای ضمنی (مثلاً کلیک‌ها) پنهان می‌شود. به‌منظور بهبود بیشتر دقت پیش‌بینی مدل‌های RNN، داشتن توانایی یادگیری علایق بلندمدت و علایق کوتاه‌مدت چنین بازخوردهای ضمنی مهم است. هم علایق کوتاه‌مدت و هم علایق بلندمدت کاربران برای توصیه اهمیت زیادی دارد [6]، اما معماری‌های RNN سنتی برای تمایز و بهره

برداری همزمان از این دو نوع علاقه طراحی نشده‌اند. در این مقاله، ما در نظر داریم این مشکل را با معرفی یک مکانیسم توجه در مدل SBRS حل کنیم[3](MSAN)، که می‌تواند علایق کاربر را به‌طورکلی و علایق فعلی وی را به‌طور همزمان در نظر بگیرد.

در چند سال گذشته، حجم فزاینده ای از توجه تحقیقاتی به چالش مسئله SBRS معطوف شده است، با توجه به فرضیه مدل‌سازی SBRSها، رویکردهای رایج را می‌توان به دو دسته تقسیم کرد: مدل‌های جهانی که بر شناسایی علایق کاربران به‌طورکلی تمرکز دارند، و مدل‌های بومی سازی شده که بر علایق موقتی کاربران تاکید دارند.

یکی از رویکرد های به‌دست آوردن علایق عمومی کاربر از طریق روش‌های فیلتر مشارکتی بر اساس کل تاریخچه خرید/کلیک کاربران است. به‌عنوان مثال، رویکرد فاکتورگیری ماتریس [7] از بردار های پنهان برای نشان دادن علایق عمومی استفاده می‌کند که از طریق فاکتورسازی یک ماتریس کاربر-آیتم متشکل از کل داده تراکنش های قبلی تخمین زده می شوند. رویکرد دیگر، روش‌های هم‌سایگی [14] نامیده می شود که سعی می‌کند بر اساس شباهت‌های آیتم‌ها که از رویدادهای مشابه آیتم‌ها در نشست‌ها محاسبه می‌شود، توصیه‌هایی ارائه کنند. مدل‌های فوق علایق عمومی یا علایق فعلی کاربران را بررسی می‌کنند. با این حال، توصیه‌کنندگان مبتنی بر علایق فعلی به ندرت تعاملات متوالی بین آیتم‌هایی را که در نشست نیستند در نظر می‌گیرند، اگرچه توصیه‌کنندگان مبتنی بر علایق عمومی در جلب سلیقه عمومی کاربران خوب عمل می‌کنند، اما به سختی می‌توانند توصیه‌های خود را با خریدهای اخیر کاربران تطبیق دهد.

در حالت ایده‌آل، یک توصیه‌کننده خوب باید بتواند رفتار متوالی را بررسی کند، و همچنین علایق عمومی کاربران را برای توصیه در نظر بگیرد، زیرا این دو عامل ممکن است با یکدیگر تعامل داشته باشند تا بر کلیک بعدی کاربران تأثیر بگذارند. بنابراین، برخی از محققان سعی کردند با در نظر گرفتن هر دو نوع علایق کاربر، مدل‌های SBRS را بهبود بخشند. در [10] یک مدل ترکیبی (FPMC) را پیشنهاد شده است که از ترکیب رویکرد فاکتورگیری ماتریس و رویکرد زنجیره مارکوف برای مدل‌سازی رفتار متوالی و علایق عمومی برای توصیه‌های بعدی استفاده می‌کند، بنابراین عملکرد بهتری نسبت به در نظر گرفتن علایق کوتاه‌مدت یا علایق بلندمدت به تنهایی به‌دست می‌آورد. با این حال، این روش فقط می‌تواند رفتارهای متوالی محلی را بین اقدامات مجاور، بدون در نظر گرفتن اطلاعات کلی منتقل شده توسط نشست، مدل کند.

در مرجع [3] برای مدل سازی داده‌های نشست، از شبکه‌های عصبی بازگشتی عمیق با واحد بازگشتی دروازه‌ای[4] استفاده کردند، که نمایش نشست را مستقیماً از کلیک‌های قبلی در نشست داده شده یاد می‌گیرد و توصیه‌هایی را برای اقدام بعدی ارائه می‌دهد. این اولین تلاش برای اعمال شبکه‌های RNN برای حل مشکل SBRS است، به لطف قابلیت مدل‌سازی متوالی ارائه شده توسط RNNها، مدل آن‌ها می‌تواند رفتار گذشته کاربران را هنگام پیش‌بینی حرکت بعدی در نظر بگیرد. در مرجع [12] یک تکنیک افزایش داده را برای بهبود عملکرد RNNها برای توصیه‌های مبتنی بر نشست پیشنهاد می‌کنند. در مدل‌های شبکه عصبی ذکر شده در بالا حالت پنهان در آخرین مرحله زمانی، حاوی اطلاعاتی در مورد دنباله با تمرکز بر روی قسمت‌های نزدیک به کلیک بعدی است، بنابراین برخی از ویژگی‌های عمومی مورد علاقه آیتم‌های دارای فاصله طولانی ممکن است فراموش شوند. برای حل این مشکل، مدل‌های مختلفی برای ثبت ارتباط بین آیتم‌ها و علایق عمومی معرفی شده‌اند. هو[5] و همکاران در [4] یک شبکه عصبی (SWIWO) برای یادگیری زمینه[6] نشست-کاربر پیشنهاد دادند. زمینه نشست را از طریق ترکیب‌کردن تمام آیتم‌های تعبیه[7] شده‌ی نشست جاری می سازد، که به هر آیتم وزن ثابتی بر اساس فاصله نسبی با آیتم هدف می‌دهد. لی[8] و همکاران در [8] یک مدل رمزگذار-رمزگشا مبتنی بر شبکه عصبی بازگشتی (NARM) پیشنهاد کرده‌اند که آخرین حالت پنهان را از RNN به‌عنوان رفتار متوالی می‌گیرد و از حالت‌های پنهان کلیک‌های قبلی برای محاسبه توجه استفاده می‌کند تا هدف اصلی (علائق عمومی) را در یک نشست معین به‌دست آورد. در مرجع [9] یک مدل اولویت حافظه کوتاه مدت (STAMP) ارائه شده است که از علاقه فعلی کاربر که با آخرین کلیک منعکس می‌شود استفاده می‌کند. هوانگ[9] و همکاران در [5] یک سیستم توصیه‌کننده با استفاده از مدل بلندمدت و کوتاه‌مدت مبتنی بر توجه (ASLM) مبتنی بر شبکه عصبی بازگشتی (RNN) پیشنهاد کرده‌اند که دارای اولویت و لایه‌های توجه است.

مدل ما یعنی MSAN ، تفاوت‌های قابل توجهی با مدل‌های ذکر شده دارد. SWIWO وزن هر آیتم را در نشست به صورت ثابت تعیین می‌کند که در عمل قابل بحث است. در MSAN، مکانیسم توجه ارائه شده می‌تواند با در نظر گرفتن صریح همبستگی بین هر کدام از کلیک‌های گذشته و آخرین کلیک، و محاسبه وزن‌های پویا برای نشست داده شده، به کاهش این تناقض کمک کند. در NARM فقط علایق عمومی را به تصویر می‌کشد در حالی که MSAN به صراحت بر علاقه جاری منعکس شده توسط آخرین کلیک تأکید می‌کند تا ویژگی های ترکیبی علایق فعلی و کلی را از کلیک های قبلی در نشست فعلی و نشست‌های قبلی در صورت وجود، به تصویر بکشد. در STAMP، تنها اطلاعاتی که از نشست می‌گیرد، آیتم‌هایی است که کاربر کلیک کرده است ولی در مدل ما از اطلاعات زمینه و اطلاعات گرفته شده از نشست‌های قبل کاربر برای ارائه توصیه بهتر و کارآمدتر استفاده می‌کند. در ASLM از شبکه عصبی بازگشتی به منظور مدل‌سازی برای ارائه توصیه‌های بعدی استفاده می‌کند، اما همان طور که در این بخش نیز به آن اشاره کردیم، معماری RNN در تشخیص تمایز بین علایق کوتاه‌مدت و علایق بلندمدت به خوبی عمل نمی‌کند به همین دلیل در مدل ما از مکانیسم توجه ارائه شده استفاده می‌شود. در اکثر مدل‌های مبتنی بر نشست فرض ناشناس بودن کاربر را در نظر می‌گیرند، زیرا در برخی از مجموعه داده‌ها، کاربر قابل شناسایی نیست و رفتارهای تاریخی مرتبط با آن‌ها را نمی‌توان به دست آورد. اما در بسیاری از مجموعه داده‌ها، اطلاعات کاربران در دسترس است و این اطلاعات در این مدل‌ها استفاده نمی‌شود. بنابراین ما یک شبکه توجه مبتنی بر نشست‌های چندگانه (MSAN) را ارائه می‌دهیم. MSAN نه تنها از اطلاعات جلسه فعلی کاربر استفاده می کند، بلکه از رفتارهای تاریخی مرتبط کاربر نیز استفاده می‌کند که می‌تواند به طور موثر دقت پیش بینی را بهبود بخشد.

دربخش ۲ی این مقاله به معرفی مدل پیشنهادی می‌پردازیم و دربخش ۳، مدل پیشنهادی را با مدل‌های پیشین مقایسه و نتایج ارزیابی ارایه می‌گردد.

۲- مدل Multi STAMP

۲-۱- توصیف نمادها

در یک سیستم توصیه‌گر مبتنی بر نشست هر نشست با $S = [s_1, s_2, \dots, s_N]$ علامت‌گذاری می‌شود که متشکل از دنباله‌ای از اقدامات (آیتم‌هایی که توسط کاربر کلیک می‌شوند) است، که s_i نشان‌دهنده یک آیتم است که در مرحله زمانی i کلیک شده‌است.

$S_t = \{s_1, s_2, \dots, s_t\}$ کـه $1 \le t \le N$، برابرست با مجموعه آیتم‌های انتخابی در نشست S تا زمان t. $V = \{v_1, v_2, \dots, v_{|V|}\}$ مجموعه‌ای از آیتم‌های منحصر به فرد را در سیستم نشان می‌دهد که واژه‌نامه آیتم نامیده می‌شود. $X = \{x_1, x_2, \dots, x_{|V|}\}$ بردارهای تعبیه‌شده را با توجه به واژه‌نامه آیتم V نشان می‌دهد. مدل ارائه شده، یک تعبیه d بعدی $x_i \in \mathbb{R}^d$ برای هر یک از آیتم‌های i در V می‌آموزد. به‌طور خاص، نماد $x_t \in \mathbb{R}^d$ نشان‌دهنده تعبیه آخرین کلیک s_t از نشست جاری S_t است. هدف مدل ما پیش‌بینی کلیک احتمالی بعدی (یعنی s_{t+1}) بر اساس نشست فعلی S_t و نشست‌های قبل از آن است. به‌طور دقیق، مدل ما به‌عنوان طبقه‌بندی ساخته و آموزش داده می‌شود که یاد می‌گیرد برای هر یک از آیتم‌ها در واژه‌نامه آیتم V امتیازی به‌دست آورد. $\hat{y} = \{\hat{y}_1, \hat{y}_2, \dots, \hat{y}_{|V|}\}$ بردار امتیاز خروجی را نشان می دهد به‌طوری‌که $\hat{y}_i$ نظیر امتیاز آیتم v_i است. پس از بدست آوردن این پیش‌بینی، عناصر $\hat{y}$ به ترتیب نزولی رتبه‌بندی می‌شوند و آیتم‌های مربوط به k امتیاز برتر برای توصیه استفاده می‌شوند.

۲-۲- مدل ارائه شده

همان‌طور که در شکل (۱) مشاهده می‌گردد مدل ارائه شده به دو قسمت کلی تقسیم شده است. یک بخش مربوط به نشست فعلی کاربر است و بخش دیگر مربوط به نشست‌های قبلی همان کاربر که با هم الحاق شده‌اند. در ابتدا سازوکار مربوط به بخش اول را مورد بررسی قرار می‌دهیم. به این منظور، به دو مورد از ورودی‌های مدل ارائه شده می‌پردازیم. مدل، دو تعبیه m_t و m_s را به‌عنوان ورودی می‌گیرد که در آن m_s نشان دهنده علایق کلی کاربر با توجه به نشست جاری است که به‌صورت زیر تعریف می‌شود:

$$m_s = \frac{1}{t}\sum_{i=1}^{t} x_i$$

(۱)

که x_i تعبیه d بعدی برای هر یک از آیتم‌ها می‌با شد. نماد m_t نشان دهنده علاقه فعلی کاربر در آن نشست است، در این مطالعه از آخرین کلیک x_t برای نشـان دادن علاقه فعلی کاربر اسـتفاده می‌شـود: $m_t = x_t$. توجه نمایید که x_t را حافظه کوتاه مدت علایق کاربر می‌نامیم.

در این مدل از یک مکانیزم توجه، به نام شـبکه توجه اسـتفاده می‌کنیم. شبکه توجه پیشنهادی از دو جزء تشکیل شده است: یک شبکه عصبی پیشخور ساده[10] (FNN) که مسئول تولید وزن توجه برای هر یک از آیتم‌ها در نشست جاری S_t است، و یک تابع ترکیبی توجه که به‌طورکلی مسئول محاسبه علایق کاربر مبتنی بر توجه است که با m_a نشان می‌دهیم. FNN مورد استفاده برای محاسبه توجه به صورت زیر تعریف می‌شود:

$$\alpha_i = W_0\, \sigma(W_1 x_i + W_2 x_t + W_3 m_s + b)$$

(۲)

که $x_i \in \mathbb{R}^d$ نشـان دهنده آیتم i-ام در $s_i \in S_t$، $x_t \in \mathbb{R}^d$ نشـان دهنده آخرین کلیک، $W_0 \in \mathbb{R}^{1\times d}$ بردار وزن، $W_1, W_2, W_3 \in \mathbb{R}^{d\times d}$ ماتریس‌های وزنی، $b_a \in \mathbb{R}^d$ بردار سـوگیری[11] و $\sigma(.)$ نشـان‌دهنده تابع سیگموئید است. α_i نشان دهنده ضریب توجه آیتم x_i در نشست جاری است.

از معادله قبل می‌توان دریافت که ضـرایب توجه آیتم‌ها در یک نشـسـت بر اساس تعبیه آیتم هدف x_i، آخرین کلیک x_t و نمایش نشست m_s محاسبه می‌شـود، بنابراین، می‌تواند ارتباط بین آیتم مورد نظر و حافظه بلند مدت/کوتاه مدت علایق کاربر را به تصویر کشد.

با توجه به [15]، اسـتدلال می‌کنیم که آیتم‌های مشـابه ممکن اسـت تأثیرات متفاوتی بر کاربران مختلف داشته باشد. به‌عنوان مثال، فرض کنید که کاربر a به‌دلیل علاقه، آیتم x را برای خود می‌خرد، در حالی که کاربر b آیتم x را به‌عنوان هدیه برای دیگران می‌خرد. در چنین حالتی منطقی اسـت که اسـتنباط کنیم آیتم x هنگام پیش‌بینی آیتم‌های بعدی، وزن متفاوتی برای کاربران a و b دارد. به‌همین منظور از ترکیب α با بردار تعبیه شناسه(آی‌دی) هر کاربر نیز استفاده می‌کنیم:

$$\alpha_i = \frac{\exp(u^T \alpha_i)}{\sum_{p\epsilon\alpha} \exp(u^T \alpha_p)}$$

(۳)

پس از بدسـت آوردن بردار ضـرایب توجه $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_t)$ نسبت به نشـسـت جاری S_t، بردار علایق کاربر m_a به‌صـورت زیر محاسـبه می‌گردد:

$$m_a = m_s + \sum_{i=1}^{t} \alpha_i x_i$$

(٤)

حال m_a که علایق کاربر مبتنی بر توجه اسـت و m_tکه علایق فعلی کاربر اسـت، با دو شـبکه MLP به‌منظور انتزاع ویژگی پردازش می‌شـوند. سـاختار شبکه سلول‌های MLP با یکدیگر یکسان هستند، با این تفاوت که تنظیمات پارامتر مسـتقلی دارند. یک MLP سـاده بدون لایه پنهان برای انتزاع ویژگی استفاده می‌شود، عملیات روی m_a به صورت زیر تعریف می‌شود:

$$h_s = f(W_a m_a + b_a)$$

(۵)

که $h_s \in \mathbb{R}^d$ نشان‌دهنده حالت خروجی، $W_a \in \mathbb{R}^{d\times d}$ ماتریس وزن و $b_a \in \mathbb{R}^d$ بردار سوگیری می‌با شد. $f(.)$ یک تابع فعال سازی غیرخطی اسـت که در این‌جا از تانژانت هایپربولیک اسـتفاده می‌کنیم. بردار حالت h_t با توجه به m_t را می‌توان مشابه h_s محاسبه کرد.

برای بخش دوم که مربوط به الحاق نشست‌های قبلی کاربر است هم مانند روندی که تا اینجا برای نشـسـت فعلی کاربر گفته‌شـد، عمل می‌کنیم با این تفاوت که در معادله (۲)، ترم مربوط به آخرین کلیک دیگر استفاده نمی‌شود.

در انتها خروجی شـبکه‌های MLP را با یکدیگر جمع می‌کنیم و سـپس به‌عنوان ورودی به یک تابع سافتمکس می‌دهیم تا خروجی y بدست آید:

$$\hat{y} = softmax(\hat{z})$$

(٦)

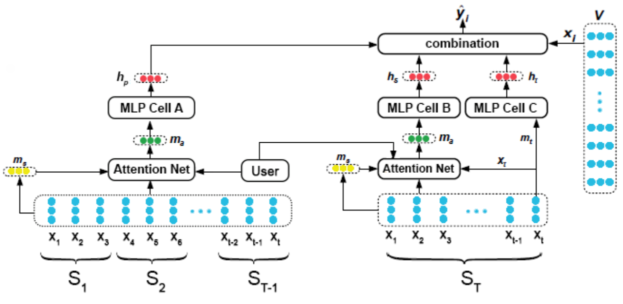
که $\hat{y} \in \mathbb{R}^{|V|}$ بردار خروجی مدل را نشـان می‌دهد که توزیع احتمال را بر روی آیتم‌های $v_i \in V$ نشان می‌دهد. هر عنصر $\hat{y}_i \in \hat{y}$ احتمال رخدادی را نشان می‌دهد که آیتم v_i به‌عنوان کلیک بعدی در این نشست ظاهر شود. برای هر نشـسـت معین $S_t \in S\ (t \in [۱, \dots, N])$ تابع ضـرر به‌عنوان آنتروپی متقاطع نتایج پیش‌بینی به‌صورت زیر تعریف می‌شود:

$$\mathcal{L}(\hat{y}) = -\sum_{i=۱}^{|V|} y_i \log(\hat{y}_i) + (۱ - y_i)\log(۱ - \hat{y}_i)$$

(۷)

که در آن y یک بردار با مقادیر صفر یا یک کدگذاری شده را نشان می‌دهد که منحصراً توسط $s_{t+۱} \in S$ فعال می‌شود. به‌عنوان مثال، اگر s_{t+1} عنصر i– امین v_i را در واژه‌نامه آیتم V نشـان دهد، سـپس $y_k = ۱$، اگر $i = k$، و $y_k = ۰$ اگر $i \neq k$. سپس یک بهینه ساز گرادیان کاهشی تصادفی[۱۲] برای بهینه‌سازی ضرر آنتروپی متقاطع انجام می‌شود.

شکل (۱)، مدل پیشنهادی این مقاله یعنی MSAN، را نشان می‌دهد.



شکل (۱) : تصویر شماتیک از مدل MSAN

۳- نتایج

۳-۱- مجموعه داده‌ها و پیش‌پردازش

به‌منظور سـنجش عملکرد مدل از سـه مجموعه داده Diginetica و onlineRetail و ThirtyMusic اسـتفاده می‌کنیم. این داده‌ها شـامل اطلاعاتی همچون تعاملات کاربران با آیتم‌ها در هر نشست، آی‌دی هر کاربر و اطلاعات زمینه هر آیتم انتخاب شده می‌باشند. خلاصه‌ای از آمار این داده‌ها در جدول ۱ آورده شده است.

جدول (۱) : آمار مجموعه دادگان

Datasets	#Users	#Items	#Actions	Avg. length
Diginetica	۵۲۱۲۸	۱۶۸۸۲	۲۴۳۲۹۷	۴،۴۶
onlineRetail	۴۲۵۷	۳۱۸۰۲	۴۰۱۶۱۱	۲۱،۳۹
TM 1/10	۴۰۲۶	۷۰۱۴۳	۲۱۲۱۴۲	۱۲،۷۰
TM 1/40	۱۰۲۴	۲۷۱۴۳	۵۳۰۲۷	۱۲،۲۹

برای پیش‌پردازش ابتدا ما نشست‌های به طول ۱ و آیتم‌هایی که کمتر از ۵ بار در هر سه مجموعه داده ظاهر می‌شوند را فیلتر می‌کنیم. سپس سطرهایی که آیتم null دارند (مخصوصا در بخش شناسه(آی‌دی) کاربر که بر اساس کاربر تقسیم‌بندی می‌شود) حذف می‌شود.

۳-۲- معیارهای ارزیابی

برای ارزیابی سیستم‌های توصیه‌گر فرمول‌های مختلفی وجود دارد که بر اساس نوع خروجی این سیستم‌ها مورد استفاده قرار می‌گیرد. همان‌طور که گفته شد، این مدل در انتها یک لیست رتبه‌بندی شده از آیتم‌ها را تولید می‌کند، به همین منظور ما از معیارهای زیر برای ارزیابی عملکرد مدل استفاده می‌کنیم:

- نرخ ضربه[۱۳] (HR): نرخ ضربه نشان‌دهنده نسبت موارد آزمایشی است که آیتم‌های به درستی توصیه‌شده در k جایگاه برتر فهرست رتبه‌بندی قرار دارند.

$$HR@k = \frac{n_{hit}}{N}$$

(۸)

که N تعداد داده‌های تست در سیستم توصیه‌گر G و n_{hit} تعداد مواردی را نشان می‌دهد که آیتم‌های مورد نظر در فهرست k تای اول قرار دارند. برخورد زمانی اتفاق می‌افتد که t در k موقعیت بالای فهرست رتبه‌بندی G ظاهر شود.

- میانگین رتبه معکوس[۱۴](MRR): میانگین رتبه‌های معکوس آیتم مورد نظر t. اگر رتبه بالاتر از k با شد، رتبه معکوس صفر می شود. MRR یک معیار نرمال شـده از محدوده [۰،۱] اسـت، ، افزایش در مقدار آن نشـان‌دهنده این اسـت که اکثر "برخوردها" در ترتیب رتبه‌بندی لیسـت توصـیه‌ها بالاتر ظاهر می‌شـوند، که نشـان‌دهنده عملکرد بهتر سـیسـتم توصیه‌کننده مربوطه است.

$$MRR@k = \frac{۱}{N}\sum_{t \in G}\frac{۱}{Rank(t)}$$

(۹)

۳-۳- روش‌های پیشین

برای تایید کارایی مدل پیشنهادی، آن را با چند نمونه از مدل‌های معروف استفاده شده در سیستم‌های توصیه‌گر مقایسه کرده‌ایم که عبارت اند از:

- FPMC: این روش ترجیحات کاربر را از طریق فاکتورسـازی ماتریس و اطلاعات متوالی از طریق زنجیره مارکوف مرتبه اول به‌طور همزمان مدل می‌کند و سـپس آن‌ها را به‌روش خطی برای توصـیه سـبد بعدی ترکیب می‌کند [10].
- NARM : یک مدل پیشـرفته مبتنی بر RNN که از مکانیسـم توجه برای گرفتن هدف اصـلی از حالت‌های پنهان اسـتفاده می‌کند و آن را با رفتار متوالی به‌عنوان نمایش نهایی برای تولید توصـیه‌ها ترکیب می‌کند [8].
- STAMP: یک مدل اولویت حافظه کوتاه مدت است که از علاقه فعلی کاربر که با آخرین کلیک منعکس می‌شود استفاده می‌کند [9].
- SR-SAN: یک سیستم توصیه‌گر مبتی بر شبکه خودتوجه برای توصیه مبتنی بر جلسـه اسـت که نه تنها قادر به یادگیری وابسـتگی‌ها در بین آیتم‌های مجاور است، بلکه قادر به یادگیری وابستگی‌ها در بین آیتم‌های راه دور بدون توجه به فاصـله آن‌ها نیز اسـت. از بردار تعبیه آخرین آیتم،

برای نشان دادن علاقه جاری و علاقه کلی به جای تعبیه کل نشست که از تعبیه علاقه فعلی و تعبیه علایق کلی تشـکیل شـده اسـت، اسـتفاده می‌کند[2].

۳-٤- جزئیات پیاده‌سازی

یادگیری مدل، با استفاده از روش بهینه‌ساز آدام و با نرخ یادگیری ۱e-۴ صورت گرفته است. اندازه بعد تعبیه برابر ۲۰۰ ، اندازه دسته[15] برابر ۴۰۰ ، تعداد دوره[16] برابر ۴۵ ، در نظر گرفته شده است. همه ماتریس-های وزن‌دهی از یک توزیع نرمال در بازه $N(۰,۰/۰۰۱^{۲})$ مقداردهی اولیه شده‌اند و همه سوگیری‌ها روی صفر تنظیم می‌شوند. همه آیتم‌های تعبیه-شده به‌طور تصادفی با توزیع نرمال $N\left(۰,۰/۰۷^{۲}\right)$ مقداردهی اولیه می‌شوند، که سپس به‌طور مشترک با پارامترهای دیگر آموزش داده می‌شوند.

۳-٥- ارزیابی مدل پیشنهادی

همان طور که گفته شد، مدل از دو بخش کلی تشکیل شده است. برای اینکه میزان تاثیر استفاده از نشست‌های قبلی را بررسی کنیم، یک مدل به نام MSAN - طراحی می‌کنیم که تنها تفاوتش با مدل MSAN در این می‌باشد که بخش الحاق شده نشست‌های قبلی را دارا نمی‌باشد. نتایج این مقایسه را در جدول ۲ مشاهده می‌کنید:

جدول (۲) : تاثیرات نشست‌های گذشته

Dataset	MSAN		MSAN -	
Measures	HR@۲۰	MRR@۲۰	HR@۲۰	MRR@۲۰
Diginetica	۶۱,۲۰	۲۹,۰۹	۶۱,۳۹	۲۹,۲۲
onlineRetail	۲۵,۸۴	۱۱,۰۱	۲۵,۰۸	۱۰,۵۹
TM ۱/۱۰	۱۸,۴۱	۱۰,۱۹	۱۶,۶۲	۸,۳۰
TM ۱/۴۰	۱۵,۰۰	۸,۷۲	۱۴,۲۰	۷,۶۷

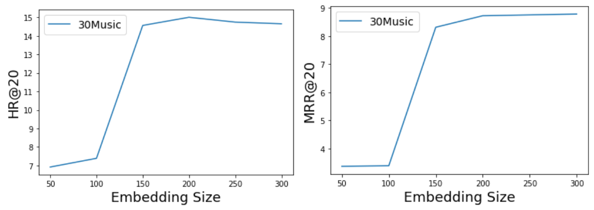
با توجه به جدول ۲، مدلی که علاوه بر نشـسـت فعلی، از نشـسـت‌های گذشـته نیز در ارائه توصـیه اسـتفاده می‌کند، نتایج بهتری نسـبت به مدلی که فقط از نشـسـت فعلی اسـتفاده می‌کند، به‌دسـت می‌آورد. در مجموعه داده‌ی Diginetica اما به این صورت نیست و نتایج تقریبا با هم برابر می‌باشد. دلیل آن، این است که در این مجموعه داده، بیش از ۹۷ در صد کاربران، تنها دارای یک نشست می‌باشند، یعنی تقریبا هر دو مدل شبیه به هم هستند، که همین باعث این نتیجه اسـت. در مجموعه داده‌ی onlineRetail حدود ۴۰ درصـد کاربران و در مجموعه داده‌ی ThirtyMusic حدود ۳۰ در صد کاربران دارای تنها یک نشست می‌باشند. پس می‌توان نتیجه گرفت، هرچه تعداد کاربران با چند نشست بیشتر باشد، تفاوت دو مدل بیشتر است.

در جدول ۳، نتایج حاصل از مدل MSAN و مدل‌هایی که در بخش ۳-۳ ارائه شده است را مشاهده می‌کنید. همان‌طور که مشاهده می شود مدل MSAN در اکثر مجموعه داده‌ها، نتیجه برتر را به دست آورده است که نشان دهنده‌ی عملکرد قابل قبول مدل MSAN می‌باشـد، اما در مجموعه داده‌ی onlineRetail مدل SR-SAN نتایج بهتری را نسبت به مدل ما به دست آورده است و دلیل آن، این است که میانگین طول هر نشست در این مجموعه داده برابر ۲۱,۳۹ می‌باشـد و همین امر باعث این نتیجه می‌باشـد. در مدل MSAN به این دلیل که از نشست‌های قبل نیز در مدل استفاده می شود، و چون در این مجموعه داده اندازه میانگین طول هر نشست نسبت به مجموعه داده‌های دیگر بیشتر می‌باشد، خروجی مدل ما در این مجموعه داده نسبت به روش SR-SAN ضعیف‌تر می‌باشد.

جدول (۳) : مقایسه MSAN با کارهای پیشین

Dataset	Diginetica	onlineRetail	TM 1/10	TM 1/40
Measures	HR@۲۰	HR@۲۰	HR@۲۰	HR@۲۰
FPMC	۲۰,۰۵	۷,۵۵	۲,۸۷	۳,۷۰
NARM	۶۰,۴۸	۲۵,۵۵	-	-
STAMP	۵۸,۰۵	۲۳,۸۵	۱۴,۴۱	۱۲,۹۰
SR-SAN	۵۴,۱۶	۳۱,۷۹	-	-
MSAN	۶۱,۲۰	۲۵,۷۰	۱۸,۴۱	۱۵,۰۰
Measures	MRR@۲۰	MRR@۲۰	MRR@۲۰	MRR@۲۰
FPMC	۷,۱۱	۲,۹۵	۱,۱۷	۲,۷۸
NARM	۲۲,۱۶	۹,۲۹	-	-
STAMP	۲۶,۹۰	۹,۹۸	۸,۶۹	۷,۷۷
SR-SAN	۲۰,۶۷	۱۲,۲۹	-	-
MSAN	۲۹,۰۹	۱۰,۸۳	۱۰,۱۹	۸,۷۲

در شـکل(۲)، اثربخشـی اندازه بعد تعبیه از ۵۰ تا ۳۰۰ را در مجموعه داده‌ی ThirtyMusic مورد بررسـی قرار می‌دهیم. برای سـادگی، اندازه بعد تعبیه کاربر و آیتم را یکسان فرض کرده‌ایم. می‌توان مشاهده کرد که ابعاد بالا برای کاربر و آیتم می‌تواند خروجی بهتری ارائه کند. در آزمایش‌ها، اندازه بعد تعبیه را به منظور توازن بین هزینه محاسباتی و کیفیت توصیه، ۲۰۰ قرار داده‌ایم.



شکل (۲) : عملکرد مدل در ThirtyMusic با اندازه بعد تعبیه مختلف

٤- نتیجه‌گیری

در این مقاله یک سیستم توصیه‌گر مبتنی بر نشست که از شبکه توجه استفاده می‌کند ارائه شد. از معایب SBRSها می‌توان به این مورد اشاره کرد که ممکن است در نشست اطلاعات لازم جهت ارائه توصیه را نداشته باشد، برای مقابله با این مسئله، در مدل MSAN ما از اطلاعات حاصل از نشست‌های قبلی کاربر، در نشست فعلی کاربر استفاده کردیم. همچنین از اطلاعات زمینه نشست هم برای ارائه بهتر بهره بردیم. نتایج حاصل از مدل نشان می‌دهد که مدل ما، نسبت به روش‌های قبل در اکثر موارد خروجی بهتری را ثبت می‌کند.

مراجع

[1] Dey, A.K. 2001. *Understanding and using context*. Personal and Ubiquitous Computing. 5, 1 (2001).

DOI:https://doi.org/10.1007/s007790170019.

[2] Fang, J. 2021. *Session-based Recommendation with Self-Attention Networks*. arXiv preprint arXiv:2102.01922. (2021).

[3] Hidasi, B., Karatzoglou, A., Baltrunas, L. and Tikk, D. 2015. *Session-based recommendations with recurrent neural networks*. arXiv preprint arXiv:1511.06939. (2015).

[4] Hu, L., Cao, L., Wang, S., Xu, G., Cao, J. and Gu, Z. 2017. *Diversifying Personalized Recommendation with User-session Context.* IJCAI (2017), 1858–1864.

[5] Huang, R., McIntyre, S., Song, M., E, H. and Ou, Z. 2018. *An attention-based recommender system to predict contextual intent based on choice histories across and within sessions*. Applied Sciences. 8, 12 (2018), 2426.

[6] Jannach, D., Lerche, L. and Jugovac, M. 2015. *Adaptation and evaluation of recommendations for short-term shopping goals*. Proceedings of the 9th ACM Conference on Recommender Systems (2015), 211–218.

[7] Koren, Y., Bell, R. and Volinsky, C. 2009. *Matrix factorization techniques for recommender systems*. Computer. 42, 8 (2009), 30–37.

[8] Li, J., Ren, P., Chen, Z., Ren, Z., Lian, T. and Ma, J. 2017. *Neural attentive session-based recommendation*. Proceedings of the 2017 ACM on Conference on Information and Knowledge Management (2017), 1419–1428.

[9] Liu, Q., Mokhosi, R., Zeng, Y. and Zhang, H. 2018. *STAMP: Short-term attention/memory priority model for session-based recommendation*. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2018).

[10] Rendle, S., Freudenthaler, C. and Schmidt-Thieme, L. 2010. *Factorizing personalized markov chains for next-basket recommendation*. Proceedings of the 19th international conference on World wide web (2010), 811–820.

[11] Ricci, F., Rokach, L. and Shapira, B. 2011. *Introduction to Recommender Systems Handbook*. Recommender Systems Handbook.

[12] Tan, Y.K., Xu, X. and Liu, Y. 2016. *Improved recurrent neural networks for session-based recommendations*. Proceedings of the 1st workshop on deep learning for recommender systems (2016), 17–22.

[13] Tang, J. and Wang, K. 2018. *Personalized top-N sequential recommendation via convolutional sequence embedding*. WSDM 2018 - Proceedings of the 11th ACM International Conference on Web Search and Data Mining (2018).

[14] Wang, P., Guo, J., Lan, Y., Xu, J., Wan, S. and Cheng, X. 2015. *Learning hierarchical representation model for nextbasket recommendation*. Proceedings of the 38th International ACM SIGIR conference on Research and Development in Information Retrieval (2015), 403–412.

[15] Ying, H., Zhuang, F., Zhang, F., Liu, Y., Xu, G., Xie, X., Xiong, H. and Wu, J. 2018. *Sequential recommender system based on hierarchical attention network*. IJCAI International Joint Conference on Artificial Intelligence (2018).

[16] Yu, F., Liu, Q., Wu, S., Wang, L. and Tan, T. 2016. *A dynamic recurrent model for next basket recommendation*. Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval (2016), 729–732.

پانویس‌ها

1 Session-based Recommender systems
2 Recurrent Neural Networks
3 Multiple Sessions Attention Network
4 gated recurrent unit
5 Hu
6 context
7 item embeddings
8 Li
9 Huang
10 feed-forward neural network
11 bias
12 stochastic gradient descent
13 Hit Rate
14 Mean Reciprocal Rank
15 batch size
16 epoch

چهارمین کنفرانس ملی انفورماتیک ایران
پژوهشگاه دانش‌های بنیادی، پردیس فرمانیه، تهران
۱۴ و ۱۵ دی ماه ۱۴۰۱

یک مدل خودتوجه تجزیه‌شده برای سیستم‌های توصیه‌گر دنباله‌ای

پگاه جعفریان [۱]، علی کتانفروش [۲]

[۱] کارشناسی ارشد، گروه علوم کامپیوتر و داده ها، دانشکده علوم ریاضی، دانشگاه شهید بهشتی، تهران
pegahjafarian1997@gmail.com

[۲] استادیار، گروه علوم کامپیوتر و داده ها، دانشکده علوم ریاضی، دانشگاه شهید بهشتی ، تهران
a_katanforosh@sbu.ac.ir

چکیده

پویایی در دنباله، از ویژگی‌های کلیدی سیستم‌های توصیه‌گر مدرن است. برای مدل سازی روندهای پویا در علایق و انتخاب‌های کاربران، از دو رویکرد زنجیره‌های مارکوف و شبکه‌های عصبی بازگشتی استفاده می‌شود. روش‌های مبتنی بر زنجیره مارکوف بر روی مجموعه داده‌های خلوت عملکرد بهتری دارند؛ اما شبکه‌های عصبی بازگشتی بر روی مجموعه داده متراکم عملکرد بهتری نشان داده اند. ما در این مقاله قصد داریم که با استفاده از ***سازوکار خودتوجه****، مزیت‌های هر دو نگرش زنجیره مارکوف و شبکه‌های عصبی بازگشتی را به دست آوریم. با استفاده از* ***جاسازی موقعیت****، مدل خودتوجه از موقعیت‌های آیتم‌های قبلی آگاه خواهد بود و با استفاده از* ***سازوکار خودتوجه تجزیه شده*** *پیش بینی آیتم های بعدی کاربر را بر اساس تعداد کمی از تعاملات‌های قبل او انجام می‌دهیم. ارزیابی روش پیشنهادی بر روی مجموعه دادگان* ***Beauty*** *،**Steam**،* ***Games*** *و* ***ML-1M*** *در مقایسه با دیگر روش‌های سیستم‌های توصیه‌گر دنباله‌ای بر روی هر دو مجموعه داده خلوت و متراکم نرخ برخورد و دقت بالاتری را نشان می‌دهد.*

کلمات کلیدی

سیستم‌های توصیه‌گر، توصیه دنباله‌ای، توصیه آیتم بعدی، سازوکار خودتوجه، مدل خودتوجه ضرب نقطه‌ای مقیاس شده، جاسازی موقعیت

۱- مقدمه

سیستم‌های توصیه‌گر[۱] بر مدل‌سازی همبستگی بین کاربران و آیتم‌ها بر اساس بازخورد تاریخی تمرکز می‌کنند؛ به عنوان مثال: بازدید از صفحه، خرید، پسند کردن یک محصول در یک فروشگاه اینترنتی و یا یک رسانه در شبکه‌های اشتراک‌گذاری. بازخورد کاربر می‌تواند صریح باشد، مانند رتبه‌بندی و یا می‌تواند ضمنی باشد، مانند: کلیک‌ها، خریدها و نظرات. هدف سیستم‌های توصیه‌گر دنباله‌ای، ترکیب‌کردن مدل‌های شخصی‌شده رفتار کاربر، بر اساس سابقه فعالیت‌ها با مفهومی از "زمینه[۲] " بر اساس تعاملات اخیر کاربران است[1,2,3]. بسیاری از سیستم‌های توصیه‌گر دنباله‌ای به دنبال مدل‌سازی ماتریس‌های انتقال حلت[۳] آیتم–آیتم به عنوان ابزاری برای کشف الگوهای متوالی در بین آیتم‌های متوالی هستند. از آنجایی که آخرین آیتم بازدید شده اغلب عامل کلیدی است که بر عملکرد بعدی کاربر تاثیر می گذارد، روش‌های مبتنی بر زنجیره مارکوف مرتبه اول عملکرد قوی، به ویژه در مجموعه داده‌های خلوت نشان می‌دهد [4,5]. به دست آوردن الگوهای مفید در دنباله‌هایی که از پویایی بالایی برخوردارند، در درجه اول، به دلیل بعد فضای ورودی است که به طور تصاعدی با تعداد تعاملات گذشته کاربر افزایش می‌یابد.

زنجیره مارکوف[۴] یک رویکرد رایج در مدل‌سازی دنباله‌ها است که در آن فرض می‌شود عمل بعدی فقط مشروط به عمل و یا عمل‌های قبلی است. این رویکرد با موفقیت برای توصیف انتقال آیتم‌های کوتاه‌برد برای توصیه استفاده می‌شود[6]. در نقطه مقابل شبکه‌های عصبی بازگشتی[۵] با خلاصه‌کردن فعالیت‌های قبلی کاربر در یک حالت پنهان، اقدام بعدی کاربر را پیش‌بینی می‌کنند [7].

کارایی هر دو رویکرد، تا حدودی به انواع خاصی از داده‌ها محدود می‌شود. روش‌های مبتنی بر زنجیره مارکوف، در مجموعه داده‌های خلوت در جایی که صرفه‌جویی در تعداد پارامترهای مدل مهم است، عملکرد بهتری را دارد؛ در حالی که، شبکه‌های عصبی بازگشتی بر روی مجموعه داده‌های متراکم بهتر عمل می‌کند.

اخیراً، یک مدل دنباله‌ای ترانسفرمر[۶] در کاربردهایی از جمله ترجمه ماشینی استفاده می‌شود. بر خلاف مدل‌های موجود که از لایه‌های بازگشتی و پیچشی[۷] استفاده می‌کنند، ترانسفرمرها صرفاً براساس تبدیلاتی موسوم به "سازوکار خودتوجه [۸]" استفاده می‌کنند. این مدل بسیار کارآمد است و قادر به کشف الگوهای معنایی بین کلمات در یک جمله است[8].

سازوکارهای توجه نشان داده است که در زمینه‌های مختلف از جمله شرح تصویر و ترجمه ماشین موثر هستند [9,10]. اخیرا، سازوکارهای توجه در سیستم‌های توصیه‌گر گنجانده شده اند[11].

در این مقاله، ما در پی استفاده از سازوکار خودتوجه برای مسائل توصیه دنباله‌ای هستیم. به طور خاص در این پژوهش ما یک مدل خودتوجه تجزیه‌شده برای توصیه دنباله‌ای (DeSASRec) ارائه می‌دهیم که به واسطه جداکردن نقش پارامترهای وزن جاسازی آیتم و جاسازی موقعیت از یکدیگر در مدل خودتوجه، مدل از لحاظ پارامتر، نسبت به SASRec [1] سبک‌تر بوده و به لحاظ کارایی، مدل قدرتمندی در پیشگویی آیتم بعدی در چارچوب توصیه‌گر دنباله‌ای است. مدل پیشنهادی، با دنباله‌های نسبتا طولانی از تعاملات کاربر آموزش داده می‌شود تا وابستگی‌های طولانی مدت بین آیتم‌ها

چهارمین کنفرانس ملی انفورماتیک ایران
پژوهشگاه دانش‌های بنیادی، پردیس فرمانیه، تهران
۱۴ و ۱۵ دی ماه ۱۴۰۱

را بیاموزد، در حالی که قادر است دانش موثری از فعالیت‌های اخیر کاربر در مجموعه داده‌های خلوت نیز به‌دست آورد. علاوه بر این، محاسبات در ماژول اصلی مدل، یعنی بلوک خودتوجه، از قابلیت موازی‌سازی مناسبی برخوردار است و در نتیجه مدل پیشنهادی سریع‌تر از شبکه‌های عصبی بازگشتی و شبکه‌های عصبی پیچشی[۹] عمل می‌کند[1].

۲- روش پیشنهادی

صورت‌بندی مسئله توصیه‌گر دنباله‌ای، به بیان ساده چنین است: دنباله‌ای از فعالیت‌های کاربر در ارتباط با مجموعه‌ای از آیتم‌ها در گذشته مفروض است، کدام آیتم محتملاً انتخاب بعدی کاربر است؟ مدل آموزش داده می‌شود به گونه‌ای که در مرحله زمانی t ام ، آیتم بعدی را بسته به t آیتم قبلی پیش‌بینی کند. در هر گام زمانی، مدل به دنبال بازشناسایی زمینه تعاملات کاربر با توجه به سابقه عملکرد کاربر است تا از آن برای پیش‌بینی آیتم بعدی استفاده کند. یک رویکرد برای بازشناسایی زمینه و پویایی الگوی تعاملات، استفاده از سازوکار خودتوجه برای تمرکز بر روی آیتم‌های مرتبط با انتخاب بعدی کاربر است. مدل پیشنهادی، یک شبکه چهار لایه به ترتیب شامل لایه‌های جاسازی، خودتوجه، شبکه پیش‌رونده نقطه‌ای و لایه پیش‌بینی است. دنباله تعاملات گذشته کاربر به عنوان ورودی به شبکه داده می‌شود. در لایه جاسازی یک بازنمایی ماتریسی از داده‌های ورودی به دست می‌آید و به لایه خودتوجه داده می‌شود. در لایه خودتوجه، از مدل **خودتوجه ضرب نقطه‌ای مقیاس‌شده**[۱۰] استفاده می‌کنیم و حاصل به شبکه پیش‌رونده نقطه‌ای داده می‌شود. مجموعه لایه خودتوجه و پیش‌رونده نقطه‌ای را در قالب یک "بلوک" در نظر می‌گیریم و آن را "بلوک خودتوجه" می‌نامیم. به منظور عمیق‌تر کردن شبکه، دو یا بیشتر تکرارهایی از این بلوک‌ها را به صورت پشت سرهم در ساختار شبکه قرار می‌دهیم. خروجی آخرین بلوک خودتوجه به لایه پیش‌بینی داده می‌شود. در لایه پیش‌بینی، یک امتیاز برای هر آیتم به دست می‌آید. جدول (۱) شرح نمادهای مورد استفاده در روابط ریاضی مدل را نشان می‌دهد.

جدول (۱) : نمادگذاری

توضیحات	نشانه گذاری
مجموعه آیتم ها و کاربران	U, I
دنباله تعاملات گذشته کاربر u	$S = (s_1, s_2, \dots, s_n)$
بعد فضای پنهان	$d \in N$
بیشترین طول دنباله	$n \in N$
تعداد بلوک های خودتوجه	$b \in N$
ماتریس جاسازی آیتم	$M \in R^{\|I\| \times d}$
ماتریس جاسازی موقعیت	$P \in R^{n\times d}$
ماتریس جاسازی ورودی	$\hat{E} \in R^{n\times d}$
جاسازی آیتم بعد از b – امین لایه خودتوجه	$S^b \in R^{n\times d}$
جاسازی آیتم بعد از b – امین شبکه پیش‌رونده	$F^b \in R^{n\times d}$

۲-۱- لایه جاسازی

در این لایه دنباله آموزش را به صورت زیر در نظر می‌گیریم :

$$S = (s_1, s_2, \dots, s_n)$$

که در این دنباله n نشان‌دهنده ماکزیمم طول دنباله است که مدل برای آن طراحی شده‌است. اگر طول دنباله بزرگتر از n باشد، تنها n آیتم اخیر را در نظر می‌گیریم. اگر طول دنبلاه کمتر از n باشد، به صورت تکراری، یک نماد قراردادی با عنوان padding را به سمت چپ دنباله اضافه می‌کنیم تا طول دنباله به n برسد. در این لایه یک ماتریس جاسازی آیتم مانند $M \in R^{|I|\times d}$ محاسبه می‌شود که d یک ابرپارامتر مدل و نشان‌دهنده بعد فضای جاسازی است. ماتریس جاسازی ورودی از چینش بردارهای جاسازی آیتم‌های دنباله در ماتریس E به دست می‌آید: $E \in R^{n\times d}$, $E_i = M_{S_i}$

همچنین بردار ثابت صفر نیز به عنوان جاسازی برای آیتم padding در نظر گرفته می‌شود[1].

جاسازی موقعیت [۱۱] : یک ماتریس جاسازی موقعیت قابل یادگیری مانند P در ماتریس جاسازی ورودی $P \in R^{n\times d}$ در نظر می‌گیریم. داریم:

$$\hat{E} = \begin{bmatrix} M_{s_1} & P_1 \\ M_{s_2} & P_2 \\ . & . \\ . & . \\ . & . \\ M_{s_n} & P_n \end{bmatrix} \qquad (۱)$$

توجه کنید رابطه (۱) متفاوت از رویکردی است که در [1]، برای ترکیب جاساز موقعیت با جاساز آیتم‌ها با استفاده از عملگر جمع صورت گرفته‌است.

۲-۲- بلوک خودتوجه

در مدل پیشنهادی از مدل خودتوجه ضرب نقطه‌ای مقیاس‌شده[8] استفاده می‌کنیم (رابطه (۲)) :

$$Attention\,(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = Softmax(\frac{\boldsymbol{QK^T}}{\sqrt{d}})\mathbf{V} \qquad (۲)$$

در رابطه (۲)، Q نشان‌دهنده درخواست[۱۲]، K نشان‌دهنده کلید[۱۳] و V نشان‌دهنده ارزش‌ها[۱۴] است. مقیاس تقسیم بر مجذور d برای جلوگیری از بزرگی بیش از حد ارزش‌ها در ضرب داخلی استفاده می‌شود، به ویژه زمانی که ابعاد خیلی بزرگ باشد [8].

در پردازش زبان طبیعی مانند ترجمه ماشینی، منظور از خودتوجهی قرار دادن $K = V$ است. اخیراً یک روش خودتوجهی پیشنهاد شده است که از همان درخواست‌ها، کلیدها و ارزش‌ها استفاده می‌کند. در روش پیشنهادی، عملیات خود توجه ماتریس جاسازی $\hat{E}$ را به عنوان ورودی می‌گیرد، آن را از طریق سه تبدیل خطی به سه فضای خطی دیگر تصویر می‌کند و آن‌ها را در رابطه (۲) قرار می‌دهد؛ طبق رابطه (۳) :

$$S = SA\,(\hat{E}\,) = Attention\,(PW^Q, EW^K, \hat{E}W^V) \qquad (۳)$$

در رابطه (۳)، ماتریس‌های وزن درخواست‌ها و کلید هر دو ماتریس‌های $d \times d$ بعدی هستند و ماتریس وزن ارزش $(2d) \times d$ بعدی است. با توجه به ماهیت دنبلاه‌ها، مدل باید هنگام پیش‌بینی آیتم $t+1$ ام، فقط t آیتم اول

را در نظر بگیرد. این نحوه از مدل‌سازی سازوکار خودتوجه مشابه مدل خودتوجه مورد استفاده در توصیه‌گر دنباله‌ای [1] است با این تفاوت که در اینجا ما نقش جاسازی‌های آیتم و موقعیت را در تناظر با پرس‌وجو و کلید جدا می‌کنیم. بدین ترتیب ماتریس W^Q ضرایب همبستگی بین موقعیت‌ها و W^K ضرایب همبستگی بین آیتم‌ها را می‌آموزند.

شبکه پیش‌رونده نقطه‌ای[۱۵] : اگر چه سازوکار خودتوجه می‌تواند همه جاسازی‌های آیتم‌های قبلی را با وزن‌های تطبیقی جمع کند، اما در نهایت همچنان یک مدل خطی است. برای بهره‌گیری از مزیت مدل‌های غیرخطی و در نظر گرفتن همبستگی‌های بین ابعاد فضای پنهان، یک لایه پیش‌رونده نقطه‌ای را برای تمام S_i ها با پارامترهای به طور یکسان اعمال می‌کنیم [1] ، طبق رابطه (۴) :

$$F_i = FFN\ (S_i) = ReLU(S_i W^1 + b^1)W^2 + b^2 \quad (۴)$$

در رابطه (۴)، ماتریس‌های W^1 , W^2 همگی $d \times d$ و b^1, b^2 بایاس را نشان می‌دهند. این نکته قابل ذکر است که هیچ‌گونه تعاملی بین S_i و S_j وجود ندارد، بدین ترتیب از نشت اطلاعات پس از عبور از لایه خودتوجه جلوگیری می‌شود.

۲-۳- انباشتن بلوک خودتوجه

بعد از اولین بلوک خودتوجه، F_i در واقع تمام جاسازی‌های آیتم‌های قبلی را در نظر می‌گیرد. با این حال، یادگیری پیچیده‌تر حالت بین آیتم‌ها از طریق تکرار بلوک‌های خودتوجه ممکن است مفید باشد[1]. به طور خاص، در این روش یک لایه خودتوجه به همراه یک شبکه پیش‌رونده نقطه‌ای را به عنوان یک "بلوک خودتوجه" در نظر می‌گیریم. حال، b- امین بلوک خودتوجه بر اساس رابطه (۵) تعریف می‌شود.

$$S^b = SA(F^{b-1})$$
$$F_i^b = FFN\left(S_i^b\right)\ , \quad \forall i \in \{1,2,\dots,n\} \quad (۵)$$

بدیهی است برای اولین بلوک، قرار می‌دهیم:

$$S_1 = \boldsymbol{S}\,, \quad F_1 = \boldsymbol{F}$$

با این حال، هنگامی که شبکه عمیق‌تر می‌شود، مشکلات شناخته شده‌ای به وجود می‌آید؛ از جمله:

- افزایش ظرفیت مدل باعث بیش‌برازش می‌شود.
- روند آموزش ناپایدار می‌شود (به دلیل ناپدیدشدن گرادیان و غیره)
- مدل‌هایی با پارامترهای بیشتر اغلب به زمان آموزش بیشتری نیاز دارد.

برای جلوگیری از مشکلات ذکر شده، دو لایه با هدف منظم‌سازی طبق رابطه (۶) به شبکه اضافه می‌شود [1].

$$g(x) = Dropout(g(layerNorm(x))) \quad (۶)$$

در رابطه (۶)، $g(x)$ می‌تواند هر یک از لایه‌های خودتوجه و یا لایه پیش‌رونده نقطه‌ای باشد. به این صورت که برای لایه g درهر بلوک، قبل از اینکه ورودی x را به g بدهیم، ابتدا بر روی آن یک نرمال‌سازی و سپس عملگر ریزش (*dropout*) را اعمال می‌کنیم و آن را به عنوان خروجی نهایی در نظر می‌گیریم.

۲-۴- لایه پیش‌بینی

بعد از b بلوک خودتوجه که به صورت سلسله مراتبی از اطلاعات آیتم‌های بازدید شده قبلی استخراج شدند، آیتم بعدی را با در نظر گرفتن t آیتم اول، براساس F_b پیش‌بینی می‌کنیم [1]، طبق رابطه (۷) :

$$r_{i,t} = F_t^b\ M_i^T \quad (۷)$$

در رابطه (۷)، طرف چپ بیانگر امتیاز آیتم i به عنوان انتخاب بعدی کاربر بر اساس t آیتم اول است و ماتریس M نشان دهنده ماتریس جاسازی آیتم است. هر قدر مقدار این امتیاز بالاتر باشد، به این معناست که مدل، آیتم i را به عنوان انتخاب محتمل‌تر کاربر پیش‌گویی می‌کند.

۲-۵- آموزش مدل

همانطور که گفته شد، دنباله تعاملات هر کاربر را به استثنای آخرین فعالیت کاربر، یک دنباله s با طول ثابت در نظر گرفتیم. خروجی مورد انتظار در گام زمانی t را با o_t نمایش می‌دهیم که طبق رابطه (۸) تعریف می‌شود.

$$o_t = \begin{cases} <pad> & if\ s_t\ is\ a\ item\ padding \\ s_{t+1} & 1 \le t < n \\ s_t & t = n \end{cases} \quad (۸)$$

مدل یک دنباله S را به عنوان ورودی دریافت می‌کند و o مولفه آخر، خروجی مورد انتظار است. بر این اساس، تابع زیان یک آنتروپی تقاطعی است که o_t را به عنوان نمونه مثبت و میانگینی از هر آیتم غیر o_t را به عنوان نمونه منفی در نظر می‌گیرد؛ طبق رابطه (۹) :

$$-\sum_{s^u \in S}\sum_{t=1}^{n}[\log(\sigma(r_{o_t,t})) + \frac{1}{|I|-1}\sum_{j\in I,\, j\notin o_t}\log\,(1-\sigma(r_{j,t}))\quad (۹)$$

شبکه پیشنهادی توسط بهینه‌ساز آدام که یک روش گرادیان کاهشی تصادفی با نرخ یادگیری تطبیقی و تکانه است، آموزش داده می‌شود.

۳- نتایج

در این بخش به ارزیابی و تحلیل نتایج مدل پیشنهادی بر روی چهار مجموعه داده می‌پردازیم. همچنین کارایی روش پیشنهادی در مقایسه با عملکرد چهار روش مطرح در حوزه توصیه‌گر دنباله‌ای ارزیابی می‌شود.

۳-۱- مجموعه دادگان

به منظور سنجیدن عملکرد مدل از چهار مجموعه داده استفاده می‌کنیم. این داده‌ها شامل اطلاعاتی همچون تعاملات کاربران با آیتم‌ها، امتیاز هر کاربر به آیتمی که با آن تعامل داشته و اطلاعات زمینه‌ای برای آیتم‌ها می‌باشد. خلاصه‌ای از دادگان محک در جدول (۲) آورده شده است.

جدول (۲) : آمار مجموعه دادگان

مجموعه دادگان	تعداد کاربران	تعداد آیتم‌ها	تعداد تعاملات	میانگین طول
Beauty	۵۲۰۲۴	۵۷۲۸۹	۴۰۰۰۰۰	۸/۸
Steam	۳۴۴۷۳۰	۱۳۰۴۴	۳۷۰۰۰۰۰	۱۲/۴
ML-1M	۶۰۴۰	۳۴۱۶	۱۰۰۰۰۰۰	۱۶۳/۵
Games	۳۱۰۱۳	۲۳۷۱۵	۳۰۰۰۰۰	۹/۳

۳-۲- معیارهای ارزیابی

برای ارزیابی سیستم‌های توصیه‌گر شاخص‌های مختلفی وجود دارد. متعارفاً، خروجی نهایی یک سیستم توصیه‌گر، یک لیست رتبه بندی‌شده از آیتم‌ها است. برای این منظور، ما از معیارهای زیر برای ارزیابی عملکرد مدل استفاده می کنیم:

- نرخ برخورد [۱۶] (Hit) : فرض کنیم لیست توصیه‌گر شامل k آیتم است. نرخ برخورد برای یک توصیه‌گر دنباله‌ای از رابطه (۱۰) به دست می‌آید.

$$Hit@k = \frac{1}{|U|}\sum_{u\in U} 1\ (R_{u,g_u} \leq k)$$

(۱۰)

در رابطه (۱۰) ، قدرمطلق U نشان‌دهنده تعداد کاربران و g_u ، آیتم مرتبط با کاربر u در آخرین گام زمانی است و R_{u,g_u} ، رتبه آیتم g_u برای کاربر u است. بر این اساس هر چه مقدار k بیشتر باشد، مقدار نرخ برخورد نیز بیشتر می شود، چرا که آیتم آزمایش شانس بیشتری برای قرار گرفتن در kتای اول لیست رتبه بندی را خواهد داشت.

- بهره تجمعی تنزل یافته نرمال سازی شده [۱۷] (NDCG) : این شاخص، کمیتی برای اندازه‌گیری "ارتباط[۱۸]" در موتورهای جستوجوگر است. بهره تجمعی به صورت مجموع بهره‌ها تا موقعیت k ام در لیست رتبه‌بندی تعریف می‌شود. برای جریمه‌کردن آیتم‌های مرتبط که در انتهای لیست قرار می‌گیرند، DCG معرفی گردید. مقدار $IDCG$ برابر با مقدار DCG به ازای رتبه‌بندی ایده‌آل تعریف می‌شود که در آن آیتم‌ها را براساس ارتباطشان تا موقعیت k ام از زیاد به کم رتبه‌بندی می‌کنیم. در نهایت $NDCG$ برابر با نرمال سازی‌شده‌ی مقدار DCG توسط $IDCG$ تعریف می‌شود. بدین ترتیب مقدار آن بدون توجه به طول ، همیشه در بازه [0 , 1] است. تعریف شاخص‌های DCG، $IDCG$ و $NDCG$ در رابطه (۱۱) آورده شده‌است.

$$DCG(k) = \sum_{i=1}^{k} \frac{G_i}{\log_2(\,i+1)}$$

$$IDCG(k) = \sum_{i=1}^{|I(k)|} \frac{G_i}{\log_2(i+1)}$$

$$NDCG(k) = \frac{DCG(k)}{IDCG(k)}$$

(۱۱)

۳-۳- مقایسه با روش های پیشین

برای ارزیابی کارایی مدل پیشنهادی، آن را با چند نمونه از مدل‌های معروف استفاده شده در سیستم‌های توصیه‌گر مقایسه کرده‌ایم که عبارت اند از :

- GRU4Rec+ : این مدل نسخه بهبود یافته از مدل GRU4Rec است که در آن از برخی توابع زیان جدید و یک استراتژی جدید برای نمونه برداری استفاده شده است که مدل GRU4Rec از GRU برای مدل کردن دنباله تعاملات کاربر برای سیستم‌های توصیه‌گر استفاده می‌کند [12].
- Caser : این مدل با بهره گیری از شبکه های عصبی پیچشی، زنجیره مارکوف از مرتبه بالا را برای توصیه دنباله‌ای استفاده می‌کند [13].
- TransRec : یک روش توصیه‌گر دنباله‌ای مرتبه اول است که هر کاربر را به صورت یک *بردار ترجمه*[۱۹] در نظر می‌گیرد تا بتواند انتقال حالت از آیتم فعلی به آیتم بعدی را مدل می‌کند [14].
- SASRec : این روش، از یک مدل زبانی با ترانسفورمر یک سویه چپ به راست استفاده می‌کند تا دنباله‌های رفتار کاربران را به تصویر بکشد [1].

۳-۴- جزئیات پیاده سازی

مدل، با استفاده از زبان برنامه نویسی پایتون و با استفاده از PyTorch پیاده‌سازی شده‌است. همچنین برای یادگیری مدل از روش بهینه ساز آدام و با نرخ یادگیری ۰/۰۰۱ صورت گرفته‌است. اندازه بعد بردار پنهان[۲۰] برابر ۱۰۰، اندازه دسته[۲۱] برابر با ۱۲۸، بیشترین طول دنباله برای مجموعه داده ML-1M برابر با ۱۰۰ و برای بقیه مجموعه داده ها برابر ۵۰ ، تعداد بلوک‌های خودتوجه برابر با ۲، مقدار ریزش[۲۲] برابر ۰/۵ و تعداد دوره[۲۳] برابر با ۳۰۰ در نظر گرفته شده‌است.

۳-۵- ارزیابی مدل پیشنهادی

همان‌طور که گفته شد، برای بهبود در عملکرد، در لایه جاسازی و همچنین در لایه خودتوجه مدل پیشنهادی تغییراتی را اعمال کرده و نام این مدل را DeSASRec نامیدیم. تفاوتی که مدل DeSASRec با مدل SASRec

دارد، این است که در لایه جاسازی و لایه خودتوجه تغییراتی اعمال شده‌است. نتایج این مقایسه در جدول (۳) آورده شده‌است.

جدول(۳) : تاثیرات تغییر در لایه های شبکه

مجموعه دادگان	SASRec		DeSASRec	
سنجه ارزیابی	NDCG@10	Hit@10	NDCG@10	Hit@10
Beauty	۰/۳۲۴۵	۰/۳۵۷۸	۰/۵۱۰۰	۰/۴۵۱۰
Steam	۰/۳۶۸۴	۰/۴۱۳۹	۰/۵۳۹۰	۰/۶۱۷۸
ML-1M	۰/۵۹۶۲	۰/۸۲۶۸	۰/۶۳۶۰	۰/۸۴۹۸
Games	۰/۵۱۴۰	۰/۶۳۶۶	۰/۶۸۰۰	۰/۷۴۲۰

با توجه به جدول (۳)، کارایی مدل DeSASRec افزایش یافته به گونه‌ای که برای مجموعه ML-1M مقدار نرخ برخورد ۳ درصد افزایش یافته و مقدار *NDCG* نیز برای این مجموعه داده ۴ درصد افزایش پیدا می‌کند. همچنین برای مجموعه داده Steam مقدار نرخ برخورد ۱۴ درصد بهبود یافته و مقدار *NDCG* برای این مجموعه داده ۱۵ درصد افزایش یافته‌است. همچنین برای بقیه مجموعه دادگان نیز این بهبود عملکرد وجود دارد.

در جدول (۴)، نتایج حاصل از مدل DeSASARec و مدل‌هایی که در بخش ۳-۳ ارائه شده‌است، آورده شده و همان‌طور که مشاهده می‌شود، مدل DeSASRec در تمام مجموعه دادگان، نتیجه بهتری را به دست آورده‌است که نشان‌دهنده عملکرد قابل قبول مدل DeSASRec می‌باشد.

جدول (۴) : مقایسه DeSASRec با کارهای پیشین

مجموعه دادگان	Beauty	Steam	Games	ML-1M
سنجه ارزیابی	Hit@10	Hit@10	Hit@10	Hit@10
GRU4Rec+	۰/۳۹۴۲	۰/۸۰۱۸	۰/۶۵۹۹	۰/۷۵۰۱
TransRec	۰/۴۶۰۷	۰/۷۶۲۰	۰/۶۸۳۸	۰/۶۴۱۳
Caser	۰/۴۲۶۴	۰/۷۸۸۴	۰/۵۲۸۲	۰/۷۹۸۶
SASRec	۰/۳۵۷۸	۰/۶۱۷۸	۰/۶۳۶۶	۰/۷۶۶۸
DeSASRec	۰/۵۱۰۰	۰/۸۱۷۸	۰/۷۴۲	۰/۸۲۹۸
سنجه ارزیابی	NDCG@10	NDCG@10	NDCG@10	NDCG@10
GRU4Re+	۰/۲۵۵۶	۰/۵۵۹۵	۰/۴۷۵۹	۰/۵۵۱۳
TransRec	۰/۳۰۲۰	۰/۴۸۵۲	۰/۴۵۵۷	۰/۳۹۶۹
Caser	۰/۲۵۴۷	۰/۵۲۸۱	۰/۳۲۱۴	۰/۵۶۱۳
SASRec	۰/۳۲۴۵	۰/۳۶۸۴	۰/۵۱۴۰	۰/۵۷۶۲
DeSASRec	۰/۴۵۱۰	۰/۵۳۹۰	۰/۶۸۰۰	۰/۶۲۶۰

۴- نتیجه گیری

در این مقاله، یک مدل خودتوجه تجزیه شده برای سیستم‌های توصیه‌گر دنباله‌ای ارائه شد. این مدل به ما این اجازه را می‌دهد که در طولانی مدت انگیزه‌های مصرف کاربر را ثبت کرده و با استفاده از سازوکار خودتوجه، پیش‌بینی‌های خود را بر اساس فعالیت‌های نسبتاً کمی انجام می‌دهیم. همچنین با تغییر در لایه‌های شبکه، این نتیجه حاصل شد که عملکرد مدل بهتر شده و در نتیجه در هر مرحله زمانی که به دنبال شناسایی آیتم‌های مرتبط با سابقه کاربر هستیم، بهترین عملکرد را نسبت به روش‌های مبتنی بر MC/CNN/RNN خواهیم داشت و در نتیجه برای پیش‌بینی آیتم بعدی، این تغییرات باعث عملکرد بهتر مدل خواهد شد.

مراجع

[1] WC.Kang, J.McAuley., “*self-attentive sequential recommendation*,” IEEE International Conference on Data Mining (ICDM), Singapore, pp. 81-91, 2018.

[2] Y. Hu, Y. Koren, and C. Volinsky, “*Collaborative filtering for implicit feedback datasets*,” in 8th IEEE International Conference on Data Mining (ICDM), Pisa, Italy, pp. 263-272, 2008.

[3] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme, “*BPR: bayesian personalized ranking from implicit feedback*,” in 25th Conference on Uncertainty in Artificial Intelligence (UAI), Montreal, Canada, pp.151-162, 2009.

[4] J. Tang and K. Wang, “*Personalized top-n sequential recommendation via convolutional sequence embedding*,” in Proceedings of the 11th ACM International Conference on Web Search and Data Mining (WSDM), Phoenix, AZ, USA, pp.565-573, 2018.

[5] R. He, W. Kang, and J. McAuley, “*Translation-based recommendation*,” in Proceedings of 11th ACM Conference on Recommender Systems (RecSys), Como, Italy, pp. 161-169, 2017.

[6] S. Rendle, C. Freudenthaler, and L. Schmidt-Thieme, “*Factorizing personalized markov chains for next-basket recommendation*,” in Proceeding of the 19th international Conference on World Wide Web (WWW), Raleigh North California, USA, pp. 811-820, 2010.

[7] B. Hidasi, A. Karatzoglou, L. Baltrunas, and D. Tikk, “*Session-based recommendations with recurrent neural networks*,” in International Conference on learning representations(ICLR), SanDiego, California, USA, pp.611-621, 2016.

[8] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “*Attention is all you need*,” in 31st Conference on Neural Information Proceeding Systems (NIPS), Long Beach, CA, USA, pp. 571-582, 2017.

[9] K. Xu, J. Ba, R. Kiros, K. Cho, A. C. Courville, R. Salakhutdinov, R. S. Zemel, and Y. Bengio, “*Show, attend and tell: Neural image caption generation with visual attention*,” in Proceeding of the 32nd International Conference on Machine Learning (ICML), Lille, France, Vol. 37, 2015.

[10] D. Bahdanau, K. Cho, and Y. Bengio, “*Neural machine translation by jointly learning to align and translate*,” in International Conference on Learning Representation (ICLR), Banff National Park, Canada, pp.567-578, 2015.

[11] J. Xiao, H. Ye, X. He, H. Zhang, F. Wu, and T. Chua, *“Attentional factorization machines: Learning the weight*

of feature interactions via attention networks," in ArXiv abs/1708.04617, 2017.

[12] B. Hidasi and A. Karatzoglou, "*Recurrent neural networks with top-k gains for session-based recommendations*," in Proceedings of the 27th ACM International Conference on Information and Knowledge Management, Torino, Italy, pp.843-852, 2017.

[13] J. Tang and K. Wang, "*Personalized top-n sequential recommendation via convolutional sequence embedding*," in Proceedings of the 11th ACM International Conference on Web Search and Data Mining (WSDM), Phoenix, AZ, USA, pp.565-573, 2018.

[14] R. He, W. Kang, and J. McAuley, "*Translation-based recommendation*," in Proceedings of 11th ACM Conference on Recommender Systems (RecSys), Como, Italy, pp. 161-169, 2017.

پانویس ها

[1] Recommender Systems
[2] Context
[3] Transition
[4] Markov Chain (MC)
[5] Recurrent Neural Networks (RNNs)
[6] Transformer
[7] Convolutional
[8] Self-Attention Mechanism
[9] Convolutional Neural Networks (CNNs)
[10] Scaled dot-Product Self-Attention
[11] Position Embedding
[12] Query
[13] Key
[14] Values
[15] Feed-forward Network
[16] Hit Rate
[17] Normalized Discounted Cumulative Gain (NDCG)
[18] Relevancy
[19] Translation Vector
[20] Latent Vector Dimensionality
[21] Batch Size
[22] Dropout
[23] Epoch

چهارمین کنفرانس ملی انفورماتیک ایران
پژوهشگاه دانش‌های بنیادی، پردیس فرمانیه، تهران
۱۴ و ۱۵ دی ماه ۱۴۰۱

دسته‌بندی ابرنقاط سه‌بعدی با استفاده از یک معماری عمیق توسعه‌یافته

ایمان احمدی[۱]، حسن ختن‌لو[۲]، معین پیغمبرزاده[۳]

[۱] کارشناسی ارشد هوش مصنوعی، گروه مهندسی کامپیوتر، دانشگاه بوعلی‌سینا ، همدان،
Iman.ai.1372@gmail.com

[۲] استاد، گروه مهندسی کامپیوتر ، دانشگاه بوعلی‌سینا ، همدان
khotanlou@basu.ac.ir

[3] کارشناسی ارشد هوش مصنوعی، گروه مهندسی کامپیوتر، دانشگاه بوعلی‌سینا ، همدان
smmoeinp@gmail.com

چکیده

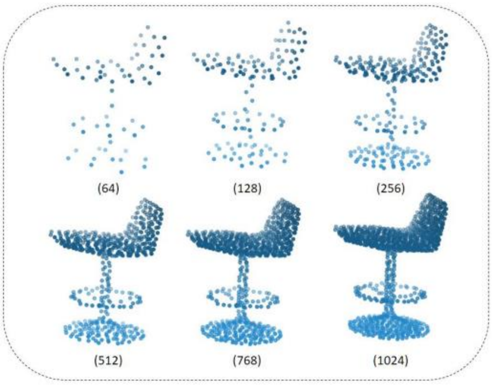
با گسترش روزافزون حوزه یادگیری عمیق ، امکان حل مسائل بسیاری که پیش از این دشوار به نظر می‌رسید، فراهم آمده است. از طرفی با تقویت ابزارها و نیز گسترش حسگرهای سه‌بعدی بینایی ماشین از جمله RGB-Dها و نیز Lidarها ، پردازش تصاویر سه‌بعدی و نیز داده‌هایی که فضای سه‌بعد را ترسیم می‌کنند، مورد توجه بیشتری قرار گرفته است. ابرنقاط سه‌بعدی از جمله داده‌های است که اجسام و پدیده‌ها را در فضای سه‌بعدی ترسیم می‌کند. در این مقاله به منظور بهبود دقت دسته‌بندی روشی ارائه شده است که با استفاده از تابع تغییر یافته KNN ویژگی‌های نسبی هر شیء را استخراج می‌کند و پس از نرمال‌سازی، با ویژگی‌های سراسری ادغام می‌کند. به این ترتیب ویژگی‌های سراسری حاصل از شبکه‌های عصبی چندلایه، بعد از اعمال هر لایه‌ی همسایگی به مجموعه ویژگی‌های جامع‌تری دسترسی پیدا می‌کنند و این ویژگی‌ها مدل را به فهم بهتری نسبت به اشیاء مورد پردازش می‌رساند. نتایج حاصل از آزمایش‌های گوناگون نشان می‌دهد این روش در مقایسه با روش‌های پیشتاز دسته‌بندی ابرنقاط از توان بسیارو کارائی بهتری برخوردار است.

کلمات کلیدی

ابرنقاط، دسته‌بندی، پردازش داده سه‌بعدی، شبکه‌های عصبی عمیق

۱- مقدمه

با توجه به کاربردهای فراوان پردازش تصاویر سه‌بعدی در بخش‌های مختلف صنعت همچون اتومبیل‌های خودران، ربات‌ها، واقعیت مجازی، سنجش‌ازدور و بسیاری زمینه‌ی دیگر، حوزه‌ی بینایی ماشین طی چندسال اخیر گرایش بسیاری به پردازش تصاویر و داده‌های سه‌بعدی پیدا کرده‌است[۱, ۲]. پیش از این برای رسیدن به درک پدیده‌های پیرامونی از تصاویر دوبعدی استفاده می‌شد که بخش زیادی از اطلاعات محیط را پوشش نمی‌داد . اما رفته‌رفته با گسترش حوزه بینایی ماشین و ظهور انواع داده‌های نوین این امکان فراهم شد که با حفظ ویژگی‌های هندسی اشیاء موجود در محیط به پردازش آن‌ها پرداخته‌شود.



شکل ۱ نمایش ابرنقطه یک صندلی از مجموعه دادگان ModelNet-40 **با تعداد نقاط متفاوت**

ابرنقاط گونه‌ی جدیدی از داده‌هاست که قادر است سطوح اجسام را به صورت دوبعدی و سه‌بعدی بازنمایی کند. آنچنان که از نامش بر می‌آید این نوع داده، حاصل قرار گرفتن مجموعه‌ای از نقاط در کنار یکدیگر است. هر ابرنقطه می‌تواند پدیده و یا مجموعه‌ای از پدیده‌ها را بازنمایی کند. در سال‌های اخیر با توجه به پیشرفت چشمگیر حسگرهای سه‌بعدی و در دسترس قرار گرفتن هر چه بیشتر آن‌ها، توجه به پردازش ابرنقاط بسیار افزایش یافته‌است. موبایل‌ها به دوربین‌های عمیق دارای قابلیت زمان پرواز مجهز شده‌اند و از طرفی ظهور انواع حسگرهای Lidar همچون Tele-15 ، BLK360 Leica و Kinect V2 دسترسی به ابرنقاط را تسهیل کرده است [۳-۵] .

مجموعه‌های ابرنقاط سه‌بعدی در ساده‌ترین حالت خود دارای سه مختصهx,y,z هستند که موقعیت مکانی هر نقطه را در فضا نمایش می‌دهد(شکل ۱) . علاوه بر آن می‌توان ویژگی‌هایی همچون رنگ، بردار نرمال و شدت نور را نیز به آن اضافه کرد که به عنوان ویژگی‌های افزوده به کار می‌روند[۶]. در مقایسه با تصاویر که ساختار پیکسل‌ها منسجم و منظم شکل گرفته است، در ابرنقاط با مجموعه‌ای از نقاط مواجه‌ایم که هیچ‌گونه نظم و انسجامی ندارند و با میزان تراکم متفاوتی در فضای سه‌بعد پراکنده‌اند. این ویژگی باعث می‌شود بخش‌های مختلف یک شیء با کیفیت متفاوتی بازنمایی شوند و در نتیجه طراحی روش و راه حلی برای پردازش ابرنقاط به سادگی صورت نپذیرد. اگرچه برای بهبود تفاوت در تراکم نقاط انواع الگوریتم‌های نمونه‌برداری تصادفی ، نمونه‌برداری توروراه و نمونه‌برداری

دورترین فاصله به کار گرفته‌شده‌اند اما پژوهشگران همچنان با این ویژگی به عنوان یک چالش برطرف نشده روبه‌رو هستند.

علاوه بر تفاوت تراکم نقاط در بازنمایی هر شیء، مجموعه‌ی نقاط سازنده‌ی هر ابرنقطه هیچ ترتیب خاصی ندارند[۷] . برخلاف تصاویر که در آن‌ها ترتیب پیکسل‌ها و روند پردازش هر یک از آن‌ها مشخص است ، ابرنقطه‌ی یک جسم می‌تواند با چینش‌های متفاوتی ذخیره شود و در عین حال جسمی یکتا را بازنمایی کند. این ویژگی ابرنقاط نیز چالش دیگریست که طراحی روش‌ها را دشوار می‌کند.

مجموعه این چالش‌ها و برخی دشواری‌های دیگر باعث شده‌اند که روش‌های معمول پردازش تصاویر دوبعدی، به سادگی قابل استفاده در پردازش ابرنقاط نباشند. به عنوان مثال شبکه‌های عصبی کانولوشنی که به عنوان یکی از روش‌های قدرتمند پردازش تصاویر دوبعدی شناخته می‌شوند، برای اعمال بر روی ابرنقاط نیاز به انجام تغییرات است . تعریف فیلترهای کانولوشنی در فضای ابرنقاط به سادگی تصاویر نیست و نیاز به تعریف همسایگی‌ها دارد[۶] . به همین دلیل پژوهشگران ناچارند برای استفاده از روش‌های کانولوشنی به جای تعریف فیلترهای ثابت به سراغ انواع الگوریتم همسایگی بروند. الگوریتم‌هایی چون KNN ، همسایگی شعاعی و نیز همسایگی استوانه‌ای از جمله این روش‌هاست[۸] .

سهم عمده‌ی این پژوهش در پردازش ابرنقاط عبارت است از :

- ارائه روشی مبتنی بر شبکه عصبی عمیق به منظور بهبود دقت دسته‌بندی ابرنقاط
- ارائه‌ی ماژولی به منظور استخراج ویژگی‌ها در همسایگی هر نقطه
- استفاده از ماژول همسایگی به طور سلسله‌مراتبی در شبکه‌های عصبی چندلایه به منظور استخراج ویژگی‌های نسبی
- ترکیب ویژگی‌های سراسری استخراج شده از شبکه‌های عصبی چندلایه و با ویژگی‌های حاصل از ماژول‌های همسایگی

در ادامه مقاله در بخش دوم به مرور کارهای پیشین می‌پردازیم. پس از آن روش پیشنهادی را در بخش سوم ارائه می‌دهیم. بخش چهارم اختصاص دارد به ارزیابی نتایج و نتیجه‌گیری در بخش پنجم آمده است.

۲- مروری بر کارهای پیشین

روش‌های گوناگونی به منظور پردازش ابرنقاط ارائه شده‌است که عمدتاً در دو دسته‌ی روش‌های کلاسیک و روش‌های مبتنی بر یادگیری عمیق قرار می‌گیرند. با توجه به اینکه در سالیان اخیر گرایش پژوهشگران بیشتر به استفاده از روش‌های مبتنی بر یادگیری عمیق بوده، در اینجا به بررسی و مرور روش‌هایی می‌پردازیم که از الگوریتم‌های یادگیری عمیق استفاده کرده‌اند.

در مواجهه با چالش‌هایی که تعدادی از آن‌ها در بخش مقدمه آمد، پژوهشگران عمدتاً سه رویکرد پیش گرفته‌اند. نخستین رویکرد تبدیل ابرنقاط سه‌بعدی به مجموعه‌ای از تصاویر دوبعدی است. با این کار می‌توان از الگوریتم‌های قدرتمند پردازش تصاویر دوبعدی استفاده کرد. دومین رویکرد، انتقال ابرنقاط سه‌بعدی به فضای مشبک‌های سه‌بعدیست. با این کار می‌توان نقاط نامنسجم را به فضای دارای ساختار مشبک‌ها برد و بخشی از چالش‌ها را حل کرد. رویکرد سوم به تحلیل و پردازش مستقیم ابرنقاط، بدون هیچگونه تبدیل و یا انتقال می‌پردازد که خود انواع گوناگونی دارد. در ادامه به بررسی روش‌های ارائه‌شده با هر یک از رویکردهای گفته‌شده می‌پردازیم.

۲-۱- روش‌های مبتنی بر بررسی چندوجهی

در این روش‌ها نماهای مختلف ابرنقاط سه‌بعدی به تصاویر دوبعدی تبدیل می‌شوند و پس از آن عمل استخراج ویژگی از تصاویر دوبعدی صورت می‌پذیرد. در انتها ویژگی‌های حاصل از تصاویر مختلف با یکدیگر تجمیع می‌شوند و عمل دسته‌بندی با تکیه بر ویژگی‌های به دست آمده انجام می‌شود. در این رویکرد روش MVCNN پیشگام است؛ از جمله ضعف‌های این رویکرد از دست رفتن بخشی از اطلاعات ضمن ایجاد تصاویر است[۹] .

۲-۲- روش‌های مبتنی بر نمایش‌های حجمی

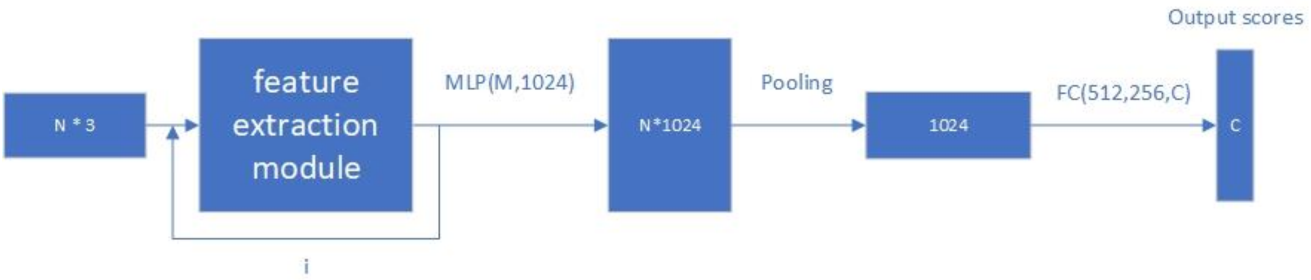
از آنجا که برای استفاده از شبکه‌های عصبی کانولوشنی نیاز به داده‌هایی با ساختار منسجم است، به کار گرفتن روش‌هایی همچون کانولوشن‌های سه‌بعدی بر روی ابرنقاط نامنظم امکان پذیر نیست. برای حل این چالش، برخی پژوهشگران، ابرنقاط سه‌بعدی را به فضای مشبک‌ها انتقال دادند تا بتوانند با تکیه بر ساختار منسجم و گسسته‌ی مشبک‌ها، فیلترهای دوبعدی را در فضای سه‌بعدی بازتعریف کنند. روش شاخص این رویکرد VoxNet است. نقدی که به این روش و روش‌های مشابه آن وارد می‌شود، بار محاسباتی و هزینه‌ی بسیار زیادیست که به سخت‌افزار تحمیل می‌کند[۱۰] .

۲-۳- روش‌های مبتنی بر تحلیل مستقیم ابرنقاط

از دست رفتن بخشی از اطلاعات و نیز بارمحاسباتی بسیار زیادِ حاصل از انتقال ابرنقاط به فضاهای دیگر ، باعث شد پژوهشگران به پردازش مستقیم ابرنقاط گرایش پیدا کنند. در این روش‌ها کوشیده می‌شود تا بدون انتقال و یا تبدیل ابرنقاط به فضایی دیگر، به طور مستقیم از نقاط ویژگی استخراج شود. از آنجایی که روش‌های این رویکرد خود به انواع مختلفی تقسیم‌بندی می‌شوند، دسته‌های مختلف این رویکرد را به تفکیک مورد بررسی قرار می‌دهیم.

۲-۳-۱- روش‌های مبتنی بر شبکه‌های عصبی چندلایه

در این روش‌ها کوشیده می‌شود نقاط به صورت منحصر به فرد مدل شوند. هر نقطه به پرسپترون چندلایه سپرده می‌شود و پس از آن با تجمیع ویژگی‌های حاصل از این نقاط، عمل دسته‌بندی صورت می‌پذیرد. روش پیشگام در این حوزه مقاله PointNet است[۱۱] .این روش به نقطه‌ی عطفی در پردازش ابرنقاط تبدیل شد. چرا که بی‌واسطه به پردازش ابرنقاط می‌پرداخت و به پاسخ خوبی می‌رسید. نقد اصلی به این روش، ناتوانی در به دست آوردن روابط میان نقاط بود. PointNet اگرچه به خوبی توانست ویژگی‌های سراسری یک ابرنقطه را استخراج کند اما در به دست آوردن ویژگی‌های محلی ناتوان بود. پس از طرح این روش، مقالات بسیاری کوشیدند که مشکل ضعف در استخراج روابط محلی را برطرف کنند. یکی از آن‌ها که



شکل ۲- معماری روش پیشنهادی که در آن N نمایانگر تعداد نقاط، i نمایانگر تعداد دفعات استفاده از لایه‌ی همسایگی ، M ابعاد ویژگی‌های حاصل از اعمال ماژول پیشنهادی و C برابر است با تعداد کلاس ها.

توسط نویسندگان PointNet ارائه شد، ++PointNet بود[۱۲] . این مقاله با استفاده از نمونه‌برداری و نیز همسایگی از روش PointNet به صورت سلسله‌مراتبی استفاده می‌کند و می‌کوشد تا ضعف‌های روش پیشین را برطرف کند. اگر چه مولفان ++PointNet توانستند با ایجاد این تغییرات میزان دقت را بهبود دهند اما همچنان روش‌های بسیاری با هدف استخراج ویژگی‌های محلی ارائه می‌شود.

۲-۳-۲- روش‌های مبتنی بر کانولوشن‌های سه‌بعدی

این روش‌ها بسته به نوع فیلتر مورد استفاده، به دو نوع گسسته و پیوسته تقسیم‌بندی می‌شوند. فیلترهای گسسته هنگامی مورد استفاده قرار می‌گیرند که ابرنقاط به فضای مشبک‌ها انتقال یافته باشد. اما منظور از روش‌های کانولوشن پیوسته، روش‌هایست که با تعریف عملگرهایی در فضای نامنظمِ ابرنقاط در پی یافتن و استخراج ویژگی‌های محلی در نواحی مختلف هستند. از جمله‌ی این روش‌ها KPConv است[۱۳] . این روش با تعریف دو نوع فیلتر می‌کوشد تا روشی کانولوشنی در فضای سه‌بعد ارائه دهد. فیلترها دو نوع هستند و از تعدادی نقطه تشکیل شده‌اند. یکی از آن‌ها از تعدادی نقطه با موقعیت فضایی ثابت و دیگری از تعدادی نقطه با قابلیت تغییر موقعیت مکانی تشکیل شده‌است. این فیلترها بر روی ابرنقاط سه‌بعدی گام برمی‌دارند و می‌کوشند تا بافت‌ها و ویژگی‌های ممکن را از طریق روندی کانولوشنی استخراج کنند و به فهم بهتری نسبت به اشیاء برسند. این روش علی‌رغم دستیابی به دقت قابل توجه نسبت به روش‌های ارائه‌شده ، بار محاسباتی بسیار زیادی دارد. از روش‌های دیگر این حوزه می‌توان به PointConv ، ConvPoint و SpiderConv اشاره کرد[۱۴-۱۶] .

۲-۳-۳- روش‌های گرافی

شبکه‌های مبتنی بر گراف، هر یک از نقاط ابر نقاط را به عنوان رأسی از گراف در نظر می‌گیرند و روابط همسایگی را با استفاده از یال‌ها تعریف می‌کنند. پس از آن عمل آموختن ویژگی‌ها در دو حوزه‌ی مختلف می‌تواند صورت گیرد. حوزه‌ی مکان و حوزه‌ی فرکانس. از جمله روش‌های مطرح این حوزه DGCNN است . این روش در فضای ویژگی‌ها گراف می‌سازد و این گراف را در هر مرحله به روزرسانی می‌کند. و در انتها با ویژگی‌های حاصل‌شده عمل دسته‌بندی را انجام می‌دهد[۱۷] .

۳- روش پیشنهادی

همانطور که پیش‌تر توضیح داده شد روش‌های پیشین توانسته‌اند ویژگی‌های سراسری و تا حدودی ویژگی‌های محلی را به خوبی استخراج کرده و به کار گیرند. در روش پیشنهادی کوشیده شده تا ضمن در نظر گرفتن ویژگی‌های سراسری و محلی ، ویژگی‌های نسبی نیز استخراج شود و در روند پردازشی مورد استفاده قرار گیرد. منظور از ویژگی‌های نسبی آن دسته از ویژگی‌هاییست که حاصل یافتن روابط بین نقاط است. روابطی که الزاماً ما بین نقاط نزدیک نیست بلکه نقاط دوردست را نیز شامل می‌شود. بخش‌های مختلف روش پیشنهادی در ادامه آمده است. معماری کامل روش پیشنهادی در شکل ۲ قابل مشاهده است.

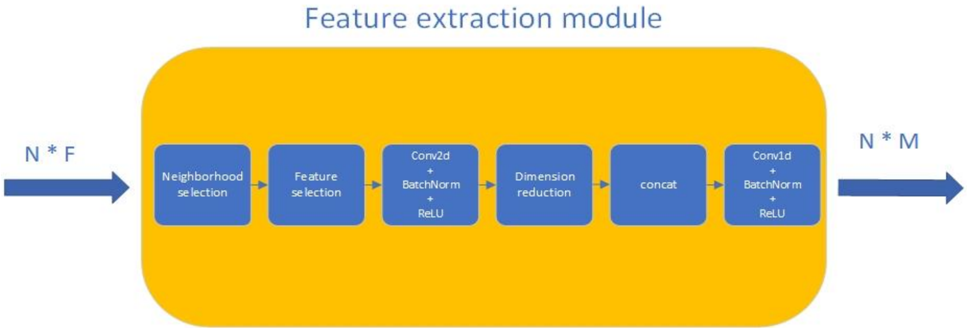
۳-۱- ماژول استخراج ویژگی

این ماژول با استفاده از شکلی توسعه یافته از الگوریتم KNN قادر است تعدادی از نقاط را به عنوان یک لایه‌ی همسایگی-در اطراف یک نقطه انتخاب کند و از آن‌ها ویژگی استخراج کند. با استفاده از این ماژول و در هر بار اعمال آن، بافت‌هایی از ابرنقاط استخراج می‌شود و در نهایت با ویژگی‌های سراسری ادغام می‌گردد. این عملیات می‌تواند به صورت سلسله‌مراتبی تکرار گردد.

$$N_{HKNN}(y_0, P, k, h) = \{x_{k,h} \in P, |y_0 - x_{k,h-1}| \leq |y_0 - x_{k,h-1}|, |y_0 - x_{k,h}| \leq |y_0 - x_{k,h-1}|, k, h \in N\} \quad (۱)$$

که در آن k تعداد همسایه‌ها، h لایه‌ی همسایگی ، y_0 نقطه‌ی موردنظر، $x_{k,h}$ hاُمین همسایگی شامل k همسایه و P مجموعه ابرنقاط می‌باشد. حاصل مجموعه‌ای از نقاط خواهد بود که برمبنای فاصله از هر نقطه به دست آمده است. بسته به اینکه این تابع در کدام مرحله از شبکه عصبی چندلایه اعمال شود ویژگی‌های نقاط گزین‌شده انتخاب خواهد شد.

در مرحله‌ی بعد نرمالسازی صورت می‌گیرد. به این معنا که مقادیر ویژگی همسایه‌ها نسبت به مقدار ویژگی نقطه‌ی مرکزی تنظیم می‌شود. و ویژگی‌های انتخاب‌شده از هر لایه همسایگی در حالت نرمال و غیرنرمال به کانولوشن ۲بعدی سپرده می‌شوند. خروجی این کانولوشن پس از انجام عملیات نرمالسازی بسته‌ای و اعمال تابع فعال‌ساز Relu برداری خواهد بود که با کاهش بعد به شبکه سپرده خواهد شد. عملیات کاهش بعد با استفاده از عمگر ماکزیمم صورت می‌پذیرد. این بردار ویژگی به ویژگی‌های سراسری



شکل ۳- ماژول استخراج ویژگی که در آن N نمایانگر تعداد نقاط، F نشاندهنده‌ی ویژگی‌های ورودی و M نشاندهنده‌ی ویژگی‌های خروجی است

حاصل از مراحل پیشین شبکه اضافه می‌شود. عمل ادغام ویژگی‌ها با استفاده از عملگر الحاق صورت می‌پذیرد. بدین گونه که ویژگی‌های نسبی به دست آمده در هر لایه‌ی همسایگی به ویژگی‌های سراسری الحاق می‌شود و ادامه پردازش آنچنان که در شکل ۳ آمده است، پیگیری می‌شود.

پس از آخرین بار استفاده از ماژول استخراج ویژگی و انجام عمیات ادغام، خروجی به لایه‌های تماماً متصل سپرده می‌شود تا عمل دسته‌بندی انجام گیرد.

۴- ارزیابی نتایج

در این بخش به بررسی و ارزیابی نتایج به دست آمده از روش پیشنهادی می‌پردازیم. برای اینکار ابتدا پایگاه داده را معرفی کرده، سپس از معیارهای ارزیابی مورد استفاده سخن می‌گوییم. جزئیات پیاده‌سازی و نیز مقایسه با دیگر روش‌ها در انتهای این بخش آمده است.

۴-۱- پایگاه داده

در این پژوهش از مجموعه داده ModelNet40 استفاده شده است. این پایگاه داده شامل ۱۲۳۱۱ ابرنقطه است که در ۴۰ کلاس تقسیم شده‌اند. از این تعداد ۹۸۴۳ ابرنقطه برای آموزش و ۲۴۶۸ ابرنقطه برای آزمون به کارگرفته می‌شود. داده‌های این مجموعه مصنوعی هستند و همگی از روی مشبک نمونه‌برداری شده‌اند. هیچگونه انسدادی در این پایگاه داده وجود ندارد و فاقد پس‌زمینه می‌باشد.

۴-۲- معیار ارزیابی

برای ارزیابی مسائل مختلف در حوزه‌ی ابرنقاط ،معیارهای گوناگونی تعریف شده است. در مورد مسئله دسته‌بندی ابرنقاط این معیارها عبارتند از *Overall accuracy* و *Mean class accuracy*. *OA* به عنوان نخستین معیار، عبارت است از دقت میانگین بر روی همه‌ی نمونه‌های تست و *mAcc* عبارت است از میانگین دقت بر روی همه کلاس‌ها.

$$mAcc = \frac{1}{|class|}\sum_{i\in class}\frac{TOTAL\ CORRECT_i}{TOTAL\ SEEN\ INSTANCE_i} \quad (۲)$$

$$OA = \frac{TOTAL\ INSTANCE}{TOTAL\ SEEN\ INSTANCE} \quad (۳)$$

که در دو عبارت فوق صورت کسرها به معنای آن دسته از نمونه‌هایی است که به درستی تشخیص داده شده‌اند و نیز مخرج کسر به معنای همه نمونه‌های بررسی شده‌اند. تعداد کلاس‌ها به صورت $|class|$ نشان داده شده است.

۴-۳- جزئیات پیاده‌سازی

این پژوهش با استفاده از زبان برنامه‌نویسی پایتون ۳.۹ و بر بستر Pytorch پیاده‌سازی شده است . این نسخه‌ی پیاده‌سازی شده بر روی دو سیستم مورد آزمایش قرار گرفته است. نخستین سیستم دارای پردازنده‌ی گرافیکی 3070 Ti با میزان حافظه‌ی اصلی ۸ گیگابایت و دومین سیستم دارای پردازنده گرافیکی 1070 Ti و با حافظه‌ی اصلی ۱۶ گیگابایت بوده است.

برای آموزش این شبکه از بهینه‌ساز ADAM استفاده شده است. آموزش، در بسته‌های ۱۶، ۲۴ و ۳۲ تایی ، با نرخ یادگیری ۰.۰۰۱ و نیز با نرخ کاهشی ۰.۰۰۰۱ صورت پذیرفته است. از طرفی تعداد همسایه‌های گوناگونی آزموده شده که از جمله آن‌ها می‌توان به مقادیر ۲۰ و ۳۲ اشاره کرد. ۲۵۰ بار روند آموزش تکرار گشته و در نهایت به نتیجه‌ی ارائه‌شده همگرا شده است.

۴-۴- مقایسه با دیگر روش‌ها

نتایج حاصل از روش ارائه‌شده در جدول ۱ آمده است. آزمایش‌ها نشان می‌دهد استفاده از دو لایه‌ی همسایگی در مقایسه با یک لایه‌ی همسایگی به نتیجه‌ی بهتری منجر می‌شود. و همینطور با افزودن سومین لایه همسایگی، دقت بهبود می‌یابد. نتایج آزمایش‌ها همچنین نشان می‌دهد که روش ارائه‌شده نسبت به بسیاری از روش‌های مطرح این حوزه عملکرد بهتری دارد.

جدول ۱ مقایسه روش ارائه شده با دیگر روش‌ها(باویژگی‌های ورودی x,y,z)

روش	mAcc-دقت میانگین(%)	OA-دقت کل(%)
PointNet[11]	86.2	89.2
FlexConv[18]	-	90.2
PointNet++[12]	-	90.7
Kd-Network[19]	88.5	91.8
PointPlaneNet[20]	90.1	92.0
PointCNN[21]	-	92.2
DGCNN[17]	90.2	92.9
Ours(one layer)	**89.6**	**92.3**
Ours(Two layer)	**90.5**	**92.6**
Ours(Three layer)	**90.7**	**93.0**

۵- نتیجه‌گیری

آنچنان که از نتایج بر می‌آید، روش ارائه شده با افزایش تعداد تکرارهای اعمال ماژول توانسته به بهبود قابل ملاحظه‌ای در هر دو معیار ارزیابی برسد. این نتایج گواه بر این امر است که با استخراج روابط میان نقاط و استفاده از ویژگی‌های هر نقطه در کنار یکدیگر مدل می‌تواند به فهم بهتری نسبت اشیاء برسد و با تکیه بر آن دقت تصمیم‌گیری را افزایش دهد. از کارهای آینده بررسی و طراحی مدل‌هاییست که قادر به استخراج بهتر روابط میان نقاط باشد و ویژگی‌های نسبی را به شکل بهینه‌تری به کار گیرد.

مراجع

[۱] S. D. Pendleton *et al.*, "Perception, planning, control, and coordination for autonomous vehicles," *Machines*, vol. 5, no. 1, p. 6, 2017.

[۲] J. W. Weingarten, G. Gruener, and R. Siegwart, "A state-of-the-art 3D sensor for robot navigation," in *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)(IEEE Cat. No. 04CH37566)*, 2004, vol. 3: IEEE, pp. 2155-2160 .

[۳] L. Tech, "Tele-15," 2020.

[۴] "Leica Geosystems.LEICA BLK360," 2016.

[۵] M. Corporation, "Kinect V2 3D Scanner," 2014.

[۶] Y. Guo, H. Wang, Q. Hu, H. Liu, L. Liu, and M. Bennamoun, "Deep learning for 3d point clouds: A survey," *IEEE transactions on pattern analysis and machine intelligence*, vol. 43, no. 12, pp. 4338-4364, 2020.

[۷] S. A. Bello, S. Yu, C. Wang, J. M. Adam, and J. Li, "Deep learning on 3D point clouds," *Remote Sensing*, vol. 12, no. 11, p. 1729, 2020.

[۸] H. Thomas. U. P. s. e. lettres. (2019). Learning new representations for 3D point cloud semantic segmentation (KPCONV) .

[۹] H. Su, S. Maji, E. Kalogerakis, and E. Learned-Miller, "Multi-view convolutional neural networks for 3d shape recognition," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 945-953 .

[۱۰] D. Maturana and S. Scherer, "Voxnet: A 3d convolutional neural network for real-time object recognition," in *2015 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, 2015: IEEE, pp. 922-928 .

[۱۱] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "Pointnet: Deep learning on point sets for 3d classification and segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 652-660 .

[۱۲] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "Pointnet++: Deep hierarchical feature learning on point sets in a metric space", *Advances in neural information processing systems*, vol. 30, 2017.

[۱۳] H. Thomas, C. R. Qi, J.-E. Deschaud, B. Marcotegui, F. Goulette, and L. J. Guibas, "Kpconv: Flexible and deformable convolution for point clouds," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 6411-6420 .

[۱۴] W. Wu, Z. Qi, and L. Fuxin, "Pointconv: Deep convolutional networks on 3d point clouds," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 96۲۱-۹۶۳۰.

[۱۵] A. Boulch, "ConvPoint: Continuous convolutions for point cloud processing," *Computers & Graphics*, vol. 88, pp. 24-34, 2020.

[۱۶] Y. Xu, T. Fan, M. Xu, L. Zeng, and Y. Qiao, "Spidercnn: Deep learning on point sets with parameterized convolutional filters," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 87-102 .

[۱۷] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon, "Dynamic graph cnn for learning on point clouds," *Acm Transactions On Graphics (tog)*, vol. 38, no. 5, pp. 1-12, 2019.

[۱۸] F. Groh, P. Wieschollek, and H. Lensch, "Flex-convolution," in *Asian Conference on Computer Vision*, 2018: Springer, pp. 105-122 .

[۱۹] R. Klokov and V. Lempitsky, "Escape from cells: Deep kd-networks for the recognition of 3d point cloud models," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 863-872 .

[۲۰] S. M. Peyghambarzadeh, F. Azizmalayeri, H. Khotanlou, and A. Salarpour, "Point-PlaneNet: Plane kernel based convolutional neural network for point clouds analysis," *Digital Signal Processing,* vol. 98, p. 102633, 2020.

[۲۱] Y. Li, R. Bu, M. Sun, W. Wu, X. Di, and B. Chen, "Pointcnn: Convolution on x-transformed points," *Advances in neural information processing systems* , vol. 31, 2018.

[۲۲] X. Liu, Z. Han, Y.-S. Liu, and M. Zwicker, "Point2sequence: Learning the shape representation of 3d point clouds with an attention-based sequence to sequence network," in *Proceedings of the AAAI Conference on Artificial Intelligence* ,۲۰۱۹ ,vol. 33, no. 01, pp. 8778-8785 .

[۲۳] X. Yan, C. Zheng, Z. Li, S. Wang, and S. Cui, "Pointasnl: Robust point clouds processing using nonlocal neural networks with adaptive sampling," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 5589-5598 .

چهارمین کنفرانس ملی انفورماتیک ایران
پژوهشگاه دانش‌های بنیادی، پردیس فرمانیه، تهران
۱۴ و ۱۵ دی ماه ۱۴۰۱

قطعه بندی خودکار دو مرحله‌ای اسبک مغز در تصاویر MRI با استفاده از شبکه های عصبی عمیق

علیرضا صادقی[۱]، حسن ختن لو[۲]، محمدرضا رضاییان[۳]

[۱] گروه مهندسی کامپیوتر، دانشگاه بوعلی سینا، همدان
a.sadeghi@eng.basu.ac.ir

[۲] گروه مهندسی کامپیوتر، دانشگاه بوعلی سینا، همدان
khotanlou@eng.basu.ac.ir

[۳] گروه مهندسی پزشکی، دانشگاه صنعتی همدان، همدان
rezaeian@hut.ac.ir

چکیده

اسبک مغزی قسمتی کوچک از دستگاه لیمبک مغز است که مرکز یادگیری می‌باشد. این ناحیه از مغز در بیماری‌هایی همچون صرع و آلزایمر دچار تغییر اندازه و حجم می‌شود. بررسی تغییر حجم و اندازه این بخش از مغز بصورت دستی کاری دشوار و زمان بر است و قطعه بندی خودکار این بخش از مغز برای بررسی و مطالعه مواردی همچون تغییر حجم و اندازه می‌تواند برای پزشکان بسیار حائز اهمیت باشد. در این مقاله روشی مبتنی بر شبکه‌های عصبی عمیق برای قطعه بندی خودکار اسبک مغزی در تصاویر MRI در دو فاز مجزا ارائه شده است که می‌تواند ناحیه اسبک مغزی را با دقت خوبی قطعه بندی کند. در فاز اول با استفاده از یک مدل شبکه عصبی پیچشی U_Net ماسک اولیه تولید می‌شود و سپس در فاز بعدی با استفاده از ماسک تولید شده در فاز اول یکبار دیگر با استفاده از مدل شبکه عصبی پیچشی U_Net که در فاز اول نیز از آن استفاده شد، اسبک مغزی قطعه بندی می‌شود. نتایج بدست آمده نشان می دهد که روش ارائه شده در مقایسه با روشهای دیگر از دقت بهتری برخوردار می‌باشد.

کلمات کلیدی

اسبک مغزی، قطعه بندی، هیپوکامپوس، تصویر MRI

۱- مقدمه

اسبک مغز[۱] بخش کوچکی از مغز پستانداران می‌باشد. اسبک یک ساختار مغزی کوچک، میانی و زیر قشری[۲] است و با حافظه بلند مدت و کوتاه مدت مرتبط است [1]. بررسی مواردی همچون تشخیص، قطعه بندی، ساختار و ابعاد اسبک مغز در تصاویر تشدید مغناطیسی[۳] بارها در پژوهش‌های مختلف مورد مطالعه و ارزیابی قرار گرفته است. اهمیت این مطالعات و پژوهش‌ها به دلیل نقش حیاتی شناسایی اسبک مغز در پاتوفیزیولوژی[۴] بیماری‌های عصبی و روانی می‌باشد.

شکل و حجم اسبک مغز می‌تواند تحت تاثیر آسیب‌های مختلف همچون بیماری آلزایمر تغییراتی داشته باشد و با توجه به سخت بودن و زمان‌بر بودن تشخیص ناحیه دقیق اسبک در تصاویر MRI، قطعه بندی و تقسیم بندی خودکار بخش اسبک مغز در تصاویر می‌تواند کمک زیادی به جامعه آسیب شناسان و پزشکان کند. به عنوان مثال از آنجایی که تشخیص سریع آلزایمر در مراحل ابتدایی می‌تواند کمک بزرگی به بیمار بکند و همچنین اولین قسمتی که در بیماری آلزایمر آسیب می‌بینند اسبک مغز می‌باشد، تشخیص و قطعه بندی ناحیه اسبک مغز در تصاویر MRI می‌تواند اطلاعاتی در مورد تغییر حجم و اندازه آن نسبت به حالت نرمال به متخصصین ارائه کند [1,2].

در این پژوهش اقدام به قطعه بندی و مشخص کردن ناحیه اسبک مغز در تصاویر MRI شده است. MRI یا روش تصویربرداری با تشدید مغناطیس یکی از روش های پیشرفته تصویربرداری پزشکی است. با استفاده از این روش میتوان از اجزا و بافت‌های درونی بدن تصویر برداری کرد و با استفاده از تصاویر اخذ شده مشکلات و بیماری های بدن را تشخیص داد. از آنجایی که تصاویر MRI می‌تواند اطلاعات بسیار ارزشمندی در مورد ساختارهای بدن مانند شکل، اندازه و محل قرار گیری ارائه دهد، این گونه تصویر برداری امروزه توانسته است توجه پژوهشگران در حوزه‌های پزشکی و تشخیص به کمک کامپیوتر[۵] را به خود جلب کند [3, 4].

تصاویر MRI که از مغز گرفته می‌شود، اطلاعات کامل و جامعی در مورد ساختار، حجم و ابعاد مغز به پزشکان و متخصصین ارائه می‌دهد، ولی از آنجایی که ناحیه اسبک مغز ابعاد کوچکی دارد و توسط قسمت‌های دیگر مغز تا حدودی پوشانده شده است، بررسی و تشخیص آن بصورت دستی کاری زمان‌بر و سخت می‌باشد [5 ,1].

با توجه به مشکلاتی که تجزیه و قطعه بندی ناحیه اسبک مغز در تصاویر MRI بصورت دستی دارد، ایجاد سیستمی که بتواند بطور خودکار این ناحیه را تشخیص دهد می‌تواند در روند افزایش سرعت، دقت و صحت قطعه بندی اسبک مغز موثر باشد.

در سالیان اخیر روش‌های قطعه بندی مبتنی بر شبکه‌های عصبی عمیق[۶] برای قطعه بندی تصاویر پزشکی بسیار مورد توجه قرار گرفته است.

تا کنون پژوهش‌های فراوانی در رابطه با قطعه بندی اسبک مغزی صورت گرفته است که در ادامه ابتدا روش و شیوه‌ی کار برخی از مهم‌ترین پژوهش‌های صورت گرفته در این حوزه و سپس مدل پیشنهادی در این پژوهش بررسی خواهد شد.

تا قبل از ظهور و رشد شبکه‌های عصبی عمیق از روش‌های سنتی پردازش تصویر برای قطعه بندی در تصاویر استفاده می‌شد. اما پس از ظهور و توسعه روش‌های یادگیری عمیق استفاده از این روش‌ها افزایش قابل توجهی

در میان پژوهشگران داشت [1]. قطعه بندی ناحیه اسبک مغزی نیز ازین قاعده مستثنی نیست و در پژوهش‌های انجام شده در سالیان اخیر استفاده از روش‌های یادگیری عمیق برای قطعه بندی این ناحیه از مغز در تصاویر MRI بارها مورد بررسی و آزمایش قرار گرفته است که در ادامه برخی از پژوهش‌های مهم انجام گرفته در سالیان اخیر مورد بررسی قرار گرفته می‌شود.

Chen و همکاران در سال ۲۰۱۷ پژوهشی در زمینه تقسیم بندی اسبک مغزی با استفاده از مدل شبکه عصبی[7] ConvNets ارائه دادند [5]. آن‌ها مدل خود را با استفاده از معیار Dice ارزیابی کرده‌اند، که این معیار، معیاری برای سنجش شباهت میان دو مجموعه داده است. در این مقاله گزارش شده است که با استفاده از این روش آن‌ها توانسته‌اند با استفاده از مدلی که ارائه دادند به Dice برابر ۰.۹ دست پیدا کنند [5].

Thyreau و همکاران مدلی به نام Hippodeep را در سال ۲۰۱۸ معرفی کردند [6]. این مدل بصورت عمومی موجود و در دسترس می‌باشد. آن‌ها در پژوهش خود از سه نوع آوری کلیدی استفاده کردند. در گام نخست مجموعه داده آموزشی متغیر و گسترده ای را در نظر گرفتند علاوه بر این برچسب های آموزشی که در این پژوهش استفاده شد را تا حدی از خروجی الگوریتم FreeSurfer استخراج کردند و در گام نهایی از داده های مصنوعی و از یک طرح تقویت داده قدرتمند استفاده کردند [6]. آن‌ها با استفاده از شبکه‌های عصبی پیچشی در منطقه مورد نظر[8] اقدام به قطعه بندی اسبک مغز کردند. پژوهشگران حاضر در این مطالعه از تصاویر MRI متعلق به بیش از ۲۰۰۰ بیمار استفاده کردند و مدلی که آنها بدست آوردند بر روی دیتاست Harp به Dice برابر ۰.۸۵ رسید [6]. Nogovitsyn و همکاران نیز روش Thyreau را مورد بررسی قرار دادند و درستی روش آن‌ها را آزمایش و تایید کردند [7].

Wachinger و همکاران مدل DeepNat را بررسی کردند [8]. DeepNat در کنار استفاده از یک روش پارامتری سازی جدید، از فیلدهای تصادفی شرطی سه بعدی برای پس‌پردازش استفاده می کند [8]. DeepNat عمدتاً بر روی مجموعه داده MALC تمرکز دارد و آن‌ها توانستند بر روی این مجموعه داده به معیار Dice برابر ۰.۸۶ دست پیدا کنند [8].

Roy و همکاران با ارائه و بررسی مدل QuickNat گام دیگری در راستای توسعه قطعه بندی اسبک مغزی برداشتند [9]. آنها از رویکرد CNN چندگانه استفاده کردند و با استفاده از این رویکرد توانستند تقسیم‌بندی سریع‌تری نسبت به DeepNat انجام دهند. مدلی که آن‌ها ارائه دادند از چندین معماری U-Net دو بعدی پیروی می کند [9].

Dinsdale و همکاران نیز مطالعاتی در این زمینه داشتند. آن‌ها از یک مدل سه بعدی CNN U-Net استفاده کردند و توانستند به Dice برابر با ۰.۸۶ برسند. این پژوهشگران نیز از پایگاه داده Harp برای ارزیابی مدل خود استفاده کردند [10].

در سال ۲۰۲۱ Carmo و همکاران اقدام به قطعه بندی اسبک مغز برای مطالعات و تحقیقات در بیماری‌های صرع و آلزایمر کردند [1]. آن‌ها از معماری U-Net و دیتاست Harp و پایگاه داده جمع آوری شده از بیماران مبتلا به صرع به نام Hcunicamp که شامل تصاویر MRI بیمارانی که اسبک مغز آن‌ها برداشته شده است نیز هست، استفاده کردند [1]. آن‌ها از سه شبکه عصبی پیچشی[9] U-Net با الهام از پژوهش انجام گرفته توسط Lucena و همکاران، در سه جهت مختلف استفاده کردند. در نهایت آن‌ها توانستند در بهترین حالت به Dice حدودا ۰.۹۱ برسند [1, 11].

همانطور که از پژوهش‌های اخیر قابل استنباط است، استفاده از شبکه‌های عصبی عمیق مخصوصا شبکه‌های عصبی پیچشی عمیق نتایج خوب و قابل توجهی ارائه داده است [1, 5, 6]. مدل U-Net از جمله مدل‌های پر استفاده در زمینه قطعه بندی تصاویر پزشکی می‌باشد. در این پژوهش نیز از شبکه عصبی پیچشی و مدل U_Net استفاده شده است.

پژوهش حاضر نسبت به پژوهش‌هایی که تا کنون در حوزه قطعه بندی اسبک مغز صورت گرفته و برخی از آن‌ها بررسی شد، دو ویژگی کلیدی دارد. اولین ویژگی کلیدی این پژوهش نسبت به دیگر پژوهش‌ها تشخیص دو مرحله‌ای و استفاده از مدل U-Net یکسان با ورودی‌های متفاوت است. و دومین ویژگی کلیدی این پژوهش استفاده از یک ماسک با الگوریتمی ابتکاری است، که برای مشخص کردن ناحیه‌ای که احتمال وجود اسبک مغز در آن مکان هست، استفاده می‌شود.

اولاف رنه برگر و همکاران وی در سال ۲۰۱۵ مدل U-Net را معرفی کردند [11]. مدلی که آن‌ها معرفی کردند یک نکته مهم داشت که باعث می‌شد میزان پیچیدگی شبکه کاهش پیدا کند. این ویژگی عدم وجود لایه‌های تمام متصل می‌باشد. آن‌ها در این مدل از یک مسیر انقباضی متوالی که در آن لایه‌های pooling به‌جای لایه‌های sampling قرار گرفته است استفاده کردند [11].

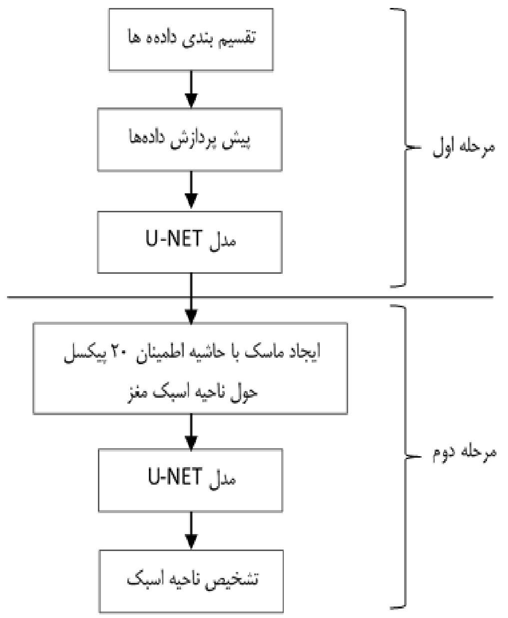
در این پژوهش نیز از مدل شبکه عصبی پیچشی U_Net در دو مرحله‌ی مختلف برای عملیات قطعه بندی ناحیه اسبک مغز در تصاویر MRI استفاده شده است. در این پژوهش از تصاویر MRI موجود در پایگاه داده Harp استفاده شده است.

در هر دو مرحله پیاده سازی شده در این پژوهش از مدل U-Net استفاده شده است. U-Net یک معماری شبکه کانولوشن برای تقسیم بندی سریع و دقیق تصاویر است که بارها برای قطعه بندی در تصاویر پزشکی از این معماری استفاده شده است.

از مهم ترین دلایلی که باعث شده تا الگوریتم U-Net در سال های اخیر در امر شناسایی اجزای تصویر و بخش بندی و قطعه بندی کردن تصاویر بخصوص در زمینه پردازش تصاویر پزشکی، مورد توجه قرار گیرد، می توان به مواردی همچون پاسخ دهی با دقت بالا، سرعت بالا فرآیند آموزش و عدم نیاز به مجموعه داده بزرگ اشاره کرد.

۲- روش پیشنهادی

روش پیشنهادی برای قطعه بندی اسبک مغز که در این پژوهش به آن پرداخته شده است، مبتنی بر شبکه‌های عصبی عمیق و الگوریتم U-Net می‌باشد. در این پژوهش مدل پیشنهادی برای قطعه بندی تصاویر اسبک مغز در تصاویر MRI ارائه شده است. مراحلی که در این پژوهش دنبال شده است در شکل (۱) قابل ملاحظه می‌باشد.



شکل (۱) : روش پیشنهادی

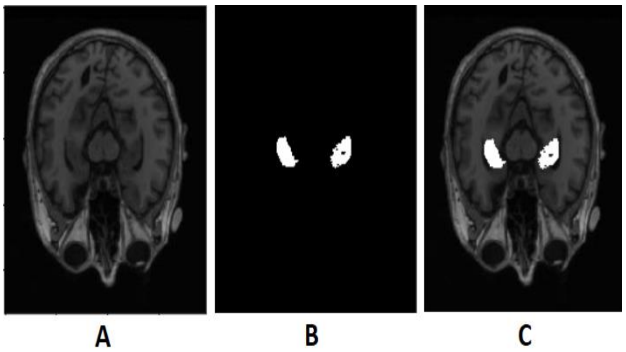
روند کلی که برای قطعه بندی و تشخیص اسبک مغزی که در این پژوهش پیش گرفته شد، در سه بخش پیاده سازی شده است که در ابتدا بطور خلاصه این مراحل بررسی خواهد شد.

در این پژوهش از پایگاه داده Harp که به صورت عمومی موجود و در دسترس می‌باشد، استفاده شده است پس از جمع آوری پایگاه داده متشکل از تصاویر MRI، در قدم اول اقدامات پیش پردازش روی تصاویر انجام می‌شود، سپس در بخش دوم مدل شبکه عصبی در دو مرحله آموزش داده می‌شود. در هر دو مرحله این بخش، از الگوریتم U-Net استفاده شده است. علاوه بر الگوریتم U-Net استفاده شده در این بخش از یک الگوریتم برای تولید ماسک نیز استفاده شده است. بعد از گذر تصاویر از مدل U-Net اولیه، این مدل برای هر تصویر یک برچسب ایجاد می‌کند که با استفاده از این برچسب و الگوریتم تولید ماسک، برای هر داده یک ماسک مستطیل شکل ایجاد می‌شود که با ضرب ماسک تولید شده برای هر داده، در تصویر MRI مربوطه داده‌های ماسک گذاری شده بدست می‌آید. این داده‌های ماسک گذاری شده فقط شامل نواحی از تصویر MRI هستند که مدل U-Net اولیه احتمال حضور ناحیه اسبک را در آن ناحیه پیش بینی کرده است. سپس این داده‌های ماسک گذاری شده یکبار دیگر با مدل U-Net آمورش داده می‌شود. در بخش نهایی نیز نوبت به تست و ارزیابی مدل نهایی با استفاده از داده‌های آموزشی می‌رسد که در این بخش مدل تولید شده نهایی، مورد بررسی و تست قرار می‌گیرد. در این بخش پس از تولید برچسب تصاویر آزمایشی، با استفاده از معیار Dice مدل مورد ارزیابی قرار گرفته و نتیجه به دست آمده با نتایج بدست آمده توسط بقیه پژوهشگران و مقالات آن‌ها مقایسه می‌شود.

۲-۱- پیش پردازش داده‌ها

در بخش اول به بررسی، انتخاب و پیش پردازش داده‌ها پرداخته شد. در این پژوهش از پایگاه داده Harp استفاده شده است. این پایگاه داده شامل ۱۳۵ تصویر T1 MRI می‌باشد که بطور عمومی در دسترس می‌باشد. این تصاویر در سه دسته جدا کلاس بندی شده اند. دسته‌ی اول این تصاویر، تصاویر MRI مغزی از افراد نرمال و سالم می‌باشد، دسته‌ی دوم تصاویر مربوط به تصاویر MRI مغزی از افرادی است که اختلال شناختی خفیفی دارند و در نهایت دسته سوم تصاویر مربوط به افرادی است که درگیر بیماری آلزایمر می‌باشند [13].

شکل (۲) نمونه‌ای از تصاویر موجود در پایگاه داده Harp می‌باشد. تصویر A در شکل (۲) یک تصویر T1 MRI است. تصاویر B و C نیز به ترتیب بیانگر ماسک تصویر و قرار گیری ماسک تصویر روی تصویر MRI است.



شکل (۲) : نمونه تصویری از پایگاه داده به همراه ماسک آن

تصویر B در شکل (۲) یک تصویر دودویی سیاه و سفید است که مشخص کننده مکان قرارگیری اسبک مغز در تصویر MRI می‌باشد. به نوعی هر تصویر MRI دارای یک ماسک سیاه و سفید است که این ماسک مشخص کننده مکان قرار گیری اسبک در تصویر می‌باشد.

۲-۲- شبکه عصبی عمیق ارائه شده

در این فاز ابتدا با استفاده از داده‌های مربوط به فاز آموزش یک مدل شبکه عصبی پیچشی U-Net آموزش داده می‌شود و سپس برچسب‌های داده‌های آموزشی با استفاده از این مدل U-Net پیش بینی خواهد شد.

پس از پیش بینی برچسب‌های داده‌های آموزشی یک ماسک ایجاد می‌شود و این ماسک در تصویر MRI اصلی ضرب می‌شود. ایجاد این ماسک از نکات قابل توجه این پژوهش است که از دلایل انجام این کار می‌توان به مواردی همچون حذف قسمت‌های اضافه تصویر MRI و تمرکز بر ناحیه اسبک مغزی و در نهایت کاهش پیچیدگی در پردازش و تولید ماسک از تصویر MRI توسط مدل U-Net اشاره کرد. برای تولید این ماسک از الگوریتمی که در ادامه بحث می‌شود، استفاده شده است.

این الگوریتم ابتدا ۴ نقطه در تصویر ماسک پیش بینی شده پیدا نموده که این چهار نقطه بیانگر چهار پیکسل در تصویر برچسب می‌باشند. از آنجایی که برچسب‌ها تصاویری سیاه و سفید هستند، ماتریس تصاویر برچسب‌ها یک بعدی بوده و فقط شامل اعداد ۰ و ۱ می‌باشد. این نقاط که به ترتیب با نام‌های minX, minY, maxX, maxY نام گذاری شده‌اند برابرند با کوچک ترین نقطه در محور افقی، کوچکتری نقطه در محور عمودی، بزرگترین نقطه در محور افقی و در نهایت بزرگترین نقطه در محور عمودی که مقدار آن‌ها در ماتریس تصویر برچسب تولید شده برابر با ۱ می‌باشد.

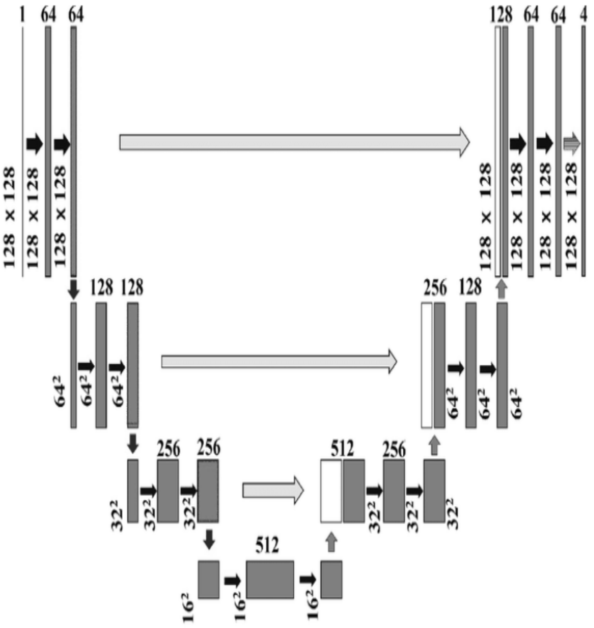
این ۴ نقطه برای هر تصویر به صورت جداگانه حساب می‌شود. این نقاط با استفاده از تصویر برچسب پیش بینی شده در مرحله‌ی اول بدست می‌آید.

به نوعی این نقاط تشکیل یک مستطیل حول مکانی که اسبک مغز تشخیص داده شده است می‌دهند. هنگام تولید این ماسک برای جلوگیری از، از دست دادن دیتا و از دست دادن بخشی از ناحیه اسبک مغز از هر چهار نقطه بدست آمده مقدار ۲۰ پیکسل به طرف اطراف تصویر، مستطیل گسترش داده می‌شود. یعنی تمام نقاط minX, minY, maxX, maxY در داخل ماسک نهایی مستطیل شکل قرار می‌گیرند.

درنهایت دو نقطه روی تصویر مشخص می‌شود که با استفاده از این دو نقطه یک مستطیل روی تصویر تشکیل داده می‌شود که هر یک از دو نقطه به ترتیب برابر با گوشه‌ی بالا سمت چپ مستطیل و گوشه‌ی پایین سمت راست مستطیل می‌باشد. بعد از تشکیل مستطیل مقدار تمامی پیکسل‌های موجود در ناحیه داخلی این مستطیل برابر با ۱ و مابقی پیکسل‌ها برابر با ۰ قرار می‌گیرد. در نهایت ماسک بدست آمده برای هر تصویر را در تصویر اصلی بصورت نقطه در نقطه ضرب کرده و تصاویر جدیدی بوجود می‌آید که فقط شامل بخشی از تصویر MRI می‌باشد که احتمال وجود اسبک مغزی در آن‌ناحیه بیشتر است و بخش‌های اضافه تصویر کاملا از تصویر MRI حذف می‌شود.

در قسمت آخر فرآیند آموزش، مدل U-Net دیگری همچون مرحله‌ی قبل ایجاد می‌شود و این بار با استفاده از تصاویر ماسک شده در فاز قبل این مدل آموزش داده می‌شود.

شکل (۳) ساختار مدل U-Net استفاده شده در این پژوهش را نمایش می‌دهد.



شکل (۳) : مدل U-NET

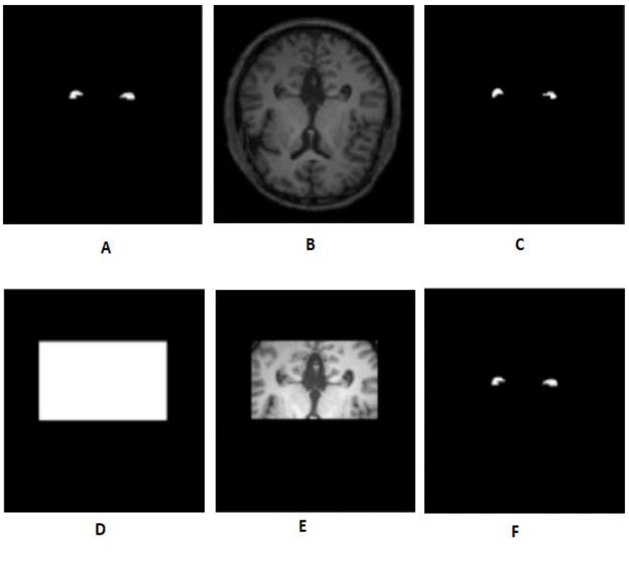
ورودی مدل U-Net که در این پژوهش از آن استفاده شده است تصاویر و برچسب‌های تصاویری است، که اندازه آن‌ها ۱۲۸ در ۱۲۸ می‌باشد. این مدل همانند شکل (۳) از دو بخش Encoder و Decoder تشکیل شده است و دارای یک مسیر گسترش‌سازی در سمت راست و یک مسیر فشرده‌سازی در سمت چپ می‌باشد. هر یک از لایه‌های پیچشی این مدل، یک تابع فعال‌سازی Relu و یک الگوریتم max-pooling دارد.

از این مدل U-Net در هر دو فاز آموزش استفاده شده است و به نوعی بعد از اتمام این بخش دو مدل آموزش دیده موجود است. یک مدل آموزش دیده شده برای مرحله‌ی اول که ورودی آن تصاویر MRI از پایگاه داده Harp است و یک مدل برای مرحله‌ی دوم که ورودی آن تصاویر MRI ماسک شده است که این تصاویر ماسک شده با استفاده از الگوریتم تولید ماسک و خروجی برچسب‌های پیش بینی شده بعد از فاز اول تولید شده است.

۳-۲- نتایج ارزیابی

برای پیاده سازی سیستم مورد نظر دیتاست به سه بخش داده‌های آموزشی، داده های اعتبار سنجی و داده‌های آزمایش به ترتیب با اندازه‌های ۷۰٪، ۱۵٪ و ۱۵٪ تقسیم بندی شده است. در این مرحله پس از جمع آوری پایگاه داده اقدام به پیش پردازش‌های اولیه روی داده‌ها برای نمایش و فراخوانی داده‌ها شده است. تصاویر پیش پردازش شده به ابعاد ۱۲۸*۱۲۸ تبدیل شده و سپس تصاویر به سه بخش داده‌های آموزشی، داده‌های تست و داده‌های ارزیابی تقسیم بندی شده است. پایگاه داده ایجاد شده در نهایت دارای مجموعه‌ای از برچسب‌ها و مجموعه‌ای از تصاویر MRI می‌باشد.

در بخش نهایی این پژوهش پس از تولید مدل آموزش دیده دو مرحله‌ای، با استفاده از داده‌های تست که از تصاویر موجود در پایگاه داده‌ی Harp بدست آمده است مدل نهایی مورد آزمایش و ارزیابی قرار می‌گیرد. برای ارزیابی مدل دو مرحله‌ای پیشنهادی از بررسی شباهت برچسب اصلی هر داده‌ی تست و برچسب پیش بینی شده توسط مدل دو مرحه‌ای استفاده می‌شود. در شکل (۴) نمونه‌ای از مراحل تست یک تصویر نمایش داده شده است.



شکل (۴): مراحل پیش بینی ناحیه اسبک در مغز

در شکل (۴) تصویر A ماسک اولیه و تصویر B همان تصویر MRI اولیه موجود در دیتاست Harp می‌باشد که بطور تصادفی انتخاب شده‌اند. تصویر C بیانگر برچسب پیش بینی شده توسط مدل U-NET در مرحله‌ی اول مدل می‌باشد. تصویر D ماسک ایجاد شده از تصویر خروجی C می‌باشد و با ضرب این ماسک در تصویر B تصویر E بدست می‌آید. همانطور که در

تصویر E مشخص است، این تصویر فقط حاوی قسمتی است که احتمال وجود اسبک مغز در آن ناحیه بسیار زیاد می‌باشد. در نهایت تصویر نهایی F خروجی مدل U-Net مرحله دوم است که از با استفاده از تصویر ماسک شده پیش بینی شده است.

در ادامه برچسب‌های پیش بینی شده داده‌های تست توسط مدل پیشنهادی را با برچسب اصلی داده‌های تست مقایسه کرده و میزان شباهت برای هر برچسب با استفاده از معیار Dice بررسی می‌شود. Dice معیاری است برای مقایسه دو تصویر و بررسی میزان شباهت دو تصویر [14]. در اکثر پژوهش‌هایی که برای قطعه بندی تصاویر انجام شده است از معیار Dice برای ارزیابی مدل استفاده می‌کنند. فرمول (۱) روش محاسبه Dice می‌باشد که در آن n به معنای تعداد کل داده‌ها، pi تصویر پیش‌بینی شده و gi بیانگر تصویر اصلی می‌باشد.

$$Dice = \frac{2\sum_{i}^{n} p_i g_i}{\sum_{i}^{n} p_i^2 + \sum_{i}^{n} g_i^2} \qquad (۱)$$

Dice به نوعی یک ابزار آماری است که شباهت بین دو مجموعه داده را اندازه گیری می‌کند [14]. در این پژوهش مجموعه داده در نظر گرفته شده تصاویر سیاه و سفید برچسب‌ها می‌باشد.

این مدل ارزیابی میزان شباهت بین دو تصویر را اندازه گیری می‌کند که همواره عددی بین ۰ تا ۱ می‌باشد و هر چه عدد به ۱ نزدیک‌تر باشد، به معنی آن است که شباهت دو تصویر به یک‌دیگر بیشتر می‌باشد. نکته قابل توجه در محاسبه Dice این است که هر دو تصویر باید از نظر ابعاد و اندازه مشابه یکدیگر باشند.

پس از محاسبه تک به تک Dice برای هر تصویر برچسب پیشبینی شده با برچسب اصلی موجود در پایگاه داده، میانگین کلی آن‌ها محاسبه شده و به عنوان معیار Dice مدل معرفی می‌شود که در این پژوهش مقدار آن برابر با ۰.۹۴ است.

در جدول (۱) نتایج بدست آمده توسط پژوهشگران و میزان Dice که به آن دست پیدا کرده‌اند آورده شده است و در ردیف آخر جدول نیز روش پیشنهادی ارائه شده توسط این پژوهش مشخص است.

جدول (۱) : مقایسه مقدار Dice پژوهش‌های مرتبط

روش	سال انتشار	میزان Dice
Thyreau و همکاران	۲۰۱۸	۰.۸۵
Wachinger و همکاران	۲۰۱۹	۰.۸۶
Dinsdale و همکاران	۲۰۱۹	۰.۸۶
Carmo و همکاران	۲۰۲۱	۰.۹۱
مدل پیشنهادی دومرحله‌ای	**۲۰۲۲**	**۰.۹۴**

همانطور که از جدول (۱) مشخص است، مدل دو مرحله ای که در این پژوهش ارائه شده است، میزان Dice بهتری روی تصاویر پایگاه داده Harp به نسبت مابقی پژوهش‌های بررسی شده کسب کرده است.

۳- نتیجه گیری

یکی از بخش‌های مهم موجود در مغز اسبک می‌باشد که در برخی بیماری‌های مرتبط با مغز این ناحیه دچار آسیب یا تغییر اندازه می‌شود. در این پژوهش سعی شد مدلی دو مرحله‌ای برای قطعه بندی اسبک مغز ارائه داده شود. در این پژوهش از مدل شبکه عصبی پیچشی U-Net در دو مرحله استفاده شد. در مرحله‌ی اول یک مدل U-Net با استفاده از داده‌های آموزش که از پایگاه داده Harp جمع آوری شده بودند آموزش داده شد و سپس در مرحله دوم با استفاده از این مدل، تصاویر ماسک شده جدیدی ایجاد شد که این تصاویر فقط روی ناحیه‌ای که احتمال وجود اسبک در آن پیش بینی شده است تمرکز دارند. در نهایت یک مدل U-Net دیگر با استفاده از این تصاویر ماسک شده آموزش داده شد.

برای تست این مدل نیز با استفاده از داده‌های تست در دو مرحله اقدام به شناسایی ناحیه اسبک مغزی شد. در قدم اول بـا استفاده از مدل از پیش آموزش دیده شده U-Net در مرحله‌ی اول، برچسب تصویر پیش بینی می‌شود و سپس در مرحله‌ی دوم یک ماسک مستطیل شکل حول نواحی که امکان وجود اسبک هست ایجاد می‌شود. سپس تصاویر تست را در این ماسک‌ها ضرب کرده و با استفاده از مدل از پیش آموزش دیده شده U-Net در مرحله دوم ناحیه اسبک مغزی در تصاویر ماسک شده شناسایی می‌شود.

در این پژوهش برای ارزیابی مدل از معیار شباهت Dice استفاده شده است. برای ارزیابی کلی این مدل مقدار Dice برای تمامی داده‌های پیش بینی شده محاسبه شد و سپس مقدار میانگین کلی این مقادیر محاسبه شد که مقدار میانگین Dice محاسبه شده کلی این مدل دو مرحله‌ای روی تصاویر پایگاه داده Harp برابر با ۰.۹۴ می‌باشد.

مراجع

[1] Carmo D, Silva B; Alzheimer's Disease Neuroimaging Initiative, Yasuda C, Rittner L, Lotufo R. Hippocampus segmentation on epilepsy and Alzheimer's disease studies with multiple convolutional neural networks. Heliyon. 2021 Feb 10;7(2):e06226. doi: 10.1016/j.heliyon.2021.e06226. PMID: 33659748; PMCID: PMC7892928.

[2] Inglese, M., Patel, N., Linton-Reid, K. et al. A predictive model using the mesoscopic architecture of the living brain to detect Alzheimer’s disease. Commun Med 2, 70 (2022). https://doi.org/10.1038/s43856-022-00133-4

[3] Liu, J., Pan, Y., Li, M., Chen, Z., Tang, L., Lu, C., & Wang, J. (2018). Applications of deep learning to MRI images: A survey. Big Data Mining and Analytics, 1(1), 1-18.

[4] Magnetic resonance imaging. (2012, February 10). In Wikipedia, The Free Encyclopedia. Retrieved 07:01, February 15, 2012, from http://en.wikipedia.org/w/index.php?title=Magnetic_resonance_imaging&oldid=476118369

[5] Chen, Y., Shi, B., Wang, Z., Zhang, P., Smith, C. D., & Liu, J. (2017, April). Hippocampus segmentation through multi-view ensemble ConvNets. In 2017 IEEE 14th international symposium on biomedical imaging (ISBI 2017) (pp. 192-196). IEEE.

[6] B. Thyreau, K. Sato, H. Fukuda, Y. Taki, Segmentation of the hippocampus by transferring algorithmic knowledge for large cohort processing, Med. Image Anal. 43 (2018) 214–228

[7] N. Nogovitsyn, R. Souza, M. Muller, A. Srajer, S. Hassel, S.R. Arnott, A.D. Davis, G.B. Hall, J.K. Harris, M. Zamyadi, et al., Testing a deep convolutional neural network for automated hippocampus segmentation in a

longitudinal sample of healthy participants, NeuroImage 197 (2019) 589–597.

[8] C. Wachinger, M. Reuter, T. Klein, DeepNAT:deep convolutional neural network for segmenting neuroanatomy, NeuroImage 170 (2018) 434–445.

[9] A.G. Roy, S. Conjeti, N. Navab, C. Wachinger, A.D.N. Initiative, et al., Quick-NAT:afully convolutional network for quick and accurate segmentation of neu-roanatomy, NeuroImage 186 (2019) 713–727.

[10] N.K. Dinsdale, M. Jenkinson, A.I. Namburete, Spatial warping network for 3Dsegmentation of the hippocampus in MRimages, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2019, pp.284–291.

[11] O. Lucena, R. Souza, L. Rittner, R. Frayne, R. Lotufo, Silver standard masks for data augmentation applied to deep-learning-based skull-stripping, in: Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on, IEEE, 2018, pp.1114–1117.

[12] Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.

[13] Frisoni GB, Jack CR. HarP: the EADC-ADNI Harmonized Protocol for manual hippocampal segmentation. A standard of reference from a global working group. Alzheimers Dement. 2015 Feb;11(2):107-10. doi: 10.1016/j.jalz.2014.05.1761. PMID: 25732924.

[14] C.H. Sudre, W. Li, T. Vercauteren, S. Ourselin, M.J. Cardoso, Generalised dice over-lap as a deep learning loss function for highly unbalanced segmentations, in: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Springer, 2017, pp.240–248.

پانویس ها

---

[1] Hippocampus
[2] Subcortical
[3] Magnetic Resonance Imaging (MRI)
[4] Pathophysiology
[5] Computer-aided diagnosis
[6] Deep Neural Network
[7] Neural Network (NN)
[8] Region of interest (ROI)
[9] Convolutional Neural Network (CNN)

چهارمین کنفرانس ملی انفورماتیک ایران
پژوهشگاه دانش‌های بنیادی، پردیس فرمانیه، تهران
۱۴ و ۱۵ دی ماه ۱۴۰۱

پتانسیل پژوهشی طرح‌های رمزنگاری متقارن در بستر رایانش ابری: عملیات مجموعه‌ای محرمانه و طرح‌های پیوند تطابق‌پذیر چندگانه

مجتبی رفیعی [۱,۲]

[۱] استادیار دانشکده ریاضی و آمار، دانشگاه اصفهان، اصفهان، ایران
m.rafiee@mcs.ui.ac.ir

[۲] پژوهشکدهٔ ریاضیات، پژوهشگاه دانش‌های بنیادی، تهران، ایران
mojtaba.rafiee@ipm.ir

چکیده

در سال‌های اخیر، اولیه‌های [۱] رمزنگاشتی متفاوتی برای طیف وسیعی از کاربردهای رایانش ابری معرفی شده‌اند که با توجه به بده‌بستان بین کارایی و امنیت قابل رده‌بندی هستند. در این مقاله، به معرفی اولیه‌های رمزنگاری متقارن جدیدی برای عملیات مجموعه‌ای محرمانه و در مرکزیت آن عملیات اشتراک محرمانه در بستر رایانش ابری می‌پردازیم. این اولیه‌ها از آنجاییکه می‌توانند در ساخت طرح‌های امن برای طیف وسیعی از عملکردها در بستر ابری مورد استفاده قرار گیرند، از اهمیت ویژه‌ای برخوردارند. کاربران با استفاده از طرح‌های معرفی شده می‌توانند مجموعه‌داده‌های خود را به صورت امن بر روی فضای ابری ذخیره نموده و به هنگام نیاز پرسمان‌های مجموعه‌ای خود را بر روی آنها اجرا و نتایج مطلوب را دریافت کنند. در این راستا، ابتدا مرور مختصری بر روی طرح‌های عملیات مجموعه‌ای محرمانه و طرح‌های پیوند تطابق‌پذیر چندگانه از حیث مفاهیم نحوی و امنیتی، کاربردها و تحلیل کارایی خواهیم داشت. در ادامه نیز با معرفی نمای کلی از پژوهش مذکور و پتانسیل‌های بالقوه پژوهشی سعی می‌گردد تا مخاطبان علاقمند در این حوزه، به نحو مناسبی جایگاه خود را برای پیشبرد پژوهش پیشرو یافته و در راستای ارتقای آن گام‌های مناسبی را بردارند.

کلمات کلیدی

رایانش ابری، برون‌سپاری داده، محرمانگی، پیوند تطابق‌پذیر، عملیات مجموعه‌ای محرمانه، اشتراک محرمانه.

۱- مقدمه

با پیشرفت فن‌آوری‌های نو ظهور، حجم اطلاعات سازمان‌ها به طور چشمگیری در حال افزایش است. این حجم وسیع و روز افزون، نگهداری و مدیریت داده‌ها را بیش از پیش پیچیده‌تر ساخته و هزینه‌های مربوطه را افزایش داده است. سازمان‌ها برای رویارویی با این چالش‌ها می‌توانند تمهیداتی چون افزودن منابع ذخیره‌سازی و به کارگیری افراد اجرایی بیشتر و یا واگذاری مدیریت داده‌های خود به یک سرویس‌دهنده خارجی را لحاظ کنند. راه‌حل اول تا حدودی مدیریت و کنترل داده‌ها را آسان می‌سازد اما نهایتا به سبب به‌کارگیری منابع ذخیره‌سازی و نیروی انسانی بیشتر، افزایش هزینه‌های سازمان را به دنبال دارد. از این‌رو، رویکرد برون‌سپاری داده و محاسبات به یک سرویس‌دهنده خارجی در جوامع تحقیقاتی، صنعتی و تجاری از جایگاه ویژه‌ای برخوردار است [۲، ۳].

با ظهور میزبانی محاسبات ابری و سرویس‌های ذخیره‌سازی، سرویس‌های برون‌سپاری داده در بستر رایانش ابری شکل تازه‌ای به خود گرفت. در این بستر، کاربران تنها به منابع مورد نیاز برای انجام کارشان دسترسی دارند و بنابراین نیازی به پرداخت هزینه برای منابع مصرف نشده شبیه به روش سنتی وجود ندارد. رایانش ابری علاوه بر اینکه باعث صرفه‌جویی در هزینه‌های سازمان می‌گردد به کاربران نیز امکان دسترسی به آخرین فن‌آوری‌های موجود به منظور نوآوری در کسب و کار را ارائه می‌دهد. امروزه سرویس‌دهندگانی نظیر گوگل، مایکروسافت، آمازون و آی.بی.ام با طیف وسیعی از خدمات نظیر ایجاد رایانامه، اشتراک اسناد، میزبانی فایل‌ها و بسیاری خدمات دیگر در بستر رایانش ابری پا به عرصه ظهور گذاشته و از محبوبیت زیادی برخوردار هستند [۳].

با وجود مزایای فوق، از آنجا که داده‌های سازمان در صورت استفاده از چنین سرویس‌هایی در قالب برون سپاری نگهداری می‌شود و تحت نظارت و کنترل مستقیم مالک داده نیست، نگرانی‌های امنیتی مطرح می‌شود. یکی از مهم‌ترین این نگرانی‌ها، حفظ محرمانگی داده برون‌سپاری شده است. رمزگذاری داده قبل از برون‌سپاری را می‌توان راهکاری برای مرتفع کردن این نگرانی دانست. با این وجود از آنجا که داده رمز شده در اختیار سرویس‌دهنده خارجی است، انجام جستجوهای درخواستی کاربر و استخراج داده مورد نظر به صورت کارا به عنوان یک چالش مطرح می‌شود. بسته به نوع عملکرد، امنیت و میزان کارایی مورد نیاز، ابزارهای رمزگذاری مختلفی برای این منظور معرفی و توسعه داده شده‌اند. در ادامه برخی از مهم‌ترین ابزارهای موجود آورده شده است.

رمزگذاری همریخت: ایده رمزگذاری همریخت (HE[۲]) آن است که اگر پیام‌های $c_۱$ و $c_۲$ که به ترتیب رمز شده پیام‌های $m_۱$ و $m_۲$ است به همراه یک تابع f داده شده باشد، آنگاه هر شخصی بدون نیاز به دانستن کلید خصوصی بتواند رمز شده $f(m_۱, m_۲)$ را محاسبه کند. در یک دسته بندی کلی و از حیث توابع قابل پشتیبانی، رمزگذاری همریخت را

[۱] Primitives

[۲] Homomorphic Encryption

می‌توان به دو دسته‌ی رمزگذاری تمام همریخت (FHE[3]) و رمزگذاری همریخت جزئی (PHE[4]) تقسیم‌بندی نمود. رمزگذاری تمام همریخت، از نظر عملکرد امکان اجرای هر تابع از پیش مشخص شده‌ای را بر روی داده رمز شده فراهم می‌کند. از نظر امنیت نیز دارای امنیت معنایی است و جز طول پیام های مورد محاسبه، نشتی دیگری ندارد. همچنین به سبب سربارهای رمزگذاری و تعریف قوی از امنیت، از کارایی پایینی برخوردار است. با اینحال رمزگذاری همریخت جزئی در مقایسه با رمزگذاری تمام همریخت از نظر عمکرد، امکان اجرای توابع محدودتری را بر روی داده رمز شده فراهم می‌کند. از نظر امنیت شبیه به رمزگذاری تمام همریخت است ولی از نظر کارایی به مراتب بهتر از رمزگذاری تمام همریخت عمل می‌کند.

رمزگذاری تابعی: ایده رمزگذاری تابعی (FE[5]) آن است که اگر پیام c که رمز شده پیام m است به همراه یک تابع f داده شده باشد، آنگاه هر فردی بتواند تابع $f(m)$ را بدون داشتن m محاسبه کند. رمزگذاری تابعی از نظر عملکرد می‌تواند امکان اجرای تابع خاص یا هر تابعی را فراهم کند. همچنین به سبب اینکه فرد در این نوع رمزگذاری، حاصل تابع و نه حاصل رمزشده تابع را به دست می‌آورد، دارای امنیت ضعیف‌تری نسبت به رمزگذاری همریخت است. از نظر کارایی نیز وابسته به تعاریف امنیتی و سربارهای رمزگذاری است، با اینحال دانش کنونی بیان کننده پایین بودن کارایی این نوع رمزگذاری می‌باشد.

محاسبات امن چندعاملی: ایده محاسبات امن چند عاملی (MPC[6])، محاسبه امن یک تابع است که ورودی‌های آن بین عامل‌های مختلفی توزیع شده است. این ابزار از نظر عملکرد، امکان اجرای هر تابعی را بین عامل‌های مختلف با ورودی‌های محرمانه فراهم می‌کند. از نظر امنیت، هر عامل جز ورودی خود، حاصل تابع و هر آنچه از آن دو استنتاج می‌شود، دانش دیگری کسب نمی‌کند. از نظر کارایی نیز پروتکل‌های ارایه شده برای پشتیبانی کلی از هر تابعی بسیار نا کار آمد بوده و خاص منظوره طراحی کردن پروتکل‌ها از نا کار آمدی آن می‌کاهد.

رمزگذاری جستجوپذیر: ایده رمزگذاری جستجوپذیر (SE[7]) در حوزه بازیابی امن اطلاعات مطرح می‌گردد، جاییکه کاربر نیاز به بازیابی اسنادی دارد که حاوی کلمه خاصی است و از نظر امنیتی نمی‌خواهد محتوای اسناد و همچنین کلمه مورد پرسمان برای دیگران فاش شود. این نوع رمزگذاری از نظر عملکرد، برای پشتیبانی از توابع خاصی طراحی می‌شوند و به سبب خاص منظوره بودن و سربارهای کم، از کارایی بالایی برخوردار هستند. همچنین با توجه به اهمیت بالای کارایی، معمولا طرح‌های ارایه شده از امنیت کمتری نسبت به ابزارهای دیگر برخوردار می‌باشند.

در این مقاله، سعی بر آن است تا به طور مختصر دو اولیه رمزنگاری جدید به نام‌های طرح عملیات مجموعه‌ای محرمانه (PSO[8]) و طرح پیوند تطابق‌پذیر چندگانه (M-Adjoin[9]) که می‌توانند هسته اصلی طیف وسیعی از پروتکل‌ها در بستر رایانش ابری باشد را معرفی کنیم. این اولیه‌های رمزنگاری از انجاییکه به منظور به کار گیری در سرویس‌های ابری طراحی و توسعه داده شده‌اند و در آنها تعادل بین امنیت و کارایی از اهمیت بالایی برخوردار است، از مفاهیم نحوی و امنیتی مطرح در رمزگذاری جستجوپذیر الگو گرفته‌اند.

۲- طرح عملیات مجموعه‌ای محرمانه در بستر رایانش ابری

فرض کنید $\mathcal{DS} = (l_i, S_i)_{i=1}^{n}$ بیانگر یک مجموعه داده شامل یک لیست از زوج برچسب/مجموعه‌هاست که در آن $l_i \in \{0,1\}^{\lambda}$ و $S_i = (w_1^i, \cdots, w_{k_i}^i)$ است. همچنین فرض کنید یک عملیات مجموعه‌ای به صورت یک سه تایی $q_{ij} = (op_{ij}, l_i, l_j)$ است که در آن l_i و l_j بیانگر دو برچسب ستون انتخابی و op_{ij} بیانگر یکی از عملیات مجموعه‌ای به قرار زیر است:

- $\cap$: بیانگر اشتراک دو مجموعه است،
- $\cup$: بیانگر اجتماع دو مجموعه است،
- $\setminus$: بیانگر تفاضل بین دو مجموعه است.

یک طرح عملیات مجموعه‌ای محرمانه (PSO)، یک طرح رمزگذاری متقارن است که کاربر را قادر می‌سازد تا مجموعه‌داده‌های $\mathcal{DS}$ خود را رمز و بر روی فضای ابری برون‌سپاری نماید و سپس به هنگام نیاز با تولید یک سه تایی op_{ij} کارگزار را قادر نماید تا نتیجه عملیات مجموعه‌ای را محاسبه و نتیجه را به او برگرداند.

مفاهیم نحوی و امنیتی این اولیه رمزنگاری برای اولین بار توسط رفیعی و خزایی در [۱] برای سناریوی ابری معرفی گردید. آنها برای مدل کردن نحو مربوط به طرح عملیات مجموعه‌ای محرمانه، تمامی عملیات مجموعه‌ای پایه مانند اشتراک، اجتماع و تفاضل را در نظر گرفته‌اند و همچنین برای مدل کردن امنیت نیز از مفاهیم امنیتی استاندارد مبتنی بر شبیه‌ساز استفاده کرده‌اند. به سبب اهمیت تعادل بین امنیت و کارایی در بستر رایانش ابری، نشتی معقول و کمینه‌ای را برای این اولیه در نظر گرفته‌اند. از آنجاییکه در ادبیات موضوع، ارایه یک ساخت برای مدل ارایه شده یک نیازمندی ضروری برای اعتباربخشی به مفاهیم نحوی و امنیتی است، آنها یک ساخت عام برای طرح عملیات مجموعه‌ای محرمانه بر اساس اولیه‌های چندتایی امن [10] و پیوند تطابق‌پذیر [11] ارایه کرده‌اند. آنها همچنین به منظور افزایش کارایی ساخت‌های قابل ارایه بر اساس مولفه‌های معرفی شده، اقدام به ارایه ساخت جدید برای طرح پیوند تطابق‌پذیر کردند که تلاش آنها منجر به کاهش چشمگیر سربارهای محاسباتی، ذخیره‌سازی و ارتباطی شده است. جدول ۱ و شکل ۱ به ترتیب

[3] Fully Homomorphic Encryption
[4] Partially Homomorphic Encryption
[5] Functional Encryption
[6] Multi-party Computation
[7] Searchable Encryption
[8] Private Set Operation Scheme
[9] Multi-Adjustable Join Scheme
[10] Tuple Set Scheme
[11] Adjustable Join Scheme

جدول ۱. ساخت‌های کاندید برای PSO

ساخت	طرح Tset	طرح Adjoin	فرض سختی
PSO I	Tset [2]	Our Adjoin	XDHV
PSO II (نامتقارن)		Adjoin [3]	matix-DDH [4] (نامتقارن)
PSO II (متقارن)		Adjoin [3]	matix-DDH [4] (متقارن)

ساخت‌های کاندید و سربارهای مربوط به ساخت‌های مختلف PSO را نشان می‌دهند.

آنها همچنین نشان دادند که طرح‌های PSO، علاوه بر اینکه می‌توانند به صورت مستقل و به تنهایی در سرویس‌های ابری مورد استفاده قرار گیرند، می‌توانند به عنوان یک بلاک اصلی در ساخت دیگر پروتکل‌های امنیتی در این بستر نیز ایفای نقش کنند. در این راستا، آنها دو عملکرد پر کاربرد در حوزه بازیابی امن اطلاعات در بستر رایانش ابری، یعنی: ۱) پرسمان‌های پیوند بر روی پایگاه‌داده رمزشده و ۲) پرسمان‌های بولی روی مجموعه اسناد رمزشده را به عنوان دو کاربرد قابل انجام از طریق طرح عملیات مجموعه‌ای محرمانه معرفی کردند و نحوه ساخت این طرح‌ها را نیز تشریح کردند.

۳- طرح پیوند تطابق‌پذیر چندگانه

با وجود مدل کردن مفاهیم نحوی و امنیتی طرح عملیات مجموعه‌ای محرمانه، همچنان این چالش باقی می‌ماند که این طرح همواره بر روی دو مجموعه قابل اعمال است و برای استفاده بر روی لیستی از مجموعه‌ها مجبور هستیم عملیات مجموعه‌ای را برای هر زوج مجموعه لحاظ و سپس نتیجه نهایی را بدست آوریم. در این حالت کارگزار به جای فهمیدن نتیجه نهایی عملیات مجموعه‌ای روی لیست درخواستی، نتایجی میانی را بدست می‌اورد که با نشتی کمینه مورد انتظار برای چنین طرح‌هایی فاصله دارد. شکل ۲، یک نمونه از نشتی مورد انتظار و نشتی‌های میانی را نشان می‌دهد.

برای مرتفع کردن این چالش، طرح‌پیوند تطابق‌پذیر چندگانه (M-Adjoin) توسط خزایی و رفیعی در [۵] معرفی شده است. با استفاده از این طرح کاربران می‌توانند مجموعه‌داده خود را به صورت امنی بر روی فضای ابری ذخیره نموده و سپس به هنگام نیاز بر روی هر لیست دلخواهی از مجموعه‌ها، اشتراک مجموعه‌ای محرمانه را محاسبه نمایند.

برای مدل کردن امنیت این اولیه، آنها ابتدا یک آزمایش امنیتی به نام M3Partition را معرفی کردند که نشتی کمینه چنین طرح‌هایی را پوشش می‌داد. در ادامه آنها به منظور افزایش کارایی با پارامتری کردن میزان نشتی، یک خانواده از مفاهیم امنیتی $\{\mathsf{M3P}_k\}$ را معرفی کردند که این خانواده در واقع فاصله بین مفهوم امنیتی ارایه شده برای پوشش اشتراک دو مجموعه یعنی $\mathsf{M3P}_۱$ و اشتراک کل لیست درخواستی یعنی مفهوم M3Partition را پر می‌کند. در این خانواده، مفهوم امنیتی $\mathsf{M3P}_۱$ دارای پایین‌ترین سطح امنیت و مفهوم امنیتی $\mathsf{M3Partition} = \mathsf{M3P}_\infty$ دارای بیشترین سطح امنیت است.

آنها به منظور اعتبار بخشی مفاهیم نحوی و امنیتی ارایه شده نیز اقدام به ارایه یک ساخت M-Adjoin با سطح امنیت M3Partition و پیچیدگی زمانی $\mathcal{O}(n^{m-۱})$ برای پیوند m ستون هر کدام از طول n نموده‌اند. همچنین ساخت دیگری که دارای امنیت $\mathsf{M3P}_k$ با پیچیدگی زمانی $\mathcal{O}\big((m-k)n^k/k\big)$ است را نیز ارایه کردند که برای کاربردهای عملی به خصوصی برای زمانیکه $k = ۱$ و نشتی بیشتری را مجاز به افشای آن هستیم، کارا و مناسب است. جدول ۲ و شکل ۳ به ترتیب مقایسه مجانبی و واقعی بین طرح‌های موجود را نشان می‌دهد.

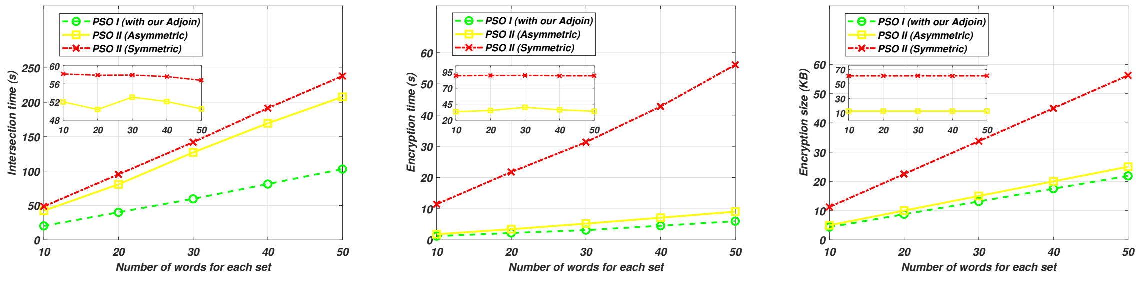
۴- روابط بین مفاهیم امنیتی طرح‌های پیوند تطابق‌پذیر چندگانه

از آنجاییکه طرح پیوند تطابق‌پذیر چندگانه به عنوان هسته اصلی طرح‌های عملیات مجموعه‌ای محرمانه است و با این هدف طراحی شده است که به عنوان یک بلاک اصلی در طراحی سایر پروتکل‌های بستر رایانش ابری به کار گرفته شود، رفیعی و خزایی در [۷] جایگاه امنیتی معرفی شده برای این اولیه را در میان سایر مفاهیم امنیتی استاندارد موجود در ادبیات موضوع مطالعه کرده‌اند. در این راستا، آنها مفاهیم استاندارد امنیتی مبتنی بر شبیه‌ساز و مبتنی بر تمایزناپذیری را برای گونه‌های تطبیقی و غیر تطبیقی به لحاظ دو تابع نشتی: MinLeak و $\mathsf{MonLeak}_k$ برای طرح‌های پیوند تطابق‌پذیر چندگانه معرفی کردند. آنها نشان داده‌اند که خانواده $\{\mathsf{M3P}_k\}_{k\in N}$ از مفاهیم امنیتی، قوی‌تر از گونه غیر تطبیقی و ضعیف‌تر از گونه تطبیقی مفاهیم امنیتی استاندارد است. علاوه بر موارد فوق، روابط میان توابع نشتی MinLeak و $\mathsf{MonLeak}_k$ را نیز مورد مطالعه و به صورت شفافی مشخص کرده‌اند. شکل ۴ تمامی روابط امنیتی مطرح برای طرح‌های پیوند تطابق‌پذیر را بر اساس تابع های نشتی MinLeak و $\mathsf{MonLeak}_k$ به صورت مجزا نشان می‌دهد.

۵- پتانسیل‌های تحقیقاتی موجود

در این بخش، سعی بر آن است تا نمای کلی از فضای پژوهشی مربوط به اولیه‌های رمزنگاری معرفی شده، ارایه شود. در ادامه و با یک دید سطح بالا، طرح‌های رمزنگاشتی ارایه شده را به همراه پتانسیل‌های پژوهشی بالقوه هر یک از آنها، مورد بحث و بررسی قرار می‌دهیم. انتظار می‌رود خواننده با مطالعه این بخش به نحو مناسبی جایگاه خود را برای پیشبرد پژوهش پیشرو یافته و در راستای ارتقای آن گام‌های مناسبی بردارد.

همانطورکه در بخش‌های قبلی اشاره شد، مرکزیت این مقاله بر روی معرفی اولیه‌های رمزنگاشتی متقارن برای پشتیبانی از عملگرهای مجموعه‌ای بر روی داده رمزشده در بستر رایانش ابری است. مطابق با پژوهش‌های انجام شده تا زمان نگارش این مقاله، یافته‌ها بر این مهم دلالت دارد که چنین عملکردی در بستر مذکور، مفهوم نو ظهوری است. با اینحال، عملکردهای یکسانی با رویکرد محاسبات امن چندعاملی وجود دارد که از منظر مفاهیم امنیتی و نحوی با رویکرد اتخاذی در این اولیه‌ها متفاوت است. از طرفی دیگر نیز عملکردهایی مشابه (و نه یکسان) با عملکرد مجموعه‌ای نیز برای کاربردهای خاص منظوره در فضای رایانش



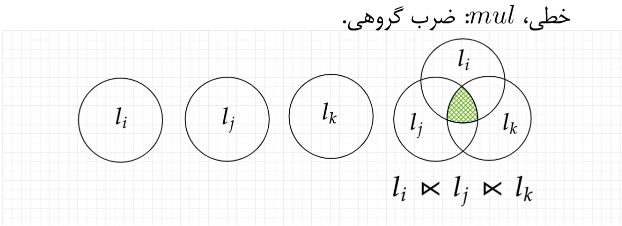
(آ) اندازه خروجی الگوریتم PSO.Enc — (ب) زمان اجرای الگوریتم PSO.Enc — (ج) زمان اجرای الگوریتم PSO.Oprate

شکل ۱. تحلیل کارایی ساخت‌های کاندید برای PSO

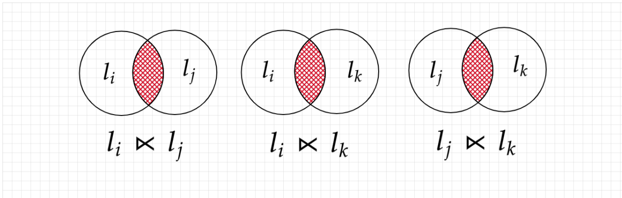
جدول ۲. مقایسه مجانبی ساخت‌های M-Adjoin

ساخت	فرض	Encod	Token	Adjust	Eval	اندازه (نشان، کلمه کد شده) تعداد عناصر گروه	امنیت
non-monotonous	MXDHV	$(tn)\ exp$	$(m)\ exp$	$(mn)\ bm$	$(n^{m-۱})\ mul$	(tn, m)	$M3P_{\infty}$
k-monotonous	MXDHV	$(tn)\ exp$	$(m)\ exp$	$\left(\left\lceil\frac{(m-۱)}{k}\right\rceil(k+۱)(n)\right)\ bm$	$\left(\left\lceil\frac{(m-۱)}{k}\right\rceil(n^k)\right)\ mul$	(tn, m)	$M3P_k$
M-Adjoin II [3]	matix-DDH [4]	$(۲tn)\ exp$	$(۲m)\ exp$	$(۲n(m-۱))\ bm$	$(۲n(m-۱))\ mul$	$(۲tn, ۲m)$	$M3P_۱$
M-Adjoin I [3]	Decision Linear [6]	$(۴tn)\ exp$	$(۴m)\ exp$	$(۴n(m-۱))\ bm$	$(۴n(m-۱))\ mul$	$(۴tn, ۴m)$	$M3P_۱$

t: تعداد ستون‌ها در پایگاه‌داده، n: ماکزیمم اندازه هر ستون در پایگاه‌داده، m: ماکزیمم طول هر پرسمان پیوند، exp: توان‌رسانی گروهی، bm: نگاشت دو خطی، mul: ضرب گروهی.



(آ) نشتی مورد انتظار



(ب) نشتی‌های میانی

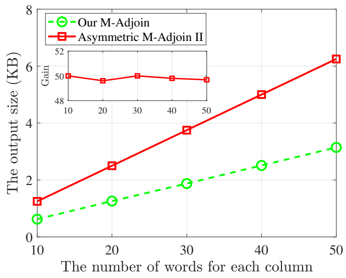
شکل ۲. نشتی مورد انتظار و نشتی‌های میانی برای سه مجموعه

ابری ارایه شده است که برای پوشش کامل انتظارات (کارایی و امنیت معقول)، همچنان فاصله زیادی وجود دارد. موارد مذکور سبب شده است تا انگیزه کافی برای ارایه چنین اولیه‌های رمزنگاشتی احساس و گام‌هایی در راستای شناسایی و مرتفع ساختن چالش‌های مذکور برداشته شود.

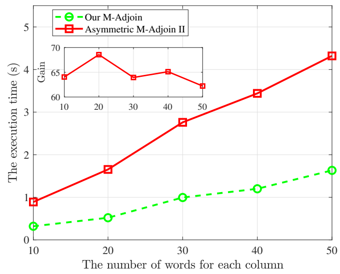
طرح‌های عملیات مجموعه‌ای محرمانه. برای عملکرد پرسمان مجموعه‌ای می‌توان مفاهیم امنیتی زیادی را تعریف کرد. با اینحال، تا کنون تنها مفهوم امنیتی مبتنی بر شبیه‌ساز در مقابل مهاجم غیر تطبیقی را که یکی از ساده‌ترین مفاهیم امنیتی در این حوزه است، معرفی شده است. لازم به ذکر است که ساخت ارایه شده برای طرح PSO در [۱] از مجموعه‌ای از طرح‌ها با سطوح امنیتی بالاتری از امنیت مبتنی بر شبیه‌ساز در مقابل مهاجم غیر تطبیقی تشکیل شده است. با این وجود برای کاستن از پیچیدگی‌ها، اتکا به مفاهیم استاندارد امنیتی موجود و حرکت در طول خط پژوهشی انتخابی، به مفهوم ساده امنیتی مذکور برای طرح PSO بسنده شده است. از این‌رو، یکی از پتانسیل‌های بالقوه برای پژوهش در این حوزه، معرفی مفاهیم امنیتی قوی‌تر از مفهوم ارایه شده است. همچنین از بعد دیگر، ارایه ساخت‌های با کارایی بالاتر می‌تواند از جنبه عملیاتی، شاخه‌ای دیگر از پتانسل پژوهشی را فرا روی پژوهشگران علاقمند در این حوزه قرار دهد.

طرح‌های پیوند تطابق‌پذیر چندگانه. برای طرح پیوند تطابق‌پذیر به عنوان قلب عملیات مجموعه‌ای (برای پرسمان‌های مجموعه‌ای چندتایی و نه لزوما دو مجموعه)، مفاهیم امنیتی مختلفی معرفی شده است. با اینحال، اگر چه مطالعات انجام شده از نظر مفاهیم امنیتی پوشش کاملی را در اختیار پژوهشگران قرار می‌دهد، اما ارایه ساخت‌هایی که برخی از مفاهیم امنیتی را ارضا نماید هنوز به عنوان یک دغدغه پژوهشی باقی مانده است. همچنین ارایه ساخت‌هایی با انعطاف‌پذیری بالاتر برای طرح یک پرسمان چندتایی با مشخص کردن ساختار نشت اطلاعات، رابطه باز باقی‌مانده بین مفاهیم امنیتی در این طرح و ارایه ساخت‌های کاراتر با مفاهیم امنیتی موجود را می‌توان برخی دیگر از پتانسیل‌های بالقوه این بخش به شمار آورد.

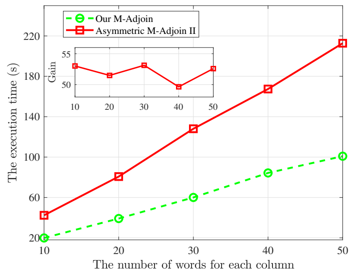
پتانسیل‌های جانبی. علاوه بر پتانسیل‌های فوق، برخی از پتانسیل‌های مستقل دیگری که برای هر اولیه رمزنگاشتی جدیدی نیز قابل طرح است،نظیر: رویکرد نگهداری داده برون‌سپاری شده، میزان اعتماد به کارگزار نگهدارنده داده و مدل اشتراکی برون‌سپاری داده را که موجب



(آ) اندازه خروجی الگوریتم Encod

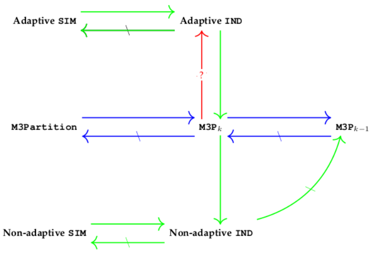


(ب) زمان اجرای الگوریتم Encod

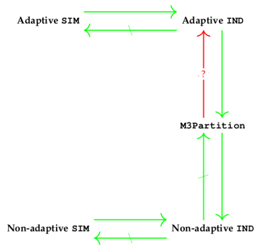


(ج) زمان پیوند برای بدست‌آوردن امنیت $M3P_1$

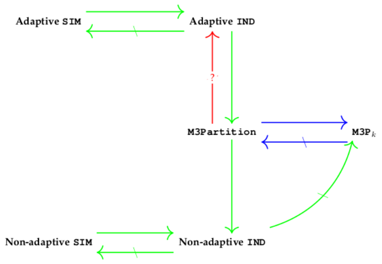
شکل ۳. تحلیل کارایی ساخت‌های M-Adjoin



(آ) روابط برای تابع نشتی $\mathsf{MonLeak}_k$



(ب) روابط با M3Partition



(ج) روابط برای تابع نشتی MinLeak

شکل ۴. روابط بین مفاهیم امنیتی طرح M-Adjoin

عملیاتی‌تر شدن هر اولیه رمزنگاشتی در بستر رایانش ابری می‌شود، می‌تواند برای مطالعه و پژوهش مورد توجه قرار گیرد.

– استراتژی نگهداری داده. داده برون‌سپاری شده بر روی کارگزار ابری را می‌توان از حیث میزان تغییرات، به دو رده کلی: ایستا و پویا تقسیم‌بندی کرد. در استراتژی ایستا پس از برون‌سپاری داده از سوی مالک داده، تغییری در محتوای داده برون‌سپاری شده ایجاد نخواهد شد و تمامی پرسمان‌های ارسالی کاربر تنها مجاز به خواندن اطلاعات درخواستی هستند. با اینحال در استراتژی پویا، علاوه بر خواندن اطلاعات درخواستی، امکان حذف، اضافه و ویرایش داده برون‌سپاری شده نیز برای کاربر وجود دارد.

طرح‌های معرفی شده در این مقاله، همگی بر پایه رویکرد استراتژی نگهداری داده ایستا مدل شده‌اند. با اینحال برای اکثر کاربردهای واقعی، امکان بروزرسانی داده‌های برون‌سپاری شده امری حیاتی تلقی می‌شود. استراتژی پویا به سبب امکان تغییر اطلاعات پس از گام برون‌سپاری، چالش‌های امنیتی به مراتب پیچیده‌تری را نسبت به استراتژی ایستا با خود به همراه دارد. از اینرو، بسط طرح‌ها و مفاهیم امنیتی ارایه شده در این مقاله برای پوشش چالش‌های مطرح در استراتژی نگهداری پویا از اهمیت ویژه‌ای برخوردار است.

– میزان اعتماد به کارگزار. کارگزار ابری را از نظر میزان اعتماد می‌توان در دو رده کلی: کارگزار صادق اما کنجکاو و کارگزار خرابکار تقسیم‌بندی کرد. کارگزار صادق اما کنجکاو از پروتکل مشخص شده تبعیت و از تغییر غیرمجاز داده‌ها اجتناب می‌کند. با این وجود، کارگزار صادق اما کنجکاو همواره سعی دارد تا دانش خود را نسبت به محتوای اصلی داده برون‌سپاری شده و همچنین پرسمان‌های درخواستی افزایش دهد. در حالت خرابکار، کارگزار علاوه بر تلاش به منظور کسب دانش پیرامون داده اصلی و پرسمان‌های درخواستی، سعی می‌کند رفتار خصمانه چون سرپیچی از اجرای درست پروتکل و تعرض (حذف یا تغییر) داده‌های برون‌سپاری شده را در دستور کار خود قرار دهد.

در تمامی طرح‌های معرفی شده در این مقاله، میزان اعتماد برای کارگزار از نوع صادق اما کنجکاو در نظر گرفته شده است. با این حال همواره مالک داده بیم آن را دارد که کارگزار نتایج درست و اصیلی را به او باز نگرداند. از اینرو حرکت به سوی افزایش اعتماد کاربران ابری همواره از اهمیت ویژه‌ای برخوردار بوده و طرح‌های با این میزان از اعتماد خواستگاه بیشتری دارند. لازم به ذکر است که برای برآورده کردن چنین سطح امنیتی نیازمند مدل کردن مفاهیم امنیتی به مراتب پیچیده‌تری نسبت به کارگزار صادق اما کنجکاو هستیم.

– مدل برون‌سپاری داده. مدل برون‌سپاری را از منظر اشتراک داده می‌توان در دو دسته: تک کاربری و چندکاربری رده‌بندی نمود. در حالت تک کاربری در واقع اشتراک داده‌ای وجود ندارد و مالک داده پس از برون‌سپاری تنها خود پرسشگر پرسمان‌های درخواستی از کارگزار است. با اینحال در حالت چندکاربری، مالک داده بعد

از برون‌سپاری می‌تواند داده خود را بین مجموعه‌ای از کاربران به اشتراک بگذارد. در ادامه کاربران مجاز می‌توانند پرسمان‌های درخواستی خود را به کارگزار ارسال و پاسخ مناسب را دریافت نمایند.

مدل برون‌سپاری لحاظ شده برای طرح‌های این مقاله از نوع تک کاربری بوده و در واقع فاقد ویژگی اشتراک داده هستند. با این وجود برای اکثر کاربردهای واقعی در بستر رایانش ابری همواره مجموعه‌ای از کاربران و نه یک کاربر وجود دارد و لذا در نظر گرفتن مباحث مربوط به شتراک داده از اهمیت بالایی برخوردار است. بنابراین، بسط طرح‌ها و مفاهیم امنیتی معرفی شده در این مقاله به نحوی که ویژگی اشتراک داده بین کاربران را نیز مهیا سازد، می‌تواند طرح‌های ارایه شده را کاربردی‌تر و چالش‌های واقعی بیشتری را پوشش دهد.

مراجع

[1] M. Rafiee and S. Khazaei, “Private set operations over encrypted cloud dataset and applications,” *The Computer Journal*, pp.1145-1162, 2020.

[2] D. Cash, S. Jarecki, C. S. Jutla, H. Krawczyk, M. Rosu, and M. Steiner, “Highly-scalable searchable symmetric encryption with support for boolean queries,” in *Advances in Cryptology - CRYPTO 2013 - 33rd Annual Cryptology Conference, Santa Barbara, CA, USA, August 18-22, 2013. Proceedings, Part I*, pp.353–373, 2013.

[3] I. Mironov, G. Segev, and I. Shahaf, “Strengthening the security of encrypted databases: Non-transitive joins,” in *Theory of Cryptography - 15th International Conference, TCC 2017, Baltimore, MD, USA, November 12-15, 2017, Proceedings, Part II*, pp.631–661, 2017.

[4] A. Escala, G. Herold, E. Kiltz, C. Ràfols, and J. L. Villar, “An algebraic framework for diffie-hellman assumptions,” in *Advances in Cryptology - CRYPTO 2013 - 33rd Annual Cryptology Conference, Santa Barbara, CA, USA, August 18-22, 2013. Proceedings, Part II*, pp.129–147, 2013.

[5] S. Khazaei and M. Rafiee, “Towards more secure constructions of adjustable join schemes,” *IEEE Transactions on Dependable and Secure Computing*, vol. 19, no. 2, pp. 1078-1089, 2022.

[6] D. Boneh, X. Boyen, and H. Shacham, “Short group signatures,” in *Advances in Cryptology - CRYPTO 2004, 24th Annual International CryptologyConference, Santa Barbara, California, USA, August 15-19, 2004, Proceedings*, pp.41–55, 2004.

[7] M. Rafiee and S. Khazaei, “Security of multi-adjustable join schemes: Separations and implications,” *IEEE Transactions on Dependable and Secure Computing*, vol.19, no.4, pp.2535-2545, 2022.

[8] E.-J. Goh *et al.*, “Secure indexes.,” *IACR Cryptology ePrint Archive*, vol.2003, p.216, 2003.

[9] Y.-C. Chang and M. Mitzenmacher, “Privacy preserving keyword searches on remote encrypted data,” in *International Conference on Applied Cryptography and Network Security*, pp.442–455, Springer, 2005.

[10] M. Chase and S. Kamara, “Structured encryption and controlled disclosure,” in *International Conference on the Theory and Application of Cryptology and Information Security*, pp.577–594, Springer, 2010.

[11] S. Kamara and T. Moataz, “Boolean searchable symmetric encryption with worst-case sub-linear complexity,” in *Advances in Cryptology - EUROCRYPT 2017 - 36th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Paris, France, April 30 - May 4, 2017, Proceedings, Part III*, pp.94–124, 2017.

[12] S.-F. Sun, J. K. Liu, A. Sakzad, R. Steinfeld, and T. H. Yuen, “An efficient non-interactive multi-client searchable encryption with support for boolean queries,” in *European symposium on research in computer security*, pp.154–172, Springer, 2016.

[13] J. Alderman, K. M. Martin, and S. L. Renwick, “Multi-level access in searchable symmetric encryption,” in *International conference on financial cryptography and data security*, pp.35–52, Springer, 2017.

چهارمین کنفرانس ملی انفورماتیک ایران
پژوهشگاه دانش‌های بنیادی، پردیس فرمانیه، تهران
۱۴ و ۱۵ دی ماه ۱۴۰۱

ارائه روشی برای ارزیابی کیفیت تصاویر محتوای صفحه با تأکید بر ویژگی‌های توصیف‌گر لبه، وزن‌دهی شده توسط نواحی چشمگیر تصویر

ریحانه مسافر[۱]، محمد حسین خسروی[۲]، حامد وحدت‌نژاد[۳]

[۱] دانشجوی کارشناسی ارشد مهندسی فناوری اطلاعات، دانشکده مهندسی برق و کامپیوتر، دانشگاه بیرجند، بیرجند،
reyhane.mosafer@gmail.com

[۲] استادیار گروه مهندسی کامپیوتر و فناوری اطلاعات ، دانشکده مهندسی برق و کامپیوتر، دانشگاه بیرجند، بیرجند،
mohokhosravi@birjand.ac.ir

[۳] دانشیار گروه مهندسی کامپیوتر و فناوری اطلاعات ، دانشکده مهندسی برق و کامپیوتر، دانشگاه بیرجند، بیرجند،
vahdatnejad@birjand.ac.ir

چکیده

با رشد روزافزون ارتباطات چند دستگاهی و رایانش ابری، ارزیابی کیفیت تصاویر محتوای صفحه، اهمیت دوچندان یافته است. در این مقاله، یک معیار جدید برای ارزیابی کیفیت این تصاویر با تأکید بر ارزیابی میزان وفاداری ویژگی‌های مبتنی بر لبه در دو تصویر مرجع و آزمون پیشنهاد شده است. روش پیشنهادی از توصیف‌گر باینری امضای تصویر بعنوان تابع وزن‌دهی برای بهبود نتایج استفاده می‌کند. این توصیفگر روشی جهت استخراج نواحی چشمگیر تصویر است که به طور قابل توجهی داده‌های مرجع را کاهش می‌دهد. معیار پیشنهادی بر روی پایگاه‌های داده SIQAD و SCID اعمال شده، نتایج ارزیابی نشان می‌دهند که این مدل در مقایسه با سایر معیارهای کارآمد ارزیابی کیفیت تصاویر محتوای صفحه، بهتر عمل کرده و سازگاری بالاتری با پیش‌بینی‌های انجام شده توسط سیستم بینایی انسان دارد.

کلمات کلیدی

تصاویر محتوای صفحه، ارزیابی کیفیت تصویر، تشخیص نواحی چشمگیر، امضای تصویر

۱- مقدمه

یکی از مهمترین موضوعات پژوهشی در حوزه پردازش تصویر و بینایی ماشین ارزیابی کیفیت تصویر (IQA[۱]) است. در این حوزه تلاش می‌شود برای یک تصویر ورودی به ماشین، بصورت خودکار امتیازی تولید شود که نشان‌دهندۀ کیفیت آن تصویر بوده حتی‌الامکان به نظر کاربر انسانی نزدیک باشد. الگوریتم‌های ارزیابی کیفیت تصویر، در سامانه‌هایی که تصاویری را از کاربر دریافت می‌کنند، می‌توانند بطور خودکار کیفیت تصاویر دریافتی را تشخیص داده، از امتیاز داده شده به تصویر برای پذیرش یا رد آن تصمیم‌گیری کنند. بعلاوه از این روش‌ها می‌توان برای ارزیابی میزان موفقیت الگوریتم‌های بهسازی تصویر یا تنظیم خودکار پارامترهای این الگوریتم‌ها نیز استفاده کرد. از دیگر کاربردهای این روش‌ها می‌توان به انتخاب خودکار و پویای الگوریتم‌های پردازشی بر اساس نحوه عملکرد آنها اشاره کرد.

روش‌های ارزیابی کیفیت تصویر را می‌توان به سه دسته روش‌های با مرجع کامل (FR[۲])، روش‌های بدون مرجع (NR[۳]) و روش‌های با مرجع کاهش‌یافته (RR[۴]) دسته‌بندی کرد. در روش‌های با مرجع کامل، اصل تصویر و تصویر مورد ارزیابی در دسترس هستند در حالیکه در روش‌های بدون مرجع فقط تصویر تحت آزمایش در اختیار است و اصل تصویر در دسترس نیست. در روش‌های با مرجع کاهش یافته نیز اصل تصویر در اختیار نیست اما ویژگی‌های توصیف کننده آن در دسترس است و می‌توان از آنها برای سنجش میزان وفاداری تصویر مورد آزمایش به تصویر اصلی استفاده کرد.

در ارزیابی کیفیت تصاویر صحنه طبیعی (NSI[۵])، ساده‌ترین و معمول‌ترین معیارها، PSNR[۶] و MSE[۷] هستند. این دو معیار اغلب در مقایسه با پیش‌بینی‌های انجام شده توسط سیستم بینایی انسان (HVS[۸]) عملکرد خوبی ندارند. برای رفع این مشکل، Wang و همکاران، معیار SSIM[۹] [1] را پیشنهاد کردند. از آن زمان، پژوهش‌های بسیاری بر پایه SSIM انجام شده است که در مراجع [2] و [3] به برخی از آنها اشاره شده است. اغلب این روش‌ها نیز برای ارزیابی کیفیت تصاویر صحنه طبیعی ارائه شده و کاربرد دارند.

پیشرفت‌های اخیر در زمینه رایانش ابری، چالش‌های جدیدی را برای ارزیابی کیفیت تصویر به وجود آورده است. اپلیکیشن‌های ابری موبایل [4] و پلتفرم‌های محاسبات از راه دور [5] ، باعث تولید محتوای عظیمی به نام تصاویر

---

[۱] Image Quality Assessment
[۲] Full Reference
[۳] No Reference
[۴] Reduced Reference
[۵] Natural Scene Images
[۶] Peak Signal to Noise Ration
[۷] Mean Square Error
[۸] Human Visual System
[۹] Structural SImilarity Measure

محتوای صفحه (SCI[۱۰]) شده‌اند. تصویر محتوای صفحه یک تصویر ترکیبی شامل قسمت‌های متنی و تصویری است. این تصاویر از طریق فایل‌های pdf، اسلایدهای PowerPoint، صفحات وب، مجلات دیجیتال و غیره تهیه می‌شوند. SCIها در سمت سرور تولید و فشرده شده و به صفحه وب متقاضی آن تصویر ارسال می‌شوند. این روند، می‌تواند باعث تخریب کیفیت این تصاویر شده و کاهش عملکرد تعاملی کاربران را در پی داشته باشد. بنابراین، پیش‌بینی دقیق کیفیت این تصاویر، نقش مهمی در برنامه‌های رایانش ابری و محاسبات از راه دور خواهد داشت.

تفاوت‌های بسیاری بین تصاویر محتوای صفحه و تصاویر صحنه طبیعی وجود دارد. تصاویر صحنه طبیعی به دلیل محدودیت‌های سنسورهای تصویر دارای نویز هستند در حالی که که تصاویر محتوای صفحه کاملا بدون نویز بوده و به طور مستقیم توسط ابزار ضبط صفحه ایجاد می‌شوند. تصاویر طبیعی گرفته شده توسط دوربین، دارای خطوط ضخیم، تنوع رنگی غنی و محتوای بافت پیچیده هستند در حالیکه تصاویر محتوای صفحه از خطوط نازک، رنگ‌های محدود و اشکال بدون عارضه تشکیل شده‌اند [6]. این تفاوت‌ها باعث ایجاد یک چالش در پژوهش‌های ارزیابی کیفیت تصویر می‌شوند.

در سال‌های اخیر، روش‌های بسیاری جهت ارزیابی کیفیت تصاویر محتوای صفحه ارائه شده است. Gu و همکاران، معیار کیفیت ناشی از ساختار (SIQM[۱۱]) را برای ارزیابی کیفیت تصاویر محتوای صفحه پیشنهاد کردند [7]. در یک مطالعه دیگر، یک مدل ارزیابی به نام مدل مبتنی بر ویژگی Gabor ارائه شده است [8] . در این پژوهش، از استراتژی جمع مبتنی بر ویژگی Gabor[۱۲]، برای ادغام نگاشت‌های تشابه تولیدشده، استفاده شده است. در [9] ، Ni و همکاران مدل ESIM را معرفی کردند که در این روش از ویژگی‌های لبه جهت ارزیابی کیفیت تصاویر محتوای صفحه استفاده شده است. یکسری از روشها برای بهبود نتایج ارزیابی کیفیت تصاویر محتوای صفحه، از مدل‌های تشخیص نواحی چشمگیر استفاده کرده‌اند. در [10]، Gu و همکاران، معیار SQMS[۱۳] را برای ارزیابی کیفیت این تصاویر پیشنهاد کردند. همچنین در یک پژوهش دیگر، روش ارزیابی کیفیت کاهش یافته ناشی از نواحی چشمگیر جهت ارزیابی کیفیت تصاویر محتوای صفحه ارائه شده است.[11] .

مدل طراحی شده برای ارزیابی کیفیت تصاویر محتوای صفحه، باید با سیستم بینایی انسان، سازگاری بالایی داشته باشد. علاوه بر این، با توجه به اینکه تصاویر محتوای صفحه، دارای لبه‌ها و خطوط تیز هستند که سیستم بینایی انسان نسبت به آن بسیار حساس است، روش ارائه شده باید بر روی استخراج ویژگی‌های لبه تمرکز کند. به همین دلیل، در این مقاله، از تعدادی از ویژگی‌های شاخص لبه مانند کنتراست، عرض و جهت لبه به منظور ارزیابی کیفیت تصاویر محتوای صفحه، استفاده شده است. ترکیب این ویژگی‌های مرتبط با لبه‌های تصویر در مقالات پیشین بکار گرفته نشده است. همچنین معیار ارائه شده در این مقاله، مبتنی بر تشخیص نواحی چشمگیر[۱۴] است. عملیات تشخیص نواحی چشمگیر نیز باعث کاهش ابعاد داده می‌شود. بنابراین در افزایش سرعت و کارآمدی ارزیابی کیفیت تصاویر، تأثیر بسزایی دارد. همچنین در پژوهش‌های قبل، saliency به عنوان یک ویژگی کیفی در نظر گرفته شده است. بیشتر مدل‌های تشخیص نواحی چشمگیر از پایین به بالا بوده و ویژگی‌های سطح پایین را در نظر می‌گیرند. بنابراین به تخریب کیفیت حساس هستند. تخریب کیفیت محسوس، سبب تغییر در ویژگی‌های سطح پایین و در نتیجه تغییر مناطق چشمگیر می‌شود. در این مقاله نیز، برای پیش‌بینی کیفیت، تغییرات مناطق چشمگیر، کمّی شده و از یک توصیف‌کننده تصویر باینری به نام امضای تصویر استفاده شده است.

ادامه این مقاله به شرح زیر است. بخش ۲ معیار پیشنهادی را ارائه می‌دهد. نتایج تجربی در بخش ۳ آورده شده است. در این بخش، معیار پیشنهادی با سایر معیارهای ارزیابی کیفیت تصویر مقایسه شده است. سرانجام، در بخش ۴ به نتیجه‌گیری و جمع‌بندی مقاله پرداخته خواهد شد.

۲- روش پیشنهادی

در این قسمت، مدل پیشنهادی برای ارزیابی کیفیت تصاویر محتوای صفحه ارائه شده است. این مدل شامل دو مرحله است: ۱- تشخیص نواحی چشمگیر ۲- ارزیابی توصیف‌گرهای لبه. در نهایت میزان وفاداری توصیف‌گرهای لبه بصورت وزندهی شده بر اساس نواحی چشمگیر تصویر بعنوان معیار ارزیابی کیفیت تصویر در نظر گرفته شده‌اند. در ادامه، این مراحل به طور کامل شرح داده شده است.

۲-۱- تشخیص نواحی چشمگیر

همانطور که در بخش قبل گفته شد، روش‌های بسیاری برای تشخیص نواحی چشمگیر وجود دارد. در این مقاله، از یک توصیف‌کننده تصویر باینری به نام امضای تصویر[۱۵]، جهت تشخیص نواحی چشمگیر استفاده شده است. این روش توسط Min و همکاران در سال ۲۰۱۷ مطرح و در کاربردهایی مانند کاهش ابعاد داده و کمّی کردن تغییرات saliency بکار گرفته شده است. فرض کنید x تصویر ورودی جهت تشخیص نواحی چشمگیر باشد.

$$x'(i,j) = a * x(ri, rj) \qquad (۱)$$

در رابطه بالا x' تصویر نمونه‌برداری شده، r نرخ نمونه‌برداری با مقدار $r = 7$، a یک فیلتر پایین‌گذر و $\otimes$ نشان‌دهنده عملیات کانولوشن است. i و j نیز به ترتیب اندیس سطرها و ستون‌های تصویر x هستند.

در مرحله بعد، تبدیل کسینوس گسسته دوبعدی (DCT[۱۶]) بر روی تصویر نمونه‌برداری‌شده انجام شده و علامت مؤلفه‌های DCT استخراج شده‌اند.

$$\hat{x} = \mathrm{sgn}(DCT(x')) \qquad (۲)$$

که $\hat{x}$ امضای تصویر x' و sgn تابع علامت با تعریف زیر است:

$$\mathrm{sgn}(x) = \begin{cases} -1 & if x < 0 \\ 1 & if x \geq 0 \end{cases} \qquad (۳)$$

سپس تصویر $\hat{x}$ به حوزه مکان تبدیل شده و تبدیل کسینوس گسسته معکوس (IDCT[۱۷]) بر روی آن اعمال می‌شود:

۱۰ Screen Content Images
۱۱ Structure-Induced Quality Metric
۱۲ Gabor-Feature Pooling Strategy
۱۳ Saliency-guided Quality Measure of SCIs
۱۴ Saliency Detection
۱۵ image signature
۱۶ Discrete Cosine Transform
۱۷ Inverse Discrete Cosine Transform

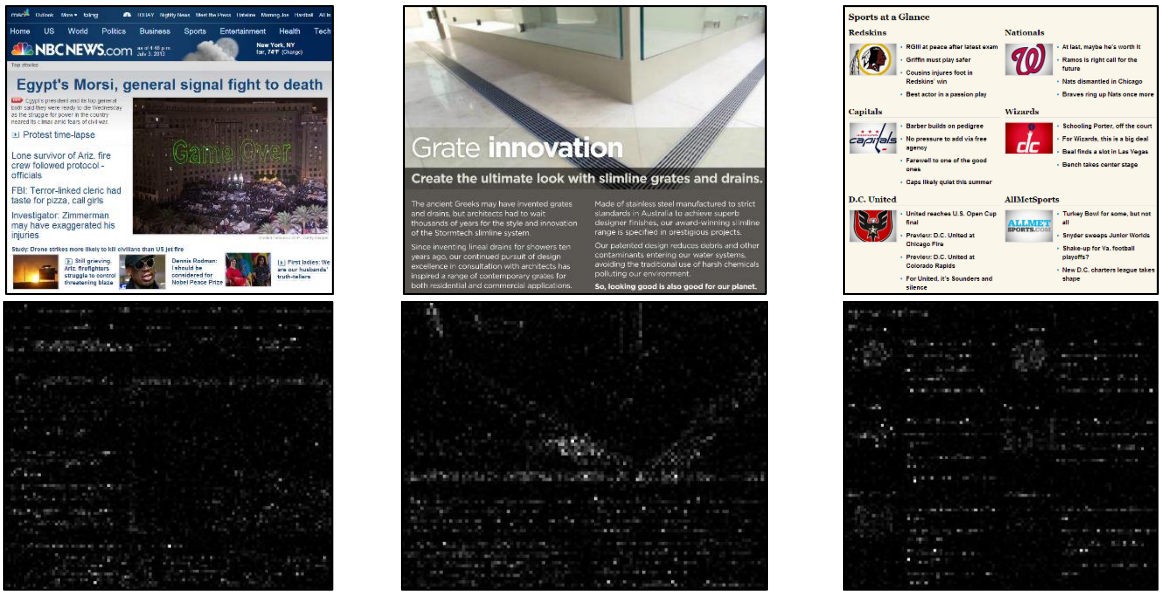
چهارمین کنفرانس ملی انفورماتیک ایران
پژوهشگاه دانش‌های بنیادی، پردیس فرمانیه، تهران
۱۴ و ۱۵ دی ماه ۱۴۰۱

$$\bar{x} = IDCT(\hat{x}) \qquad (۴)$$

در نهایت، نگاشت چشمگیری برای تصویر x از طریق رابطه (۵) محاسبه می‌شود:

$$m = \bar{x} \circ \bar{x} \qquad (۵)$$

در این معادله ∘ نشان‌دهنده ضرب هادامارد یا درایه‌ای است.



شکل (۱) : توصیف‌گر امضای تصویر در تصاویر محتوای صفحه. ردیف اول: SCIهای اصلی. ردیف دوم:نگاشت‌های نواحی چشمگیر متناظر

شکل شماره ۱ نتیجه استخراج توصیف‌گر امضای تصویر در چند تصویر نمونه را نشان می‌دهد. همانگونه که ملاحظه می‌شود نواحی چشمگیر هر تصویر با دقت مناسبی استخراج شده‌اند.

۲-۲- توصیف‌گرهای پیشنهادی لبه

در این مرحله، ویژگی‌هایی که در ارزیابی کیفیت تصاویر محتوای صفحه بیشترین تأثیر را دارند شناسایی و در دو تصویر مرجع و تخریب‌شده مورد مقایسه قرار می‌گیرند. همانطور که پیشتر نیز گفته شد ویژگی‌های مبتنی بر لبه از مهمترین ویژگی‌های توصیف‌کننده در تصاویر محتوای صفحه هستند. در روش پیشنهادی ابتدا اندازه گرادیان (GM[18]) بین تصاویر مرجع و تخریب‌شده، محاسبه شده است. سپس نگاشت‌های دیگری از ویژگی‌های لبه مانند کنتراست لبه، عرض لبه و جهت لبه، بین تصاویر مرجع و تخریب‌شده استخراج و مورد ارزیابی و مقایسه قرار گرفته‌اند.

۲-۲-۱- محاسبه اندازه گرادیان

اندازه گرادیان جهت اندازه‌گیری تخریب ساختار، به صورت زیر محاسبه می‌شود:

$$G = \sqrt{g_x^2 + g_y^2} \qquad (۶)$$

برای محاسبه اندازه گرادیان نیز از عملگر Scharr استفاده شده است. به این صورت که از دو کرنل 3×3 استفاده شده که با تصویر اصلی کانوالو می‌شوند و تقریب مشتقات (در جهت افقی و عمودی) محاسبه می‌شود:

$$g_x = h_x \otimes A = \frac{1}{16}\begin{bmatrix} +3 & 0 & -3 \\ +10 & 0 & -10 \\ +3 & 0 & -3 \end{bmatrix} \otimes A$$

$$g_y = h_y \otimes A = \frac{1}{16}\begin{bmatrix} +3 & +10 & +3 \\ 0 & 0 & 0 \\ -3 & -10 & -3 \end{bmatrix} \otimes A \qquad (۷)$$

سپس برای اندازه‌گیری تفاوت در نگاشت‌های اندازه گرادیان بین تصویر مرجع r و تصویر تخریب‌شده d، از معیار تشابه مشترک[19] استفاده می‌شود.

$$GM = \frac{2G_r \cdot G_d + c}{G_r^2 + G_d^2 + c} \qquad (۸)$$

که G_r و G_d، به ترتیب، نگاشت‌های اندازه گرادیان برای تصاویر مرجع و تخریب‌شده و c یک ثابت مثبت است.

۲-۲-۲- محاسبه نگاشت‌های Three Edge-Feature

همانطور که قبلا هم گفته شد، تصاویر محتوای صفحه دارای لبه‌ها و خطوط تیز هستند. در این مقاله نیز، برخی از ویژگی‌های لبه مانند کنتراست، عرض و جهت لبه به منظور ارزیابی کیفیت تصاویر، استخراج شده‌اند.

نگاشت کنتراست لبه (ECM[20]) و نگاشت عرض لبه (EWM[21])

از نظر ریاضی، مدلسازی یک step edge ایده‌آل، در مرکز $x = x_0$، را می‌توان به صورت زیر بیان کرد.

$$u(x; b, c, x_0) = c \cdot U(x - x_0) + b \qquad (۹)$$

[18] Gradient Magnitude
[19] common similarity metric
[20] Edge Contrast Map
[21] Edge Width Map

که در آن U تابع unit step، c کنتراست لبه و b شدت روشنایی محلی در $x = x_0$ را نشان می‌دهد.

یک مدل لبه در تصویر محتوای صفحه ایجاد شده مانند $s(x;b,c,w,x_0)$ می‌تواند با کانوالو کردن مدل لبه ایده‌آل با کرنل گاسین $g(x;w)$ به دست آید:

$$\begin{aligned} s(x;b,c,w,x_0) &= u(x;b,c,x_0) \otimes g(x;w) \\ &= u(x;b,c,x_0) \otimes \frac{1}{\sqrt{2\pi w^2}} \exp\left(\frac{-x^2}{2w^2}\right) \\ &= b + \frac{c}{2}\left[1 + erf\left(\frac{x-x_0}{\sqrt{2}w}\right)\right] \end{aligned} \quad (۱۰)$$

که در آن $\otimes$ نشان‌دهنده عملیات کانولوشن، w انحراف استاندارد تابع هموارسازی کرنل گاسین است که برای مدلسازی عرض لبه استفاده می‌شود و erf تابع خطا است.

هر لبه از طریق یک مجموعه سه پارامتری نشان داده می‌شود: کنتراست c، عرض w و شدت روشنایی b . مسئله اصلی، نحوه تعیین این سه پارامتر برای هر مدل لبه است. این پارامترها، با استفاده از اطلاعات پیکسل محلی مرکز لبه به دست می‌آیند. برای این کار، سیگنال لبه $s(x;b,c,w,x_0)$ با مشتق کرنل گاسین $g'(x;w)$ کانوالو می‌شود.

$$d(x;c,w,\sigma_d,x_0) = \frac{c}{\sqrt{2\pi(w^2+\sigma_d^2)}} \exp\left[\frac{-(x-x_0)^2}{2(w^2+\sigma_d^2)}\right] \quad (۱۱)$$

از آنجا که سه پارامتر c، w و b برای حل مسئله لازم است، پس باید سه معادله وجود داشته باشد. برای این منظور $d(x;c,w,\sigma_d,x_0)$ در سه مکان $x = 0, a, -a$ نمونه‌برداری شده که در آن a فاصله نمونه‌برداری است. براساس سه مکان فوق، سه معادله $d_1 = d(0;c,w,\sigma_d,x_0)$، $d_2 = d(a;c,w,\sigma_d,x_0)$ و $d_3 = d(-a;c,w,\sigma_d,x_0)$ به دست می‌آید. سپس سه پارامتر، به شکل زیر محاسبه می‌شوند.

$$\begin{aligned} c &= d_1 \cdot \sqrt{\frac{2\pi a^2}{\ln(l_1)}} \cdot l_2^{\frac{1}{4a}} \\ w &= \sqrt{\frac{a^2}{\ln(l_1)} - \sigma_d^2} \\ b &= s(x_0) - \frac{c}{2} \end{aligned} \quad (۱۲)$$

که در آن $l_1 = d_1^2/(d_2 d_3)$ و $l_2 = d_2/d_3$.

نگاشت جهت لبه (EDM)

جهت لبه یکی دیگر از ویژگی‌های لبه است که می‌تواند برای ارزیابی کیفیت تصاویر محتوای صفحه مفید باشد. همانطور که قبلا گفته شد، اندازه گرادیان برابر است با:

$$G(x,y) = |G_H| + |G_V| \quad (۱۳)$$

که G_H و G_V مشتقات جهت‌دار در جهت‌های افقی و عمودی هستند که به صورت زیر محاسبه می‌شوند:

$$\begin{aligned} G_H &= I(x+1,y) - I(x,y) \\ G_V &= I(x,y+1) - I(x,y) \end{aligned} \quad (۱۴)$$

که $I(x,y)$ شدت روشنایی تصویر محتوای صفحه ورودی است.

نگاشت جهت لبه با کانوالو کردن کرنل L_1 در جهت l، برای $l = (0,...,11)$ به دست می‌آید.

$$EDM_l(x,y) = L_1 \otimes G(x,y) \quad (۱۵)$$

کرنل کانولوشن L_1 با چرخش کرنل کانولوشن L_0 با زاویه $l \times \frac{\pi}{12}$ به دست می‌آید. L_0 یک ماتریس با اندازه 27×27 است که در امتداد جهت افقی کار می‌کند. در این ماتریس، تمام ورودی‌های ردیف میانی مقدار ثابت ۱ و بقیه ورودی‌ها مقدار ۰ دارند.

در نهایت $EDM(x,y)$ نهایی، بر اساس دوازده $EDM_l(x,y)$، به دست می‌آید.

$$\begin{aligned} EDM(x,y) &= n\frac{\pi}{12} \\ n &= \arg\max\{EDM_l(x,y)\} \end{aligned} \quad (۱۶)$$

در مرحله بعد، اندازه‌گیری تشابه لبه برای سه ویژگی لبه (کنتراست، عرض و جهت)، با استفاده از رابطه زیر به دست می‌آید.

$$s_\phi(\phi_r,\phi_d,T_\phi) = \frac{2\phi_r\phi_d + T_\phi}{\phi_r^2 + \phi_d^2 + T_\phi} \quad (۱۷)$$

که در آن ϕ می‌تواند ECM، EWM یا EDM باشد و r و d به ترتیب، تصاویر محتوای صفحه مرجع و تخریب‌شده را نشان می‌دهند. همچنین T_ϕ یک ثابت مثبت برای جلوگیری از بی ثباتی است.

سپس با ترکیب این سه نگاشت، اندازه‌گیری تشابه کلی لبه محاسبه می‌شود.

$$s(x,y) = [ECS(x,y)]^\alpha \cdot [EWS(x,y)]^\beta \cdot [EDS(x,y)]^\gamma \quad (۱۸)$$

ECS، EWS و EDS به ترتیب، نگاشت‌های تشابه کنتراست، عرض و جهت لبه و α، β و γ سه ثابت مثبت برای تنظیم اهمیت نسبی نگاشت‌های تشابه هستند که در این پژوهش $\alpha = \beta = \gamma = 1$ در نظر گرفته شدند.

سپس به جای اینکه برای محاسبه نمره کیفیت، از اندازه‌گیری تشابه کلی لبه استفاده شود، یک فرآیند وزن‌دهی در نظر گرفته می‌شود. برای به دست آوردن این پارامتر، از اطلاعات عرض لبه استفاده می‌شود.

$$W_m(x,y) = \max(EWM_r, EWM_d) \quad (۱۹)$$

که در آن EWM_r و EWM_d به ترتیب، نگاشت‌های عرض لبه تصویر مرجع و تخریب‌شده هستند.

۲-۳- معیار نهایی

در نهایت، نگاشت تشابه بین نواحی چشمگیر تصویر مرجع و تخریب‌شده به صورت زیر به دست می‌آید.

$$VSS = \frac{2VS_r \cdot VS_d + c}{VS_r^2 + VS_d^2 + c} \quad (۲۰)$$

که در آن VS_r و VS_d به ترتیب، نگاشت‌های نواحی چشمگیر تصاویر مرجع و تخریب‌شده هستند. در ادامه از یک تابع وزن‌دهی برای وزن دادن به نواحی چشمگیر تصویر استفاده می‌شود.

$$weight(x,y) = \max(VS_r, VS_d) \quad (۲۱)$$

در نهایت، اندازه گرادیان (GM)، تشابه کلی لبه (s(x,y))و تشابه بین نگاشت‌های نواحی چشمگیر (VSS) با هم ترکیب شده و نمره ارزیابی کیفیت نهایی بر اساس رابطه (۲۲) محاسبه می‌شود.

$$SM(x,y) = \frac{\sum_{(x,y)} GM(x,y)\cdot s(x,y)\cdot W_m(x,y)\cdot VSS(x,y)\cdot weight(x,y)}{\sum W_m(x,y)\cdot weight(x,y)}$$

(۲۲)

۳- ارزیابی روش پیشنهادی

در این بخش مقاله، ابتدا به معرفی دو پایگاه داده برای تصاویر محتوای صفحه پرداخته شده است. سپس روش پیشنهادشده، با برخی روش‌های ارائه‌شده برای ارزیابی کیفیت تصاویر محتوای صفحه مقایسه شده است.

۳-۱- معرفی پایگاه داده

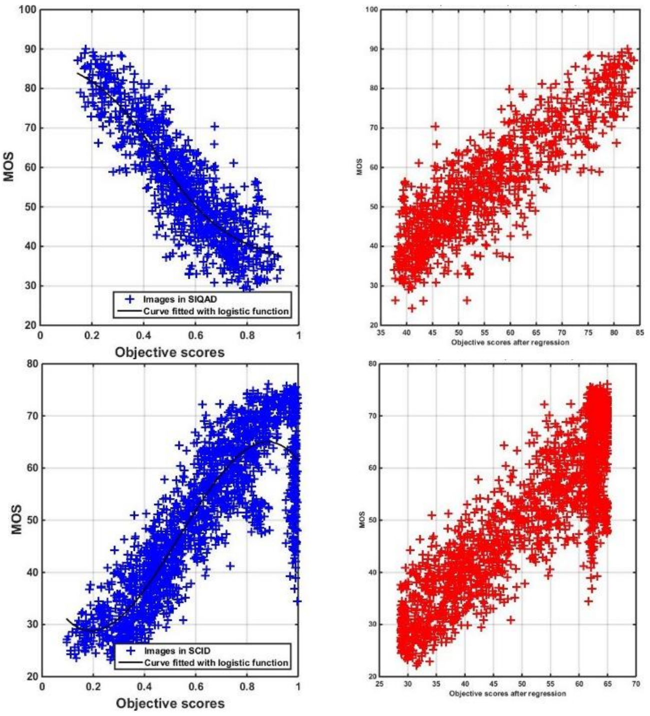
در این قسمت، دو پایگاه داده برای تصاویر محتوای صفحه معرفی شده است: پایگاه داده SIQAD[۲۲] [12] و SCID[۲۳] [9]. در ادامه، این پایگاه‌های داده تعریف شده‌اند.

پایگاه داده SIQAD شامل ۲۰ تصویر مرجع و ۹۸۰ تصویر تخریب‌شده است. تصاویر مرجع از طریق فایل‌های pdf، اسلایدهای PowerPoint، مجلات دیجیتال و غیره جمع‌آوری شده‌اند. برای تولید تصاویر تخریب‌شده نیز از ۷ نوع خرابی مختلف استفاده شده است. نویز گاسین (GN[۲۴])، تاری گاسین (GB[۲۵])، تاری حرکت (MB[۲۶])، تغییر کنتراست (CC[۲۷])، فشرده‌سازی JPEG و JPEG2000 و کدگذاری مبتنی بر تقسیم‌بندی لایه‌ها (LSC[۲۸]). برای هر نوع خرابی نیز، ۷ سطح تخریب تنظیم شده است.

پایگاه داده SCID شامل ۴۰ تصویر مرجع و ۱۸۰۰ تصویر تخریب‌شده است. این تصاویر در اندازه ثابت ۱۲۸۰×۷۲۰ برش داده شده‌اند. پایگاه داده SCID شامل ۹ نوع خرابی است و برای هر نوع خرابی ۵ سطح تخریب ایجاد شده است. این ۹ نوع خرابی عبارتند از: نویز گاسین (GN)، تاری گاسین (GB)، تاری حرکت (MB)، تغییر کنتراست (CC)، فشرده‌سازی JPEG و JPEG2000، تغییر اشباع رنگ (CSC[۲۹])، کدگذاری ویدئویی با بازده بالا (HEVC-SCC[۳۰]) و کمی‌سازی رنگ با dithering (CQD[۳۱]).

۳-۲- ارزیابی عملکرد

برای ارزیابی عملکرد الگوریتم، از چهار معیار ارزیابی استفاده شده است: ضریب همبستگی خطی Pearson (PLCC[۳۲]) جهت اندازه‌گیری دقت پیش‌بینی، ضریب همبستگی مرتبه Spearman (SROCC[۳۳])و ضریب همبستگی مرتبه Kendall (KROCC[۳۴]) به‌منظور تخمین یکنواختی پیش‌بینی و جذر میانگین مربع خطا (RSME[۳۵]). شکل ۲ نمودار پراکندگی نتایج SROCC معیار پیشنهادی نسبت به امتیازات انسانی (MOS[۳۶]) برای دو پایگاه داده SIQAD و SCID را نشان می‌دهد. همبستگی بالای قابل مشاهده در این نمودارها بر سازگاری بالا بین امتیازات انسانی و امتیازات منتج از معیار پیشنهادی تأکید میکند و این نشان‌دهندۀ توان بالای پیشگویی این معیار است.



شکل (۲) : نمایش پراکندگی MOS نسبت به نتایج حاصل از اعمال الگوریتم پیشنهادی بر روی پایگاه‌های داده SIQAD و SCID

در این قسمت، روش پیشنهادی با الگوریتم‌های پیشرفته ارزیابی کیفیت تصویر موجود مقایسه شده است. الگوریتم‌های ارزیابی کیفیت تصویر که با روش پیشنهادی مقایسه شده‌اند عبارتند از: PSNR، SSIM [1]، MSSSIM [2]، FSIM [13]، IWSSIM [3]، VIF [14]، VSI [15]، SIQM [7]، ESIM [9] و SQMS [10]. در این میان سه روش آخر روش‌های ارزیابی کیفیت تصاویر محتوای صفحه و باقی روش‌ها برای ارزیابی کیفیت همه تصاویر پیشنهاد شده‌اند. جدول ۱ نتایج مقایسه الگوریتم پیشنهادی با الگوریتم‌های گفته شده را در پایگاه‌های داده SIQAD و SCID نشان می‌دهد.

---

[۲۲] Screen Image Quality Assessment Database
[۲۳] Screen Content Image Database
[۲۴] Gaussian Noise
[۲۵] Gaussian Blurring
[۲۶] Motion Blurring
[۲۷] Contrast Change
[۲۸] Layer-Segmentation based Compression
[۲۹] Color Saturation Change
[۳۰] HEVC Screen Content Compression
[۳۱] Color quantization with dithering
[۳۲] Pearson Linear Correlation Coefficient
[۳۳] Spearman Rank-Order Correlation Coefficient
[۳۴] Kendall Rank-Order Correlation Coefficient
[۳۵] Root Mean Squared Error
[۳۶] Mean Opinion Score

جدول (۱) : مقایسه عملکرد الگوریتم پیشنهادی با سایر الگوریتم‌های IQA بر روی پایگاه‌های داده SIQAD و SCID

	Criteria	PSNR	SSIM	MSSSIM	FSIM	IWSSIM	VIF	VSI	SIQM	ESIM	SQMS	Pro.
SIQAD	PLCC	0.5858	0.7561	0.6161	0.5888	0.6519	0.8083	0.5543	0.8520	0.8788	**0.8872**	0.8734
	SROCC	0.5570	0.7566	0.6115	0.5824	0.6546	0.8069	0.5381	0.8450	0.8632	**0.8803**	0.8554
	KROCC	0.4209	0.5583	0.4547	0.4253	0.4977	0.6082	0.3874	0.6521	0.6742	**0.6936**	0.6672
	RMSE	11.6008	9.3680	11.2744	11.5700	108550	8.4282	11.9152	7.4936	6.8310	**6.6039**	6.9716
SCID	PLCC	0.7622	0.7343	0.7579	0.7718	0.7877	0.8179	0.7694	0.8304	0.8630	0.8557	**0.8798**
	SROCC	0.7512	0.7146	0.7407	0.7550	0.7714	0.7969	0.7621	0.8086	0.8478	0.8320	**0.8626**
	KROCC	0.5512	0.5180	0.5378	0.5541	0.5726	0.6055	0.5648	0.6168	0.6516	0.6429	**0.6711**
	RMSE	9.1682	9.6133	9.2400	9.0040	8.7243	8.1479	9.0456	7.8910	7.1552	7.3276	**6.7322**

با توجه به نتایج به دست آمده در جدول ۱، می توان دریافت که الگوریتم پیشنهادی بر روی پایگاه داده SCID، نسبت به سایر الگوریتم‌های ارزیابی کیفیت تصویر، عملکرد بهتری داشته و سازگاری بالاتری با ادراک بصری انسان دارد. همچنین این روش، نسبت به خرابی‌های تغییر اشباع رنگ، کدگذاری ویدئویی با بازده بالا و کمی‌سازی رنگ با dithering حساس‌تر بوده و در مورد این خرابی‌ها نیز بهتر عمل کرده است. در خصوص عملکرد روش پیشنهادی بر روی پایگاه داده SIQAD نیز می‌توان ادعا کرد که فقط روش SQMS نسبت به روش پیشنهادی عملکرد بهتری دارد. روش ESIM نیز نتایجی مشابه و نزدیک به روش پیشنهادی را از خود نشان می‌دهد.

۴- نتیجه‌گیری

در این مقاله، یک روش ارزیابی کیفیت تصاویر محتوای صفحه بر اساس ویژگی‌های شاخص مستخرج از لبه‌های تصویر و وزن‌دهی شده بر اساس نواحی چشمگیر پیشنهاد شده است. آزمایشات بر روی دو پایگاه داده تصاویر محتوای صفحه SIQAD و SCID نشان می‌دهد که عملکرد الگوریتم پیشنهادی در مقایسه با ۱۱ الگوریتم IQA دیگر بهتر بوده و سازگاری بالاتری با پیش‌بینی‌های انجام‌شده توسط سیستم بینایی انسان دارد.

مراجع

[1] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE transactions on image processing, vol. 13, no. 4, pp. 600-612, 2004.

[2] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003, 2003, vol. 2: Ieee, pp. 1398-1402.

[3] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment," IEEE Transactions on image processing, vol. 20, no. 5, pp. 1185-1198, 2010.

[4] Y. Lu, S. Li, and H. Shen, "Virtualized screen: A third element for cloud–mobile convergence," Ieee Multimedia, vol. 18, no. 2, pp. 4-11, 2011.

[5] H. Shen, Y. Lu, F. Wu, and S. Li, "A high-performanance remote computing platform," in 2009 IEEE International Conference on Pervasive Computing and Communications, 2009: IEEE, pp. 1-6.

[6] T. Lin, P. Zhang, S. Wang, K. Zhou, and X. Chen, "Mixed chroma sampling-rate high efficiency video coding for full-chroma screen content," IEEE Transactions on Circuits and Systems for Video Technology, vol. 23, no. 1, pp. 173-185, 2012.

[7] K. Gu, S. Wang, G. Zhai, S. Ma, and W. Lin, "Screen image quality assessment incorporating structural degradation measurement," in 2015 IEEE International Symposium on Circuits and Systems (ISCAS), 2015: IEEE, pp. 125-128.

[8] Z. Ni, H. Zeng, L. Ma, J. Hou, J. Chen, and K.-K. Ma, "A Gabor feature-based quality assessment model for the screen content images," IEEE Transactions on Image Processing, vol. 27, no. 9, pp. 4516-4528, 2018.

[9] Z. Ni, L. Ma, H. Zeng, J. Chen, C. Cai, and K.-K. Ma, "ESIM: Edge similarity for screen content image quality assessment," IEEE Transactions on Image Processing, vol. 26, no. 10, pp. 4818-4831, 2017.

[10] K. Gu et al., "Saliency-guided quality assessment of screen content images," IEEE Transactions on Multimedia, vol. 18, no. 6, pp. 1098-1110, 2016.

[11] X. Min, K. Gu, G. Zhai, M. Hu, and X. Yang, "Saliency-induced reduced-reference quality index for natural scene and screen content images," Signal Processing, vol. 145, pp. 127-136, 2018.

[12] H. Yang, Y. Fang, and W. Lin, "Perceptual quality assessment of screen content images," IEEE Transactions on Image Processing, vol. 24, no. 11, pp. 4408-4421, 2015.

[13] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," IEEE transactions on Image Processing, vol. 20, no. 8, pp. 2378-2386, 2011.

[14] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," IEEE Transactions on image processing, vol. 15, no. 2, pp. 430-444, 2006.

[15] L. Zhang, Y. Shen, and H. Li, "VSI: A visual saliency-induced index for perceptual image quality assessment," IEEE Transactions on Image processing, vol. 23, no. 10, pp. 4270-4281, 2014.

چهارمین کنفرانس ملی انفورماتیک ایران
پژوهشگاه دانش‌های بنیادی، پردیس فرمانیه، تهران
۱۴ و ۱۵ دی ماه ۱۴۰۱

بهبود کیفیت تبدیل سری زمانی به گراف پدیداری نفوذپذیر محدود، با ایجاد نسخه وزن‌دار و پیش‌بینی پیوند

الهام امیری[۱]، صادق سلیمانی[۲]، پرهام مرادی[۳]، آسو مفاخری[۴]

[۱] دانشجوی کارشناسی ارشد رشته الگوریتم‌ها و محاسبات گروه کامپیوتر دانشکده مهندسی دانشگاه کردستان
elham.amiri@uok.ac.ir
[۲] عضو هیات علمی گروه کامپیوتر دانشکده مهندسی دانشگاه کردستان
s.sulaimany@uok.ac.ir
[۳] عضو هیات علمی گروه کامپیوتر دانشکده مهندسی دانشگاه کردستان
p.moradi@uok.ac.ir
[۴] دانشجوی دکتری رشته مهندسی نرم‌افزار گروه کامپیوتر دانشکده مهندسی دانشگاه کردستان
aso.mafakheri@uok.ac.ir

چکیده

تحلیل سری‌های زمانی از منظر شبکه‌های پیچیده بسیاری از دانشمندان را مورد توجه قرار داده است. در این مقاله، بر اساس الگوریتم‌های گراف پدیداری، روش جدیدی برای ساخت شبکه(گراف) وزن‌دار از سری های زمانی پیشنهاد شده است. اولین مرحله تبدیل نوع داده سری‌های زمانی به گراف پدیداری نفوذپذیر محدود(نسخه قوی‌تری از گراف پدیداری طبیعی) است. سپس از سه روش وزن‌دار کردن فاصله اقلیدسی، تانژانت زاویه دید و فاصله زمانی برای ساخت گراف وزن‌دار استفاده شده است. در مرحله بعد با اعمال پیش‌بینی پیوند، سعی در بهبود هرچه بیشتر کیفیت گراف شده است. برای سنجش شباهت در پیش‌بینی پیوند از معیارهای شباهت همسایگان مشترک، اتصال ترجیحی و ضریب جاکارد در نسخه وزن‌دار پیشنهاد و ارزیابی شده است.

*برای اثبات اعتبار روش پیشنهادی، سه مجموعه داده سری زمانی **Taiex**، فروش خانه و فروش شامپو اتخاذ شده است که در آن از معیار مبتنی بر پیش‌بینی پیوند **AUC** برای ارزیابی عملکرد گراف وزن‌دار استفاده می‌شود. نشان داده شده است که در گراف وزن‌دار ساخته شده با روش‌های پیشنهادی و اعمال پیش‌بینی پیوند، مقدار **AUC** حداکثر تا ۰/۹۶۳۳۶ افزایش می‌یابد و تا حد زیادی بهتر از گراف بدون وزن به دست آمده توسط نظریه گراف پدیداری نفوذپذیر محدود عمل می‌کند.*

کلمات کلیدی

سری زمانی، گراف پدیداری نفوذپذیر محدود، فاصله اقلیدسی، تانژانت زاویه دید، فاصله زمانی، پیش‌بینی پیوند

۱- مقدمه

سری‌های زمانی[۱] مجموعه‌ای از نقاط داده هستندکه به ترتیب زمانی مرتب می‌شوند. معمولاً یک سری زمانی، دنباله‌ای است که در نقاط متوالی با فاصله مساوی از زمان گرفته می‌شود. بنابراین توالی داده‌های زمان، گسسته است. یکی از رایج‌ترین روش‌های تبدیل داده سری‌زمانی، استفاده از گراف پدیداری (VG)[2] است که برای اولین بار در سال ۲۰۰۸ مطرح شد[1].

گراف پدیداری در بسیاری از زمینه‌ها مانند تحلیل اقتصاد[2]،زمین‌شناسی[3]، زیست‌شناسی[4]، گردشگری[5] و غیره به کار رفته است.اما حاصل اعمال الگوریتم گراف پدیداری بر داده‌های سری زمانی، گرافی ساده است و دقت انجام چنین کاری نیز صد درصد نیست.

گراف پدیداری نفوذپذیر محدود (LPVG)[3] ، به عنوان یک نسخه قوی‌تر با مقاومت بیشتر نسبت به نویز در سال ۲۰۱۲ ارائه شد[6]. روش LPVG برای تجزیه و تحلیل سیگنال‌های رادیویی[7]، تشخیص راه رفتن برای درمان بالینی در پزشکی[8] و رفتار جریان دینامیکی دو فازی گاز-مایع[9] با موفقیت به‌کار رفته است. با این حال گراف حاصل شده از تبدیل سری زمانی توسط گراف پدیداری نفوذپذر محدود نیز گرافی ساده است. گراف ساده ممکن است بیانگر ماهیت اصلی داده‌های سری زمانی نباشد، پس طی پژوهشی که در سال ۲۰۱۸ توسط ژو و همکاران ارائه شد ثابت گردید وزن‌دار کردن گراف می‌تواند تاثیر بسزایی در کیفیت گراف داشته باشد[10]. با توجه به امکان بهبود کیفیت گراف از طریق وزن‌دار کردن، آشکار می‌شود که دیگر روش‌های وزن‌دهی هم مورد بررسی قرار گیرند. انتخاب بهترین روش می‌تواند موضوع جذابی از منظر علاقمندان به حیطه گراف باشد. از این رو، روش‌های دیگر برای بهبود کیفیت این گراف‌ها پس از وزن‌دار کردن، مورد توجه قرار می‌گیرند. در این مقاله به تاثیر پیش‌بینی پیوند در جهت بهبود گراف اشاره شده است و نتایج مورد تحلیل و بررسی قرار می‌گیرند.

چهارمین کنفرانس ملی انفورماتیک ایران
پژوهشگاه دانش‌های بنیادی، پردیس فرمانیه، تهران
۱۴ و ۱۵ دی ماه ۱۴۰۱

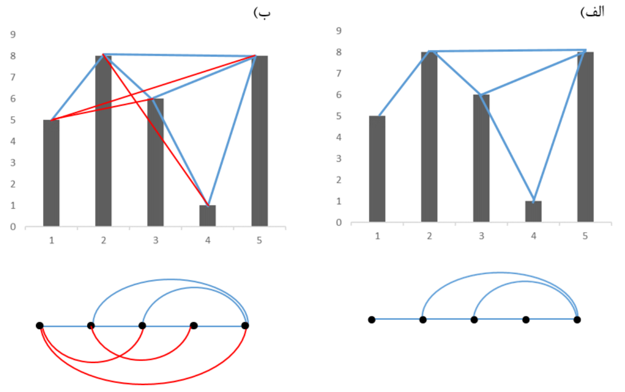
ساختار مقاله به شرح زیر خواهد بود: بخش دوم تعریفی از گراف‌های پدیداری و نوع خاص‌تری از آنها یعنی گراف پدیداری نفوذپذیر محدود ارائه شده است. در بخش سوم روش‌هایی از بهبود کیفیت گراف که در این کار پژوهشی مورد استفاده قرار گرفته بیان شده است. در بخش چهار معیار ارزیابی گراف معرفی شده است و در بخش پنجم به شرح مراحل روش پیشنهادی پرداخته شده است. بخش ششم به معرفی مجموعه داده‌ها اختصاص داده شده است. نتایج و تفاسیر در بخش هفت و نتیجه‌گیری در بخش هشتم آورده شده است.

۲- گراف پدیداری

گراف پدیداری به عنوان پلی بین سری‌های زمانی و شبکه در نظر گرفته می‌شود که در آن هر نقطه دارای یک زمان و یک مقدار است. گراف پدیداری در انواع مختلفی وجود دارد. گراف پدیداری طبیعی(NVG)، اولین نوع معرفی شده از گراف‌های پدیداری است. با چیدن متوالی نقاط بر محور زمان(t)، می‌توان مقدار هر کدام را توسط یک میله عمودی(y) نشان داد. در ایجاد گراف پدیداری از سری زمانی، در صورتی که سر میله‌ها دارای دید مستقیم باشد، یک یال بین دو نقطه به عنوان دو رأس در گراف ترسیم می‌شود. به زبان ریاضی، شرط لازم برای ایجاد اتصال(یال) بین دو سر میله (t_i,y_i) و (t_j, y_j) با وجود میله میانی (t_n,y_n) صدق کردن در معادله(۱) است.

$$y_n < y_j + (y_i - y_j)\frac{t_j - t_n}{t_j - t_i} \qquad (۱)$$

کوچکتر بودن y_n از حاصل جمع y_j و بخش دوم سمت راست معادله(۱) تضمین عدم تقاطع y_i و y_j با میله میانی y_n است. این موضوع مستقل از بزرگتر بودن y_j از y_i یا برعکس است. برای درک بهتر می‌توان y_i را به سمت چپ معادله برد و صورت کسر تفاضل زمانی را بر مخرج آن تقسیم کرد. درشکل(۱) قسمت الف نمونه‌ای از تبدیل یک سری زمانی ساده به گراف پدیداری طبیعی را مشاهده می‌کنیم.



شکل(۱): نمونه‌ای از تبدیل سری زمانی به گراف پدیداری. الف) تبدیل سری زمانی به NVG. ب)تبدیل سری زمانی به LPVG با فرض l=1

۱-۲- گراف پدیداری نفوذپذیر محدود

گراف پدیداری نفوذپذیر محدود اولین بار در سال ۲۰۱۲ توسط تینگ تینگ ژو و همکاران به منظور بهبود گراف پدیداری طبیعی و افقی (درحالت ساده) ارائه شد[6]. مزایای آن، توانایی بالقوه در تشخیص سری‌های زمانی مختلف و مقاومت بالا در برابر نویز به ویژه برای داده‌های ناخالص است. در این روش، شرط ایجاد یال میان دو میله (گره) از سری زمانی، با صرف نظر از تعداد تعیین شده‌ای از موانع در نظر گرفته می‌شود به این صورت که به بیان ریاضی، شرط ایجاد یال میان دو گره (t_i, y_i) و (t_j, y_j) در معادله(۲) بیان شده است.

$$y_{i+l} < y_j + (y_i - y_j)\frac{t_j - (t_i + t_l)}{t_j - t_i}, \; l < j - i \qquad (۲)$$

که در آن l تعیین کننده تعداد میله‌هایی است که خط عبوری از ارتفاع y_i به y_j می‌تواند آن را قطع کند. اگر l=0 باشد آنگاه LPVG=NVG است و اطلاعات کمتری از شبکه مستخرج می‌شود. در شکل(۱) قسمت ب نمونه‌ای از تبدیل یک سری زمانی ساده به گراف پدیداری نفوذپذیر محدود را در مقایسه با گراف پدیداری طبیعی مشاهده می‌کنید.

دسته بندی شکل(۲) شامل انواع این گراف‌ها و زیرگراف‌هایی هست که تا کنون مورد بررسی پژوهشگران بسیاری قرار گرفته‌اند. طبق بررسی‌های گذشته، گراف پدیداری طبیعی (NVG)[4] یا همان (VG) قادر به تفکیک سری‌های زمانی با صفات خاص نیست. از طرفی LPVG ها بسیار امیدوارکننده هستند. از آنجا که LPVG ها چندین ویژگی از سری زمانی را به ارث می‌برند و قادر به تشخیص تفاوت بین سری‌های تصادفی و بی نظم هستند، محل انشعابات معکوس در سیستم‌های دینامیکی آشفته را تشخیص می‌دهند. علاوه بر این، LPVG ها تحمل خوبی نسبت به نویز از خود نشان می‌دهند که یک مزیت در مقایسه با NVG ها است . با این حال، لازم است به انتخاب پارامتر l توجه شود تا اطلاعات زیادی در گراف‌ها ادغام نشود و نویز را نیز با خود به همراه نیاورد. به همین دلیل، در این کار تحقیقاتی که وابسته به کیفیت گراف‌های ایجاد شده است، ترجیحا از نوع LPVG بهره می‌بریم.

شکل(۲): دسته بندی الگوریتم‌های گراف پدیداری

۳- بهبود کیفیت گراف

هرچه روش بهتری برای تبدیل سری زمانی به گراف پدیداری استفاده شود، گراف معادل، خصوصیات بیشتری از سری زمانی را در خود خواهد داشت. لذا نفس ایجاد با کیفیت‌تر گراف پدیداری یک خواسته و چالش علمی محسوب می‌شود. برای پاسخ به این چالش تا کنون ایده‌هایی برای بهبود کیفی گراف ارائه شده است. در ادامه به بررسی روش‌های مورد استفاده در این مقاله می‌پردازیم.

۳-۱- وزن‌دار کردن گراف

وزن‌دار کردن گراف می‌تواند یک ایده ثمربخش باشد. لذا تاکنون نسخه وزن‌دار گراف پدیداری طبیعی نیز عرضه و کارایی آن به اثبات رسیده است. در حالت عمومی‌تر برای اختصاص وزن به یال‌های ایجاد شده از تانژانت زاویه دید استفاده می‌شود.

در پژوهشی متفاوت ژو و همکارانش برای وزن دهی به گراف پدیداری از ترکیب خطی اپراتورهای IOWA[5] و VGA[6] بهره بردند[10]. این اپراتورها گره‌ها را در گراف پدیداری وزن‌دار می‌کردند که برای اختصاص وزن گره‌ها به یال‌ها از دو راهبرد میانگین‌گیری[۷] و جاذبه[۸] استفاده کردند. آن‌ها گراف حاصل شده را با معیار AUC مورد سنجش قرار دادند. حداکثر میزان بهبود روش در مقاله مذکور AUC=0.7226 بود که همواره ملزم به بهبود است، چراکه میدانیم در معیار AUC هر چه به عدد ۱ نزدیکتر باشیم نتیجه بهتر و بهبود کیفیت گراف ملموس‌تر است. پس از آن، روش‌های متعددی برای وزن‌دار کردن گراف پدیداری پیشنهاد شد. ژنگ و همکاران در پژوهش اخیر خود دو روش ساده اما کارآمد علاوه بر تانژانت زاویه دید، برای وزن‌دهی به گراف پدیداری ارائه دادند[11]. در یک جمع‌بندی روش‌های کاربردی که در پژوهش‌های متعدد بیشترین استفاده را برای وزن‌دار کردن گراف حاصل از خانواده الگوریتم‌های گراف پدیداری داشته‌اند در جدول(۱) قابل مشاهده است.

جدول(۱): فرمول‌های پایه وزن‌دهی به یال‌های حاصل از LPVG

عنوان روش	فرمول‌ها
فاصله اقلیدسی	$w_{ij} = \sqrt{(y_i - y_j)^2 + (t_i - t_j)^2}$
تانژانت زاویه دید	$w_{ij} = arctan\frac{y_j - y_i}{t_j - t_i}$
فاصله زمانی	$w_{ij} = \lvert t_i - t_j \rvert$

۳-۲- پیش‌بینی پیوند

راهکارهای دیگری را نیز می‌توان با شناخت از محاسبات و تحلیل‌های شبکه برای ایجاد گراف پدیداری بهتر، پیشنهاد داد. مثلا یک ایده برای افزایش کیفیت گراف پدیداری می‌تواند پیش‌بینی پیوند[۹] باشد. پیش‌بینی پیوند یک چالش اصلی شبکه اجتماعی است که در پایه‌ای‌ترین تعریف، به امتیازدهی یال‌های غیرموجود در گراف، برای یافتن محتمل‌ترین ارتباطات می‌پردازد[12]. شبکه مورد نظر در پیشگویی پیوند می‌تواند ساده، وزن‌دار و جهت‌دار و... باشد. پیش‌بینی پیوند در شبکه‌های اجتماعی، همکاری علمی، زیستی و پزشکی و... مورد توجه بوده است[12]. در بالاترین سطح دسته‌بندی روش‌های پیش‌بینی پیوند، می‌توان آن‌ها را به دو دسته با ناظر[۱۰] و بدون ناظر[۱۱] کرد[13]. معمولاً پذیرفته شده است که اگر دو گره به نحوی شباهت یا مجاورت بیشتری داشته باشند، بهم متصل می‌شوند.

اغلب، شاخص‌های همسایگان مشترک[۱۲](CN)، اتصال ترجیحی[13](PA) و ضریب جاکارد[۱۴](JC) را با توجه به سادگی و اثربخشی آن به عنوان معیار ارزیابی شباهت در گره‌ها به‌کار می‌برند که جزییات ویرایش وزن‌دار آن‌ها در جدول(۲) آمده است. پیش‌بینی پیوند الزاما تنها برای یافتن یال‌های از قلم افتاده یا محتمل برای تشکیل نیست، بلکه می‌توان از آن برای یافتن یال‌هایی که در آینده نزدیک حذف خواهند شد نیز بهره برد. بدین ترتیب از نظر استراتژی حذف یا اضافه کردن یال، پیش‌بینی پیوند را به دو دسته مثبت و منفی تقسیم می‌کنند که منظور از پیش‌بینی پیوند منفی، حذف یال‌های ناهنجار یا اضافی است[13]. حذف یال‌های غیر محتمل، نویز یا ناهنجاری، نیز می‌تواند به افزایش بیشتر کیفیت شبکه مورد بررسی منجر شود.

جدول(۲): فرمول‌های پایه امتیازدهی به یال‌های محتمل در روش پیش‌بینی پیوند وزن‌دار. w(x,y) بیانگر وزن یال بین x و y است و N(x) به مجموعه همسایگان رأس x اشاره دارد.

عنوان معیار شباهت	فرمول‌ها برای گراف وزن‌دار
همسایگان مشترک	$CN(x,y) = \sum_{u \in N(x) \cap N(y)} w(x,u) + w(y,u)$
اتصال ترجیحی	$PA(x,y) = \sum_{u \in N(x)} w(u,x) * \sum_{v \in N(y)} w(v,y)$
ضریب جاکارد	$JC(x,y) = \frac{\sum_{z \in (N(x) \cap N(y))} w(x,z) + w(y,z)}{\sum_{u \in N(x)} w(u,x) + \sum_{v \in N(y)} w(v,y)}$

۴- ارزیابی گراف

معیار AUC معیار ارزیابی قدرتمند در نظر گرفته در این زمینه است. در این آزمون، تعدادی یال به عنوان یال‌های مورد بررسی از گراف حذف می‌شود. به یال‌های حذف شده از گراف و یال‌هایی که از ابتدا نیز در گراف وجود نداشتند، یک امتیاز به صورت تصادفی داده می‌شود. محاسبه AUC با فرمول معاله(۳) انجام می‌شود[10].

$$AUC = \frac{m_1 + m_2}{m} \qquad (۳)$$

که در آن m تعداد کل مقایسات بین امتیازهای داده شده به یال‌ها است و m_1 تعداد دفعاتی است که امتیاز یال‌های حذف شده از امتیاز یال‌های ناموجود بیشتر باشد و m_2 تعداد دفعاتی است که یال حذف شده و یال ناموجود امتیاز برابری داشته باشند.

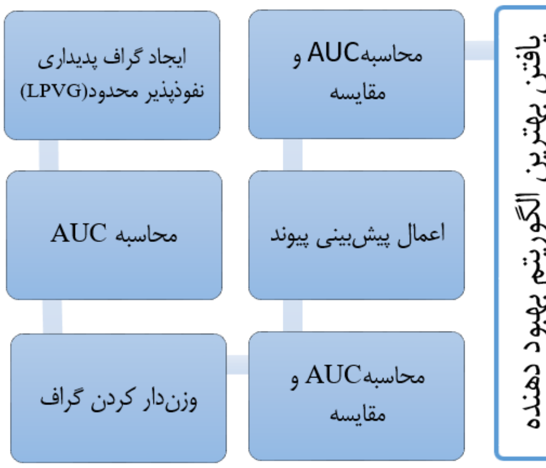
۵- روش پیشنهادی

مراحل و روش کار پیشنهادی برای بهبود هرچه بیشتر کیفیت گراف به صورت شکل(۳) است. یعنی:

گام اول: گراف پدیداری نفوذپذیر محدود را از سری‌های زمانی مورد بررسی استخراج می‌کنیم و با استفاده از معیار AUC مورد سنجش قرار می‌دهیم.

گام دوم: وزن‌دارکردن گراف با استفاده از روش‌های وزن‌دهی مذکور و سنجش دوباره گراف. این مرحله از تنوع خاصی در انتخاب روش وزن‌دار کردن گراف برخوردار است.

گام سوم: اعمال پیش‌بینی پیوند مثبت و منفی روی هر سه گراف ایجاد شده از سه روش وزن‌دهی. در این مرحله یال‌های جدیدا اضافه شده پس از اعمال پیش‌بینی پیوند مثبت به گراف، بدون وزن هستند. برای تخصیص وزن به یال‌های جدیدا اضافه شده می‌توان از روش‌های متعددی استفاده کرد، اما در این مقاله با توجه به اینکه هریک از گراف‌ها از قبل طبق فرمول خاصی وزن‌دهی شده بودند، به یال‌های جدید نیز طبق همان فرمول، وزن اختصاص می‌دهیم. همچنین در پیش‌بینی پیوند مثبت، به تعداد ۰/۱ از یال‌های محتمل به گراف اضافه و ۰/۱ از یال‌های غیرمحتمل و نویز از گراف حذف می‌کنیم.



شکل(۳): روش و مراحل انجام کار

۶- مجموعه داده‌ها

در این بخش به معرفی مجموعه داده‌های مورد استفاده در این مقاله، قبل و بعد از تبدیل به گراف پدیداری نفوذ پذیر محدود، می‌پردازیم.

- مجموعه داده بورس تایوان (Taiex) حاوی ۴۸۹ راس(میله های سری زمانی) طی ۲ سال ۲۰۱۸و ۲۰۱۹ جمع آوری شد.
- مجموعه داده قیمت خانه که از ۲۰۰۷/۰۹/۳۰ تا ۲۰۱۹/۰۹/۳۰ جمع‌آوری شده و شامل ۳۴۷ راس(میله های سری زمانی) است.
- مجموعه داده فروش شامپو که در طی ۳ سال به صورت ماهانه جمع‌آوری شده و شامل ۳۷ راس(میله های سری زمانی) میباشد.

اطلاعات لازم جهت آزمون و تحلیل روش پیشنهادی از مجموعه داده‌های سری زمانی بعد از تبدیل به گراف پدیداری نفوذپذیر محدود، در جدول(۳) قرار داده شده است.

۷- نتایج و تفسیر

پس از جمع‌آوری داده‌ها و اعمال مراحل ذکر شده در جهت بهبود کیفی گراف پدیداری نفوذپذیر محدود، تمامی مراحل پیاده سازی و اجرا شد. از قیاس مقدار عددی AUC در هر مرحله و چیدومان جداول و نمودارهای نهایی به این نتیجه می‌رسیم که روش ما در بهبود کیفیت گراف عملکرد خیلی خوبی داشته است. البته لازم به ذکر است که در این مقاله برای اولین بار از روش وزن‌دار کردن گراف پدیداری نفوذپذیر محدود در جهت بهبود کیفی آن استفاده شده است و تا کنون مورد مشابهی یافت نشده. لذا این بهبود در این مرحله بسیار چشمگیر بوده چرا که توانسته است مقدار عددی AUC را به اندازه ۹/۶ هم بالا ببرد که در نوع خود قابل توجه است. البته باید در نظر گرفت که این مقدار گاها به انتخاب مجموعه داده، روش وزن‌دار کردن گراف و معیار شباهت استفاده شده وابسته است.

چنانچه در جدول(۴) در مجموعه داده Taiex مشاهده می‌کنید نتایج تقریبا با بهبود مواجه شده اما به طور مثال در روش تانژانت زاویه دید با معیار شباهت PA این مقدار کاهش یافته است. در داده‌های دیگر نیز مشاهده می‌شود که روش تانژانت زاویه دید و معیار شباهت PA نتایج چنان که باید بهبود نیافته است. مجموعه داده فروش خانه تا حدی امیدوار کننده است چرا که در اکثر موارد چه در روش وزن‌دار کردن و چه در روش پیش‌بینی پیوند بهبود حاصل شده است(جدول(۵)).

با این حال در مجموعه داده کوچک فروش شامپو در جدول(۶) مشاهده می‌کنیم که دراکثر موارد و هر دو روش، ایده ما به درستی عمل نکرده است. همانطور که در جدول(۳) می‌بینیم این مجموعه داده با تعداد راس ۳۷ و تعداد یال‌های ۲۳۲ حتی بعد از اعمال پیش‌بینی پیوند نیز خوب عمل نکرده است.

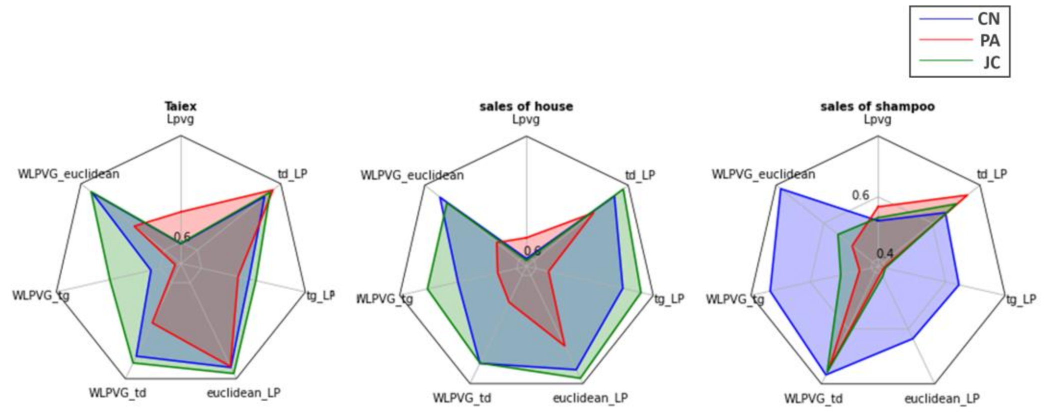
این عدم بهبود احتمالا به نوع مجموعه داده بستگی دارد. در مقایسه میان اطلاعات موجود از مجموعه داده‌ها در جدول(۳) متوجه می‌شویم تراکم گراف حاصل از مجموعه داده فروش شامپو نسبت به دو مجموعه داده Taiex و فروش خانه بسیار زیاد است. احتمال می‌رود گراف دارای تراکم بیشتر(چگال‌تر) برای روند بهبود و محاسبات ما به خوبی عمل نکند.

در جداول (۴) تا (۶)، W نماد نسخه وزن‌دار از معیار مورد نظر(CN,PA,JC) روی گراف وزن‌دار است و LP نمایانگر اعمال پیش‌بینی پیوند روی گراف موردنظر با معیار مورد نظر(CN,PA,JC) است. همچنین برای خوانایی جدول، بهترین عملکرد در هر سطر زیرخط‌دار و در هر ستون پررنگ شده است. لذا خانه‌هایی از جدول که محل تلاقی این دو نشانه هستند نمایانگر بهترین الگوریتم برای آن مجموعه داده است. به طور مثال در جدول(۴) برای مجموعه داده Taiex، روش وزن‌دار کردن فاصله اقلیدسی با معیار JC و اعمال پیش‌بینی پیوند بهترین عملکرد را دارد.

در شکل(۴) نمودار راداری(عنکبوتی) همه روش‌ها و میزان اثربخشی آن‌ها را مشاهده می‌کنید. این نمودار در انتخاب بهترین الگوریتم بهبود دهنده بسیار موثر است. باتوجه به نمودار مشاهده می‌کنیم در معیارهای شباهت، PA عملکرد خوبی نسبت به دو معیار CN وJC نداشته است. همچنین مجموعه داده فروش خانه به صورت کلی نسبت به دو مجموعه داده دیگر بهبود بیشتری داشته است. این درحالی است که مقدار AUC در بهترین عملکرد برابر با ۰/۹۶۳۳۶ متعلق به مجموعه داده Taiex است.

جدول(۳): ویژگی‌های مجموعه داده‌های مورد بررسی پس از تبدیل به LPVG

مجموعه داده‌ها	تعداد راس‌ها	تعداد یال‌ها	میانگین درجه	بیشترین درجه	چگالی شبکه	قطر شبکه
Taiex	۴۸۹	۷۸۳۳	۳۲/۰۳۶۸	۱۴۸	۰/۰۶۵۶۴۹۲	۶
فروش خانه	۳۴۷	۵۳۷۳	۳۰/۹۶۸۳	۱۴۹	۰/۰۸۹۵۰۳۷	۵
فروش شامپو	۳۷	۲۳۲	۱۲/۵۴۰۵	۲۷	۰/۳۴۸۳۴۸	۳



شکل(٤): نمودار راداری از تمامی نتایج حاصل شده از مراحل روش. در این نمودار LPVG حالت ساده گراف و WLPVG حالت وزن‌دار(euclidean ، tg وtd به ترتیب روش‌های وزن‌دهی فاصله اقلیدسی، تانژانت زاویه دید و فاصله زمانی می‌باشد). همچنین LP نمایانگر اعمال پیش‌بینی پیوند روی هریک از گراف‌های وزن‌دار است.

جدول(٤): نتایج اجرای الگوریتم پیشنهادی بهبود دهنده روی مجموعه داده Taiex

LPVG	CN		PA		JC	
	۰/۵۹۴۸		۰/۷۰۹۳		۰/۵۶۶۲	
WLPVG	W-CN	LP-CN	W-PA	LP-PA	W-JC	LP-JC
فاصله اقلیدسی	۰/۹۳۳۲۸	۰/۹۳۸۹۶	۰/۷۳۸۷۲	۰/۹۳۵۲	۰/۹۳۷۹۲	۰/۹۶۳۳۶
تانژانت زاویه دید	۰/۶۳۳۹۲	۰/۷۷۵۵۲	۰/۵۴۴۳۲	۰/۷۳۵۲	۰/۷۸۶۲۴	۰/۸۰۰۹۶
فاصله زمانی	۰/۸۹۴۲۴	۰/۹۰۹۵۲	۰/۷۶۱۱۲	۰/۹۴۸۸	۰/۹۲۱۱۲	۰/۹۳۵۹۲

جدول(۵): نتایج اجرای الگوریتم پیشنهادی بهبود دهنده روی مجموعه داده فروش خانه

LPVG	CN		PA		JC	
	۰/۵۹۲۸		۰/۶۵۷۴		۰/۵۸۷۱۹	
WLPVG	W-CN	LP-CN	W-PA	LP-PA	W-JC	LP-JC
فاصله اقلیدسی	۰/۹۱۱۶۸	۰/۹۲۳۸۴	۰/۶۸۶۸۸	۰/۸۴۲۲۴	۰/۸۸۴	۰/۹۵۳۴۴
تانژانت زاویه دید	۰/۷۸۶۷۲	۰/۸۷۴۸۸	۰/۶۶۰۳۲	۰/۶۳۹۰۴	۰/۸۸۳۶۸	۰/۹۳۲۸
فاصله زمانی	۰/۹۰۲۰۸	۰/۹۱۷۵۲	۰/۶۹۲۴۸	۰/۸۳۴۸۸	۰/۹۰۱۲	۰/۹۵۲۳۲

جدول(٦): نتایج اجرای الگوریتم پیشنهادی بهبود دهنده روی مجموعه داده فروش شامپو

LPVG	CN		PA		JC	
	۰/۵۲۳۴		۰/۵۶۸۱۹		۰/۵۳۳۲	
WLPVG	W-CN	LP-CN	W-PA	LP-PA	W-JC	LP-JC
فاصله اقلیدسی	۰/۷۷۶۶۴	۰/۶۳۴۳۲	۰/۴۸۲۸۸	۰/۳۹۹۵۲	۰/۵۱۱۴۴	۰/۴۱۱۹۲
تانژانت زاویه دید	۰/۷۳۲۴۸	۰/۶۴۳۴۴	۰/۴۳۷۶	۰/۳۹۵۸۴	۰/۴۹۹۸۴	۰/۴۰۰۵۶
فاصله زمانی	۰/۷۶۳۱۲	۰/۶۵۳۶۸	۰/۷۳۷۱۲	۰/۷۴۴۰۸	۰/۷۵۳۴۴	۰/۶۹۸۸

۸- نتیجه گیری

کاربردهای گراف پدیداری و کاربردهای آن به سرعت در حال تسری به حیطه‌های مختلف علمی و کاربردی است. بدین ترتیب، الگوریتم‌ها و روش‌هایی که منجر به ایجاد گراف‌های پدیداری با کیفیت‌تر شوند، کمک بزرگی به توسعه این زمینه مستعد علمی می‌کنند. لذا تلاش‌های پژوهشی در راستای بهبود کیفیت ایجاد گراف پدیداری نیز اهمیت زیادی پیدا کرده‌اند.

در این مقاله سعی شد تا با وزن‌دار کردن گراف حاصل شده از یکی از الگوریتم‌های گراف پدیداری یعنی گراف پدیداری نفوذ پذیر محدود، تا حد قابل توجهی در بهبود کیفیت گراف اثرگذار باشیم. سپس با اعمال روش بعدی که پیش‌بینی پیوند مثبت و منفی بود نیز توانستیم ابعاد کمتر بهبود یافته از طریق وزن‌دار کردن را نیز تا حدی بهبود دهیم.

در ادامه این کار پژوهشی همواره روش‌هایی برای وزن‌دار کردن انواع گراف پدیداری یافت خواهند شد که چه بسا بهبودی بیشتر و گسترده‌تر در نوع داده سری زمانی به همراه داشته باشند. همچنین اعمال دیگر روش‌های بهبود گراف، غیر مبتنی بر پیش‌بینی پیوند، همواره می‌توانند پژوهشگران زیادی را به سمت خود جذب کنند.

مراجع

[1] L. Lacasa, B. Luque, F. Ballesteros, J. Luque, and J. C. Nuño, “From time series to complex networks: The visibility graph,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 105, no. 13, Apr. 2008, doi: 10.1073/pnas.0709247105.

[2] P. F. Dai, X. Xiong, and W. X. Zhou, “Visibility graph analysis of economy policy uncertainty indices,” *Phys. A Stat. Mech. its Appl.*, vol. 531, Oct. 2019, doi: 10.1016/j.physa.2019.121748.

[3] S. Azizzadeh-Roodpish and C. H. Cramer, “Visibility Graph Analysis of Alaska Crustal and Aleutian Subduction Zone Seismicity: An Investigation of the Correlation between b Value and k–M Slope,” *Pure Appl. Geophys.*, vol. 175, no. 12, Dec. 2018, doi: 10.1007/s00024-018-1947-1.

[4] M. A. Zervou, E. Doutsi, P. Pavlidis, and P. Tsakalides, “Structural classification of proteins based on the computationally efficient recurrence quantification analysis and horizontal visibility graphs,” *bioRxiv*, 2021, doi: 10.1101/2020.10.23.350736.

[5] S. S. Lebe, M. Mulej, and R. Baggio, “Complex tourism systems: a visibility graph approach,” *Kybernetes*, vol. 43, no. 3, 2014, doi: 10.1108/K-12-2013-0266.

[6] T. T. Zhou, N. De Jin, Z. K. Gao, and Y. Bin Luo, “Limited penetrable visibility graph for establishing complex network from time series,” *Wuli Xuebao/Acta Phys. Sin.*, vol. 61, no. 3, Mar. 2012, doi: 10.7498/aps.61.030506.

[7] Q. Xuan, J. Zhou, K. Qiu, D. Xu, S. Zheng, and X. Yang, “CLPVG: Circular limited penetrable visibility graph as a new network model for time series,” *Chaos An Interdiscip. J. Nonlinear Sci.*, vol. 32, no. 1, p. 13130, 2022.

[8] L. Song, C. Xiang, H. Guo, and S. Chen, “Activity recognition using multiplex limited penetrable visibility graph,” *J. Mech. Behav. Biomed. Mater.*, vol. 124, p. 104880, 2021.

[9] Z.-K. Gao, M.-X. Liu, W.-D. Dang, C. Ma, L.-H. Hou, and X.-L. Hong, “Multilayer limited penetrable visibility graph for characterizing the gas-liquid flow behavior,” *Chem. Eng. J.*, vol. 407, p. 127229, 2021.

[10] P. Xu, R. Zhang, and Y. Deng, “A novel visibility graph transformation of time series into weighted networks,” *Chaos, Solitons and Fractals*, vol. 117. pp. 201–208, 2018, doi: 10.1016/j.chaos.2018.07.039.

[11] M. Zheng, S. Domanskyi, C. Piermarocchi, and G. I. Mias, “Visibility graph based temporal community detection with applications in biological time series,” *Sci. Rep.*, vol. 11, no. 1, Dec. 2021, doi: 10.1038/s41598-021-84838-x.

[12] A. Kumar, S. S. Singh, K. Singh, and B. Biswas, “Link prediction techniques, applications, and performance: A survey,” *Physica A: Statistical Mechanics and its Applications*, vol. 553. Sep. 2020, doi: 10.1016/j.physa.2020.124289.

[13] S. Sulaimany, M. Khansari, and A. Masoudi-Nejad, “Link prediction potentials for biological networks,” *Int. J. Data Min. Bioinform.*, vol. 20, no. 2, 2018, doi: 10.1504/IJDMB.2018.093684.

پانویس ها

1 Time series
2 Visibility Graph
3 Limited Penetrable Visibility Graphs
4 Natural Visibility Graph
5 Induced ordered weighted averaging
6 visibility graph aggregation
7 Averaging
8 Gravity
9 Link prediction
10 Supervised
11 Unsupervised
12 Common Neighbor
13 Preferential Attachment
14 Jaccard's Coefficient

بررسی تأثیر نظرات کاربران در مورد اخبار در قیمت ارز دیجیتال با استفاده از مدل زبانی برت

سهیل راهساز[۱]، زهرا زجاجی [۲]

[۱] گروه مهندسی نرم افزار، دانشکده کامپیوتر، دانشگاه اصفهان، اصفهان
soheilrahsaz.sr@eng.ui.ac.ir
[۲] استادیار گروه مهندسی نرم افزار، دانشکده کامپیوتر، دانشگاه اصفهان، اصفهان
z.zojaji@eng.ui.ac.ir

چکیده

امروزه استفاده از رمزارزها در بسیاری از کشورها فراگیر شده است و روزبه‌روز محبوبیت بیشتری کسب می‌کنند. رمزارز که در واقع یک پول دیجیتالی غیرمتمرکز است، به‌سادگی قابلیت ارسال به سراسر جهان است و این مسئله موجب استقبال سرمایه‌گذاران شده است. بازارهای مالی متعددی جهت انجام معاملات ارزهای دیجیتال در بستر بلاک‌چین ایجاد شده‌اند. در قرن اخیر نیز تلاش‌های گوناگونی برای پیش‌بینی قیمت در بازارهای مالی صورت‌گرفته است. ارزش‌گذاری بنیادین، استفاده از تحلیل تکنیکال، تحلیل آماری، هوش مصنوعی و... از جمله این روش‌ها است. مطالعات مختلفی نشان داده اند که پیش بینی قیمت ارز دیجیتال صرفا با استفاده از تاریخچه قیمت با دقت قابل قبولی قابل انجام نیست بلکه رویدادهای سیاسی اجتماعی و اقتصادی تاثیر چشمگیری در تغییرات قیمت خواهند داشت. هدف از پژوهش حاضر بهره‌گیری از یادگیری عمیق به منظور پیش‌بینی قیمت ارزهای دیجیتال بر مبنای اخبار مرتبط منتشر شده است. قدم اول این پژوهش جمع آوری و سازمان دهی مجموعه‌داده ای است که تغییرات قیمت را متناسب با اخبار ثبت می کند. به دلیل در دسترس نبودن داده برای ایجاد مدل پیش‌بینی قیمت، ابتدا داده‌ها از حدود ۷۵۰۰۰ خبر پیرامون ۴۰۰ رمزارز برتر در طول یک بازه ٤ ساله از منبع خبری مرتبط به نام کریپتوپنیک جمع‌آوری شدند که پس از پاک‌سازی ۱۲۲٦۷ خبر نهایی جهت ساخت مدل از میان آنها انتخاب شد. سپس با استفاده از تحلیل احساسات روی متن اخبار، مدلی از متن خبر به مثبت یا منفی بودن تاثیر آن در قیمت ساخته شد و در مرحله بعد با مقایسه این اخبار با قیمت واقعی روزانه ارزها، تأثیر واقعی اخبار روی قیمت بررسی گردید. در نهایت مدل برت بادقت خوبی توانست مثبت و یا منفی بودن خبر را تشخیص دهد. همچنین مشخص شد که همبستگی اخبار با تغییرات قیمت شدید حدود ۰٫۹ است ولی در موارد که تغییرات قیمت کمی رخ داده است همبستگی سوی اخبار با تغییرات قیمت حدود ۰٫٦ است.

کلمات کلیدی

پیش بینی قیمت، ارز دیجیتال، تحلیل بنیادی، مدل زبانی برت، تحلیل احساسات

۱- مقدمه

طی سال‌های اخیر، رمزارزها به یکی از مهم‌ترین مباحث مالی، اقتصادی و فناوری در دنیا تبدیل شده‌اند و کارشناسان و متخصصان متعددی بر این باورند که این ارزها، بخش جدانشدنی از آینده بشریت به‌حساب می‌آیند که به زودی شاهد نفوذ آن ها به بازارهای معتبر جهانی خواهیم بود [۱]. رمزارزها به‌صورت داده های دیجیتال مبادله شوند و نوعی پول دیجیتالی غیرمتمرکز هستند. این ارزهای دیجیتال به‌سادگی قابلیت ارسال به سراسر جهان را دارد، این مسئله سبب استقبال سرمایه‌گذاران شده است و خریدوفروش رمزارزها را به یک تجارت بسیار بزرگ مبدل کرده است. مجموع ارزش رمزارزهای دنیا بیش از ۱٫۵ تریلیون دلار است و روزانه حدود ۲۰۰ میلیارد دلار رمزارز خریدوفروش

چهارمین کنفرانس ملی انفورماتیک ایران
پژوهشگاه دانش‌های بنیادی، پردیس فرمانیه، تهران
۱۴ و ۱۵ دی ماه ۱۴۰۱

می‌شود که نشانگر جذابیت این بازار برای سرمایه‌گذاران است. رمزارزها همانند ارزهای سنتی ارزشی دارند و می‌توان آن را خریدوفروش کرد. اما در نحوه‌ی خلق و نوع استفاده، تفاوت‌های زیادی با ارزهای سنتی دارند. موفق‌ترین و گسترده‌ترین شکل پول دیجیتال تا به امروز، بیت‌کوین است. بااین‌حال تاکنون و دولت‌های ایالات متحده آمریکا، جمهوری فدرال آلمان و جمهوری خلق چین بر کالا بودن بیت‌کوین تأکید دارند و پول بودن آن را تأیید نکرده‌اند. اما بااین‌حال در چند سال گذشته، پول‌های دیجیتال مخصوصاً رمزارز به ایجاد الگوهای جدید برای معاملات اقتصادی و به‌عنوان ابزارهای جایگزین سرمایه، جذابیت‌های فراوانی را ایجاد کرده‌اند. این سیستم‌های نقدی الکترونیکی، یک فناوری کاملاً جدید را معرفی می‌کنند. زنجیره بلاکی که امنیت بسیار بالایی را برای فرایند معاملات و کنترل آنها بدون هیچ مرجع نظارتی وابسته و از طریق اشتراک اطلاعات در یک شبکه همتا به همتا ایجاد می‌کند. این بازار شامل بیش از یک هزار رمزارز است که همان فن آوری زنجیره بلاکی را استفاده می‌کنند.

در قرن اخیر نیز تلاش‌های گوناگونی برای پیش‌بینی قیمت در بازارهای مالی صورت‌گرفته است. ارزش‌گذاری بنیادین، استفاده از تحلیل تکنیکال، تحلیل آماری، هوش مصنوعی و... از جمله این روش‌ها است. در دنیای کنونی، معاملات خودکار سهم بزرگی از معاملات در دنیا را به خود اختصاص داده است. این‌گونه معاملات با پیش‌بینی بازاهای مالی قادر به ایجاد سیگنال‌های خریدوفروش می باشند. دقت پیش بینی در این کاربردها برای دستیابی به سود مناسب بسیار بااهمیت است. عدم دخالت انسان از یک سو باعث کاهش ریسک عملیاتی و دخالت احساسات در معاملات می‌شود و از سوی دیگر سرعت تصمیم‌گیری‌ها را افزایش می دهد. همچنین با گسترش بازارهای مالی و افزایش رقابت در آن، استفاده از سیستم‌های معاملاتی سنتی بسیار ناکارآمد شده است و نیاز به سیستمی که بتواند با سرعت و مقیاس وسیع‌تری به تحلیل داده‌ها و تصمیم‌گیری بپردازد بیش از پیش احساس می‌شود.

فراهم شدن داده‌های آزمایشی زیاد و قدرتمند شدن زیرساخت‌های محاسباتی در سال‌های اخیر سبب شده است تا پیشرفت زیادی در الگوریتم‌ها و تئوری‌های یادگیری ماشین ایجاد شود. در فرضیه‌ی بازارهای کارا بیان می‌شود که سرمایه‌گذاران اکثرا آگاه و منطقی بوده و نسبت به اخبار و اطلاعات جدید وارده به بازار، واکنش صحیح و منطقی نشان می‌دهند. بدین ترتیب قیمت‌های سهام نسبت به اطلاعات جدید، واکنشی سریع، کامل و بدون تورش داشته و در هر لحظه از زمان، نشان‌دهنده‌ی ارزش ذاتی و واقعی آن است [۲]. در فرضیه‌ی بازارهای کارا بیان می‌شود که سرمایه‌گذاران اکثرا آگاه و منطقی بوده و نسبت به اخبار و اطلاعات جدید وارده به بازار، واکنش صحیح و منطقی نشان می‌دهند. بدین ترتیب قیمت‌های سهام نسبت به اطلاعات جدید، واکنشی سریع، کامل و بدون تورش داشته و در هر لحظه از زمان، نشان‌دهنده‌ی ارزش ذاتی و واقعی آن است [۲]. بر اساس مطالعه‌ای که توسط موسسه تحقیقات اقتصادی آمریکا (AIER) انجام شده است، اخبار و احساسات تأثیرگذار در سطح جهانی می‌تواند نوسانات زیادی در قیمت بیت‌کوین ایجاد کند[۳].

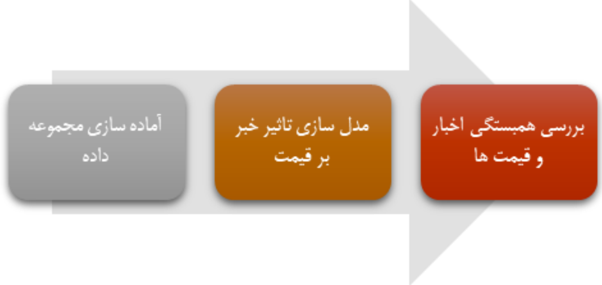
اخبار و سیگنال‌ها موجود و قابل استفاده برای پیش بینی قیمت ها به چند دسته تقسیم می‌شوند:

- سیگنال‌های خبری: سیگنال‌هایی که توسط معامله‌گرها ارائه می‌شوند که این معامله‌گرها می‌توانند حرفه‌ای و یا مبتدی باشند. این سیگنال‌ها بیشتر در شبکه‌های تلگرام و دیسکورد ارسال می‌شوند و بخش زیادی از این افراد این سیگنال‌ها را می‌فروشند و دسترسی مجانی به آن‌ها ممکن نیست. فقط بخش کوچکی از این سیگنال‌ها مجانی ارائه می‌شوند که برای این پروژه این حجم از داده کافی نیست.
- نظرات کاربران: این نظرات بیشتر در شبکه‌های ردیت و توییتر ارسال می‌شوند و صرفاً نظر شخصی کاربر در مورد یک یا چند ارز هستند.
- اخبار مربوط به ارزها: خبرهای مربوط به ارزها توسط سایت‌های خبرگزاری بزرگ منتشر می‌شوند و این اخبار بیشتر در مورد فعالیت‌های شرکت مربوط به ارزها و یا شرکت‌های مرتبط با ارزها هستند و یا حاوی اطلاعاتی در مورد سیاست‌های جدید و یا نظر افراد معروف در مورد ارزها هستند.

در این پژوهش از دسته سوم این داده‌ها، یعنی اخبار معتبر، استفاده شده است. تحقیقات معدودی بر روی پیش بینی قیمت ارزهای دیجیتال بر مبنای تحلیل احساسات کاربران درتوییتر انجام شده که می تواند به پژوهش های مانند [۴و۵و۶] اشاره کرد. اما بررسی تاثیر اخبار بر قیمت ارزهای دیجیتال زمینه بسیار جدیدی است که در پژوهش های بسیار کمی مانند [۷] به ان پرداخته شده است. در ادامه مقاله در بخش ۲ به بیان روش پیشنهادی می پردازیم . سپس در بخش ۳ نتایج به دست آمده را بیان و تحلیل می نماییم و در پایان در بخش ۴ نتیجه گیری مقاله را ارائه می نماییم.

۲- روش پیشنهادی

شکل (۱) قدم های روش پیشنهادی را نشان می دهد. در قدم اول اخبار در مورد رمزارزها از سایت های معتبر جمع آوری می شود. همچنین نظرات کمی کاربران در مورد تاثیر مثبت یا منفی هر خبر بر قیمت وبه طور کلی میزان اهمیت هر خبر ثبت می گردد. در قدم بعدی با توجه به برچسب های بدست امده برای اخبار، مدلی پیش بینی کننده بر اساس متن خبر آموزش می یابد تا بتواند با داشتن متن اخبار، تاثیر خبر را بر قیمت پیش بینی کند. در این مرحله از مدل برت به عنوان طبقه بند استفاده شده است. در قدم اخر همبستگی تاثیر اخبار بر قیمت و مقادیر واقعی قیمت بررسی و مطالعه می شود.



شکل (۱) : قدم های روش پیشنهادی

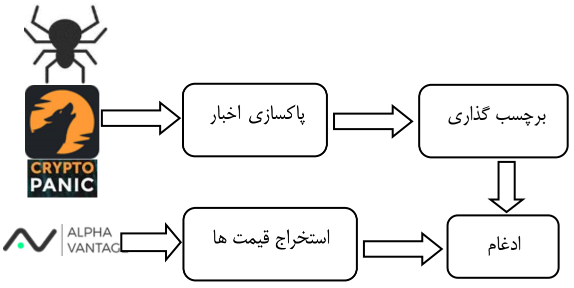
۲-۱- آماده سازی مجموعه داده

مراحل ساخت مجموعه داده در شکل زیر مشخص شده است. ابتدا باید اخبار مربوط به ارزهای دیجیتال از یک سایت معتبر جمع‌آوری شوند. پس از جمع‌آوری این داده‌ها، اخبار پاک‌سازی می‌شوند و برچسب‌گذاری می‌شوند. مراحل آماده سازی در شکل (۲) نشان داده شده است.

- **جمع آوری اخبار**

پلتفرم کریپتوپنیک یک سایت تجمیع اخبار ارز دیجیتال است که از تمامی سایت‌های خبری ارز دیجیتال، اخبار را استخراج می‌کند و به‌صورت

دسته‌بندی‌شده به کاربران نشان می‌دهد. این سایت به کاربران خود نیز قابلیت ثبت نظر و رأی را هم ارائه می‌کند و کاربران می‌توانند باتوجه‌به برداشت خودشان از خبر، به یک خبر رأی مثبت و یا منفی بدهند و همچنین میزان اهمیت خبر را نیز مشخص کنند. هر خبر می‌تواند مربوط به یک یا چند ارز دیجیتال باشد و یا مربوط به ارز خاصی نباشد و یک خبر کلی در مورد کل بازار ارز دیجیتال باشد.



شکل (۲) : آماده سازی مجموعه داده

باتوجه به جامع بودن داده‌های این سایت و همچنین رأی‌های ثبت شده روی اخبار، برای این پروژه از این پلتفرم استفاده خواهد شد. برای این کار از رابط برنامه‌نویسی کاربردی رسمی این سایت استفاده خواهد شد. رابط برنامه‌نویسی کاربردی یک واسط نرم‌افزاری است که به دو برنامه امکان برقراری ارتباط را می‌دهد. هر بار که یک صفحه از سایت بارگذاری می‌شود و یا اطلاعاتی دریافت می‌شود. خوشبختانه پلتفرم کریپتوپنیک یک رابط کاربری مجانی در اختیار برنامه‌نویسان قرار می‌دهد که برای جمع‌آوری این اخبار کافی است. سایت کریپتوپنیک متأسفانه فقط آخرین ۲۰۰ خبر را در رابط برنامه‌نویسی کاربردی خود ارسال می‌کند. در صفحه اصلی سایت، با استفاده از فیلترها و انتخاب ارز، حداکثر ۱۰۰۰ خبر را به کاربر نمایش می‌داد. به همین جهت، علاوه بر رابط برنامه‌نویسی سایت کریپتوپنیک، نیاز به یک خزشگر و پردازش اطلاعات html است. برای ۲۵ ارز اول در لیست با ۳ فیلتر مختلف، اطلاعات صفحه html ذخیره شد در نهایت با استفاده از کتابخانه Jsoup در جاوا، این اطلاعات html پردازش شد و به این صورت، اخبار تا ۳ سال قبل برای ارزها جمع‌آوری شد.

اطلاعات اخبار شامل عنوان، تاریخ و زمان خبر، آدرس خبر، منبع خبر، ارزهای دیجیتال مرتبط و رأی‌های ثبت شده کاربران برای هر خبر در پایگاه داده ذخیره می‌شود. جدول شماره ۱ آمار مربوط به اخبار به دست آمده را نمایش می دهد که مربوط به تعداد کل اخبار و خبرهای تک ارزی و چند ارزی می باشد.

جدول (۱): آمار اخبار به دست آمده

تعداد کل اخبار	۷۶۵۴۶
تعداد اخبار منحصربه‌فرد از ارز	۶۶۱
تعداد اخبار مربوط به تک ارز	۶۳۳۳۴
تعداد اخبار مربوط به چند ارز	۱۲۵۵۱

همچنین در جدول ۲ نام و تعداد اخبار مربوط به ۲۰ ارز با بیشتر تعداد خبر به دست امده آورده شده است.

جدول (۲): نام و تعداد خبر برای ۲۰ ارز اول

رتبه ارز	نام اختصاری ارز	تعداد اخبار
۱	BTC	۵۰۹۳
۲	ETH	۳۶۰۳
۵	BNB	۷۶۲
۶	ADA	۱۹۳۱
۷	XRP	۱۸۷۱
۱۰	DOGE	۱۷۵۶
۱۳	TRX	۶۷۵
۱۸	MATIC	۹۳۸
۲۰	LTC	۴۶۶

- **پاک سازی اخبار**

هدف اصلی از پاک‌سازی، حذف اخبار تکراری، اخبار غیر مهم و خبرهای مربوط به ارزهایی است که اطلاعات قیمت آن‌ها در دسترس نیست. باتوجه‌به اینکه اخبار هم از رابط برنامه‌نویسی دریافت شده بودند و هم از صفحات html پارس شده بودند، بعضی از اخبار تکراری در جدول ذخیره شده بودند که فقط تفاوت کوچکی در عنوان خبری آن‌ها بود. به‌عنوان‌مثال یک فاصله در اول و یا آخر عنوان قرار گرفته بود. با تطبیق دادن تاریخ و زمان خبرها و منبع آن‌ها، اخبار تکراری حذف شدند. باتوجه‌به اینکه اطلاعات قیمت‌ها برای تمام ارزها در سایت‌های مجانی وجود ندارد، خبرهایی که برای آن‌ها اطلاعات قیمت نیست نیز حذف شدند. سپس خبرهایی که هیچ رأی مثبت و یا منفی ای ندارند از جدول حذف شدند. باتوجه‌به اینکه استیبل کوین‌ها تغییرات قیمت فوق‌العاده ناچیزی دارند، نیازی به بررسی تغییرات قیمت این ارزها وجود ندارد. تمامی اطلاعات قیمت این استیبل کوین‌ها از جدول حذف شد تا در مراحل بعدی مورد بررسی قرار نگیرند. در مرحله بعد اخبار قدیمی تر از داده‌های قیمت روز هر ارز، حذف شدند. در نهایت ۱۲۲۶۷ خبر در جدول باقی ماند.

- **برچسب گذاری**

در این مرحله، یک امتیاز بین ‎-۱ تا ۱ به هر خبر داده می‌شود. امتیاز مثبت به معنای تأثیر مثبت روی قیمت ارز و امتیاز منفی به معنای تأثیر منفی روی قیمت ارز است. محاسبه این امتیاز بر اساس رأی‌های ثبت شده کاربران در سایت مطابق رابطه (۱) انجام می‌شود.

$$\frac{\text{تعداد رأی مثبت} - \text{تعداد رأی منفی}}{\text{مجموع رأی ها}} \qquad (۱)$$

با این روش، میزان مثبت بودن یا منفی بودن خبر نیز مشخص می‌شود. هرچه قدر مطلق امتیاز به ۱ نزدیک‌تر باشد، به این معناست که کاربران اتفاق‌نظر بیشتری روی مثبت بودن یا منفی بودن خبر داشتند.

- **استخراج قیمت ها**

برای پیداکردن تأثیر مثبت و منفی بودن اخبار روی قیمت، به تاریخچه قیمت ارزها تا ۲ سال پیش‌نیاز است. داده قیمت‌های ارزها به‌صورت نمودار شمعی در بازه‌های ۱ دقیقه، ۵ دقیقه، ۱۵ دقیقه، ۳۰ دقیقه، ۶۰ دقیقه و روزانه و هفتگی و ماهانه موجود است. اما در طرح‌های مجانی سایت‌ها، فقط ۱۰۰۰ رکورد داده ارسال می‌شود و برای دسترسی به اطلاعات قدیمی، بهترین انتخاب قیمت‌های روزانه است.

منابع زیادی برای دریافت این داده‌ها وجود دارد. برای این پروژه از سایت alphavantage.co استفاده شد. برای استفاده از رابط برنامه‌نویسی این سایت، به یک حساب در این سایت نیاز است. داده های به دست آمده حاوی ۱۰۰۰ رکورد قیمت از تاریخ روز تا ۱۰۰۰ روز قبل است. اطلاعات هر روز در جدول (۳) آمده است.

جدول (۳) اطلاعات استخراج شده برای قیمت

نام پارامتر	شرح
open	قیمت در اول روز در ساعت ۰۰:۰۰
high	بالاترین قیمت در روز
Low	پایین‌ترین قیمت در روز
close	آخرین قیمت در روز در ساعت ۲۳:۵۹:۵۹
volume	تعداد ارز معامله شده در روز

به‌ازای تمامی ارزها، این لیست قیمت‌ها در مجموعه داده ذخیره شد. اما برای بخشی از ارزها، داده‌های قیمت به‌صورت مجانی در این سایت وجود نداشت.

۲-۲- مدل سازی تاثیر خبر

باتوجه‌به اینکه اخبار ارزهای دیجیتال بر روی عملکرد کاربران تأثیر می‌گذارد و در نهایت باعث می‌شود که آن‌ها بخرند یا بفروشند، می‌توان یک مدل از متن خبر به مثبت یا منفی بودن خبر ایجاد کرد تا در نهایت در پیش‌بینی تغییر قیمت ارزها به ما کمک کند.

به معنای تشخیص و دسته‌بندی نظراتی است که در قالب متن بیان شده‌اند. این عملیات مشابه تحلیل احساسات[1] است که با استفاده از پردازش زبان طبیعی انجام می‌گیرد تا در نهایت مشخص شود که موضعگیری یک خبر در مورد یک ارز به‌خصوص، مثبت یا منفی است. در این مقاله از مدل زبانی برت[2] [8] به منظور تحلیل احساسات اخبار مربوط به ارزهای دیجیتال استفاده شده است. برت یک مدل متن‌باز است که برای پردازش زبان طبیعی به کار گرفته می‌شود. این مدل توسط گوگل در سال ۲۰۱۹ معرفی شده است و همواره به‌روزرسانی می‌شود. این الگوریتم که بر پایه یادگیری عمیق است، مدلی از زبان محاوره و روزمره ارائه می دهد که کمک می‌کند تا حتی کوچک‌ترین بخش‌های زبان محاوره مثل حروف اضافه را مانند یک انسان به‌خوبی درک کند.

ابتدا بخشی از اخبار برای آموزش انتخاب می‌شود. برای این منظور، خبرهایی انتخاب می‌شوند که شرایط زیر را داشته باشند:

- قدر مطلق امتیاز آن از ۰،۳ بیشتر باشد.
- حداقل ۴ رأی مثبت و یا منفی و یا مهم داشته باشد.
- حداقل یک رأی مهم داشته باشد.

از ۸۴۶۰ خبر با این ویژگی، ۵۰٪ آن به‌صورت تصادفی جدا شده و در یک لیست قرار می‌گیرند. برای برچسب هر خبر، اگر امتیاز خبر بزرگ‌تر از ۰ بود، عدد ۱ به معنای مثبت بودن خبر و اگر امتیاز خبر کوچک‌تر از ۰ بود، عدد ۰ به معنای منفی بودن خبر اختصاص داده می‌شود.

از اخبار انتخاب شده برای آموزش مدل از ۸۰٪ آن استفاده می‌شود و از ۲۰٪ دیگر، برای تست مدل استفاده می‌شود. برای تست مدل از کتابخانه sklearn در پایتون استفاده می‌شود.

- **تنظیم پارامترها**

مقادیر بهینه پارامترها که با آزمایش و خطا به دست امده در جدول ۴ آمده است.

جدول (٤) مقادیر پارامترهای مدل‌سازی

پارامتر	مقدار
حداکثر طول توکن	۶۴
تعداد کلاس ها	۲
درصد اعتبار سنجی	۲۰
تعداد تناوب	۴
نرخ یادگیری	۲e-۵
اندازه batch آموزش	۳۲
اندازه batch اعتبارسنجی	۶۴

حداکثر طول توکن‌ها برابر با ۶۴ قرار می‌گیرد. چرا که بیشتر از این تعداد نیز در بین خبرها وجود ندارد. تعداد کلاس ها برابر با ۲ قرار می‌گیرد که معادل برچسب مثبت و منفی است. مقدار پارامتر درصد اعتبارسنجی مشخص می‌کند چه حجمی از داده برای اعتبار سنجی مدل و جلوگیری از بیش برازش مدل استفاده شود. تعداد تناوب برابر است با تعداد دفعاتی که آموزش مدل تکرار می‌شود. اندازه batch بر اساس مقدار حداکثر طول تنظیم می‌شود و پیشنهاد می‌شود این مقدار برای آموزش روی ۳۲ قرار بگیرد. پارامتر نرخ یادگیری در شبکه‌های عصبی استفاده می‌شود که مقدار معمولاً بین ۰ و ۱ است. این پارامتر سرعت تطبیق مدل به مسئله را کنترل می‌کند. در اینجا مقدار 2e-5 یا 0.00002 تنظیم شده است.

۲-۳- تحلیل همبستگی اخبار و قیمت ها

در این مرحله همبستگی مثبت یا منفی بودن اخبار با مقادیر واقعی قیمت بدست می آید و برای سناریوهای مختلف تحلیل می گردد.
برای بررسی تأثیر اخبار بر روی قیمت‌ها در روز، برای هر ارز در هر روز اخبار تجمیع می‌شوند. اخبار تاثیر گذار مطابق آنچه در قسمت ۲-۲ گفته شد انتخاب شدند. از آن جایی که ممکن است در یک روز خبرهای متعددی راجع به یک ارز یکسان منتشر شده باشد، برای به دست اوردن موضع گیری کلی اخبار راجع به یک ارز در هر مقطع زمانی نیاز است که اخبار تجمیع شوند. تجمیع اخبار از طریق رابطه (۲) انجام می شود .

$$\frac{\sum \text{وزن خبر} * \text{امتیاز خبر}}{\sum \text{وزن خبر}} \qquad (۲)$$

وزن خبر در این قسمت تعداد رأی‌های مهم در نظر گرفته می‌شود. به‌خاطر اینکه اخباری که مهم‌تر هستند، تأثیر بیشتری در امتیاز خبر داشته باشند. به‌عنوان‌مثال یک خبر منفی که ۵ رأی مهم دارد تأثیر کمتری در مقابل یک خبر مثبت با ۱۰۰ رأی مهم دارد.

در مرحله بعد، داده مربوط به تغییر قیمت نیز در مقابل این جدول قرار می‌گیرد. در هر روز قیمت یک ارز یا صعودی و یا نزولی است. برای صعودی عدد ۱ و برای نزولی عدد ۱- به‌عنوان تغییر قیمت در روز در نظر گرفته می‌شود. سپس ضریب همبستگی امتیاز اخبار در روز با تغییر قیمت در روز محاسبه می شود.

۳- نتایج تجربی

۱-۳- معیارهای ارزیابی

معیارهای زیر برای ارزیابی قسمت مدل سازی تاثیر خبر بر قیمت از روش پشنهادی مورد استفاده قرار گرفتند:

- Accuracy: تعداد اخباری که تاثیر مثبت یا منفی ان ها درست پیش‌بینی شده است تقسیم بر کل اخبار. بالا بودن این مقدار به معنای این است که حجم بیشتری از پیش‌بینی‌ها درست بوده است.
- Recall: تعداد اخبار مثبت درست پیش‌بینی شده تقسیم بر کل اخبار مثبت موجود. هرچه این مقدار بیشتر باشد، قدرت مدل برای پیش‌بینی موارد مثبت بیشتر است.
- Precision: تعداد اخبار مثبت درست پیش‌بینی شده تقسیم بر کل اخباری که مثبت پیش‌بینی شده اند. بالا بودن این معیار، ارزش موارد پیش‌بینی‌شده مثبت مدل را نشان می‌دهد و هرچه این مقدار بیشتر باشد، موارد مثبت پیش‌بینی‌شده ارزش بیشتری دارند.
- F1 score: ترکیبی از مقدار Recall و Precision که از رابطه (۳) محاسبه می‌شود.

$$F1\ score = \frac{Precision\ \times Recall}{Precision + Recall} \quad (۳)$$

معمولاً در مدل‌ها با بالابردن Recall، Precision کاهش میابد و برعکس. به همین علت نیاز به معیاری است که بتواند دو مدل را به طور کلی باهم مقایسه کند و این معیار F1 Score است. اگر مقصود از پژوهش تنها استفاده از این مدل برای سیگنال‌دهی خرید ارز باشد، معیار precision یا دقت ارجح‌تر است. به این خاطر که اگر از بین اخبار مثبت، تعداد بیشتری درست پیش‌بینی شود، در نتیجه معامله‌گر خرید بهتری خواهد داشت.

همچنین برای تحلیل تاثیر اخبار بر قیمت ها از معیار ضریب همبستگی استفاده شده که در رابطه (۴) آمده است.

$$r = \frac{\sum(x-m_x)(y-m_y)}{\sqrt{\sum(x-m_x)^2\sum(y-m_y)^2}} \quad (۴)$$

که در این رابطه x متغیر تصادفی نشان دهنده تاثیر پیش بینی شده برای خبر، y تغییر قیمت واقعی رمزارز و m_x و m_y به ترتیب میانگین تاثیر پیش بینی شده و قیمت واقعی می باشند. مقدار ضریب همبستگی بین ۱- و ۱ متغیر است.

۲-۳- کارایی مدل پیش بینی کننده تاثیر خبر

نتایج به دست امده از مدل سازی تحلیل تاثیر خبر در جدول زیر امده است:

جدول (۵): نتایج تست مدل برت

Accuracy	۰٫۸۴۲
F1 score	۰٫۹۰۲
Recall	۰٫۹۳۲
Precision	۰٫۸۷۵

همانطور که از جدول (۵) بر می آید مدل برت کارایی بسیار خوبی بر پیش بینی تاثیر مثبت یا منفی ارز داشته است و توانسته به F1-score معادل ۰٫۹ برسد.

۳-۳- بررسی تاثیر اخبار بر قیمت رمزارز

در این بخش سناریوهای مختلفی برای بررسی تاثیر اخبار بر قیمت مرزارز طراحی شده است که در ادامه به تفصیل بیان می شوند.

- **سناریو ۱**

در این سناریو همبستگی اخبار تاثیر گذار که در بخش ۲-۳ معرفی شدند با تغییرات قیمت واقعی رمزارز بدست امده است. در این قسمت تعداد ۸۸۰۸ خبر که هر یک مختص به یک ارز و یک روز است، در مجموعه داده وجود دارد و مقدار همبستگی با تغییرات قیمت ۰٫۵۴ به دست آمد.

- **سناریو ۲**

در این سناریو تاثیر اخبار بعد از ساعت ۲۲:۳۰ برای تاریخ روز بعد محاسبه می شود. باتوجه‌به اینکه بعد از انتشار اخبار یک بازه زمانی طول می‌کشد تا به گوش کاربران برسد و در قیمت تأثیر بگذارد، احتمالاً اخباری که از ساعت ۲۲:۳۰ به بعد منتشر می‌شوند، تأثیر خود را در روز بعد می‌گذارند. بعد از انجام این تغییر، ضریب همبستگی به مقدار ۰٫۵۶۵۸ افزایش می‌یابد.

- **سناریو ۳**

باتوجه‌به اینکه تغییرات زیر ۱ درصد تقریباً برابر با ثابت ماندن قیمت یکسان است، این داده‌ها را از جدول حذف می‌کنیم. با این تغییر، ضریب همبستگی به مقدار ۰٫۵۷ افزایش می‌یابد.

- **سناریو ٤**

تقریباً ۸۰٪ اخباری که در مورد ارزهای دیجیتال منتشر می‌شوند، اخبار مثبت هستند. به این دلیل که عمده این اخبار توسط خود شرکت‌های این ارزها منتشر می‌شوند و کماکان همه سعی دارند تا جو مثبتی را برای این بازار حفظ کنند و این اخبار منفی هستند که تأثیر خیلی بیشتری از اخبار مثبت دارند. به‌نحوی‌که دیدن یک خبر منفی در روز، می‌تواند کاربران زیادی را بترساند و این کار باعث کاهش قیمت ارز در روز شود.

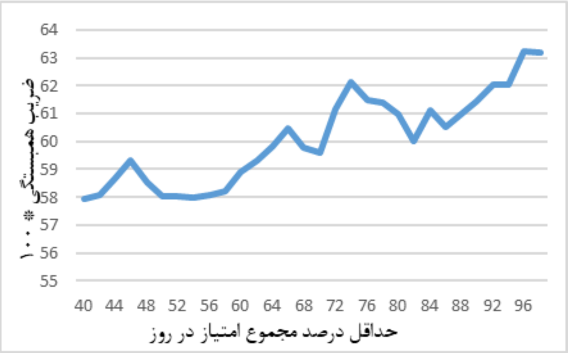
با حفظ ۱ درصد برای حداقل درصد تغییر قیمت در روز، در فرمول تجمیع امتیاز اخبار یک تغییر ایجاد می‌کنیم و برای اخبار منفی، علاوه بر وزن مهم

بودن، وزن ۵ را در نظر می‌گیریم. به معنای اینکه اخبار منفی ۵ برابر بیشتر از اخبار مثبت تأثیر دارند و یک خبر منفی به‌تنهایی می‌تواند بازار را تغییر دهد. با این تغییر ضریب همبستگی به ۰٫۶ افزایش پیدا می‌کند.

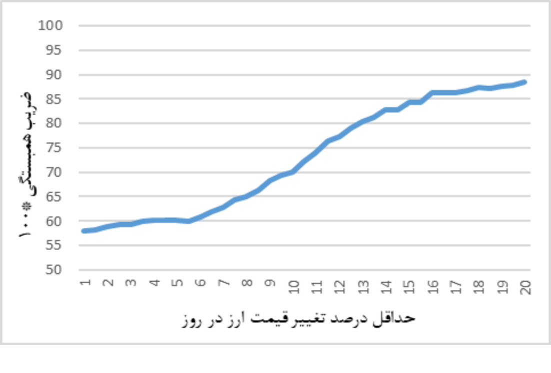
- **سناریو ۵**

در روزهایی که اخبار ضدونقیض وجود دارد و یا امتیاز خبرها به‌اندازه کافی بالا نیست، می‌توان گفت که برداشت واضحی نمی‌توان از اخبار داشت و اخبار به‌اندازه کافی برای تغییر قیمت در روز قوی نیستند. بنابرین اخباری که امتیازی زیر ۰٫۴ دارند جهت گیری قوی راجع به قیمت ندارند و از داده ها حذف می شوند. با انجام این تغییر، ضریب همبستگی به عدد ۰٫۵۸ می رسد. با افزایش این پارامتر از ۴۰ درصد به ۹۸ درصد، ضریب همبستگی نیز تا ۰٫۶۳ افزایش پیدا می‌کند که در شکل (۳) آمده است.

همچنین با ثابت نگه‌داشتن مقدار حداقل امتیاز در روز روی ۰٫۴ و افزایش حداقل درصد تغییریافته قیمت ارز در روز، ضریب همبستگی تا ۰٫۹ افزایش پیدا می‌کند. باتوجه‌به افزایش دقت نسبت به حداقل تغییر قیمت ارز در روز، برداشت می‌شود که در روزهایی که تغییرات قیمت شدیدتر بوده است، اخبار مثبت یا منفی هم راستا با تغییر قیمت حتماً وجود داشته است. نتیجه این تغییرات در نمودار شکل (۴) مشهود است.



شکل (۳): نمودار تغییر همبستگی نسبت به حداقل امتیاز اخبار در روز



شکل(٤): نمودار تغییر همبستگی نسبت به حداقل درصد تغییر قیمت در روز

٤- نتیجه گیری

در این پژوهش، هدف اصلی جمع‌آوری یک مجموعه‌داده از اخبار ارزهای دیجیتال است که با تکنیک‌های خزش و استفاده از رابط برنامه‌نویسی، با موفقیت نزدیک به ۱۳ هزار خبر پاک‌سازی‌شده معتبر جمع‌آوری شد. پس از جمع‌آوری اخبار و برچسب‌گذاری آن‌ها، مدلی آموزش داده شد که برای تشخیص موضع گیری اخبار و تاثیر آن بر قیمت مورد استفاده قرار می گیرد. سپس با استفاده از این مدل، تأثیر مثبت و منفی بودن اخبار روی قیمت واقعی رمزارز بررسی شد. در نهایت مشاهده شد که در صورتی که اخباری که جزو اخبار کلان اقتصادی باشند و تغییر قیمت شدید باشد همبستگی خبر و تغییر قیمت تا ۰٫۹ می رسد. بااین‌حال، بین بعضی از ویژگی‌های اخبار و نتیجه نهایی رابطه مستقیم دیده می‌شود: (۱) هر خبر زمان می برد تا تاثیرش را در بازار بگذارد پس اخباری که آخر شب منتشر می شوند، تاثیر خود را روز بعد می گذارند. (۲) اخبار مهم‌تر تأثیر بیشتری در جابه‌جایی قیمت ارزها دارند. (۳) اخبار منفی نسبت به اخبار مثبت، تأثیر بیشتری دارند.

مراجع

[۱] اسماعیلی، و، (۱۴۰۰)، ارائه سیستم معاملات الگوریتمی برای رمزارزها بر پایه یادگیری ماشین، پایان‌نامه کارشناسی ارشد مهندسی صنایع گرایش سیستم‌های مالی، دانشگاه مهندسی صنایع و سیستم‌ها گروه مدیریت سیستم و بهره‌وری، دانشگاه تربیت‌مدرس، تهران، ایران.

[۲] مقدس، م (۱۳۹۵) بررسی ارتباط اخبار با قیمت سهام بانک‌های خصوصی و دولتی خصوصی شده در ایران، پایان‌نامه کارشناسی ارشد اقتصاد نظری، دانشکده علوم اقتصادی و اداری، دانشگاه فردوسی، مشهد، ایران.

[3] Gulker, M. Bitcoin’s largest Price Changes Coincide with Major News Events about the Cryptocurrency. Available online: https: //www.aier.org/article/bitcoins-largest-price-changes-coincide-with-major-news-events-about-the-cryptocurrency/

[4] Ye, Zi, Yinxu Wu, Hui Chen, Yi Pan, and Qingshan Jiang. "A Stacking Ensemble Deep Learning Model for Bitcoin Price Prediction Using Twitter Comments on Bitcoin." *Mathematics* 10, no. 8 (2022): 1307.

[5] Pant, D.R.; Neupane, P.; Poudel, A.; Pokhrel, A.K.; Lama, B.K. Recurrent neural network based bitcoin price prediction by Twitter sentiment analysis. In Proceedings of the 2018 IEEE 3rd International Conference on Computing, Communication and Security (ICCCS), Kathmandu, Nepal, 25–27 October 2018; Volume 1, pp. 128–132.

[6] Li, T.R.; Chamrajnagar, A.S.; Fong, X.R.; Rizik, N.R.; Fu, F. Sentiment-based prediction of alternative cryptocurrency price fluctuations using gradient boosting tree model. Front. Phys. 2019, 7, 98.

[7] Farimani, Saeede Anbaee, Majid Vafaei Jahan, Amin Milani Fard, and Seyed Reza Kamel Tabbakh. "Investigating the informativeness of technical indicators and news sentiment in financial market price prediction." *Knowledge-Based Systems* 247 (2022): 108742.

[8] Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. "Bert: Pre-training of deep bidirectional transformers for language understanding." *arXiv preprint arXiv:1810.04805* (2018).

پانویس ها

[1] Sentiment Analysis

[2] BERT, Bidirectional Encoder Representations from Transformers

چهارمین کنفرانس ملی انفورماتیک ایران
پژوهشگاه دانش‌های بنیادی، پردیس فرمانیه، تهران
۱۴ و ۱۵ دی ماه ۱۴۰۱

بهبود عملکرد محاسباتی در تبدیل موجک گسسته با استفاده از سیستم اعداد مانده‌ای

زهراسلطانی[۱] ، محمد صالحی[۲]، حمیدرضا احمدی‌فر[۳]
[۱] دانشجوی کارشناسی ارشد،گروه مهندسی کامپیوتر، دانشکده فنی، دانشگاه گیلان، رشت، zahra.soltani75brn@gmail.com
[۲] استادیارگروه مهندسی کامپیوتر، دانشکده فنی، دانشگاه گیلان، رشت، edu.salehi@gmail.com
[۳] استادیارگروه مهندسی کامپیوتر، دانشکده فنی، دانشگاه گیلان، رشت، ahmadifar@guilan.ac.ir

چکیده

یکی از روش‌های به‌کاررفته برای بهبود عملکرد محاسباتی در سیستم‌های خاص منظوره، استفاده از سیستم اعداد مانده‌ای (Residue Number System) است. در عملیات تبدیل موجک گسسته از محاسبات پیمانه‌ای برای افزایش سرعت استفاده شده است. در این مقاله انتخاب پیمانه‌های مناسب در پیاده‌سازی سخت‌افزاری عملیات تبدیل موجک گسسته، منجر به بهبود سرعت مدار گردید. نتایج سنتز مدار پیشنهادی در FPGA نشان می‌دهد که میزان تاخیر، در مقایسه با روش ارائه شده در روش قبلی ۵۴/۷درصد بهبود یافته است و در مقابل میزان مساحت (LUT) و توان مصرفی به ترتیب ۶۵/۵ و ۴۶/۸درصد افزایش داشته است.

کلمات کلیدی

تبدیل موجک گسسته، سیستم اعداد مانده‌ای، محاسبات پیمانه‌ای، بهبود عملکرد محاسباتی.

۱- مقدمه

تبدیل موجک یکی از روش‌های پرکاربرد در عملیات پردازش سیگنال مانند فیلترها، فشرده‌سازی، رمزنگاری و کدگذاری است [1]. پیاده‌سازی آن معمولاً با استفاده از ساختارهای بانک فیلتر پاسخ ضربه محدود (FIR)[۱] انجام می شود. فیلترهای FIR در مقایسه با فیلترهای پاسخ ضربه نامحدود (IIR)[۲] به توان بیشتری برای عملیات محاسباتی نیاز دارند که باعث می‌شود در برخی از کاربردها، سربار محاسباتی و توان مصرفی بالایی داشته باشد[۲]. از این رو، بسیاری از پردازنده‌های سیگنال دیجیتال واحدهای سخت‌افزاری ویژه‌ای را برای ساخت فیلترهای FIR کارآمد ارائه داده‌اند. روش دیگر برای بهینه‌سازی فیلترهای موجود، استفاده از سیستم عددی مانده‌ای است. در این سیستم که انتشار رقم نقلی فقط در کانال‌های مجزا رخ می‌دهد، محاسبات جمع، تفریق و یا ضرب بصورت موازی و با پهنای بیتی کمتری در مقایسه با محاسبات دودویی انجام می‌شود. بنابراین امکان افزایش سرعت و کاهش توان مصرفی مهیا می‌گردد.

در [3] نویسندگان یک ساختار ترکیبی از سیستم عددی مانده‌ای و موجک Finite-field ارائه داده‌اند که صرفا ساختار کلی مدار به کمک سیستم اعداد مانده‌ای پیشنهاد شده ولی به مجموعه پیمانه‌ای مشخصی اشاره نشده است. با این حال با ذکر یک مثال عددی نشان داده شده که این سیستم می‌تواند در پردازش تصویر به کار رود. در [4] ازمجموعه سه پیمانه‌ای $\{2^n-1, 2^n, 2^n+1\}$ و $n = 3$ برای افزایش سرعت محاسبات بانک‌های فیلتر CDF97 در تبدیل موجک گسسته استفاده شد. نتایج سنتز آن در FPGA مدل xc6vlx75t نشان می‌دهد که سرعت انجام محاسبات در مقایسه با روش دودویی، ۲۸٪ بهبود یافته است. در [5] همین مجموعه سه‌پیمانه‌ای را برای نشان‌دادن افزایش کارایی محاسبات تبدیل موجک گسسته به‌کارگرفتند و در سنتزآن در ASIC نشان دادند که سرعت انجام محاسبات ۷٪ و مساحت آن ۳/۵در صد بهتر از روش دودویی است. در [6] نویسندگان نشان دادند که استفاده از سیستم عددی مانده‌ای در پیاده‌سازی فیلترهای لبه، باعث بهبود ۴۹درصدی در کارایی و ۳۱درصدی در توان مصرفی می‌گردد. در [7] نویسندگان مجموعه سه‌پیمانه‌ای $\{2^{n-1}-1, 2^n, 2^n-1\}$ را به‌کاربردند تا تاثیر استفاده از سیستم عددی مانده‌ای در محاسبات فیلترهای تصویر خاکستری را نشان دهند. در [8] از سیستم عددی مانده‌ای برای پیاده سازی فیلتر تبدیل موجک گسسته برای پردازش تصاویر سه بعدی پزشکی استفاده شده است که در مقاله مدارهای محاسبات پیمانه‌ای دیده نمی‌شود و صرفا به فرم کلی پیمانه‌ها که بصورت 2^n-1 بوده، اشاره شده است. در این مقاله از مجموعه سه‌پیمانه‌ای $\{2^{n+1}-1, 2^n, 2^n-1\}$در پیاده‌سازی بانک فیلتر CDF97 تبدیل موجک گسسته استفاده شده است. این مجموعه در مقایسه با مجموعه سه پیمانه‌ای به‌کاررفته در [4 و 5] فقط در پیمانه‌ی 2^n+1 تفاوت دارد. محاسبات در پیمانه‌های به فرم $2^n-\delta$ همواره ساده‌تر از پیمانه‌های به فرم $2^n+\delta$ است.

در ادامه مقاله و در بخش دوم تبدیل موجک و کاربردهای آن معرفی می‌شود. در بخش سوم سیستم عددی مانده‌ای و نحوه‌ی انجام محاسبات در آن شرح داده خواهد شد. در بخش چهارم روش پیشنهادی همراه با مدارهای پیمانه‌ای طراحی شده برای آن ارائه شده و در ادامه نتایج سنتز آن در FPGA نشان داده می‌شود و در بخش پنجم خلاصه و نتیجه‌گیری مقاله ارائه می‌گردد.

[۱] Finite Impulse Response

[۲] Infinite Impulse Response

۲- تبدیل موجک

هر تابع متناوب را می‌توان برحسب مجموع نامتناهی از توابع پایه سینوسی و کسینوسی (و یا تابع‌نمایی متناوب مختلط) نوشت. سال‌ها بعد از کشف این خاصیت شگفت‌انگیز توابع متناوب، این ایده تحت عنوان تبدیل فوریه به سایر توابع نیز تعمیم داده شد. پس از این تعمیم بود که تبدیل فوریه به‌عنوان ابزاری کارآمد در محاسبات کامپیوتری وارد گردید[1]. تبدیل فوریه، کاربرد بسیار زیادی در ریاضیات، مهندسی و علم فیزیک دارد. یکی از شاخه‌های تبدیل فوریه، تبدیل فوریه گسسته است که به‌صورت اختصاری با نماد DFT نمایش داده می‌شود[1]. ضعف تبدیل فوریه در مشخص‌سازی موقعیت زمانی موجود در سیگنال است. این کار در تبدیل موجک امکانپذیر است.

در تبدیل موجک (Wavelet)، یک سیگنال به سیگنال‌های مقیاس شده از موجک اصلی مادر تجزیه می‌شوند. واژه موجک به معنای موج کوچک است که در برخی ترجمه‌ها، تعبیر موجک برای آن آورده شده است. دلیل استفاده از واژه کوچک، محدود بودن و کوتاه بودن تابع پنجره می‌باشد. علت استفاده از واژه موج نیز به دلیل ماهیت نوسانی این تابع است. واژه مادر نیز به این منظور به کاربرده می‌شود که تمامی نسخه‌های انتقال‌یافته و مقیاس شده، همگی ازروی یک تابع اولیه به دست می‌آیند که اصطلاحاً موجک مادر نامیده می‌شود[1].

به طورکلی تبدیل موجک به دو دسته تبدیل موجک پیوسته و تبدیل موجک گسسته تقسیم می شود که در این مقاله از تبدیل موجک گسسته استفاده می‌شود. تبدیل موجک گسسته نسبت به نوع پیوسته آن نیاز به حجم محاسباتی کمتری دارد و بنابراین زمان پردازش سیگنال راکاهش می‌دهد . بعلاوه تعداد ضرایب تبدیل موجک گسسته نسبت به انبوه ضرایب حاصل از تبدیل موجک پیوسته کمتر بوده و تحلیل رفتار این ضرایب در شرایط مختلف پدیده‌های گذرا، آسان‌ترمی‌شود[9].

تبدیل موجک ابزار قدرتمندی ا ست که برای تحلیل سیگنال‌های گذرا کاربرد داشته و قادر به بیان ویژگی‌های سیگنال در حوزه زمان و فرکانس است. توابع مقیاس و مادر تبدیل موجک به صورت زیر تعریف می‌شوند[1].

$$\psi(t) = \sqrt{2}\, g_n\, \psi\, [2t-n] \qquad (1)$$

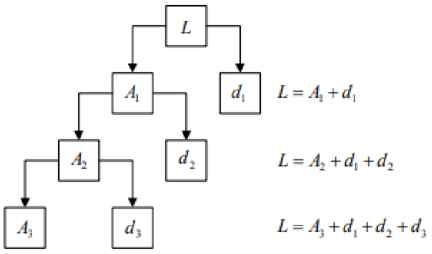
$$\varphi(t) = \sqrt{2}\sum_{n} h_n\, \varphi[2t-n] \qquad (2)$$

$\varphi(t)$ تابع مقیاس و $\psi(t)$ تابع موجک پایه (موجک مادر) برای یک سری از ضرایب موجک به صورت $\{h_a\}\ = H$ و$G = \{g_n\}$ هستند که در آن $g_n =$ $(-1)^n h_{1-n}$است (G معکوس H است). هر تابع مقیاس و پایه، یک سری ضرایب فیلترهایی با پاسخ ضربه محدود فرکانس بالا و فرکانس پایین هستند. این ضرایب برای ساختن یک‌سری توابع به صورت$\varphi_{j.k}(t)$ و $\psi_{j.k}(t)$ و که بر اساس توابع موجک پایه ساخته می شوند، به کار می‌روند. این توابع به ترتیب توابع مقیاس‌دهی شده و انتقال یافته توابع$\varphi(t)$و$\psi(t)$ هستند و به صورت زیر تعریف می‌شوند:

$$\psi_{j.k}(t) = 2^{1/2} \sum_{k} d_{j.k}\, \psi\, [2^j t - k] \qquad (3)$$

$$\varphi_{j.k}(t) = 2^{1/2} \sum_{k} c_{j.k}\, \varphi[2^j t - k] \qquad (4)$$

اگر در این روابط k و j اعداد صحیح بوده، توابع پایه توسط فاکتور 2^j مقیاس‌دهی و با واحدهای زمانی k انتقال داده می‌شود.
تبدیل موجک گسسته ، طی سطوح متوالی، سیگنال را به چندین مؤلفه فرکانسی تجزیه می‌کند. مطابق شکل (۱)، در سطح اول، سیگنال به دو مؤلفه فرکانس پایین و فرکانس بالا تجزیه می‌شود که در آن L سیگنال اصلی، A_i تقریب و d_i جزئیات است[10].



شکل (۱): نمایش درختی تجزیه موج گسسته[10]

حجم زیاد محاسبات جمع و ضرب در تبدیل موجک در برخی کاربردها می‌تواند باعث کاهش کارایی و افزایش زمان پاسخ سیستم گردد. بسیاری از پردازنده های سیگنال دیجیتال، مدارهای سخت افزاری ویژه ای را برای ساخت فیلترهای FIR کارآمد ارائه می‌دهند. روش دیگر برای بهینه‌سازی فیلترهای موجود، بسط بلوک‌های حسابی با استفاده از سیستم‌های اعداد مانده-ای(Residue Number System(RNSاست[12] که در ادامه شرح داده خواهد شد.

۳- سیستم اعداد مانده‌ایی

یکی از روش‌های پرکاربرد در تسریع عمل‌های حسابی جمع، تفریق وضرب به ویژه در پردازنده‌های خاص منظوره، بهره‌گیری از سیستم اعداد مانده‌ای است. محاسبات در چند مسیر موازی، متناظر با پیمانه‌های RNS، انجام می‌شوند. عملوندهای دودویی، به چند عملوندکوچک‌تر متناظر با کانال‌های محاسباتی تبدیل می‌شوند و انجام عملیات حسابی روی دو عدد بزرگ به همان عملیات با اعدادی به مراتب کوچک‌تر تبدیل می‌شود. پس از انجام عملیات حسابی، نتایج با تبدیل معکوس به نظام عددی دودویی برگردانده می‌شود. عمل تبدیل از دودویی به مانده‌ای (مبدل مستقیم) و عملیات حسابی در کانال‌ها، موازی و سریع انجام می‌شوند. در این روش می‌توان سرعت محاسبات را به نحو چشمگیری افزایش داد. البته تنها مزیت این نظام، افزایش سرعت نیست بلکه مزایای دیگری مانند توان مصرفی پایین و کاهش مساحت تراشه را نیز می‌توان نام برد[12,13].

پایه و ا ساس سیستم اعداد مانده‌ای تعدادی عدد دوبه‌دو نسبت به هم اول است که مجموعه پیمانه سیستم را تشکیل می‌دهند و به‌صورت $\{m_1, m_2, \dots m_n\}$ نمایش داده می‌شوند. محدوده نمایش سیستم اعداد مانده‌ای که با DR (Dynamic Range) نشان داده می‌شود، بیانگر محدوده اعدادی ا ست که قابل نمایش در سیستم هستند و قابلیت ورود به سیستم توسط مبدل مستقیم را دارند. با فرض اول‌بودن پیمانه‌ها نسبت به هم، حاصل‌ضرب آن‌ها $(M = m_1 \times m_2 \times \dots \times m_n)$ نشان‌دهنده‌ی DR است یعنی بازه‌ی (0, M]. همچنین مبتنی براین تعریف، خروجی مبدل معکوس نیز در این محدوده قرار می‌گیرد. محاسبات در کانال‌های موازی منجر به بالارفتن

سرعت و کاهش توان مصرفی می شود. رابطه (5) ساختار عمومی عملیات مانده‌ای را نشان می‌دهد که عملگر ■ می‌تواند هر یک از عمل‌های جمع، تفریق و ضرب باشد[12].

$$x■y = (|x_{k-1}■y_{k-1}|_{m_{k-1}} \dots |x_0■y_0|_{m_0}) \qquad (5)$$

۳-۱- جمع و تفریق مانده‌ایی

در محاسبات پیمانه‌ای بسته به این‌که از چه پیمانه‌هایی استفاده شود، ساختار انجام عملیات تغییر می‌کند. در حالت کلی می‌توان پیمانه‌ها را به سه فرم درنظرگرفت که عبارتند از 2^n، $2^n-\delta$ و $2^n+\delta$. محاسبات در پیمانه 2^n تفاوتی با محاسبات n بیتی با حذف رقم نقلی ندارد. در دو پیمانه‌ی دیگر نیز اگر $\delta = 1$ باشد، معمولا محاسباتی ساده‌تر از سایر مقادیر دارد، به همین دلیل در مجموعه‌های سه پیمانه‌ای، اغلب این حالت گزینه‌ی مناسب‌تری است. در این بخش به معرفی محاسبات پیمانه‌ای در دو حالت $\delta = \pm 1$ می‌پردازیم.

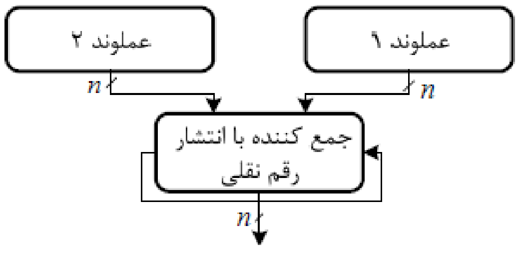
۳-۱-۱جمع و تفریق در پیمانه $2^n - 1$

جمع دو عدد در پیمانه‌ی $2^n - 1$ را می‌توان مطابق رابطه (6) نشان داد. طبق رابطه (7)، زمانی‌که حاصل عملیات بیشتر از پیمانه باشد یا در واقع رقم نقلی درخروجی حاصل شود، کافی است که مقدار ۱ به آن اضافه گردد.

$$|A+B|_{2^n-1} = \begin{cases} A+B, & A+B < 2^n-1 \\ A+B+1, & A+B \geq 2^n-1 \end{cases} \qquad (6)$$

$$|2^n c_n|_{(2^n-1)} = |(2^n-1+1)c_n|_{(2^n-1)} = c_n \qquad (7)$$

بدین ترتیب هر رقم نقلی با ارزش 2^nرا باید به بیت کم‌ارزش‌تر در حاصل، اضافه کرد، شکل (۲) ساختار کلی عملیات جمع در این پیمانه را نشان می‌دهد[11].



شکل (۲): حاصل به پیمانه $2^n - 1$ [11]

در عمل تفریق مطابق با رابطه (8) کافی است که مقدار مکمل یک، از عملوند دوم را با عملوند اول جمع کرد.

$$|A-B|_{2^n-1} = |A+\bar{B}-2^n+1|_{2^n-1} = |A+\bar{B}|_{2^n-1} \qquad (8)$$

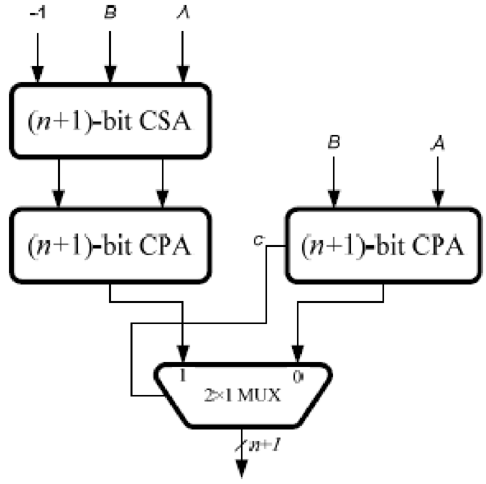
۳-۱-۲جمع و تفریق در پیمانه $2^n + 1$

رابطه (9) نحوه‌ی انجام عمل جمع در پیمانه‌ی $2^n + 1$ را نشان می‌دهد. عمل جمع در این پیمانه کمی پیچیده‌تر از پیمانه‌ی $2^n - 1$ است، زیرا مطابق با رابطه (10) رقم نقلی خروجی در عمل جمع، بصورت منفی به حاصل اضافه می‌شود[11].

$$|A+B|_{2^n+1} = \begin{cases} A+B, & A+B < 2^n+1 \\ A+B-1, & A+B \geq 2^n+1 \end{cases} \qquad (9)$$

$$|2^n c_n|_{(2^n+1)} = |(2^n+1-1)c_q|_{(2^n-1)} = -c_n \qquad (10)$$

منفی شدن بیت نقلی خروجی علت اصلی پیچیده شدن عملیات در این پیمانه است. شکل (۳) مدار انجام عملیات جمع در این پیمانه را نشان می‌دهد. در بخش چپ شکل، ابتدا کاهش عمق به دو برای سه ورودی A، Bو 1- بصورت $n+1$ بیتی و با استفاده از CSA[۳] انجام شده و سپس بر اساس وجود یا عدم وجود رقم نقلی در جمع‌کننده (CPA[۴]) حاصل جمع انتخاب می‌گردد.



شکل (۳): مدار جمع با انتشار رقم نقلی به پیمانه‌ی $2^n + 1$ [11]

با توجه به نحوه‌ی محاسبه‌ی حاصل جمع/تفریق پیمانه‌ای می‌توان بررسی کرد که چه میزان این نوع محاسبات در تعیین ضرایب فیلتر در DWT می‌تواند بهبود ایجاد کند. در ادامه به تشریح نحوه‌ی استفاده از سیستم اعداد مانده‌ای در DWT پرداخته شده است.

۴- روش پیشنهادی

در [4] از مجموعه سه پیمانه‌ای $\{2^n-1, 2^n, 2^n+1\}$ برای بهبود عملکرد محاسباتی و کاهش مساحت مدار انجام عملیات تبدیل موجک گسسته استفاده شد. در این مقاله مجموعه سه پیمانه‌ای $\{2^n-1, 2^n, 2^{n+1}-1\}$ انتخاب شده است. تفاوت این دو مجموعه در یک پیمانه است که در [4] از 2^n+1 و در این مقاله از $2^{n+1}-1$ استفاده می‌شود. علت این انتخاب وجود مدارهای انجام عملیات جمع و ضرب ساده‌تر در پیمانه‌هایی است که فرم $2^n-\delta$ را دارند. در پیمانه‌های به فرم $2^n+\delta$ در انجام عملیات جمع و ضرب در صورت

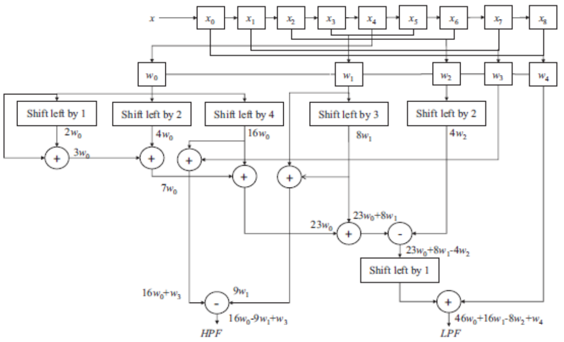
---

۳ Carry Save Adder

۴ Carry Propagation Adder

وقوع سرریز، مقدار منفی بازگردانده می‌شود روابط (8) و (10) علت این موضوع را نشان می‌دهند. بنابراین در پیمانه 2^n+1 برای ادامه انجام عملیات، نیازمند مدارهای اضافی برای تبدیل آن به مقدار مثبت هستیم در حالی‌که در پیمانه 2^n-1 چنین بخشی نیاز نیست. البته باید توجه داشت که در پیمانه‌ی $2^{n+1}-1$ انجام عملیات با یک بیت بیشتر از پیمانه 2^n+1 صورت می‌گیرد که باعث می‌شود انتشار رقم نقلی در عمل جمع پیمانه‌ای، یک بیت بیشتر باشد. اما در ادامه به کمک شبیه‌سازی و سنتز مدارها در FPGA نشان داده خواهد شد که در مجموع از عملکرد بهتری برخوردار است.

همانطور که در بخش قبل اشاره شد، در پیاده‌سازی تبدل موجک گسسته می‌توان از بانک فیلتر CDF97 با ضـرایب عددی صـحیح اسـتفاده نمود که شـامل دو بخش بالاگذر و پایین‌گذراسـت. در شـکل (۵) نحوه‌ی محاسـبه‌ی ضـرایب فیلتر بدون انجام عملیات ضـرب (و فقط با عملیات شـیفت، جمع و تفریق) نشان داده شده است[2]. رابطه (11) نحوه‌ی محاسبه ضرایب w_i را نشـان می‌دهد. در پیاده‌سـازی ضـرایب به کمک RNS کافی اسـت که بعد از انجام تبدیل مسـتقیم برای اعداد ورودی به فیلتر، در هر پیمانه بصـورت مجزا محاسـبات مربوط به بخش‌های بالاگذر و پایین‌گذر انجام گردد و در انتها به کمک تبدل معکوس، معادل دودویی آن به دسـت آید که در ادامه جزئیات آن شرح داده خواهد شد.



شکل (۵): تولید ضرایب فیلتر lowpass و highpass [4]

$$w_n = (x_i - n) + (x_i + n),\ \ 0 \le n \le 8 \qquad (11)$$

۴-۱- محاسبات پیمانه‌ایی تولید بخش‌های بالا و پایین‌گذر

محاسبات پیمانه‌ای برای اعداد صحیح امکان‌پذیر است، به همین دلیل در پیاده‌سازی ضرایب فیلتر از ضرایب صحیح که به روش نسبت دوتایی از داده‌های فرمت ممیزشناور استخراج شد، استفاده شده است [4]. جدول (۱) این ضرایب را نشان می‌دهد. پهنای محاسبات در پیمانه‌های $2^n-1, 2^n, 2^{n+1}-1$ به ترتیب ۸، ۸ و ۹ بیت ($n=8$) است. بنابراین بعد از انجام عملیات پیمانه‌ای و تبدیل معکوس مقدار نهایی ۲۵ بیتی حاصل خواهد شد. همانگونه که در بخش قبل‌تر در خصوص نحوه انجام عملیات جمع و تفریق پیمانه‌ای اشاره شد، محاسبات می‌تواند در سه کانال موازی انجام شود (مطابق روابط (12) و (13)).

$$\mathrm{LP}F(X) = \begin{cases} |X|_{2^n-1} \\ |X|_{2^{n+1}-1}\ , n = 8 \\ |X|_{2^n} \end{cases}$$

$$= \begin{cases} |46W_0 + 16W_1 - 8W_2 + W_4|_{2^n-1} \\ |46W_0 + 16W_1 - 8W_2 + W_4|_{2^{n+1}-1} \\ |46W_0 + 16W_1 - 8W_2 + W_4|_{2^n} \end{cases} \qquad (12)$$

$$HPF(X) = \begin{cases} |X|_{2^n-1} \\ |X|_{2^{n+1}-1}, n = 8 \\ |X|_{2^n} \end{cases}$$

$$= \begin{cases} |16W_0 - 9W_1 + W_3|_{2^n-1} \\ |16W_0 - 9W_1 + W_3|_{2^{n+1}-1} \\ |16W_0 - 9W_1 + W_3|_{2^n} \end{cases} \qquad (13)$$

یعنی کلیه‌ی عملیات جمع و شیفت در سه کانال موازی و با پهنای بیتی کمتر از روش دودویی انجام می‌شوند. از طرفی هر عمل شیفت پیمانه‌ای در پیمانه‌های به فرم 2^n-1 با یک شیفت چرخشی به چپ (CSL[۵]) جایگزین می‌شود. به عنوان مثال در عدد n بیتی $A = a_{n-1}a_{n-2} \dots a_1a_0$ عمل $|4A|_{2^n-1}$ مطابق رابطه (14) انجام می‌گردد.

$$|4A|_{2^n-1} = |4(a_{n-1}a_{n-2} \dots a_1a_0)|_{2^n-1} = 2CSL(A) = a_{n-3}a_{n-4} \dots a_1a_0a_{n-1}a_{n-2} \qquad (14)$$

جدول ۱: ضرایب فیلتر که به روش نسبت دوتایی برای بانک فیلتر CDF97 به دست آمده است[4].

k	High Filter(g_k)	Lowpass Filter (h_k)
0	46/64	1
±1	16/64	-9/16
±2	-8/64	0
±3	0	1/16
±4	1/64	

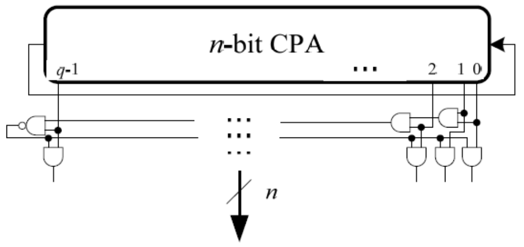
بنابراین در پیاده‌سازی سخت‌افزاری ضرایب فیلتر، عمل ضرب $2^i w_j$ بدون هزینه انجام می‌گردد. با درنظرگرفتن ضرایب به فرم $2^i w_j$ در دو بخش بالا و پایین‌گذر، باید عملیات جمع/تفریق پیمانه‌ای صورت گیرد. شکل‌های (۷) و (۸) مدارهای محاسباتی هر یک از دو بخش فیلتر را در پیمانه‌ی 2^n-1 نشان می‌دهند.

در بخش پایین‌گذر برای محاسبه $23w_0 + 8w_1$ به جای ضرب و جمع از $(16+4+2+1)\mathrm{w}_0 + 8\mathrm{w}_1$ اسـتفاده شـده تا دیگر نیازی به عمل ضرب نبا شد. از طرفی برای کاهش تاخیر محا سبه مجموع، پس از چرخش به چپ (به تعداد مورد نیاز در هر ضـریب)، عملیات کاهش عمق به دو سـطر با اسـتفاده از CSA پیمانه‌ای انجام می‌شـود. به دلیل پیمانه‌ای بودن محاسـبات طبق رابطه (7)، رقم نقلی خارج‌شـده، دوباره از بیت کم‌ارزش‌تر وارد مرحله‌ی بعد می‌گردد که به همین دلیل به آن CSAپیمانه‌ای گفته می‌شـود. عملیات کاهش عمق به دو در سـه مرحله صـورت می‌گیرد که در انتها، عمل جمع پیمانه‌ای انجام می شود. سپس باید عمل تفریق $(23\mathrm{w}_0 + 8\mathrm{w}_1) - 4\mathrm{w}_2$ صورت گیرد که در پیمانه‌ی مذکور طبق رابطه‌ی (8) فقط کافی‌ست که مقدار

---

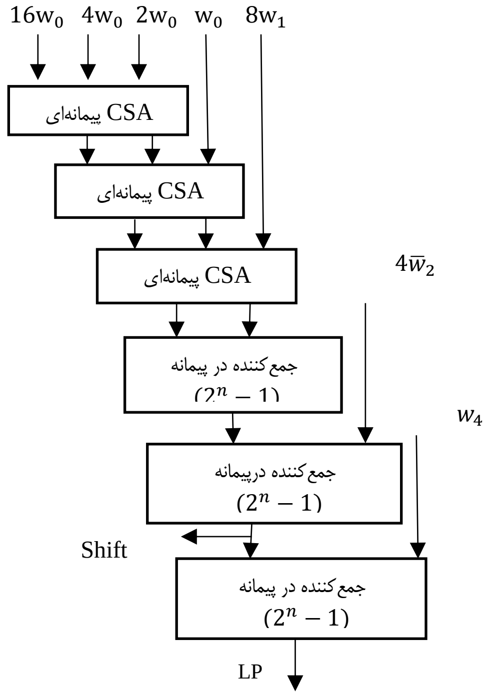
۵ Cyclic Shift Left

مکمل یک از $4w_2$ با بخش دیگر جمع شـــود. در این عمل نیز یک مزیت در مقایسه با پیمانه‌ی 2^n+1 (به کار رفته در [4]) دیده می‌شود که عمل تفریق بسیار ساده‌تر است و فقط نیاز به گیت Not دارد. در ادامه و قبل از انجام عمل جمع با w_4 ابتدا یک شـیفت به چپ انجام می‌شـود تا رابطه‌ی $46w_0+16w_1-8w_2$ به دســـت آید که بدون نیاز به ســـخت‌افزار و با سیم‌بندی بصورت چرخش به چپ قابل پیاده‌سازی است و در انتها، عمل جمع پیمانه‌ای با w_4 انجام می‌گردد. یکی از معایب محاسبات در این پیمانه مشابه با عمل تفریق مکمل یک، وجود دو نمایش برای عدد صـــفر اســـت که برای تشـخیص آن از مدار شـکل (۶) کمک گرفته شـده اسـت. مطابق شـکل در صـــورتی‌که مقدار خروجی مدار جمع‌کننده، عدد 2^n-1 باشـــد، به کمک رشته‌ای از گیت‌های منطقی AND، آن‌را به صفر تبدیل می‌کنیم.

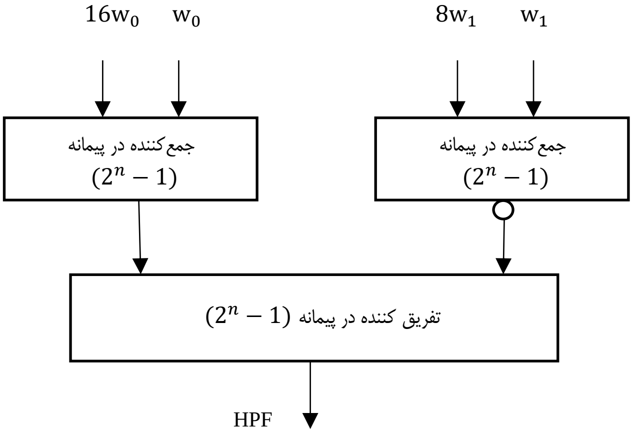


شکل (٦): مدار تصحیح‌کننده‌ی حاصل‌جمع در پیمانه‌ی 2^n-1 [15]

در بخش بالا گذر، ابتدا دو عمل جمع بطور همزمان برای دو بخش $8w_1+w_1$ و $16w_0+w_3$ انجام می‌شـود و سـپس برای تفریق حاصـل جمع‌ها از یکدیگر، از مدار تفریق‌کننده‌ی در پیمانه‌ی 2^n-1 استفاده می‌شود که مشـــابه با بخش پایین‌گذر، تنها نیاز به گیت Not برای مکمل‌کردن مقدار $9w_1$ دارد. لازم به ذکر است که مدار تشخیص‌دهنده‌ی صفر در انتهای این بخش نیز به‌کارمی‌رود.

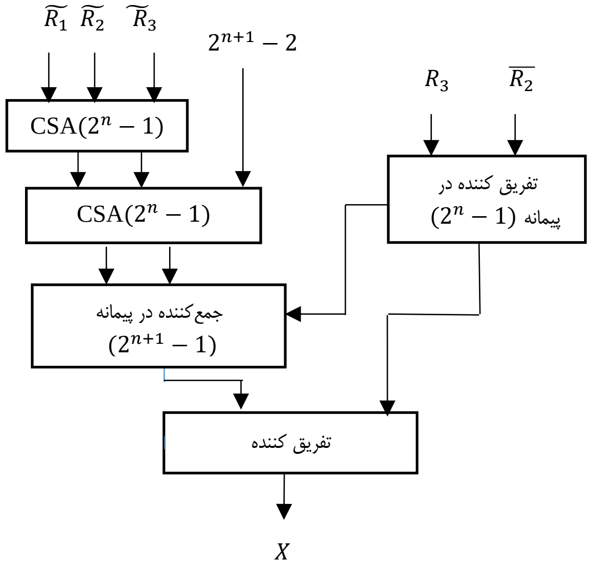


شکل (۷): مدار محاسبات در پیمانه‌ی 2^n-1 برای LPF



شکل (۸): مدار محاسبات در پیمانه‌ی 2^n-1 برای HPF

در پایان باید معادل دودویی ضرایب به‌دست آیند که برای این‌کار از مدار تبدل معکوس مطابق با شکل (۹) استفاده شد[14] تا حاصل ۲۵ بیتی نهایی به دســـت آید. در این شـــکل مقادیر $R_1=|X|_{2^{n+1}-1}$، $R_2=|X|_{2^n}$ و $R_3=|X|_{2^n-1}$ هسـتند و به ترتیب $n+1$، n و n بیتی می‌باشـند. از طرفی $\widetilde{R_1}=\overline{r_{1(n-2)}}\,\overline{r_{1(n-3)}}\dots\overline{r_{11}}\,\overline{r_{10}}\,\overline{r_{1(n)}}\,\overline{r_{1(n-1)}}$ ، $\widetilde{R_2}=$ $0\,\overline{r_{20}}\dots\overline{r_{2(n-2)}}\,\overline{r_{2(n-1)}}$ و $0\,\overline{r_{30}}\dots\overline{r_{3(n-2)}}\,\overline{r_{3(n-1)}}=\widetilde{R_3}$ فرض شده است. لازم به ذکر است که زمان انجام عملیات تبدیل معکوس تنها یک بار در کل محاســـبات ضـــرایب منظور می‌گردد زیرا این کار را می‌توان موازی با محاسبات سایر اعداد ورودی انجام داد.



شکل (۹): مدار معکوس کننده [14]

٤-٢- مقایسه و ارزیابی نتایج سنتز در FPGA

پیاده سازی سخت افزاری مدار پیشنهادی در FPGA برای مدل خروجی xc6vlx75t انجام شده است. نتایج شبیه‌سازی در جدول (۲) نشان داده شده است. در این جدول محاسبات دودویی[4]، محاسبات پیمانه‌ای ارائه شده در [4] و روش پیشنهادی قرار داده شده‌اند که همگی در همین مدل خروجی FPGA شبیه‌سازی شده‌اند. مقدار توان مصرفی، میزان LUT مورد نیاز و تاخیر مدارها نشان می‌دهد که در روش پیشنهادی ۶۴/۷ درصددرمقایسه با محاسبات دودوی و ۵۴/۷ درصد در مقایسه با [4]، بهبود در تاخیر مدار ایجاد شده است و در مقابل به ترتیب ۵۵/۱ و ۶۵/۵ درصد افزایش مساحت و ۶/۶ و ۴۶/۸ درصد افزایش توان مصرفی درمقایسه با دودویی و [4] وجود دارد.

جدول (۲) : نتایج شبیه سازی در FPGA

	Power(mW)	LUTs	Delay(ns)
روش دودویی [4]	۱۳/۹۹	۶۵۴	۹/۹۱۸
روش پیمانه‌ای [4]	۷/۹۶	۵۰۲	۷/۷۲۸
روش پیشنهادی	**١٤/٩٨**	**١٤٨٥**	**٣/٤٩٧**

۵- نتیجه گیری

تبدیل موجک گسسته کاربردهای فراوانی در پردازش تصویر دارد. یکی از مشکلات این تبدیل، وجود تعداد زیاد محاسبات جمع و ضرب است. روش‌های مختلفی برای بهبود عملکرد محاسباتی تبدیل موجک گسسته ارائه شده که یکی از آن‌ها استفاده از سیستم اعداد مانده‌ای است. در این سیستم امکان اجرای موازی عملیات جمع، تفریق و ضرب وجود دارد. یعنی می‌توان در کانال‌های موازی و با تعداد بیت کم‌تر در هر کانال در مقایسه با محاسبات دودویی، عملیات حسابی مذکور را انجام داد. معمولا از مجموعه سه پیمانه‌ای $\{2^n, 2^n - 1, 2^n + 1\}$ در چنین مواردی برای بهبود عملکرد محاسباتی استفاده می‌شود[4]. مشکل این مجموعه، وجود پیمانه‌ی $2^n + 1$ است که باعث سخت‌تر شدن محاسبات در مقایسه با پیمانه‌ی $2^n - 1$ می‌شود. در این مقاله از پیمانه‌ی $2^{n+1} - 1$ استفاده شده تا سرعت انجام محاسبات در سه پیمانه به هم نزدیک‌تر شود. بدین ترتیب سرعت مدار کل نیز افزایش پیدا می‌کند. نتایج شبیه‌سازی و سنتز مدار پیشنهادی در تراشه FPGA مدل xc6vlx75t نشان می‌دهد که میزان تاخیر، در مقایسه با روش ارائه شده در [4]، ۵۴/۷درصد بهبود یافته و در مقابل، میزان مساحت (LUT) و توان مصرفی به ترتیب ۶۵/۵ و ۴۶/۸درصد افزایش یافته است.

مراجع

[1] Gonzalez, R.C., Woods R., *Digital Image Processing*, University of Tennessee, 4th edition, Pearson Publications, 2018.

[2] Ballesteros L, Dora M., and Juan M. Moreno A., "*Wavelet-denoising on hardware devices with Perfect Reconstruction, low latency and adaptive thresholding*", Computers and Electrical Engineering Journal, Vol. 39(4), pp. 1300-1311, 2013.

[3] Chervyakov, N.I., Lyakhov P.A., and Babenko M.G., "*Digital filtering of images in a residue number system using finite-field wavelets*." Automatic Control and Computer Sciences Vol. 48 (3), pp. 180-189, 2014.

[4] Kong, Y., Safari, A., Niras, C. V.," *A low-cost architecture for DWT filter banks in RNS applications*", IEEE International Symposium on Integrated Circuits (ISIC) – Singapore, pp. 448-451, 2014, doi: 10.1109/ISICIR.2014.7029566.

[5] Safari, Azadeh, Cheecottu Vayalil Niras, and Yinan Kong, "*Power-performance enhancement of two-dimensional RNS-based DWT image processor using static voltage scaling*", Integration Journal, Vol. 53 pp. 145-156, 2016.

[6] Chervyakov N., Lyakhov P., Kaplun D., Butusov D. and Nagorov N.," *Analysis of the Quantization Noise in Discrete Wavelet Transform Filters for Image Processign*", Journal of Electronics, Vol. 135 (7), Aug. 2018, DOI: 10.3390/electronics7080135.

[7] Chervyakov N., Lakhov P., Kalita D. and Shulzhenko K.S.," *Effect of RNS Dnamic Range on Grayscale Image Filtering*", IEEE XV International Symposium of Problems of Redundancy in Information and Control Systems (REDUNDANCY), pp. 33-37, 2016, doi: 10.1109/RED.2016.7779323.

[8] Chervyakov, N. I., Lyakhov, P. A., Nagornov, N. N., Valueva, M. V., & Kaplun, D. I.," *High-Performance Hardware 3D Medical Imaging using Wavelets in the Residue Number System*", 9th Mediterranean Conference on Embedded Computing *(MECO)*,pp. 1-4, 2020, doi: 10.1109/MECO49872.2020.9134123.

[9] Johansson, E., *Wavelet Theory and some of its Applications*, Ph.D. Thesis, Luleå tekniska universitet, 2005.

[10] Eissa M.M., "*A novel digital directional transformer protection technique based on wavelet packet*", IEEE Trans. Power Delivery,Vol. 20, No. 3, pp. 1830-1836, July 2005, doi: 10.1109/TPWRD.2005.848646.

[11] H. Ahmadifar, *Balanced Moduli Selection for Efficient Reverse Conversion in Residue Number System*, PhD Thesis, Shahid Beheshti University, 2013.

[12] Parhami B., *Computer Arithmetic:Algorithms and Hardware Designs*, Oxford University Press, New York, 2nd ed., 2010.

[13] Mohan A.PV., *Residue Number Systems: Algorithms and Architectures*, New York, Kluwer Academic Publishers, 2002

[14] Hiasat, A., "*An Efficient Reverse Converter for the Three-Moduli Set* $\{2^{n+1} - 1, 2^n, 2^n - 1\}$ ", IEEE Transactions on Circuits and Systems II: Vol. 64, no. 8, pp. 962-966, Aug. 2017, doi: 10.1109/TCSII.2016.2608335.

Jaberipur G. and Ahmadifar H., "*A ROM-less Reverse RNS Converter for Moduli Set* $\{2^n \pm 1, 2^n \pm 3\}$", IET Computers & Digital Techniques, 2013, doi: 10.1049/iet-cdt.2012.0148.

چهارمین کنفرانس ملی انفورماتیک ایران
پژوهشگاه دانش‌های بنیادی، پردیس فرمانیه، تهران
۱۴ و ۱۵ دی ماه ۱۴۰۱

طراحی سیستم معاملاتی نیمه خودکار برای اعلام سیگنال‌های معاملاتی بایننس‌کوین

امیرحسین برادران[۱]، معصومه بهلولی[۲]، محمدرضا جاهد مطلق[۳]

[۱] دانشجوی کارشناسی ارشد کنترل، دانشکده برق، دانشگاه علم و صنعت ایران، تهران
a_baradaran@cmps2.iust.ac.ir
[۲] دانشجوی کارشناسی ارشد هوش مصنوعی، دانشکده فنی و مهندسی، دانشگاه الزهرا، تهران
m.bohlouli@student.alzahra.ac.ir
[۳] استاد، دانشکده کامپیوتر، دانشگاه علم و صنعت ایران، تهران
jahedmr@iust.ac.ir

چکیده

ظهور بازار ارزهای دیجیتال نحوه انجام معاملات آنلاین را تغییر داده و فرصت سرمایه‌گذاری جدیدی را فراهم کرده است. هنر پیش‌بینی قیمت در این بازار برای بسیاری از محققان و تحلیلگران کار دشواری بوده است. امروزه اکثر فعالین بازار به‌دلیل مزایای ذاتی همچون پردازش مقادیر زیاد داده، عدم احساس ترس یا طمع و پیش‌بینی قیمت بازار از طریق هوش مصنوعی، به‌دنبال استفاده از معاملات الگوریتمی در این بازار هستند. بنابراین باتوجه به افزایش محبوبیت ارزهای دیجیتال تقاضای زیادی برای ابزارهای معاملاتی خودکار و نیمه خودکار به‌جهت ایجاد سیگنال‌های معاملاتی برای خرید و فروش وجود دارد. بایننس‌کوین ارز دیجیتال صرافی بایننس، که در حال حاضر بزرگ‌ترین صرافی ارز دیجیتال در جهان است. هدف اصلی این مقاله ایجاد یک سیستم معاملاتی نیمه خودکار برای ارز دیجیتال بایننس‌کوین است، که کارایی مناسبی در ارائه سیگنال‌های معاملاتی دارد. این سیستم معاملاتی ازطرفی به‌وسیله پردازش متن با کتابخانه TextBlob به بررسی احساسات موجود در اخبار مربوط به بایننس‌کوین می‌پردازد و ازطرف دیگر با شبکه عصبی حافظه طولانی کوتاه مدت، قیمت روز آینده بایننس‌کوین را پیش‌بینی می‌کند و درنهایت با تحلیل موارد به‌دست آمده سیگنال معاملاتی مبنی‌بر خرید، فروش یا نگهداری ارائه می‌دهد.

کلمات کلیدی

ارزهای دیجیتال، بایننس‌کوین، معاملات الگوریتمی، سیگنال معاملاتی، پردازش زبان طبیعی، تحلیل احساسات، پردازش متن، شبکه حافظه طولانی کوتاه مدت، پیش‌بینی قیمت، جذر میانگین مربعات خطا

۱- مقدمه

بایننس‌کوین با نماد اختصاری BNB یک ارز دیجیتال یا شکلی از دارایی دیجیتال است[1] که با ارزش بازار حدود ۴۷.۲۶ میلیارد دلار، در رتبه ۵ بازار قرار داشته و سهم ۴.۲۴ درصدی از کل بازار را در اختیار دارد و حجم مبادلات روزانه آن ۱.۰۲ میلیارد دلار است که با هدف جذب سرمایه برای گسترش و توسعه صرافی بایننس، ایجاد شده‌اند. بایننس‌کوین به‌دلیل داشتن صرافی اختصاصی و امن نسبت به بیت‌کوین که برای معامله به یک صرافی مورد اعتماد نیاز دارد، از نوعی برتری برخوردار است[2]. تعداد بیت‌کوین حداکثر ۲۱ میلیون واحد است اما ارز دیجیتال بایننس‌کوین ۲۰۰ میلیون واحد دارد، بنابراین ارز دیجیتال بایننس کوین از بیت کوین به لحاظ حجم کلی، در آینده سبقت خواهد گرفت. فعالین بازار همواره به‌دنبال روش‌هایی هستند که با کمترین خطا بهترین نقاط ورود و خروج را برای یک ارز دیجیتال همچون بایننس‌کوین پیدا کنند. اما این کار مستلزم صرف زمان زیادی است و در مواردی خطای قابل توجهی به‌دلیل خطای انسانی همراه دارد. بنابراین معاملات الگوریتمی توجه بسیاری از سرمایه‌گذاران را به خود جلب کرده است. معاملات الگوریتمی بسته به اهدافی که از آن‌ها بهره‌گیری می‌شود، از یک سری الگوریتم‌ها برای تصمیم‌گیری استفاده می‌کنند تا فرصت‌های معاملاتی مناسبی را برای کاربر ایجاد کند. یک نمونه از اهداف معاملات الگوریتمی، کمک به معامله‌گران برای پیداکردن سیگنال‌های خرید و فروش در بازارها (مشاور) است.

Ferdiansyah و همکارانش با به‌کارگیری شبکه عصبی حافظه طولانی کوتاه مدت توانسته‌اند قیمت بیت‌کوین را برای روز آینده پیش‌بینی کنند، اما مقدار جذر میانگین مربعات خطا در بهترین حالت برای این مدل برابر با ۲۸۸.۵۹۸۶۶ شده است[3].

Hans و همکارش با ترکیب ابزارهای رایج تجزیه و تحلیل احساسات مانند VADER و TextBlob در یک مدل واحد توانسته‌اند تحلیل احساسات مربوط به ارزهای دیجیتال همچون بیت‌کوین را از طریق توییتر به‌دست آورند. اما مدل آن‌ها صرفا با بررسی داده‌های موجود در توییتر به بررسی وضعیت قیمت پرداخته است[4].

بنابراین در مطالعات انجام پذیرفته صرفا به پیش‌بینی قیمت و یا تحلیل اخبار موجود در بازار به‌صورت جداگانه پرداخته شده است. یکی از ویژگی‌های منحصر به فرد این مقاله ایجاد یک سیستم معاملاتی هوشمند که به‌صورت همزمان با پیش‌بینی قیمت و نیز تحلیل اخبار بازار از نظر احساسی به بررسی وضعیت بایننس‌کوین پرداخته و سیگنال خرید یا فروش را صادر می‌کند. لازم به ذکر است که جذر میانگین مربعات خطا برای شبکه عصبی حافظه طولانی کوتاه مدت این سیستم معاملاتی برابر ۱۵.۴۶۱ به‌دست آمد که در مقایسه با کار مشابه انجام شده، به‌طور قابل توجهی کم‌تر شده است.

۲- روش‌ها

هدف اصلی این پژوهش ایجاد یک سیستم معاملاتی نیمه خودکار به‌جهت اعلام سیگنال‌های معاملاتی بایننس‌کوین است. برای آموزش و ارزیابی از دو سری مجموعه داده برحسب نیاز این سیستم استفاده شده است. در این بخش، ابتدا به تشریح بانک داده‌گان مورد استفاده در این پژوهش پرداخته و متعاقب آن، سیستم پیشنهادی تشریح می‌گردد.

۲-۱- تحلیل احساسات

دیدگاه هر یک از سرمایه‌گذاران نسبت به تغییر روند بایننس‌کوین، در آینده بر تصمیم‌گیری آن‌ها هنگام معامله تاثیر می‌گذارد. در بلندمدت، ممکن است ارزش این ارز باشد که قیمت را تعیین می‌کند، اما در کوتاه مدت احساس و رفتارهای هیجانی موجود باعث به‌حرکت درآوردن قیمت می‌شود. این مسئله می‌تواند برای سرمایه‌گذاران به‌جهت پیدا کردن نقاط ورود جذاب، فرصت‌های سرمایه‌گذاری ایجاد کند[5].

به‌طور کلی منظور از احساسات بازار، اندازه‌گیری کیفی از نگرش و احساس کلی سرمایه‌گذاران نسبت به بازارهای مالی یا یک بخش خاص از بازار است[6]. احساسات مثبت و منفی باعث حرکت قیمت و همچنین ایجاد فرصت‌های معاملاتی و سرمایه‌گذاری برای معامله‌گران فعال و سرمایه‌گذاران بلندمدت می‌شود.

بنابراین در این مقاله، پردازش زبان طبیعی به‌جهت تشخیص احساسات و هیجانات مردم از طریق اخبار، برای کمک به پیش‌بینی آینده بایننس‌کوین به‌کار گرفته شده است.

۲-۲- مجموعه داده اخبار بایننس‌کوین

برای جمع‌آوری مجموعه داده اخبار مورد نیاز بایننس‌کوین از سایت گوگل نیوز (https://news.google.com) که یک سایت خبری رایانه‌ای است[7] و سرفصل‌های منابع خبری در سراسر جهان را به‌صورت روزانه جمع‌آوری می‌کند، استفاده شده است. اما دریافت اخبار از گوگل نیوز در زبان پایتون نیاز به یک رابط برنامه‌نویسی دارد تا به‌وسیله آن به این سایت متصل شده و اخبار مورد نظر دریافت گردد. در این مقاله کتابخانه GoogleNews که یک رابط برنامه‌نویسی پشتیبانی شده[8] توسط گوگل نیوز در پایتون مورد استفاده قرار گرفته، و دارای ویژگی جستجو اخبار براساس محدوده تاریخ و کلمات کلیدی است. بنابراین براساس جدول (۱) برای انجام هرچه دقیق‌تر تحلیل احساسات، بااستفاده از کلمات کلیدی [BNB, Binance Coin, BNB-USD] متن صفحه اول مقاله‌های مربوط به اخبار منتشر شده در روز فعلی مورخ ۲۰۲۲/۱۰/۲۷ به‌صورت کامل دریافت می‌شود. در این مقاله از اخبار روزهای گذشته استفاده نشده است زیرا اخبار روزهای گذشته تاثیر خود را در همان تاریخ بر روی قیمت می‌گذارد.

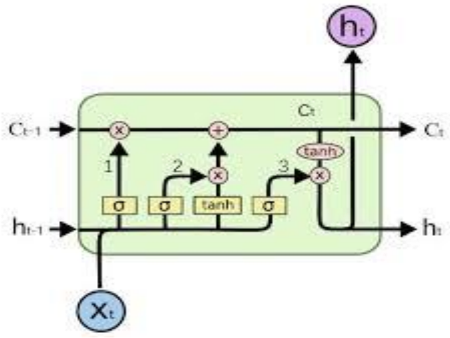
۲-۳- شبکه عصبی حافظه طولانی کوتاه مدت

حافظه طولانی کوتاه مدت نوع خاصی از شبکه‌های عصبی بازگشتی هستند که قادر به یادگیری وابستگی‌های طولانی مدت هستند و مشکل حافظه بلندمدت شبکه عصبی بازگشتی را حل می‌کند یعنی به خاطر سپردن اطلاعات برای مدت زمان طولانی، رفتار پیش فرض آن‌ها است.

جدول (۱) : داده‌های جمع‌آوری شده از GoogleNews

شماره	تاریخ	زمان	عنوان	مقالات	پیوند
1	2022-10-27 18:32:36	1 hours ago	Central Bank Of Khazakhstan Integrates Its CBDC …	The National Bank Of Khazakhstan (NBK) has agre…	https://gadgettendency.com/central-bank-of-kaz...
2	2022-10-27 18:32:36	1 hours ago	BNB Price Prediction: New Bullish Cycle to New...	Binance Coin (BNB) price has increased since f...	https://beincrypto.com/bnb-price-prediction-ne...
…	…	…	…	…	…
26	2022-10-27 13:32:38	6 hours ago	No Regret For Binance and TRON Investors Turni…	Binance Coin and TRON coins dipped for the las...	https://www.analyticsinsight.net/no-regre-for...

این شبکه، مدل را بااستفاده از پس انتشار آموزش می‌دهد[9]. براساس شکل (۱)، در ساختار شبکه حافظه طولانی کوتاه مدت نشان داده شده است که مفهوم اصلی این شبکه حالت سلولی و دروازه‌های همراهش است. درواقع حالت سلولی مانند آزادراهی عمل می‌کند که اطلاعات را در طول زنجیره‌ی توالی جلو می‌برد. بنابراین به حالت سلولی می‌توان به‌عنوان حافظه‌ی شبکه نگاه کرد. دروازه‌ها نیز اطلاعات را در حالت سلولی به‌روز نگه می‌دارند. این دروازه‌ها شبکه‌های عصبی مختلفی هستند که تصمیم می‌گیرند چه اطلاعاتی به حالت سلولی وارد شوند و در طول آموزش شبکه یاد می‌گیرند چه اطلاعاتی باید حفظ یا فراموش شوند. به‌این شکل، شبکه اطلاعات مهم را در طول زنجیره‌ی توالی عبور می‌دهد تا خروجی مدنظر را داشته باشیم. شبکه حافظه طولانی کوتاه مدت سه دروازه تحت عنوان‌های دروازه فراموشی، دروازه ورودی و دروازه خروجی دارد[10].



شکل (۱) : ساختار شبکه عصبی حافظه طولانی کوتاه مدت [13]

دروازه فراموشی تصمیم می‌گیرد کدام اطلاعات حفظ یا فراموش شود. در رابطه (۱) اطلاعات ورودی گام جدید به‌همراه اطلاعات حالت نهان گام

قبلی به این دروازه وارد می‌شوند و از تابع سیگموئید عبور می‌کنند[11]. خروجی این تابع عددی در بازه [۰ ۱] است. اگر عدد خروجی به صفر نزدیک‌تر باشد به‌معنی فراموشی اطلاعات و هر قدر به ۱ نزدیک‌تر باشد به‌معنی حفظ اطلاعات است.

$$f_t=\sigma(W_f.[h_{t-1},x_t]+b_f) \quad (۱)$$

که در آن f_t دروازه فراموشی، σ تابع سیگموئید، $x_t \in R^d$ بردار ورودی به واحد حافظه طولانی کوتاه مدت، h_t بردار حالت نهان که به‌عنوان بردار خروجی واحد حافظه طولانی کوتاه مدت نیز شناخته می‌شود و $W \in R^{h \times d}$ و $b \in R^h$ ماتریس‌های وزن و پارامترهای بردار بایاس که در طول آموزش به‌دست می‌آیند، به‌طوری‌که بالانویس‌های d و h به‌ترتیب به تعداد ویژگی‌های ورودی و تعداد واحدهای نهان اشاره دارند.

دروازه ورودی برای به‌روزرسانی مقادیر موجود در حالت سلولی آماده‌سازی شده است. در رابطه (۲) اطلاعات ورودی گام جدید، به‌همراه اطلاعات حالت نهان گام قبلی، به این دروازه وارد می‌شوند و از تابع سیگموئید عبور می‌کنند تا این تابع تصمیم بگیرد کدام حذف یا به‌روزرسانی شوند. در این فرآیند نحوه انجام کار به این صورت انجام می‌پذیرد که اطلاعات نزدیک به صفر حذف می‌گردد و اطلاعات نزدیک به ۱ به‌روزرسانی می‌شوند. همچنین در رابطه (۳) اطلاعات ورودی گام جدید، به‌همراه اطلاعات حالت نهان گام قبلی، به تابع تانژانت هایپربولیک وارد می‌شوند تا مقادیرشان در بازه [۱ ۱-] قرار بگیرد. درنهایت خروجی تابع سیگموئید و تانژانت هایپربولیک با هم ضرب می‌شوند تا تابع سیگموئید تصمیم بگیرد چه مقادیری از خروجی تابع تانژانت هایپربولیک باید حفظ شوند[12].

$$i_t=\sigma(W_i.[h_{t-1},x_t]+b_i) \quad (۲)$$

$$\hat{C}_t=\tanh(W_C.[h_{t-1},x_t]+b_C) \quad (۳)$$

که i_t دروازه ورودی و $\hat{C}_t$ بردار فعال سازی ورودی سلول است.

حال با اطلاعات خروجی از دروازه ورودی و دروازه فراموشی مقادیر حالت سلولی به‌روزرسانی می‌شود. به‌این صورت که ابتدا مقدار حالت سلولی در خروجی گیت فراموشی ضرب نقطه‌ای می‌شود. در این قسمت هر مقداری در حالت سلولی که با مقدار نزدیک به صفر خروجی دروازه فراموشی ضرب شود، فراموش می‌شود. بعد از این مرحله مقدار حالت سلولی با خروجی دروازه ورودی جمع نقطه‌ای می‌شود که مقدار حالت سلولی با مقادیر جدیدی که شبکه تشخیص داده است، به‌روزرسانی شوند. در پایان این مرحله مقادیر حالت سلولی به‌روزرسانی می‌شود[13].

دروازه خروجی درنهایت تصمیم می‌گیرد که حالت نهان گام بعدی چه مقداری باشد. حالت نهان، اطلاعات ورودی‌های قبلی را همراه خودش دارد. بنابراین در رابطه (۴) ابتدا اطلاعات ورودی گام جدید به‌همراه اطلاعات حالت نهان گام قبلی به تابع سیگموئید وارد می‌شوند.

در نهایت در رابطه (۵) مقدار به‌روزرسانی شده حالت سلولی به تابع تانژانت هایپربولیک وارد می‌شود سپس خروجی این تابع و رابطه (۴) با هم ضرب می‌شود تا تصمیم گرفته شود حالت نهان چه اطلاعاتی را با خودش به گام بعدی ببرد. درنهایت حالت سلولی جدید و حالت نهان جدید به گام زمانی بعدی منتقل می‌شوند[12].

$$O_t=\sigma(W_O.[h_{t-1},x_t]+b_O \quad (۴)$$

$$h_t=O_t * \tanh(C_t) \quad (۵)$$

که O_t دروازه خروجی و C_t بردار حالت سلولی است.

۲-۴- داده‌های تاریخی قیمت بایننس‌کوین

برای پیش‌بینی قیمت با شبکه عصبی حافظه طولانی کوتاه مدت، نیاز به داده‌های تاریخی قیمت بایننس‌کوین است. بنابراین داده‌های قیمتی مورد نیاز از طریق سایت یاهو فایننس (https://finance.yahoo.com) جمع‌آوری شده است[14].

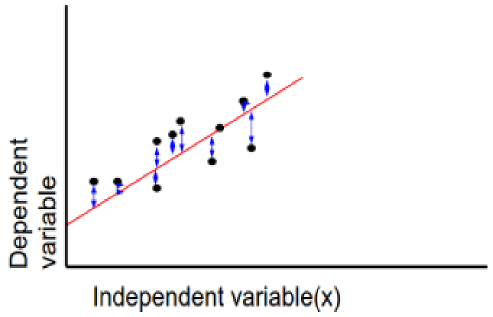
یاهو فایننس یک رسانه مالی، که اخبار مالی و داده‌هایی ازجمله قیمت سهام، بیانیه‌های مطبوعاتی، گزارش‌های مالی، محتوای اصلی و همچنین برخی از ابزارهای آنلاین را برای مدیریت مالی شخصی ارائه می‌دهد. بنابراین در جدول (۲) برای آموزش مدل قیمت‌گذاری بایننس‌کوین، داده‌های تاریخی قیمت در دوره زمانی روزانه از تاریخ ۲۰۱۷/۱۱/۰۹ تا تاریخ ۲۰۲۲/۱۰/۲۷ استفاده شده است.

جدول (۲) : داده‌های قیمتی بایننس‌کوین

حجم	تعدیل‌شده آخرین	آخرین	کم‌ترین	بیشترین	آغازین	تاریخ
۱۹۱۹۲۲۰۰	۱.۹۹	۱.۹۹	۱.۸۹	۲.۱۷	۲.۰۵	۲۰۱۷-۱۱-۰۹
۱۱۱۵۵۰۰۰	۱.۷۹	۱.۷۹	۱.۶۴	۲.۰۶	۲.۰۰	۲۰۱۷-۱۱-۱۰
...	...	...	...	...	...	...
۱۱۱۰۲۴۱۹۲۰	۲۸۹.۰۷	۲۸۹.۰۷	۲۸۷.۲۷	۲۹۲.۹۸	۲۹۰.۴۰	۲۰۲۲-۱۰-۲۷

۲-۵- تحلیل کارایی

برای تجزیه و تحلیل کارایی در پیش‌بینی قیمت با شبکه عصبی حافظه طولانی کوتاه مدت از جذر میانگین مربعات خطا[1] استفاده شده است که براساس شکل (۲) تفاوت بین مقادیر پیش‌بینی شده و مقادیر مشاهده شده یا میانگین مرتبه دوم این تفاوت‌ها است. جذر میانگین مربعات خطا یک معیار ارزیابی است که برای مقایسه خطای پیش‌بینی مدل‌های مختلف برای یک مجموعه داده خاص مورد استفاده قرار می‌گیرد[9].



شکل (۲) : محاسبه ارزش جذر میانگین مربعات خطا [9]

۳- نتایج و بحث

روش پیشنهادی این مقاله بااستفاده از واحد پردازش گرافیکی (GPU) که توسط Collaboratory توسعه یافته، پیاده سازی شده است. باتوجه به این که اعتماد مردم در گسترش بازار ارزهای دیجیتال نقش دارد، احساسات عمومی تأثیر زیادی بر آینده سرمایه‌گذاری بازار ارزهای دیجیتال دارد. بنابراین تحلیل احساسات نقش اساسی در تعیین روند قیمت بایننس کوین دارد. به‌همین جهت برای تحلیل احساسات اخبار موجود در جدول (۱) از کتابخانه TextBlob استفاده شده است. این کتابخانه یکی از ابزارهای بسیار کامل برای پردازش داده‌های متنی و تحلیل سطح جمله در زبان پایتون است که یک مجموعه داده را به‌عنوان ورودی می‌گیرد سپس جملات را برای بررسی

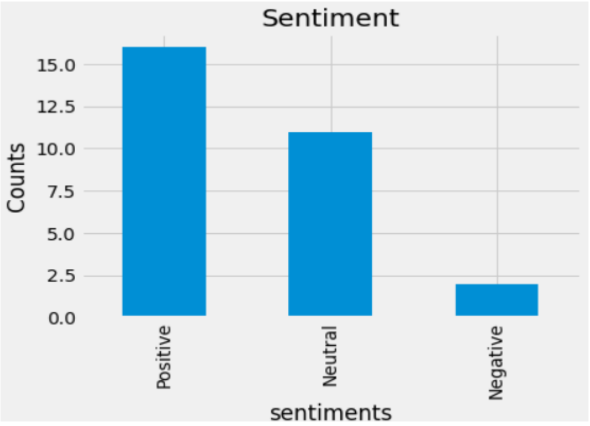
دسته‌بندی می‌کند و انجام وظایف رایج پردازش زبان طبیعی از جمله تحلیل احساسات و موارد دیگر ارائه را می‌دهد.

قطبیت [2] و ذهنیت یک جمله توسط TextBlob به‌عنوان خروجی برگردانده می‌شود. محدوده قطبیت در بازه [۱ ۱−] است، که اعداد منفی برای اخبار منفی و ۰ برای اخبار خنثی و اعداد مثبت برای اخبار مثبت به‌کار می‌رود. از کلمات منفی برای تغییر قطبیت یک جمله استفاده می‌شود و نیز برچسب‌های معنایی در TextBlob به تجزیه و تحلیل دقیق کمک می‌کنند. خروجی ذهنیت نیز در محدوده عددی [۱ ۰] قرار می‌گیرد که درجه نظر شخصی و اطلاعات واقعی در یک متن با خروجی ذهنیت سنجیده می‌شود که این خروجی هرچه به ۰ نزدیک باشد نشانگر یک عبارت واقعی و اگر به ۱ نزدیک باشد نشانگر یک عبارت ذهنی است. در برخی موارد به‌دلیل ذهنی بودن متن، به جای اطلاعات واقعی، حاوی نظرات شخصی است. در این مقاله باتوجه به این‌که منبع جمع‌آوری اخبار گوگل نیوز است و در آن نظرات شخصی کاربران جای ندارد به‌همین جهت صرفا از خروجی قطبیت استفاده می‌شود. بنابراین خروجی قطبیت اخبار بایننس کوین در جدول (۳) آورده شده است.

در شکل (۳) براساس خروجی قطبیت به‌دست آمده از جدول (۳)، اخبار از نظر احساسی به سه دسته مثبت، خنثی و منفی تفکیک شد و تعداد هر کدام به‌دست آمد، که در روز جاری تعداد اخبار مثبت و خنثی منتشر شده بسیار بیشتر از اخبار منفی است درنتیجه احتمال رشد قیمت بیشتر است.

جدول (۳) : خروجی قطبیت برای مجموعه اخبار بایننس‌کوین

قطبیت	پیوند	مقالات	عنوان	زمان	تاریخ
0.000	https://gadgettendency.com/central-bank-of-kaz...	The National Bank Of Khazakhstan (NBK) has agre…	Central Bank Of Khazakhstan Integrates Its CBDC …	1 hours ago	2022-10-27 18:32:36
0.068	https://beincrypto.com/bnb-price-prediction-ne...	Binance Coin (BNB) price has increased since f...	BNB Price Prediction : New Bullish Cycle to New...	1 hours ago	2022-10-27 18:32:36
	…	…	…	…	…
0.166	https://www.analyticsinsight.net/no-regre-for...	Binance Coin and TRON coins dipped for the las...	No Regret For Binance and TRON Investors Turni…	6 hours ago	2022-10-27 13:32:38



شکل (۳) : نتیجه تحلیل احساسات اخبار بایننس‌کوین

از نظر تکنیکالی، پیش‌بینی قیمت بایننس‌کوین با تکنیک‌های رایج پیش‌بینی در بورس مشابه است. از طرف دیگر، وقتی در معاملات برای پیدا کردن نقاط ورود و خروج صرفا از تکنیک تحلیل احساسات استفاده شود، ممکن است این استراتژی به‌تنهایی کارساز نباشد و نتایج مطلوب دریافت نگردد.

الگوریتم‌های شبکه عصبی مصنوعی به موضوعی داغ در فضای هوش مصنوعی تبدیل شده است. آن‌ها از یک حلقه تکرارشونده از داده‌ها، اطلاعات و شناسایی الگو از طریق تعدادی چرخه برای تعیین نتیجه پیش‌بینی شده استفاده می‌کنند. به‌همین جهت برای بالا بردن کارایی الگوریتم معاملاتی، رویکرد دیگر به‌کار برده شده در طراحی سیستم معاملاتی این مقاله، پیش‌بینی قیمت روز آینده بایننس‌کوین بااستفاده از شبکه عصبی حافظه طولانی کوتاه مدت است. از آن‌جایی که داده‌های تاریخچه مالی بایننس‌کوین متوالی است، این امر حافظه طولانی کوتاه مدت را به یک الگوریتم عالی برای استخراج الگوهای مناسب برای مدل‌سازی و پیش‌بینی قیمت تبدیل می‌کند.

این شبکه عصبی بر روی بستر tensorflow و keras پیاده سازی شده و قیمت آخرین معامله از مجموعه داده به دست آمده در جدول (۲) مورد استفاده قرار گرفته است. با تقسیم داده‌های قیمتی به داده‌های آموزش و آزمایش می‌توان اثرات اختلاف داده‌های واقعی و پیش‌بینی شده را به حداقل رساند و ویژگی‌های مدل را بهتر درک کرد. پس از این‌که یک مدل بااستفاده از مجموعه آموزشی پردازش شد، دقت مدل با پیش‌بینی در مقابل مجموعه آزمایش مورد بررسی قرار می‌گیرد. به‌همین جهت ۸۰ درصد از مجموعه داده قیمتی برای داده‌های آموزش و ۲۰ درصد باقی‌مانده برای داده‌های آزمایش در نظر گرفته شده است. درنهایت این شبکه به‌گونه‌ای تنظیم شده است که هربار بااستفاده از داده قیمتی ۱۵۰ روز گذشته، قیمت آخرین معامله روز بعد را پیش‌بینی می‌کند.

پیدا کردن مدل درست نیازمند مهارت بالایی است زیرا ساخت یک مدل با لایه‌های درست و تعیین پارامترهای مناسب برای هر کدام نیاز به چندین بار تست و اصلاح دارد. مدلی که برای این مقاله در نظر گرفته شده، کاملاً استاندارد است. این مدل از نوع ترتیبی است. لایه‌های اصلی از نوع حافظه طولانی کوتاه مدت هستند که با تعداد واحدهای متفاوت مقداردهی شدند. این مدل از ۲ لایه حافظه طولانی کوتاه مدت و ۲ لایه متراکم [3] تشکیل شده است. لایه اول حافظه طولانی کوتاه مدت، بازگشتی و شامل ۵۰ نورون است. لایه دوم حافظه طولانی کوتاه مدت نیز شامل ۵۰ نورون، اما به‌دلیل این‌که لایه بعدی از نوع حافظه طولانی کوتاه مدت نمی‌باشد پس این لایه بازگشتی

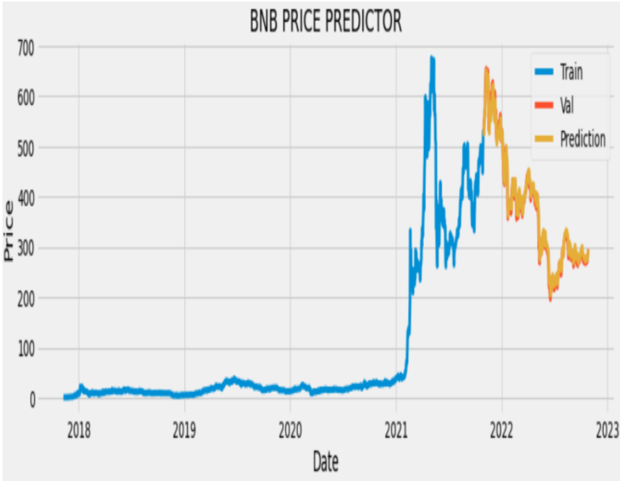
نیست. همچنین لایه دیگر از نوع متراکم، شامل ۲۵ نورون است. درنهایت آخرین لایه نیز از نوع متراکم، اما به‌دلیل این‌که یک خروجی وجود دارد پس شامل ۱ نورون است. برای کامپایل کردن مدل نیز، تابع زیان از نوع میانگین مربعات خطا و همچنین بهینه‌ساز از نوع گرادیان کاهشی adam در نظر گرفته شده است. مطابق با جدول (۴) برای متناسب کردن داده‌ها نیاز به تعیین مقدار دوره[4] و اندازه دسته[5] است. بنابراین همان‌گونه که قابل مشاهده است مقادیر مختلف برای آن‌ها منجر به نتایج متفاوت جذر میانگین مربعات خطا می‌شود. اما به‌دلیل آن‌که هرچه مقدار جذر میانگین مربعات خطا کمتر باشد عملکرد مدل بهتر می‌شود، بنابراین دوره و اندازه دسته‌ایی که منجر به جذر میانگین مربعات خطا کمتر می‌شود، در نهایت انتخاب می‌گردد.

جدول (٤) : نتایج جذر میانگین مربعات خطا

جذر میانگین مربعات خطا	اندازه دسته	دوره	شماره
۱۷.۸۲۸	۱	۱۰۰	۱
۲۱.۷۸۲	۱	۲۰۰	۲
۲۴.۵۸۶	۱	۴۰۰	۳
۱۷.۹۳۴	۱۸	۱۰۰	۴
۱۵.۴۶۱	۱۸	۲۰۰	۵
۳۶.۶۳۵	۱۸	۴۰۰	۶
۴۰.۶۰۰	۵۰	۱۰۰	۷
۲۵.۶۰۸	۵۰	۲۰۰	۸
۴۴.۴۴۴	۵۰	۴۰۰	۹

مطابق با نتایج به‌دست آمده در جدول (۴) کمترین مقدار به‌دست آمده برای جذر میانگین مربعات خطا برابر ۱۵.۴۶۱ است، که براساس مقادیر ۲۰۰ و ۱۸ که به ترتیب برای دوره و اندازه دسته در نظر گرفته شد، به دست آمد.

براساس شکل (۴) که از پیش‌بینی قیمت بااستفاده از شبکه عصبی حافظه طولانی کوتاه مدت به‌دست آمده است، خط آبی و قرمز به‌ترتیب داده‌های آموزش و آزمایش هستند که کل مجموعه داده تاریخی قیمت بایننس‌کوین را تشکیل می‌دهند و خط زرد نتیجه برای پیش‌بینی قیمت است که شبکه با کمترین خطا داده‌های آزمایش و همچنین قیمت روز آینده را پیش‌بینی کرده است.



شکل (٤) : نمودار پیش‌بینی قیمت بایننس‌کوین

همان‌طور که گفته شد در این مقاله هدف استفاده از شبکه عصبی حافظه طولانی کوتاه مدت، پیش‌بینی قیمت یک روز آینده بایننس‌کوین است بنابراین در جدول (۵) قیمت آخرین معامله روز آینده بایننس‌کوین پیش‌بینی شد.

جدول (٥) : نتیجه پیش‌بینی قیمت

آخرین	تاریخ
۲۹۴.۱۷۶	۲۰۲۲-۱۰-۲۸

در جدول (۶)، برای قیمت دو روز اخیر و همچنین قیمت روز آینده که شبکه عصبی حافظه طولانی کوتاه مدت پیش‌بینی کرده است بازده تجمعی قیمت محاسبه شده است تا وضعیت قیمت تحلیل شود. باتوجه به بازده تجمعی قیمت که برابر ۰.۰۱۳ شد و این مقدار از ۰.۰۱ بزرگتر است، نشان از نوسان مثبت دارد. عدد ۰.۰۱ به این دلیل در نظر گرفته شد که عددی بزرگتر از صفر است و این اطمینان را می‌دهد که نوسانی مثبت، هرچند کوچک در حال تشکیل شدن است. اما در صورتی‌که بازده تجمعی قیمت عددی کوچکتر از صفر شد نشان از نوسان منفی دارد.

جدول (٦) : محاسبه بازه تجمعی قیمت

بازده تجمعی	آخرین	تاریخ
NaN	۲۹۰.۳۷۶	۲۰۲۲-۱۰-۲۶
-۰.۰۰۴	۲۸۹.۰۷۹	۲۰۲۲-۱۰-۲۷
۰.۰۱۳	۲۹۴.۱۷۶	۲۰۲۲-۱۰-۲۸

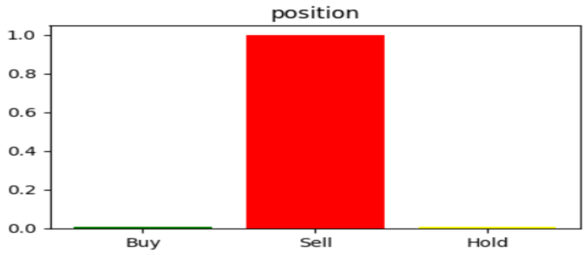
در آخر این سیستم معاملاتی، نتایج به‌دست آمده از تحلیل احساسات و پیش‌بینی قیمت صورت گرفته با شبکه عصبی حافظه طولانی کوتاه مدت که منجر به محاسبه بازده تجمعی قیمت شد را به‌صورت همزمان مورد بررسی قرار می‌دهد تا وضعیت روزانه بایننس‌کوین را تحلیل نماید و هدف اصلی این مقاله که اعلام سیگنال معاملاتی به‌صورت خرید، فروش یا نگهداری بود را انجام دهد.

لازم به ذکر است که این سیستم معاملاتی از آن جهت نیمه خودکار است که صرفا سیگنال خرید یا فروش را اعلام می‌نماید و درنتیجه معامله‌گر را از تحلیل‌های انسانی بی‌نیاز می‌کند اما تصمیم گیرنده نهایی در جهت اجرای سیگنال‌های صادر شده و همچنین مدیریت سرمایه، شخص انسان است.

بنابراین با تحلیل احساساتی که بر روی اخبار بایننس‌کوین در روز جاری انجام گرفت، تعداد اخبار مثبت بیشتر از اخبار منفی است و ازطرف دیگر بازده تجمعی قیمت بزرگتر از ۰.۰۱ به‌دست آمد. بنابراین همان‌گونه که در شکل (۵) قابل مشاهده است سیستم برای روز جاری سیگنال خرید اعلام نموده است.

شکل (٥) : سیگنال خرید

اما برای سایر روزهای معاملاتی برحسب شرایط بازار حالت‌های دیگری ممکن است که رخ دهد. براساس شکل (۶) اگر در یک روز معاملاتی، تحلیل احساسات اخبار صورت گرفته به‌گونه‌ای باشد که تعداد اخبار منفی بیشتر از اخبار مثبت باشد و همچنین بازده تجمعی قیمت نیز عددی کوچکتر از صفر به‌دست آمد آن‌گاه سیستم سیگنال فروش صادر می‌کند.



شکل (٦) : سیگنال فروش

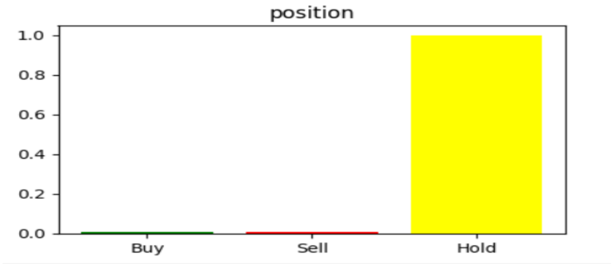
در آخرین حالت نیز اگر در یک روز معاملاتی سیستم به یکی از سه تحلیل زیر برسد:

۱) تعداد اخبار مثبت بیشتر از اخبار منفی باشد اما بازده تجمعی قیمت عددی کوچکتر از صفر به‌دست آمد.

۲) تعداد اخبار منفی بیشتر از اخبار مثبت باشد و بازده تجمعی قیمت عددی بزرگتر از ۰.۰۱ به‌دست آمد.

۳) بازده تجمعی قیمت، عددی در بازه [۰.۰۱ ۰] به‌دست آمد.

آن‌گاه همان‌گونه که در شکل (۷) قابل مشاهده است، سیستم سیگنال نگهداری صادر می‌کند.



شکل (۷) : سیگنال نگهداری

٤- نتیجه گیری

بازار ارزهای دیجیتال به‌دلیل ظهور فناوری اطلاعات و ورود سرمایه زیاد ازطرف سرمایه‌گذاران به سرعت در حال رشد است. این امر محققان را تشویق می‌کند تا بااستفاده از تکنیک‌های جدید به دنبال ایجاد روش‌هایی باشند تا سیگنال‌های معاملاتی را با حداقل خطا بیابند. بنابراین در این مقاله یک سیستم معاملاتی نیمه خودکار پیشنهاد شد که اخبار روزانه را از طریق سایت گوگل نیوز دریافت می‌کند سپس بااستفاده از کتابخانه TextBlob به تحلیل احساسات موجود در اخبار بایننس‌کوین می‌پردازد. قابلیت دیگری نیز برای این سیستم معاملاتی درنظر گرفته شده که داده‌های تاریخی قیمت بایننس‌کوین از طریق سایت یاهو فایننس جمع‌آوری می‌کند، آن‌گاه این داده‌های قیمتی در شبکه عصبی حافظه طولانی کوتاه مدت مورد استفاده قرار می‌گیرد تا از این طریق قیمت روز آینده بایننس‌کوین پیش‌بینی شود. در آخر با جمع‌بندی نتایج حاصل شده از تحلیل احساسات اخبار و پیش‌بینی قیمت صورت گرفته با شبکه عصبی حافظه طولانی کوتاه مدت، این سیستم سیگنال معاملاتی را به‌صورت خرید، فروش یا نگهداری اعلام می‌کند.

مراجع

[1] Qureshi, M., Ahmed, N., *"Forecasting Cryptocurrencies using the Classical Time Series Approach"*, KASBIT Business Journal, Vol.15, pp. 15-27, 2022.

[2] Anson, M., *"Initial exchange offerings: The next evolution in cryptocurrencies"*, The Journal of Alternative Investments, Vol.23, pp. 110-121, 2021.

[3] Ferdiansyah, F., Othman, S. H., Radzi, R. Z. R. M., Stiawan, D., Sazaki, Y., Ependi, U., *"A lstm-method for bitcoin price prediction: A case study yahoo finance stock market"*, 2019 International Conference on Electrical Engineering and Computer Science (ICECOS), pp. 206-210,2019.

[4] Hans, A., Mohadikar, K. R., *"A Coherent Approach to Analyze Sentiment of Cryptocurrency"*, Cyber Security, Privacy and Networking. Springer, Singapore, pp. 155-164, 2022.

[5] Nemes, L., Kiss, A., *"Prediction of stock values changes using sentiment analysis of stock news headlines"*, Journal of Information and Telecommunication, Vol. 5, pp. 375-394, 2021.

[6] Aslam, N., Rustam, F., Lee, E., Washington, P. B., Ashraf, I.,*"Sentiment Analysis and Emotion Detection on Cryptocurrency Related Tweets Using Ensemble LSTM-GRU Model"*, IEEE Access,Vol. 10, pp. 39313-39324, 2022.

[7] Farimani, S. A., Jahan, M. V., Fard, A. M., Haffari, G., *"Leveraging latent economic concepts and sentiments in the news for market prediction"*, In 2021 IEEE 8th International Conference on Data Science and Advanced Analytics (DSAA), pp. 1-10, 2021.

[8] Barua, R., Maity, R., Minj, D., Barua, T., Layek, A. K., *"F-NAD: An application for fake news article detection using machine learning techniques"*, In 2019 IEEE Bombay Section Signature Conference (IBSSC), pp. 1-6, 2019.

[9] Roondiwala, M., Patel, H., Varma, S., *"Predicting stock prices using LSTM"*, European Journal of Business and Management Research, Vol. *6*, pp. 1754-1756, 2017.

[10] Yadav, A., Jha, C. K., Sharan, A., *"Optimizing LSTM for time series prediction in Indian stock market"*, Procedia Computer Science, Vol. 167, pp. 2091-2100, 2020.

[11] Tan, J. P., Ramos, A. L. A., Abante, M. V., Tadeo, Lansigan, R. R., *"A Performance Review of Recurrent Neural Networks Long Short-Term Memory"*, In 2022 3rd International Conference for Emerging Technology (INCET), pp. 1-5, 2022.

[12] Yu, Y., Si, X., Hu, C., Zhang, J., "*A review of recurrent neural networks: LSTM cells and network architectures"*, Neural computation, Vol. 31, pp. 1235-1270, 2019.

[13] Pour, E. S., Jafari, H., Lashgari, A., Rabiee, E., Ahmadisharaf, A., *"Cryptocurrency Price Prediction with Neural Networks of LSTM and Bayesian Optimization"*, European Journal of Business and Management Research, Vol. 2, pp. 20-27, 2022.

[14] Samek, W., Montavon, G., Lapuschkin, S., Anders, C. J., Müller, K. R., *"Explaining deep neural networks and beyond: A review of methods and applications"*, European Journal of Business and Management Research, Vol. 109, pp. 247-278, 2021.

پانویس‌ها

[1] Root Mean Square Error
[2] Polarity
[3] Dense
[4] Epoch
[5] Batch Size

چهارمین کنفرانس ملی انفورماتیک ایران
پژوهشگاه دانش‌های بنیادی، پردیس فرمانیه، تهران
۱۴ و ۱۵ دی ماه ۱۴۰۱

ارائه‌ی بهینه‌ساز بر پایه‌ی گرادیان کاهشی کسری چند متغیره و توزیع شده برای شبکه‌های عصبی

محمد اخوان انوری۱، امیرحسین هادیان رسنان۲، کوروش پرند۳

۱ پژوهشگر، گروه علوم کامپیوتر، پژوهشگاه دانش‌های بنیادی، تهران،
mohammad.akhavan@ipm.ir
۲ پژوهشگر، گروه علوم کامپیوتر، پژوهشگاه دانش‌های بنیادی ، تهران
a.hadianrasanan@ipm.ir
۳ استاد، گروه علوم کامپیوتر و داده‌ها، دانشگاه شهید بهشتی ، تهران
k_parand@sbu.ac.ir

چکیده

توجه به استفاده از مشتقات کسری در حوزه هوش مصنوعی در سال‌های اخیر افزایش یافته است. علت این توجه عملکرد بهتر در همگرایی نسبت به الگوریتم‌های کلاسیک بهینه‌سازی استفاده شده، است. یکی از معیارهای مقایسه عملکرد الگوریتم‌های بهینه‌سازی، استفاده از آن‌ها روی شبکه‌های عصبی است. قابلیت اصلی مشتقات کسری، امکان محاسبه مشتق، در نقاط شکستگی در منحنی‌ها است. ما با استفاده از ویژگی‌های حسابان کسری، عملکرد این الگوریتم‌ها را بهبود داده‌ایم. با ارائه دو الگوریتم جدید در این حوزه، سرعت همگرایی را نسبت به سرعت الگوریتم‌های پیشین افزایش چشمگیری داده‌ایم. با بهبودهای انجام شده، ضمن رقابتی بودن دقت الگوریتم‌های ارائه شده با الگوریتم‌های کلاسیک، سرعت همگرایی شبکه‌های عصبی تا دو و نیم برابر افزایش یافته‌اند.

کلمات کلیدی

شبکه‌های عصبی، حسابان کسری، بهینه سازی، گرادیان کاهشی.

۱- مقدمه

بهینه‌سازی در واقع فرآیند بهبود دادن یک مساله تا جای ممکن است. امروزه، با توجه به افزایش جمعیت، صنعتی شدن زندگی‌ها و حجیم شدن داده‌های تولیدی، نیاز به بهینه‌سازی در تمامی مسائل وجود دارد. استفاده از بهینه سازی فقط منحصر به ریاضیات نبوده و در اکثر مساله‌های روزمره نیز کاربرد دارد. به عنوان مثال، می‌توان در حوزه صنعت به خودروسازی‌ها و شرکت‌های هواپیماسازی و در حوزه تکنولوژی میتوان به عنوان مثالی ساده به نقشه‌های مسیریاب اشاره کرد که تلاش برای یافتن کوتاه‌ترین و همچنین خلوت‌ترین مسیر ممکن را دارند. در ادامه به بررسی تاریخچه و روند توسعه الگوریتم‌های بهینه‌سازی برپایه گرادیان می‌پردازیم.

یکی از اصلی‌ترین روش‌های بهینه سازی که با آغاز رشد هوش مصنوعی به محبوبیت فراوانی دست یافت گرادیان کاهشی نام دارد. گرادیان کاهشی در سال ۱۸۴۷ توسط آگوستین لویس[۱] معرفی شد [1]. همچنین خواص همگرایی این الگوریتم برای مسائل بهینه‌سازی غیرخطی اولین بار در سال ۱۹۴۴ توسط هسکل کری[۲] مورد مطالعه قرار گرفت که این شروعی بر سلطه‌ی روش گرادیان کاهشی در حوزه هوش مصنوعی شد [2][3]. این الگوریتم با استفاده از روش گرادیان‌گیری (محاسبه شیب خط) به پیدا کردن کمینه در تابع‌ها می‌پردازد. فرمول کلی گرادیان به شرح زیر است:

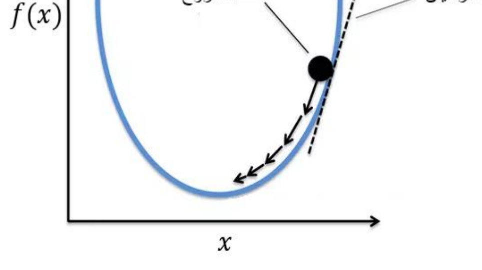
$$Slope = \frac{f(x+\Delta x) - f(x)}{\Delta x},$$

که در اینجا f تابع پیوسته و Δx تغییرات در مقدار x است.

الگوریتم گرادیان کاهشی[۳] برای پیدا کردن مینیمم تابع استفاده می‌شود. این الگوریتم از یک نقطه تصادفی روی منحنی تابع شروع می‌کند و در راستای جهت منفی گرادیان حرکت می‌کند تا به یک کمینه محلی یا سراسری دست پیدا کند [4]. فرمول محاسبه گرادیان روی تابع f به شرح زیر است:

$$\theta_{t+1} = \theta_t - D_x f(x),$$

که در آن f تابع مورد نظر، D_x عملگر گرادیان روی متغیر x و θ نقطه‌ی روی منحنی در قدم $t+1$ و t است. برای سادگی در ادامه عملگر گرادیان روی تابع $f(x)$ در قدم t را به صورت g_t نشان می‌دهیم. شکل ۱ شمای کلی عملیات گرادیان‌گیری روی تابع f را نشان می‌دهد.



شکل (۱): نحوه عملکرد گرادیان و همگرایی به نقطه مینیمم.

این الگوریتم به علت سرعت پایین همگرایی و همچنین گرفتارشدن در کمینه محلی، کارآمدی لازم برای بهینه سازی تابع‌های با پیچیدگی بالا را دارا نبوده و مقیاس‌پذیری پایینی دارد. به همین منظور، با

اضافه‌کردن پارامترِ ثابتِ سرعت یادگیری[۴]، مشکلات گرادیان کاهشی کلاسیک بهبود یافته است. با کاهش سرعت یادگیری، سرعت همگرایی کاهش می‌یابد اما از احتمال عدم همگرایی الگوریتم هم کم می‌شود. فرمول محاسبه گرادیان در شکل با اضافه شدن ضریب η تغییر می‌یابد.

$$\theta_{t+1} = \theta_t - \eta * g_t.$$

توانایی‌های منحصر به فرد شبکه‌های عصبی، آن را به یکی از پدیده‌های قرن ۲۱ تبدیل کرده است. برای مثال شبکه‌های عصبی این توانایی را دارند که بدون دخالت نیروی انسانی و نشان دادن اینکه برای تشخیص یا به دست آوردن برچسب به چه ویژگی‌هایی بلید توجه کرد، ویژگی‌ها را از تصاویر استخراج کنند. برای سادگی می‌توان به شبکه عصبی به چشم یک تابع نگاه کرد که یک ورودی می‌گیرد و یک یا چند خروجی دارد. اما مساله اصلی چگونگی آموزش وزن‌ها برای بهبود این تابع و به دست آوردن خروجی‌های مورد نظر است. اینجاست که روش‌های بر پایه گرادیان نقش بسیار مهمی در مساله بهینه‌سازی شبکه‌های عصبی دارند. دانشمندان با استفاده از معرفی یک تابع هزینه برای مقایسه خروجی تابع(شبکه) و آنچه که مورد انتظار است توانستند معیاری برای مقایسه و میزان خوب برازش‌یافتن تابع‌ها به دست آورند. در ادامه با استفاده از این تابع و کمک گرفتن از الگوریتم گرادیان کاهشی همراه با بازنشر میزان خطا در شبکه، به آموزش وزن‌ها در راستای بهبود دقت و کم شدن هزینه پرداختند [5].

با گذشت سال‌ها و پیشرفت چشم‌گیر تکنولوژی، همراه با بزرگ شدن مجموعه داده‌ها مشکل اصلی الگوریتم گرادیان کاهشی که سرعت همگرایی بود، به آن بازگشت و روی مجموعه داده‌گان بزرگ سرعت همگرایی و عملکرد این الگوریتم به شدت کاهش یافت. برای حل این مساله گرادیان کاهشی تصادفی[۵] مطرح شد. در آن به جای محاسبه‌ی گرادیان برای کل مجموعه داده در هر قدم، به محاسبه گرادیان یک داده تصادفی از مجموعه داده پرداخته شد که تقریب خوبی برای محاسبه‌ی گرادیان مجموعه داده به حساب می‌آید. این تغییر کمک کرد تا الگوریتم گرادیان کاهشی تصادفی در مجموعه داده‌های بزرگ عملکرد بهتر و همگرایی سریع‌تری نسبت به الگوریتم گرادیان کاهشی داشته باشد [6].

از آنجایی که ابعاد داده‌ها رو به افزایش بود مشکل جدیدی در مسائل با داده‌های با فضای بعدی بالا به وجود آمد؛ اگر مقدار سرعت یادگیری را بسیار کوچک انتخاب کنیم مدت زمان لازم برای بهینه شدن بسیار طولانی می‌شود و اگر سرعت یادگیری را بزرگ انتخاب کنیم، حتی با مدت زمان اجرای طولانی نیز ممکن است الگوریتم به دقت مورد نظر نرسد. همچنین با توجه به ابعاد بالای فضای مساله‌ها ممکن است که الگوریتم روی یک بعد بهینه شده باشد اما روی بعد دیگر هنوز به مقدار بهینه خود نرسیده باشد و به دلیل ذکر شده انتخاب یک سرعت یادگیری مشخص برای همه ابعاد مشکل‌زا خواهد بود [7]. در اینجا روش جدیدی با نام آداگراد[۶] ارائه شد. در این روش با تغییر در پارامتر سرعت یادگیری و اضافه کردن یک ضریب در مخرج این پارامتر که بر اساس مقدار گرادیان تعیین می‌شود کمک شایانی به ایجاد سرعت یادگیری متغیر بر اساس هر بعد صورت گرفت. بنابراین اگر در یک بعد گرادیان وزن‌ها بالا باشد سرعت یادگیری کاهش می‌یابد و اگر در بعد دیگری گرادیان وزن‌ها پایین باشد سرعت یادگیری افزایش می‌یابد [8] [9].

در این میان به نظر رسید که استفاده از گشتاور در الگوریتم‌های گرادیانی می‌تواند به سرعت و بهبود عملکرد آن‌ها در فضاهای مختلف کمک کند [10]. الگوریتم بعدی که برای بهبود عملکرد ارائه شد آر-ام-اس پراپ[۷] نام دارد [11]. این الگوریتم کمک زیادی در راستای ارائه سرعت همگرایی زیاد به خانواده الگوریتم‌های گرادیانی کرد. ایده اصلی استفاده از فرمول گشتاور برای تبدیل یک ضریب به گرادیان اصلی بود. برای مثال منحنی سه بعدی را در نظر بگیرید که برای رسیدن به اپتیمم باید در راستای x حرکت زیادی انجام دهد. این الگوریتم با اضافه کردن ضریب رادیکال توان دوم گرادیان در مخرج گرادیان کمک به حرکت سریع‌تر گرادیان به سمت صحیح اپتیمم میکند به این صورت که در راستای y گرادیان زیاد باعث می‌شود مقدار توان دوم گرادیان بزرگ‌تر شود و مقدار ضریب گرادیان کاهش پیدا کند و در راستای x چون مقدار گرادیان کمتر است حرکت سریع‌تر خواهد بود [12][13].

در نهایت معروف‌ترین و پر استفاده‌ترین الگوریتم در شبکه‌های عصبی عمیق و کانولوشنی عمیق الگورتیم آدام[۸] است. این الگوریتم با ترکیب الگوریتم آر-ام-اس پراپ و اضافه کردن فرمول اصلی گشتاور به آن به بهترین نتیجه در خانواده الگوریتم‌های گرادیانی دست می‌یابد. فرمول الگوریتم آدام مقدار گشتاور V_{dw} تقسیم به $(1-\beta_1^t)$ در قدم t و همچنین تقسیم مقدار ضریب الگوریتم آر-ام-اس پراپ بر $(1-\beta_2^t)$ در قدم t است. در این روش چهار متغیر β_1 (متغیر گشتاور) و β_2 (متغیر آر-ام-اس پراپ)، ε و در نهایت متغیر α را داریم که در اکثر مواقع به عنوان هایپر پارامتر تغییر می‌کند [14][15].

در ادامه این مقاله در بخش الگوریتم پیشنهادی به عنوان پیشنیاز به بررسی مفاهیم حسابان کسری و تلفیق آن با گرادیان خواهیم پرداخت. سپس به معرفی روش‌های پیشنهادی و جزئیات ریاضیاتی آن‌ها خواهیم پرداخت. پس از آن در قسمت نتایج عددی، به معرفی معماری شبکه عصبی، مجموعه داده‌گان مورد استفاده، نحوه پیش‌پردازش داده‌ها و در نهایت به بررسی نتایج به دست آمده و مقایسه روش‌های پیشنهادی با الگوریتم‌های پیشین خواهیم پرداخت.

۲- الگوریتم پیشنهادی

۲-۱- پیشنیازها

در سال‌های اخیر تلاش‌هایی برای تلفیق حسابان کسری و الگوریتم گرادیان کاهشی در راستای بهبود سرعت همگرایی و همچنین رفع مشکل اصلی الگوریتم‌های گرادیان کاهشی صحیح، که در منحنی‌های دارای شکستگی عملا کاربردی نخواهند داشت، شده است. در این راستا، ابتدا به معرفی روش‌های حسابان کسری و سپس بررسی الگوریتم‌های ارائه شده در بستر گرادیان کسری می‌پردازیم [16][17].

در سال‌های پیشین روش‌های زیادی برای مشتق‌های کسری ارائه شده. ما در این جا به بررسی دو روش اصلی که بیشترین استفاده را در ارائه گرادیان‌های کسری داشته‌اند می‌پردازیم. از معروف‌ترین گرادیان‌های کسری میتوان به کاپوتو[۹] و ریمان لیویل[۱۰] اشاره کرد که فرمول آن به شرح زیر است:

$$ {}^{RL}_{c}D^{\alpha}_{x} f(x) = \frac{1}{\Gamma(n-\alpha)} \frac{d^n}{dx^n} \int_c^x \frac{f(\tau)}{(x-\tau)^{\alpha-n+1}} d\tau, $$

$$ {}_{c}^{C}D_{x}^{\alpha}f(x) = \frac{1}{\Gamma(n-\alpha)}\int_{c}^{x}\frac{f^{(n)}(\tau)}{(x-\tau)^{\alpha-n+1}}d\tau, $$

که در آن n تعداد قدم‌ها و α ضـریب کسـری اسـت [18]. از آنجایی که در فرمول ابتدایی این روش‌ها انتگرال وجود دارد و در واقع انتگرال خاصـیت حافظه طولانی و غیرمحلی را دارد، امکان اجرای این مدل‌ها در کامپیوترهای امروزی دشـوار می‌شـود به همین دلیل تابع f(x) را با اسـتفاده از سـری تیلور تقریب میزنند که فرمول‌های بالا به فرم زیر تغییر پیدا می‌کنند[19][20].

$$ {}_{c}^{RL}D_{x}^{\alpha}f(x) = \sum_{i=0}^{+\infty}\binom{\alpha}{i}\frac{f^{(i)}(x)}{\Gamma(i+1-\alpha)}(x-c)^{i-\alpha}, $$

$$ {}_{c}^{C}D_{x}^{\alpha}f(x) = \sum_{i=n}^{+\infty}\binom{\alpha-n}{i-n}\frac{f^{(i)}(x)}{\Gamma(i+1-\alpha)}(x-c)^{i-\alpha}, $$

که در آن $\binom{p}{q} = \frac{\Gamma(p+1)}{\Gamma(q+1)\Gamma(p-q+1)}, p \in \mathbb{R}, q \in \mathbb{N}$ ، $f^{(i)}$ مشتق مرتبه i ام، c ثابت و α ضریب کسری است[20].

در سال‌های پیشین روش‌های ارائه شده در حوزه گرادیان کاهشی کسری روی شبکه‌های عصبی بسیار محدود بودند. علیرغم اینکه تلاش‌های انجام شده همچنان از مشکلاتی نظیر سرعت همگرایی، دقت همگرایی و مشکل حافظه بزرگ برخوردار هستند، روش‌هایی برای رفع مشکل حافظه بزرگ و سرعت همگرایی ارائه شده است. این روش‌ها شامل ارائه α به صورت یک تابع به جای یک عدد کسری و ارائه ضریب سرعت همگرایی به صورت یک تابع به جای یک عدد می‌شوند [21][22][23] .

حال ما برای محاسبه مشتق با روش کاپوتو و ارائه بهینه‌ساز کسری از فرمول ساده‌سازی شده آن در مقاله جی لی[۱۱] [24]استفاده می‌کنیم. این روش از روش اصلی تندترین نزول استفاده می‌کند که یکی از روش‌های قدیمی در مسائل بهینه‌سازی بدون محدودیت است [25].

$$ x^{(k+2)} = x^{(k+1)} - \eta_k * f'\left(x^{(k+1)}\right) * \left|x^{(k+1)} - x^{(k)}\right|^{1-\alpha}, $$

که در آن η_k سرعت یادگیری در مرحله k ام، $x^{(k)}$ متغیر در مرحله k ام است.

ما روش جدیدی در راستای بهبود همگرایی الگوریتم‌های کسری ارائه داده‌ایم. در واقع ویژگی مهمی که باعث تمایز بین دو گرادیان صـحیح و اصلی می‌شـود اعمال گرادیان به صـورت غیرمحلی است. با این کار به جای محاسبه‌ی گرادیان روی هر نقطه در روش اول، روی دو نقطه و در روش دوم روی بازه‌ای از منحنی گرادیان اعمال می‌شـود و این ویژگی عامل اصـلی افزایش سرعت همگرایی است.

۲-۲- روش اول:

ما با اسـتفاده از ویژگی غیرمحلی بودن گرادیان‌های کسـری میتوانیم به جای یک نقطه، گرادیان را روی دو نقطه حسـاب کنیم. روش پیشـنهادی ما اسـتفاده از دو ضـریب به جای یک ضـریب در الگوریتم‌های گرادیان کسـری اسـت. به این صـورت که در فرمول اصـلی محاسـبه گرادیان متغیرها برای هر دو متغیر کسـری α_1 و α_2 محاسـبه می‌شـود و هنگام اعمال گرادیان درواقع β_1 و β_2 (که $0<\beta<1$ و $\beta_1 + \beta_2 = 1$) به عنوان ضـریب در گرادیان‌ها اعمال می‌شود. در این روش به جای محاسبه‌ی یک گرادیان، روی دو نقطه گرادیان اعمال خواهد شـد و این روش می‌تواند به سـرعت همگرایی الگوریتم کمک بسـیاری کند. در ادامه فرمول روش ارائه شـده را مشـاهده می‌کنید.

$$ d_1 = D_x^{\alpha_1}f(x), $$

$$ d_2 = D_x^{\alpha_2}f(x), $$

$$ \theta_{t+1} = \theta_t - \eta * (\beta_1 * d_1 + \beta_2 * d_2), $$

که در آن β_1 و β_2 ابرمتغیرها هسـتند که می‌توان به عنوان تاثیر هر گرادیان در تغییرات در نظر گرفت.

۳-۲- روش دوم:

در این روش ما با اسـتفاده از ایده روش اول و بسـط آن به این نتیجه رسـیدیم که میتوانیم به جای استفاده از ۲ یا چند ضـریب کسـری، روی فضای ضرایب کسری انتگرال بگیریم. فرمول اصلی به صورت زیر است:

$$ \int_{\alpha_{min}}^{\alpha_{max}} D_x^{\alpha} f(x)\ d\alpha $$

این روش در واقع به جای محاسـبه گرادیان روی یک متغیر کسـری α ، گرادیان را روی بازه‌ای از متغیرهای کسری α_{min} تا α_{max} محاسبه می‌کند. در ادامه به بررسـی نتایج عددی به دسـت‌آمده در مقایسـه این روش‌ها می‌پردازیم.

۳- نتایج عددی

۱-۳- معماری شبکه

شبکه عصبی استفاده شده به شرح زیر است:

Model: “sequential”

Layer (type)	Output Shape	Param #
dense (Dense)	(None, 10)	110
dense_1 (Dense)	(None, 10)	110

Total params: 220 Trainable params: 220
Non-trainable params: 0

جدول ۱

هر سه لایه اول (ورودی)، پنهان و آخر (خروجی)، شامل ۱۰ نورون است. برای تابع فعال‌ساز لایه میانی تابع یکسـوساز[۱۲] و برای تابع فعال‌ساز لایه خروجی بیشینه‌هموار[۱۳] استفاده شـده اسـت. علت اسـتفاده از این معماری، سـاده بودن انجام عملیات ضـرب‌های ماتریسـی و محاسـبه گرادیان روی وزن‌های شـبکه عصبی است.

تابع هزینه یکی از مهم‌ترین معیارها برای یادگیری شـبکه‌های عصبی است. وظیفه تابع هزینه مقایسه خروجی شبکه با خروجی مطلوب است. در واقع عملگر گرادیان روی تابع هزینه محاسـبات خود را انجام می‌دهد و بهینه‌سـازی روی فضـای داده‌ها با اسـتفاده از تابع هزینه انجام می‌شـود. انواع مختلف تابع هزینه وجود دارد که با توجه به نوع ورودی‌های داده شده به شبکه تابع درسـت انتخاب می‌شـود. برای تابع هزینه نیز آنتروپی متقابل طبقه‌ای[۱۴]

انتخاب شده است. علت این انتخاب نوع برچسب داده‌ها است که از نوع دسته‌بندی ۱۰ گانه است. فرمول تابع هزینه ذکر شده به شرح زیر است:

$$L = -\sum_{i=0}^{n} t_i \log(p_i)\ , for\ n\ classes,$$

که در اینجا t_i برچسب واقعی و p_i احتمال به دست آمده از تابع فعال ساز بیشینه‌هموار برای کلاس iام است. هدف ما بهینه سازی این تابع روی ورودی‌های گرفته و مقایسه خروجی با آنچه که مطلوب خروجی بوده، است. با استفاده از این موضوع، گرادیان روی تابع هزینه را محاسبه می‌کنیم و مقدار گرادیان را از مقادیر وزن‌های شبکه با ضریب سرعت یادگیری کم می‌کنیم. فرمول محاسبه گرادیان روی وزن‌های شبکه عصبی به شرح زیر است:

$$w_{new} = w_{old} - l * DL(w_{old}),$$

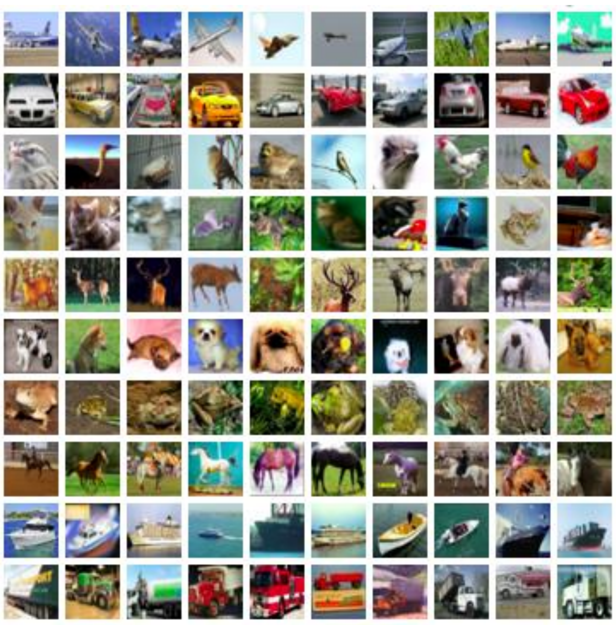
که در آن w_{old} وزن‌های فعلی شبکه، l سرعت یادگیری و L تابع هزینه شبکه و D عملگر گرادیان است.

۲-۳- مجموعه دادگان

مجموعه داده‌گان انتخابی اول، تصویر دست نویس اعداد انگلیسی امنیست[۱۵] است که با تلاش لیکون[۱۶] جمع آوری شده است [26]. این مجموعه داده‌گان از ۷۰ هزار داده تشکیل شده است. هر داده دارای ابعاد ۲۸ پیکسل در ۲۸ پیکسل سیاه و سفید است. این داده به صورت ۶۰ هزار داده آموزشی و ۱۰ هزار داده تست تقسیم‌بندی می‌شود. در این مجموعه، به ازای هر عکس برچسب مربوطه به آن عکس نیز موجود است. از آنجایی که تصاویر این مجموعه داده سیاه و سفید است، تنها یک کانال از داده موجود است. هر پیکسل از صفر(پیکسل سیاه) تا ۲۵۵(پیکسل سفید) است. مساله اصلی این مجموعه داده، تشخیص عدد نوشته شده در تصویر است. این مجموعه داده یکی از اصلی‌ترین معیارهای مقایسه‌ای عملکرد الگوریتم‌ها در حوزه یادگیری ماشین است. نمونه ای از تصاویر به شرح زیر است:

شکل (۲): قسمتی از مجموعه دادگان امنیست.

مجموعه دادگان انتخابی دوم تصاویر طبیعی از وسایل نقلیه و حیوانات(هواپیما، ماشین سواری، پرنده، گربه، گوزن، سگ، غورباقه، اسب، کشتی و کامیون) معروف به سیفار–۱۰[۱۷]، زیر مجموعه‌ای از ۸۰ میلیون تصویر کوچک ارئه شده توسط دانشگاه تورنتو است. این مجموعه داده به نسبت پیچیدگی بسیار بالاتری نسبت به مجموعه داده امنیست دارد چرا که از تصاویر رنگی(سه کاناله) و ۳۲ پیکسل در ۳۲ پیکسل است. همچنین تصاویر موجود در این مجموعه داده به نسبت تصاویر دست‌نویس بسیار پیچیده‌تر است. بخشی از تصاویر در زیر آمده است:



شکل (۳): قسمتی از مجموعه دادگان سیفار–۱۰.

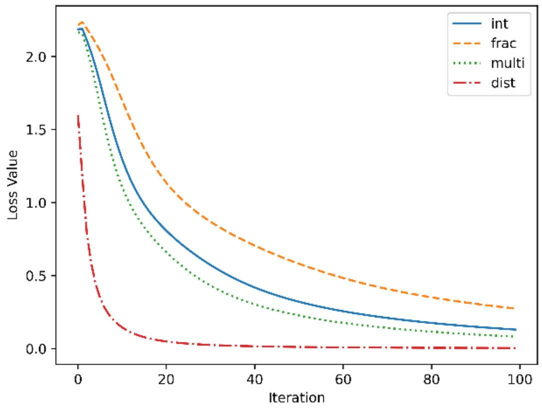
۳-۳- پیش پردازش داده‌ها

به علت این که داده‌ها از نظر ابعاد بزرگ هستند، برای سادگی محاسبات، با استفاده از الگوریتم تحلیل مؤلفه‌های اصلی[۱۸] در کتابخانه سایکیت–لرن[۱۹] اقدام به کاهش ابعاد فضای مساله کرده‌ایم. این الگوریتم به وسیله یک تبدیل خطی متعامد، داده را به دستگاه مختصات جدید می‌برد، به‌طوری‌که یک مرتب سازی بر اساس واریانس داده‌ها انجام داده و با توجه به تعداد مولفه‌های انتخاب شده آن‌ها را به ترتیب بر روی محور مختصات قرار می‌دهد. بر همین اساس مجموعه داده‌هایی را که بیشترین تاثیر در واریانس دارند حفظ می‌کند. تعداد متغیرهای انتخاب شده برای یادگیری ۱۰ است. این برازش بر روی داده‌های تست نیز انجام شده است. در نهایت روی تمامی الگوریتم‌ها از همین داده برای پردازش و انجام محاسبات شبکه و گرادیان استفاده می‌کنیم. همچنین برچسب داده‌ها را با استفاده از روش کدبندی وان هات[۲۰] تغییر داده‌ایم. این کدبندی به این نحو عمل می‌کند که با توجه به تعداد برچسب‌های یکتای مجموعه داده‌گان(در مجموعه داده‌گان امنیست ۱۰) یک آرایه پر شده از عدد صفر درست می‌کند و جایگاه معادل آن برچسب را ۱ می‌گذارد به عنوان مثال برچسب عدد ۵ به آرایه [۰، ۰، ۰، ۰، ۱، ۰، ۰، ۰، ۰، ۰] تبدیل می‌شود.

۴-۳- نتایج اصلی

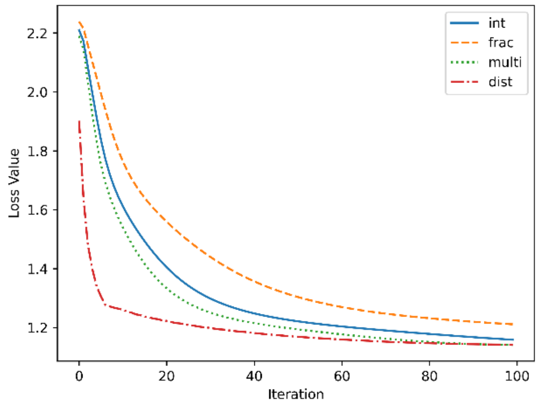
قدم اول، هدف پیاده‌سازی الگوریتم‌های کسری مطرح شده در مقدمه همین مقاله به منظور استفاده آن‌ها به عنوان بهینه‌ساز برای شبکه‌های عصبی است. نکته‌ی اصلی این است که پیاده‌سازی‌های انجام شده قبلی در این زمینه وجود نداشته و ما از ابتدا به برنامه‌نویسی تابع‌های بهینه‌ساز کسری پرداخته‌ایم. همچنین از آنجایی که امکان اضافه کردن تابع بهینه‌ساز کسری به کتابخانه‌های موجود در زبان برنامه نویسی پایتون وجود نداشت، اقدام به بازطراحی تابع‌های لازم برای برازاندن[۲۱]، پیشبینی[۲۲]، پیش‌خور[۲۳] و تابع هزینه[۲۴] مدل شبکه عصبی کردیم. بازطراحی تابع‌های مذکور این امکان را به ما می‌دهد تا فارغ از کتابخانه‌های تنسورفلو و کراس بتوانیم یک اجرای کامل از شبکه‌های عصبی را داشته باشیم و در انتهای هر پیش‌خورانی داده بتوانیم

هزینه مدل را محاسبه و بر اساس آن گرادیان تابع هزینه را روی وزن‌ها اعمال کنیم.
در ادامه به بررسی نتایج به دست آمده می‌پردازیم:



شکل (۴): نتایج به دست آمده روی مجوعه دادگان امنیست و مقایسه بین الگوریتم‌ها.

ابتدا به معرفی نمودار میپردازیم. شکل ۴ نمودار هزینه شبکه عصبی ذکر شده است. خط آبی رنگ(خط ممتد) نمودار هزینه شبکه با بهینه‌ساز گرادیان صحیح است. خط نارنجی(خط چین) نمودار هزینه شبکه با بهینه ساز گرادیان کسری با $\alpha=۰/۹$. خط سبز رنگ(نقطه چین) نمودار هزینه شبکه با روش پیشنهادی اول بهینه ساز گرادیان کسری دوتایی، با $\alpha_1=۰/۹$ و $\alpha_2=۱/۱$. در نهایت خط قرمز رنگ(نقطه-خط) نمودار هزینه شبکه با روش پیشنهادی دوم بهینه‌ساز گرادیان کسری توزیع شده در بازه $\alpha_1=۰/۹$ تا $\alpha_2=۱/۱$. تمامی الگوریتم‌ها با سرعت یادگیری ۰/۰۳ انجام شده است. همانطور که در شکل شماره ۴ مشاهده می‌شود، روند تمامی روش‌ها به طور کلی نزولی است. نمودار گرادیان صحیح در ابتدا با شیب سریع رو به پایین بوده اما پس از چند مرحله‌ی انجام عملیات گرادیان، شیب هزینه شبکه کاهش پیدا می‌کند. همانطور که در نمودار مشخص است، گردایان کسری از گرادیان صحیح عملکرد بدتری داشته و دلیل آن این است که انتخاب α مناسب برای به دست آوردن نتیجه بهتر کار دشواری است. همچنین مقدار اولیه وزن‌های شبکه تاثیر زیادی در روند انتخاب α می‌گذارند. از آنجا که روش اول ارائه شده در این مقاله در برگیرنده دو متغیر α است، تاثیرگذاری مقدار اولیه‌ی وزن‌ها پایین‌تر است و همانطور که در شکل ۴ مشاهده می‌کنید، روند کاهش سرعت میزان خطا در روش گرادیان کسری دو متغیره، بیشتر از هر دو روش گرادیان صحیح و گرادیان کسری با یک متغیر است. اما روش دوم که گرادیان کسری توزیع شده است نشان می‌دهد که سرعت همگرایی این روش از همه روش‌های مطرح شده بالاتر است. همچنین همانطور که مشاهده می‌کنید، در ۱۰۰ مرحله گرادیان‌گیری، بهترین عملکرد را در راستای کاهش میزان خطای شبکه به همراه داشته است. از نظر میزان نرخ کاهش هزینه شبکه عصبی روش ارائه شده اول نسبت به روش گرادیان کاهشی صحیح در ۵۰ قدم ابتدایی ۱۷ درصد بهبود و نسبت به روش گرادیان کاهشی کسری ۵۴ درصد بهبود عملکرد داشته اما میزان نرخ کاهش هزینه شبکه عصبی در روش دوم ارائه شده نسبت به گرادیان کاهشی صحیح ۵/۸ برابر بهبود و نسبت به گرادیان کاهشی کسری ۷/۵ برابر بهبود یافته است.



شکل (۵): نتایج به دست آمده روی مجوعه دادگان سیفار-۱۰ و مقایسه بین الگوریتم‌ها.

همانطور که در شکل ۵ مشاهده می‌کنید روند نتایجی که الگوریتم‌ها روی مجموعه داده‌ی امنیست داشته‌اند دوباره تکرار شده است و این نشان می‌دهد که روش‌های جدید ارائه شده از پایداری خوبی برخوردار هستند. همانطور که ذکر شد مجموعه داده سیفار-۱۰ از تصاویر رنگی ۳۲ پیکسل در ۳۲ پیکسل وسایل نقلیه و حیوانات تشکیل شده است و این موضوع پیچیدگی زیادی روی فضای مساله ایجاد کرده است. با این حال روش‌های ارائه شده‌ی اول و دوم دارای سرعت همگرایی بالاتری نسبت به روش‌های معمول گرادیان کاهشی کسری و گرادیان کاهشی صحیح داشته. همانطور که در شکل ۵ مشاهده می‌کنید روش‌های ارائه شده در ۱۰۰ مرحله گرادیان‌گیری، بهترین عملکرد را در راستای کاهش میزان خطای شبکه به همراه داشته است. از نظر میزان نرخ کاهش هزینه شبکه عصبی در مجموعه داده‌ی دوم، روش ارائه شده اول نسبت به روش گرادیان کاهشی صحیح در ۵۰ قدم ابتدایی ۳ درصد بهبود و نسبت به روش گرادیان کاهشی کسری ۱۳ درصد بهبود عملکرد داشته استاما میزان نرخ کاهش هزینه شبکه عصبی در روش دوم ارائه شده نسبت گرادیان کاهشی صحیح ۷۶ درصد بهبود و نسبت به گرادیان کاهشی کسری ۱/۲ برابر بهبود یافته است.

۴- نتیجه‌گیری

ما با ارائه دو روش جدید در خانواده الگوریتم‌های گرادیانی کسری توانستیم که سرعت همگرایی و همگرایی بیشتر نسبت به الگوریتم‌های پیشین را بهبود بخشیم. همچنین در نتیجه‌ی این دستاوردها، بهبود عملکرد کلی شبکه‌های عصبی در فضاهای عمومی برای حل مسائلی با پیچیدگی‌های فضایی بیشتر مشاهده شده است. قدم بعدی ارائه تابع هزینه‌های دارای شکستگی و استفاده از روش‌های ارائه شده برای حل این‌گونه مساله‌ها است. امید این است که با افزایش استفاده از این نوع بهینه‌سازها بتوانیم در راستای توسعه و افزایش عملکرد این نوع الگوریتم‌ها قدم برداریم.

مراجع

[1] Lemaréchal C, "*Cauchy and the Gradient Method*"
[2] Curry, Haskell B, "*The Method of Steepest Descent for Non-linear Minimization Problems*".
[3] Akilov, G. P.; Kantorovich, L. V, "*Functional Analysis*".

[4] Sebastian Ruder, "*An overview of gradient descent optimization algorithms*".
[5] Shalev-Shwartz, Ben-David, "*Understanding Machine Learning: From Theory to Algorithms*".
[6] Nikhil Ketkar, "*Stochastic Gradient Descent*".
[7] Yu Cheng, I Diakonikolas, R Ge, M Soltanolkotabi, "*High-dimensional robust mean estimation via gradient descent*".
[8] A Lydia, S Francis, "*Adagrad—an optimizer for stochastic gradient descent*".
[9] Duchi, John Hazan, Elad; Singer, Yoram, "*Adaptive subgradient methods for online learning and stochastic optimization*".
[10] N Qian, "*On the momentum term in gradient descent learning algorithms*".
[11] Hinton, Geoffrey, "*Lecture 6e rmsprop: Divide the gradient by a running average of its recent magnitude*".
[12] DV Babu, C Karthikeyan, A Kumar, "*Performance analysis of cost and accuracy for whale swarm and rmsprop optimizer*".
[13] M Huk, "*Stochastic optimization of contextual neural networks with RMSprop*".
[14] A Tato, R Nkambou, "*Improving adam optimizer*".
[15] Kingma, J Ba, "*Adam: A Method for Stochastic Optimization*".
[16] Y Wei, Y Kang, W Yin, Y Wang, "*Design of generalized fractional order gradient descent method*".
[17] G Yang, B Zhang, Z Sang, J Wang, H Chen, "*A Caputo-type fractional-order gradient descent learning of BP neural networks*".
[18] ZA Khan, NI Chaudhary, S Zubair, "*Fractional stochastic gradient descent for recommender systems*".
[19] Y Wei, Y Kang, W Yin, Y Wang, "*Generalization of the gradient method with fractional order gradient direction*".
[20] J Liu, R Zhai, Y Liu, W Li, B Wang, L Huang. "*A quasi fractional order gradient descent method with adaptive stepsize and its application in system identification*".
[21] S Khan, I Naseem, MA Malik, R Togneri, M Bennamoun, "*A fractional gradient descent-based rbf neural network*".
[22] C.J.Zuñiga AguilaraJ.F.Gómez-AguilarbV.M.Alvarado-MartínezaH.M.Romero-Ugalde, "*Fractional order neural networks for system identification*".
[23] Y Chen, G Zhao, "*A Caputo-type fractional-order gradient descent learning of deep BP neural networks*".
[24] J. Liu , X. Xu , X. Cui, "*An accelerated nonmonotone trust region method with adaptive trust region for unconstrained optimization*".
[25] Y Chen, Q Gao, Y Wei, Y Wang, "*Study on fractional order gradient methods*".
[26] Y LeCun, "*The MNIST database of handwritten digits*".

پانویس‌ها

[1] Augustin-Louis Cauchy
[2] Haskell Curry
[3] Gradient Descent
[4] Learning Rate
[5] Stochastic Gradient Descent
[6] AdaGrad
[7] RMSProp
[8] Adam
[9] Caputo
[10] Riemann–Liouville
[11] J. Liu
[12] Relu
[13] Softmax
[14] Categorical Cross-Entropy
[15] MNIST
[16] Lecun
[17] CIFAR-10
[18] PCA
[19] Scikit-Learn
[20] One Hot Encoding
[21] Fit
[22] Predict
[23] Feed-Forward
[24] Loss Function

تعیین سطح دانشجویان در آموزش مجازی با تحلیل سری زمانی جریان کلیک

حدیث عزیزی[۱]، صادق سلیمانی[۲] و پرهام مرادی[۳]

[۱] دانشجو ارشد مهندسی نرم افزار، دانشکده مهندسی کامپیوتر، دانشگاه کردستان، سنندج
hadis.azizi@uok.ac.ir
[۲] استادیار تخصصی بیوانفورماتیک، دانشکده مهندسی کامپیوتر، دانشگاه کردستان، سنندج
s.sulaimany@uok.ac.ir
[۳] دانشیار تخصصی یادگیری ماشین، دانشکده مهندسی کامپیوتر، دانشگاه کردستان، سنندج
p.moradi@uok.ac.ir

چکیده

داده‌های سری زمانی آموزش الکترونیک مانند داده‌های جریان کلیک و داده‌های بایگانی اهمیت زیادی در رابطه با کشف الگوی‌های رفتاری دانشجویان آن سامانه‌ها دارد. بیشتر پژوهش‌های انجام شده تاکنون بر داده‌های جریان کلیک، مبتنی بر روش‌های غیر شبکه‌ای و مرتبط با الگوریتم‌های یادگیری ماشین بوده است. با توجه به اهمیت الگوریتم‌های گراف پدیداری و نقش آن در تحلیل داده‌های سری زمانی، با نگاشت آن به حیطه داده‌های آموزشی می‌توان از این پتانسیل بهره برد. هدف این پژوهش گردآوری میانگین کلیک دانشجویان ممتاز و ضعیف براساس زمان‌های فعالیت در سامانه و تبدیل به گراف پدیداری طبیعی و بررسی با معیارهای تحلیل شبکه مانند مرکزیت درجه، مرکزیت نزدیکی و توزیع درجه برای تشخیص تمایز آن‌هاست. داده‌های پژوهش از سامانه OULAD که حاوی اطلاعات تعامل کاربران با سامانه آموزشی است، انتخاب شده است. نتایج نشانگر این است که بالاتر بودن مرکزیت درجه، مرکزیت نزدیکی و توزیع درجه متمایز کننده دانشجو ممتاز از ضعیف و تایید کننده عدم شباهت رفتاری دانشجویان براساس میانگین کلیک‌هایشان در سامانه است. میانگین دقت این معیارها با تحلیل جداساز خطی و الگوریتم‌های درخت تصمیم، جنگل تصادفی و KNN بعد از ۱۰ بار اجرا به ترتیب ۶۳٪، ۶۰٪ و ۵۷٪ است، که نشانگر دقت مطلوب این معیارها در دسته‌بندی است.

کلمات کلیدی

داده‌های سری زمانی آموزش الکترونیکی، داده‌های جریان کلیک، گراف پدیداری طبیعی، تحلیل گراف پدیداری، شبکه‌های پیچیده، تحلیل شبکه

۱- مقدمه

داده‌های سری زمانی آموزش الکترونیک مانند داده‌های جریان کلیک[۱] و داده‌های بایگانی[۲] سامانه‌های آموزشی تاکنون توسط الگوریتم‌ها و روش‌های متنوعی در حیطه داده‌کاوی آموزشی مانند الگوریتم‌های یادگیری ماشین مورد مطالعه قرار گرفته است. خروجی تحلیل این داده‌ها، باعث نتایج مفیدی جهت سازمان دهی مجدد وب سایت‌های آموزشی و ارائه اصلاحات و پیشنهادات برای بهبود فرایند آموزش و کشف الگوهای نهان در این داده‌ها بوده است[1]. از سوی دیگر ماهیت سری زمانی داده‌های جریان کلیک، مستعد استفاده از روش‌هایی است که می‌تواند آن را به فضای تحلیل گراف نگاشت کند و سبب ایجاد ارزش افزوده و انجام تحلیل‌های جدید شود. این مهم با کمک گراف پدیداری میسر است و تاکنون مورد بررسی قرار نگرفته است. پس از تبدیل داده‌های سری زمانی جریان کلیک به گراف پدیداری، می‌توان از روش‌های تحلیل گراف پدیداری به منظورهای مختلف بهره برد. روش‌های تحلیل گراف پدیداری تاکنون در حیطه‌های مختلف علمی مانند پزشکی، اقتصاد، زمین‌شناسی و غیره مورد مورد بهره‌برداری قرار گرفته است.

در این مقاله هدف آن است که برای اولین بار، ایده گراف پدیداری در داده‌های آموزش الکترونیکی مورد استفاده قرار گیرد. بدین منظور ابتدا با در اختیار داشتن اطلاعات داده‌های آموزش الکترونیکی و جریان کلیک مرتبط با کاربران، نسبت به پالایش و استخراج جریان کلیک مطلوب اقدام خواهد شد. سپس دو سری زمانی مربوط به الگوی کلیک دانشجویان ممتاز و ضعیف ایجاد می‌شود و پس از تبدیل به گراف پدیداری طبیعی، از منظر تحلیل گراف پدیداری مورد بررسی قرار می‌گیرد.

ساختار پژوهش به صورت خلاصه به این شرح است: ابتدا در فصل ۲، مروری بر تحلیل داده‌های سری زمانی آموزش الکترونیک کرده و به زمینه‌های کاربردی مطرح شده در زمینه تحلیل گراف پدیداری پرداخته می‌شود. سپس در بخش ۳ به شرح مجموعه داده و معرفی روش پژوهش اشاره می‌شود و مفهوم گراف پدیداری را شرح داده و معیارهای تحلیل شبکه‌های پیچیده معرفی می‌شود. در بخش ۴ نتایج تحلیل گراف پدیداری ارائه می‌شود. نتایج موجود برای دانشجویان ممتاز و دانشجویان ضعیف با پارامترهای تحلیل شبکه پیچیده و الگوریتم‌های دسته‌بندی گوناگون مقایسه و تحلیل می‌شوند. درگام آخر در بخش ۵ به نتیجه گیری و کارهای آینده اشاره خواهد شد.

۲- پیشینه پژوهش

۲-۱- مروری بر تحلیل داده‌های سری زمانی آموزش الکترونیک

تاکنون پژوهش‌های متنوعی در زمینه تحلیل داده‌های آموزشی مانند مسئله پیش‌بینی عملکرد دانش‌آموزان با استفاده از یک مدل حافظه کوتاه‌مدت عمیق[۳] با استفاده از مجموعه داده‌های تجزیه و تحلیل یادگیری دانشگاه صورت گرفته است[2]. در پژوهشی دیگر در سال ۲۰۲۱ یک مدل پیش‌بینی پیشنهاد شده که با الگوریتم‌های مختلف یادگیری ماشین و یادگیری عمیق آموزش داده می‌شود تا رفتار یادگیری دانش‌آموزان را با توجه به متغیرهای مطالعه آنها مشخص کند، تا مشکلاتی را که دانش‌آموزان در معرض خطر با

آن‌ها مواجه هستند، تجزیه و تحلیل کند، و متعاقباً، مربیان با مداخله به موقع، مشکل را تسهیل کنند[3]. در مطالعه بعدی با تحلیل داده‌های جریان کلیک برای سنجش، درک و حمایت از یادگیری خودتنظیمی[۴] در دوره‌های آنلاین استفاده می‌شود، این پژوهش از داده‌های جریان کلیک جمع آوری شده از یک سیستم مدیریت یادگیری، برای اندازه‌گیری دو جنبه یادگیری خودتنظیمی استفاده می‌کند: مدیریت زمان و تنظیم تلاش. که درک فرآیند یادگیری خودتنظیمی و شناسایی دانش‌آموزانی که ممکن است با شرکت در کلاس‌های آنلاین به خوبی به آنها خدمت نکنند، را راهنمایی می‌کند[4]. در مطالعه‌ای در سال ۲۰۲۰ به تحلیل بصری داده‌های جریان کلیک ویدیویی پرداخته شد و پیش‌بینی عملکرد فراگیران با استفاده از مدل‌های یادگیری عمیق در دوره‌های MOOC[۵] بررسی شد[5]. در پژوهشی دیگر در سال ۲۰۲۲ با تحلیل داده‌های جریان کلیک توسط الگوریتم‌های یادگیری ماشین، به بررسی الگوهای رفتاری دانشجویان مرتبط با موفقیت و شکست آن‌ها در وظایف تعاملی پرداخته شده است[6].

۲-۲- مروری بر تحلیل گراف پدیداری

تحلیل گراف پدیداری به عنوان ابزاری مفید برای تحلیل سری‌های زمانی در زمینه‌های متنوع پژوهشی به کار می‌رود و نتایج مفیدی در حیطه‌های مختلفی مانند پزشکی، زیست شناسی، اقتصادی، زمین شناسی، مدیریت آب و پردازش تصویر و غیره تولید می‌کند. برخی از خروجی‌های تحلیل این نوع بررسی بازده بازار با راهکارهای استراتژیک از طریق تحلیل ارزهای دیجیتال مثل تحلیل سوابق قیمت بیت کوین جهت درک تعمیق این بازارپرداخته شده است[7]. استفاده از تحلیل گراف پدیداری با پارامتر PSVG[6]، برای داده‌های سری زمانی ضربان قلب برای افراد مبتلا به CHF[7] و افراد سالم، نشان می‌دهد که گراف پدیداری افراد سالم به صورت قابل توجهی PSVG بالاتری نسبت به گراف پدیداری افراد سالم دارا هستند و می‌تواند به عنوان یک ابزار بالینی جدید و مفید برای طبقه‌بندی خطر حمله قلبی عمل کند[8].

۳- روش تحقیق

با توجه به بررسی مقالات مرتبط با روش‌های تحلیل داده‌های سری زمانی آموزش الکترونیک و اهمیت بالای تحلیل داده‌های جریان کلیک، مشخص است براساس بررسی‌های ما، تاکنون دید شبکه‌ای به این نوع داده‌ها صورت نگرفته است. نوآوری این پژوهش عبارت است از نگاشت داده‌های جریان کلیک آموزش الکترونیک به گراف پدیداری برای نخستین بار و تحلیل الگوی میانگین کلیک دانشجویان ممتاز و ضعیف براساس پارامترهای متمایز کننده تحلیل شبکه‌های پیچیده است. در ادامه به جزییات داده‌ها و روش پژوهش پرداخته می‌شود.

۳-۱- شرح مجموعه داده و روش پژوهش

در این پژوهش از مجموعه داده متن باز OULAD[8] که شامل داده‌هایی درباره دوره‌ها، دانشجویان و تعامل آنها با محیط یادگیری مجازی برای هفت دوره انتخابی (موسوم به ماژول) است، استفاده کردیم. از فایل studentAssessment که شامل نتایج ارزیابی دانشجویان است استفاده کردیم و ستون‌های مرتبط آن مانند شناسه دانشجو، تعداد روزهایی که از شروع ارائه ماژول می‌گذرد و نمره‌ای که تا آن روز مشخص برای آن دانشجو خاص ثبت شده است، را استخراج کردیم. سپس از فایل studentVle که حاوی اطلاعاتی در مورد تعاملات هر دانشجو با گزینه‌های موجود در محیط یادگیری مجازی است، استفاده کردیم و ستون‌های مرتبط آن مانند شناسه دانشجو، تاریخ تعامل دانشجو با وب سایت آموزشی و مجموع کلیک‌های هر دانشجو که تا آن روز مشخص ثبت شده است، را در نظر گرفتیم. شروع دوره آنلاین از روز صفرم است، همچنین قابل توجه است که تاریخ تعامل و کلیک دانشجو با وب سایت آموزشی قبل از شروع دوره آموزشی نیز ثبت شده است. یعنی تعداد روزهایی که به شروع دوره آنلاین زمان باقی مانده است اما دانشجو وارد سامانه شده و کلیک کرده است که با تاریخ منفی در مجموعه داده ذخیره شده است.

جدول (۱): شرح مجموعه داده انتخاب شده

ستون انتخابی	شرح فایل	فایل انتخابی از مجموعه داده OULAD
id_student, date_submitted, score	حاوی نتایج ارزیابی دانشجویان	studentAssessment
id_student, date, sum_click	حاوی اطلاعات تعامل دانشجویان با سامانه	studentVle

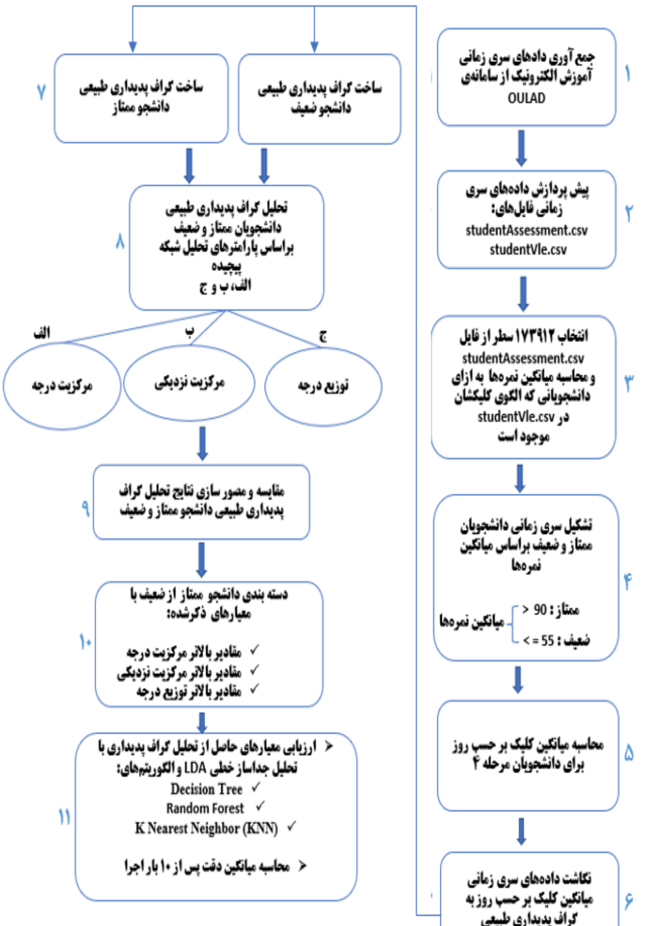
براساس جدول (۱)، ابتدا از فایل studentAssessment تمام سطرها را جدا کرده و نمرات دانشجویان و زمان فعال در سامانه، را در نظر گرفته و میانگین نمرات دانشجویانی که الگوی کلیک‌شان در فایل studentVle موجود بود را محاسبه کردیم. با استفاده از میانگین نمرات هر دانشجو، دانشجویان را دسته‌بندی کرده و از بزرگ به کوچک مرتب کردیم. سپس دانشجویان ممتاز را با فرض میانگین نمرات بالاتر از ۹۰ و دانشجویان ضعیف را با فرض میانگین نمرات پایین‌تر از ۵۵ را به عنوان دانشجویان ضعیف در نظر گرفتیم، سپس برای هر دو سری دانشجویان ممتاز و ضعیف گراف پدیداری طبیعی‌شان را به ازای میانگین کلیک‌شان بر حسب تعداد روزهایی که از شروع ارائه ماژول اندازه گیری می‌شود را ساختیم.

در واقع سری زمانی از میانگین کلیک این دو سری دانشجویان از ۲۴ روز مانده به شروع کلاس‌های آنلاین، تا اخرین روز کلیک‌شان در سامانه آنلاین تشکیل داده شده و معیارهای تحلیل شبکه را برای دو گراف پدیداری ساخته شده تحلیل و بررسی کردیم. جزییات گام به گام روند انجام پژوهش در شکل(۱) به صورت دقیق‌تر در ۱۱ گام توصیف شده است.

در گام ۱ ابتدا داده‌های سری زمانی از سامانه آموزش الکترونیکی OULAD جمع آوری شده و داده‌های مرتبط با پژوهش را ذخیره کرده و در گام ۲ روی آن پیش پردازش انجام شده است. در گام ۳ تمام سطرها را جدا کرده و میانگین نمرات دانشجویان را محاسبه کرده سپس به فایل دیگر studentVle، مراجعه کرده که حاوی تعاملات دانشجو با سامانه آنلاین است، و الگوی کلیک دانشجویانی که معدل‌شان را محاسبه کردیم، پیدا کرده و آن دسته از دانشجویانی که الگوی کلیک‌شان موجود نبود، حذف شده‌اند.

سپس در گام ۴ دو سری زمانی از دانشجویان براساس معدل محاسبه شده تشکیل دادیم و دانشجویانی که میانگین نمراتشان بالاتر از ۹۰ بود را به عنوان سری دانشجویان ممتاز و دانشجویانی که میانگین نمراتشان پایین تر از ۵۵ بود را به عنوان سری دانشجویان ضعیف در نظر گرفتیم و در گام ۵

میانگین کلیک دانشجویان گام ۴ ام را بر حسب روز محاسبه کردیم. در گام ۶ و ۷ گراف پدیداری طبیعی میانگین کلیک بر حسب روز را برای هر دو سری دانشجویان برتر و ضعیف رسم کردیم سپس در گام ۸ گراف پدیداری حاصل شده برای هر دوسری دانشجویان ممتاز و ضعیف را براساس معیارهای تحلیل شبکه پیچیده عنوان شده مانند مرکزیت درجه، مرکزیت نزدیکی و توزیع درجه تحلیل کردیم. نتایج تحلیل دو گراف پدیداری طبیعی حاصل شده براساس پارامترهای متمایز کننده در این تحلیل مانند مرکزیت درجه، مرکزیت نزدیکی، توزیع درجه که مشخص کننده دانشجویان ممتاز است، را با تحلیل جداساز خطی[9] و الگوریتم‌های درخت تصمیم[10]، جنگل تصادفی[11] و KNN[12] پس از ۱۰ بار اجرا ارزیابی کردیم که دقت دسته‌بندی و جداسازی دانشجو ممتاز از ضعیف را با دقت مطلوبی در گام ۱۱ نتیجه گرفتیم.

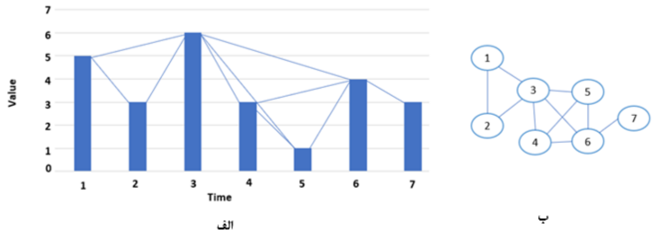


شکل (۱) : فرایند گام به گام تمایز دانشجویان ممتاز از ضعیف با کمک نگاشت الگوی جریان کلیک به گراف پدیداری و استفاده از تحلیل گراف پدیداری

۳-۲- آشنایی با گراف پدیداری

الگوریتم‌های گراف پدیداری ساده نخستین بار توسط لاکاسا و همکارانش در سال ۲۰۰۸ معرفی شد. گراف طبیعی یا ساده با جایگذاری نقاط متوالی سری زمانی بر روی محور زمان که هر نقطه نیز داری مقدار است که به صورت میله‌های عمودی نشان داده می‌شود، مدل می‌شود. چنانچه اگر سر این میله‌های عمودی از هم قابل رویت باشد (میله‌ای بین خط دید مستقیم آن‌ها حایل نباشد) یک یال بین آن دو نقطه گراف رسم می‌شود. به عبارت دیگر شرط رسم یال، قابل مشاهده بودن سر میله‌ها از همدیگر با وجود میله میانی است. به زبان ریاضی شرط ایجاد اتصال (یال) بین دو سرمیله (t_i,y_i) و (t_j,y_j) با وجود میله میانی (t_n,y_n) صدق کردن در معادله (۱) است[9].

$$y_n < y_j + (y_i - y_j)\frac{t_j - t_n}{t_j - t_i} \qquad (۱)$$



شکل (۲) : الف) نمودار میله‌ای سری زمانی (نقاط زمانی بر محور افقی، گره‌های گراف هستند و خطوط متصل کننده مقادیر، یال‌های گراف هستند که با شرط صدق کردن در رابطه ۱ ترسیم شده‌اند.) ب) گراف حاصل از نگاشت سری زمانی

گراف رسم شده در شکل (۲) گراف پدیداری طبیعی بدون جهت، بدون وزن و همبند است. انواع دیگری از گراف پدیداری برحسب روش ساخت آن وجود دارد که می‌توان گراف پدیداری افقی[10] و یا گراف پدیداری نفوذپذیر محدود را نام برد[11]. در این پژوهش از گراف پدیداری طبیعی استفاده خواهد شد.

۳-۳- معیارهای تحلیل شبکه پیچیده

بعد از ساخت گراف‌های پدیداری براساس میانگین کلیک‌های دو دسته دانشجوی ممتاز و ضعیف بر حسب تعداد روزهایی که از شروع ارائه ماژول می‌گذرد، مقایسه توسط معیارهای مرکزیت درجه، مرکزیت نزدیکی، توزیع درجه انجام خواهد شد و خروجی این تحلیل‌ها با جزییات بیشتر بررسی و مصورسازی می‌شود. انتخاب این معیارها به ترتیب براساس کارکرد مشابه در حیطه‌های متفاوت غیر آموزشی [12]،[13]، [14] بوده است.

۳-۳-۱- مرکزیت درجه

در یک گراف مرکزیت درجه[13] با تعداد کل پیوندهای مستقیم با سایر گره‌ها اندازه گیری می‌شود. طوری که در رابطه (۲) داریم :

$$C'd(Ni) = \frac{\sum_{j=1}^{n} xij}{(n-1)\,(n-2)} \quad (i \neq j) \qquad (۲)$$

$C'd$ به معنی تعداد پیوندهایی است که مستقیماً به گره N متصل می‌شوند و n به معنای تعداد کل گره‌ها در شبکه کانونی است[15].

۳-۳-۲- مرکزیت نزدیکی

مرکزیت نزدیکی [۱۴]، به معنای اندازه‌گیری فاصله‌های مجموع یک گره نسبت به گره‌های دیگر است، اگر طول کوتاه‌ترین مسیرهای گره N با سایر گره‌های شبکه کوچک باشد، گره N مرکزیت نزدیکی بالایی دارد. این نشان دهنده راحتی و سهولت اتصالات بین گره متمرکز و سایر گره‌ها است. همان طور که در رابطه (۳) مشخص است: $\sum_{j=a}^{n} d(Ni, Nj)$ به معنای تعداد کل گام‌ها از گره N به سایر گره‌های شبکه است[15].

$$C'c(Ni) = \frac{n-1}{\sum_{j=a}^{n} d(Ni, Nj)} \quad (i \neq j) \qquad (۳)$$

۳-۳-۳- توزیع درجه

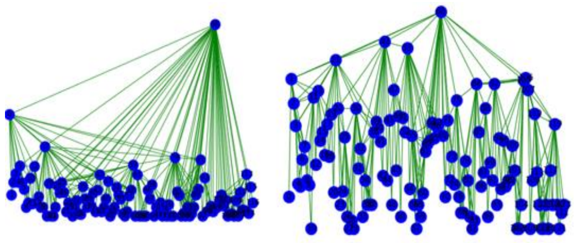
در شبکه‌های پیچیده توزیع درجه [۱۵] معیار مهمی برای تحلیل انواع شبکه است. احتمال توزیع درجه گره با درجه K به صورت P_K نشان داده می‌شود و به صورت رابطه (۴) تعریف می شود:

$$P(K) = \frac{Nk}{n} \qquad (۴)$$

که در آن N_k تعداد گره‌های با درجه k است و n تعداد کل گره‌ها است[16].

۴- نتایج و بحث

در این قسمت تلاش خواهد شد تا یافته‌های پژوهش از دیدگاه‌های مختلف عرضه و تحلیل شود. تصویر کلی گراف با تعداد گره کمتر بر حسب شکل (۳) است و آمارهای اولیه این گراف‌ها طبق جدول (۲) می‌باشد:



شکل (۳) : گراف پدیداری طبیعی حاصل از نگاشت سری زمانی میانگین کلیک دانشجویان بر حسب روز
الف) دانشجویان ضعیف ب) دانشجویان ممتاز

جدول (۲): الف) شاخصه‌های اصلی شبکه ایجاد شده براساس گراف پدیداری میانگین کلیک دانشجویان ضعیف و ممتاز

Student Group	Node Count	Edge Count	Mean Degree
High_level	289	30386	210.28
Low_level	292	28119	192.59
Student Group	**Diameter**	**Max Degree**	**Transitivity**
High_level	2	273	0.82
Low_level	3	288	0.81

جدول (۲): ب) ادامه‌ی شاخصه‌های اصلی شبکه

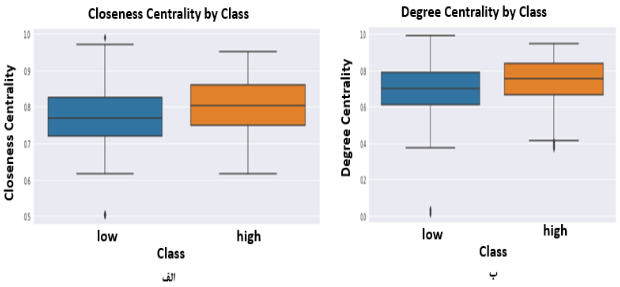
Student Group	Density	Shannon Entropy
High_level	0.73	6.76
Low_level	0.66	6.60
Student Group	**Average Clustering Coefficient**	**Average Shortest Path Length**
High_level	0.8132	1.26
Low_level	0.8130	1.33

در این قسمت گراف پدیداری دانشجویان ممتاز و ضعیف را براساس میانگین کلیک‌شان بر حسب روز در سامانه آموزش آنلاین OULAD ساختیم و معیارهای تحلیل شبکه که در جدول (۲) قسمت الف و ب نشان داده شده است، را تحلیل و بررسی کردیم. تحلیل گراف پدیداری طبیعی میانگین کلیک دانشجویان ممتاز نسبت به دانشجویان ضعیف، مرکزیت درجه، مرکزیت نزدیکی و توزیع درجه بالاتری را نشان می‌دهد. که در شکل‌های شماره (۴) و (۵) این اختلاف نشان داده شده است. به نظر می‌رسد دانشجویان ضعیف در تعامل با سامانه آموزش الکترونیکی کم‌دقت‌تر و کم حوصله‌تر هستند و با تمرکز کمتری کلاس را دنبال می‌کنند و تمایلی به چک کردن تمرینات آپلود شده در طول کلاس آنلاین ندارند و بیشتر تمرکزشان بر ثبت حضور در سامانه و گاها رها کردن کلاس آنلاین بدون توجه به مقررات است. همانطور که در جدول (۲) قسمت الف ستون سوم مشخص است، گراف پدیداری طبیعی دانشجویان ممتاز دارای میانگین درجه بالاتری می‌باشد، چندین دلیل می‌تواند در شکل گیری این نتیجه موثر باشد. با احتمال زیاد الگوی رفتاری کلیک دانشجویان ممتاز سریع است یعنی دانشجویان سطح بالا با مشاهده تمرینات درسی و پیداکردن مسیر محتوای آموزشی مانند مشاهده ویدیوها و بارگزاری پاسخ تکالیف، فکر می‌کنند و سپس سریع کلیک می‌کنند و می‌توان الگوی رفتاری آن‌ها را تحلیلی و سرعت کلیک آن‌ها را سریع به شمار آورد. دلیل دیگر می‌تواند آن باشد که دانشجویان ضعیف زمانشان را برای کلیک بر روی قسمت‌هایی از سامانه سپری می‌کنند که در ارزیابی نهایی نمرات آن‌ها خیلی تاثیر گزار نیست. به عنوان مثال این دسته دانشجویان کلیک‌هایشان را در بخش‌های مرتبط با تغییر عکس پروفایل، ویرایش نام کاربری و سایر اطلاعات شخصی مانند چک کردن پیام‌های داخل چت باکس خود و غیره انجام می‌دهند. همچنین از آنجایی که هر دو گروه دانشجویان میانگین نمرات آنها متفاوت است، ضریب شباهت جاکارد میانگین کلیک دانشجویان ضعیف و ممتاز، عدد نزدیک به صفر یعنی ۰.۰۱۹ است، یعنی اینکه این دانشجویان علاوه بر این که نمرات مشابهی نداشتند، رفتارهای مشابهی از نظر میانگین کلیک در سامانه نیز نداشته‌اند وگرنه در صورت شباهت بالای این دو گروه بایستی ضریب شباهت جاکارد عددی نزدیک به ۱ می‌شد.

همان طور در شکل (۴) الف مشاهده می‌شود، مرکزیت نزدیکی دانشجویان با بر چسپ high که با نمودار جعبه‌ای نارنجی رنگ در سمت راست مشخص است، بیانگر آن است که گراف دانشجویان ممتاز دارای مقادیر بالاتری نسبت به دانشجویان ضعیف است. یعنی این که دانشجویان سطح بالا و پایین الگوی رفتاری یکسانی از نظر تعامل و کلیک با سامانه نداشتند که نشان دهنده این است پویای مراجعات دانشجویان ممتاز به سامانه آنلاین در رابطه با حل

تمارین و حضور در سامانه آنلاین و پیگیری تکالیفی که در حین کلاس آنلاین معلم به آن‌ها برای فهم بیشتر مسئله می‌سپارد، نسبت به دانشجویان سطح پایین، بیشتر بوده است هر چند ممکن است که هردو گروه دانشجویان با انگیزه‌های یکسانی شروع به فعالیت در سامانه می‌کنند و تمایل دارند که فعالیت در سامانه را به اتمام برسانند، اما دانشجویان سطح پایین چون کم دقت‌تر هستند و در روند حل تکالیف همراه با آزمون و خطا عمل می‌کنند، ممکن است در فرایند پاسخ دهی به تکالیف سامانه را ترک کرده باشند.

همان طور در شکل (۴) ب مشاهده می‌شود، مرکزیت درجه دانشجویان با بر چسب high دارای مقادیر بالاتری نسبت به دانشجویان ضعیف است. یعنی این که دانشجویان سطح بالا در حین دوره آنلاین تمایل بیشتری به تماشای مجدد فایل‌های ضبط شده و محتوای درسی داشته‌اند هم چنین آن‌ها تمایل بیشتری به متوقف کردن فایل‌های ضبط شده برای یادداشت‌برداری نکات درسی مهم و کلیک بیشتر داشته‌اند.

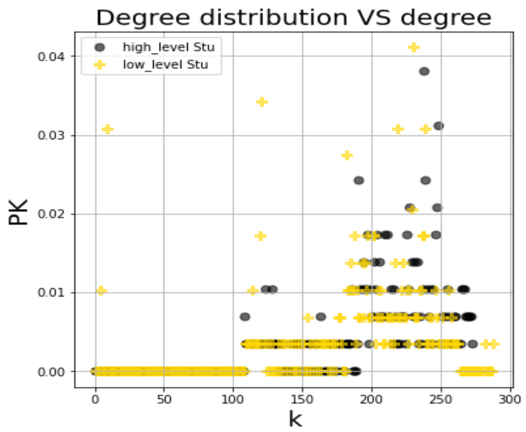


شکل (٤) : الف) مرکزیت نزدیکی گراف پدیداری طبیعی میانگین کلیک دانشجو ممتاز و ضعیف

ب) مرکزیت درجه گراف پدیداری طبیعی میانگین کلیک دانشجو ممتاز و ضعیف

با بررسی شکل (۵) می‌توان دریافت که به صورت کلی روند توزیع درجه برحسب درجه گراف پدیداری میانگین کلیک هر دو سری دانشجو در روزهای مختلف، تقریبا مشابه است به صورت کلی میانگین دانشجویان ممتاز در اکثر حالات از میانگین دانشجویان ضعیف بیشتر کلیک کرده‌اند شاید به خاطر اینکه دانشجویان ممتاز سازگاری بیشتری با محتوای درسی سخت آپلود شده دارند و زودتر تکالیف و روند دستورالعمل‌های موجود در سامانه آموزشی مانند تماشای ویدیوها، بازکردن مسیر فایل‌ها و محتوای آموزشی و نرم افزارهای کمک آموزشی را درک کرده‌اند. در حالی که دانشجویان ضعیف‌تر تمایل بیشتر به رد کردن صفحات کلاس آنلاین داشته‌اند. همچنین بیشتر به سمت عدم طی کردن گام به گام مراحل کار و نصب مثلا یک نرم افزار آموزشی جدید و چالش برانگیز بوده‌اند. از طرفی عدم بررسی گزینه‌های جدید و فایل-های کمک درسی و دکمه‌های مربوط با این مسئله خود سبب تعداد کمتر کلیک توسط دانشجویان ضعیف بوده است.

در نمودار (۵) همان طور که مشخص است، نحوه‌ی عملکرد میانگین کلیک هر دو گروه دانشجویان برای درجه‌های بین ۰ تا ۱۰۰ با هم دیگر همپوشانی داشته است، طوری که نقاط کلیک این دو نمودار همدیگر را کاملاً پوشانده‌اند. یعنی این طور برآورد می‌شود هر دو گروه دانشجویان با انگیزه یکسانی شروع به شرکت در کلاس‌های آنلاین کرده‌اند. از طرفی افزایش درجات گروهها برای مقادیر بیشتر از ۲۱۰ تا ۳۰۰ برای هر دو گروه تایید کننده این قضیه است که روند تغیرات توزیع درجه در طی زمان برای هر دو سری دانشجویان تقریبا نزدیک بهم شده است اما دانشجویان ممتاز با کمی اختلاف بیشتر در میانگین کلیک روبه رو هستند و این کلیک‌ها نشانگر آن است که با پیشرفت دوره تحصیلی هردو گروه دانشجویان سعی می‌کنند که در سامانه فعالیت کنند و پیگیر تکالیف باشند اما به دلیل عدم دقت و تمرکز دانشجویان سطح پایین در حل صحیح تکالیف، نتیجه نهایی نمرات آن‌ها متفاوت بوده است.



شکل (۵) : توزیع درجه بر حسب درجه گراف پدیداری طبیعی میانگین کلیک دانشجو ممتاز و ضعیف

برای ارزیابی اینکه معیارهای متمایز کننده در تحلیل گراف پدیداری مانند مرکزیت درجه، مرکزیت نزدیکی و توزیع درجه، با چه دقتی کار دسته‌بندی دانشجو ممتاز از ضعیف را انجام می‌دهند از تحلیل جداساز خطی و الگوریتم‌های دسته‌بندی موجود در جدول (۳) استفاده شده است، که پس از۱۰بار اجرا، میانگین دقت دسته‌بندی و جداسازی دانشجو ممتاز از ضعیف، برای معیارهای معرفی شده توسط تحلیل گراف پدیداری دقت مطلوبی را نشان می‌دهد.

جدول (۳): میانگین دقت الگوریتم‌های دسته‌بندی استفاده شده در تحلیل جداساز خطی پس از ۱۰ بار اجرا

میانگین دقت	نوع الگوریتم دسته‌بندی
۶۳٪	Decision Tree
۶۰٪	Random Forest
۵۷٪	K Nearest Neighbor (KNN)

۵- نتیجه گیری و کارهای آینده

در این مقاله ابتدا داده‌های جریان کلیک سامانه‌های آموزشی دانشجویان را پیش پردازش کرده و به شکلی که برای تحلیل مناسب باشد تبدیل کرده، سپس آن را به گراف پدیداری طبیعی نگاشت کرده و در نهایت با استفاده از معیارهای تحلیل شبکه پیچیده مانند مرکزیت درجه، مرکزیت نزدیکی و توزیع درجه تحلیل کردیم. خروجی نتایج تایید کننده این است که بالاتر بودن مرکزیت درجه، مرکزیت نزدیکی و توزیع درجه می‌تواند معیار مهمی در دسته بندی دانشجویان ممتاز و ضعیف از نظر میانگین کلیک در سامانه باشد و دقت این ادعا با الگوریتم‌های گوناگون تحلیل شده است که میانگین دقت مطلوبی رانشان داده است. برای توسعه پژوهش، می‌توان داده‌های بایگانی را با گراف پدیداری وزن‌دار جهت بررسی شباهت‌های رفتاری کاربران بررسی

کرد. و با انوع دیگر گراف پدیداری مانند افقی و نفوذپذیر محدود برای سنجش کیفیت نمره و میزان ترک تحصیل دانشجویان پیش‌بینی انجام داد.

مراجع

[1] K. Dobashi, “Real-Time Web Scraping for Analyzing Moodle Course Material Clickstream,” in *2021 19th International Conference on Information Technology Based Higher Education and Training (ITHET)*, 2021, pp. 1–10.

[2] N. R. Aljohani, A. Fayoumi, and S.-U. Hassan, “Predicting at-risk students using clickstream data in the virtual learning environment,” *Sustainability*, vol. 11, no. 24, p. 7238, 2019.

[3] M. Adnan *et al.*, “Predicting at-risk students at different percentages of course length for early intervention using machine learning models,” *Ieee Access*, vol. 9, pp. 7519–7539, 2021.

[4] Q. Li, R. Baker, and M. Warschauer, “Using clickstream data to measure, understand, and support self-regulated learning in online courses,” *Internet High. Educ.*, vol. 45, p. 100727, 2020.

[5] A. A. Mubarak, H. Cao, W. Zhang, and W. Zhang, “Visual analytics of video-clickstream data and prediction of learners’ performance using deep learning models in MOOCs’ courses,” *Comput. Appl. Eng. Educ.*, vol. 29, no. 4, pp. 710–732, 2021.

[6] E. Ulitzsch, V. Ulitzsch, Q. He, and O. Lüdtke, “A machine learning-based procedure for leveraging clickstream data to investigate early predictability of failure on interactive tasks,” *Behav. Res. Methods*, pp. 1–21, 2022.

[7] K. Liu, T. Weng, C. Gu, and H. Yang, “Visibility graph analysis of Bitcoin price series,” *Phys. A Stat. Mech. its Appl.*, vol. 538, p. 122952, 2020, doi: 10.1016/j.physa.2019.122952.

[8] A. Bhaduri, S. Bhaduri, and D. Ghosh, “Visibility graph analysis of heart rate time series and bio-marker of congestive heart failure,” *Physica A*, vol. 482, pp. 786–795, 2017, doi: 10.1016/j.physa.2017.04.091.

[9] L. Lacasa, B. Luque, F. Ballesteros, J. Luque, and J. C. Nuno, “From time series to complex networks: The visibility graph,” *Proc. Natl. Acad. Sci.*, vol. 105, no. 13, pp. 4972–4975, 2008.

[10] B. Luque, L. Lacasa, F. Ballesteros, and J. Luque, “Horizontal visibility graphs: Exact results for random time series,” *Phys. Rev. E*, vol. 80, no. 4, p. 46103, 2009.

[11] Z. T.-T. J. Ning-De and G. Bin, “Limited penetrable visibility graph for establishing complex network from time series,” *Acta Phys. Sin.*, vol. 61, no. 3, 2012.

[12] N. Makaram and R. Swaminathan, “Analysis of dynamic contractions from biceps brachii muscle using surface electromyography signals and multiscale visibility graph features,” in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2019, pp. 2653–2656.

[13] T. Plocoste, R. Carmona-Cabezas, F. J. Jiménez-Hornero, E. G. de Ravé, and R. Calif, “Multifractal characterisation of particulate matter (PM10) time series in the Caribbean basin using visibility graphs,” *Atmos. Pollut. Res.*, vol. 12, no. 1, pp. 100–110, 2021.

[14] J. Gómez-Gómez, R. Carmona-Cabezas, E. Sánchez-López, E. Gutiérrez de Ravé, and F. J. Jiménez-Hornero, “Analysis of Air Mean Temperature Anomalies by Using Horizontal Visibility Graphs,” *Entropy*, vol. 23, no. 2, p. 207, 2021.

[15] J. Zhang and Y. Luo, “Degree Centrality, Betweenness Centrality, and Closeness Centrality in Social Network,” vol. 132, no. Msam, pp. 300–303, 2017, doi: 10.2991/msam-17.2017.68.

[16] F. Liu, N. Wang, and D. Wei, “Analysis of chinese stock market by using the method of visibility graph,” *Open Cybern. Syst. J.*, vol. 11, pp. 36–43, 2017, doi: 10.2174/1874110X01711010036.

پانویس‌ها

---

1 Clickstream Data
2 Log
3 Long Short-Term Memory
4 Self-Regulated Learning
5 Massive Open Online Course (MOOC)
6 The Power of Scale-freeness in Visibility Graph (PSVG)
7 Congestive heart failure (CHF)
8 https://analyse.kmi.open.ac.uk/open_dataset
9 Linear Discriminant Analysis (LDA)
10 Decision Tree
11 Random Forest
12 K Nearest Neighbor (KNN)
13 Degree Centrality
14 Closeness Centrality
15 Degree Distribution

چهارمین کنفرانس ملی انفورماتیک ایران
پژوهشگاه دانش‌های بنیادی، پردیس فرمانیه، تهران
۱۴ و ۱۵ دی ماه ۱۴۰۱

تقریب پارامترهای مدل‌های تصادفی: رویکرد شبکه‌های عصبی پیچشی

امیرحسین ابراهیمی[۱]، هدی وفایی صفت[۲]، دارا رحمتی[۳ (*)]

[۱] پژوهشگر، مرکز پردازش سریع، پژوهشگاه دانش‌های بنیادی، تهران
Amirh.ebrahimi1377@gmail.com

[۲] پژوهشگر، مرکز پردازش سریع، پژوهشگاه دانش‌های بنیادی، تهران
Hoda.vafaeisefat@gmail.com

[۳] استادیار، دانشکده مهندسی و علوم کامپیوتر، دانشگاه شهید بهشتی، تهران
d_rahmati@sbu.ac.ir

چکیده

امروزه مدل‌های زیادی وجود دارند که هرکدام رفتار یک فرایند طبیعی را توصیف می‌کنند. برای بررسی این رفتارها از مدل‌های ریاضی مانند معادلات دیفرانسیل معمولی کمک گرفته می‌شود. این مدل‌ها معمولاً بسیار پیچیده و شامل تعداد زیادی از پارامترهای نامعلوم هستند و تابع درست‌نمایی آن‌ها از لحاظ محاسباتی قابل حل نیست. تحت این شرایط تقریب پارامترها بدلیل پیچیدگی و نبودن داده‌های کافی کار نچندان ساده‌ای می‌باشد. در شبکه‌های عصبی روش‌های مختلفی برای حل تقریب محاسبات بیزی وجود دارد. مانند یادگیری ماشین، یادگیری عمیق و ... . در این مقاله با استفاده از یک نوع معماری خاص در شبکه‌های عصبی به نام شبکه‌های عصبی رزنت که با رفتار غیر-خطی مدل‌ها سازگارتر است، عملکرد مناسبی در تخمین پارامترهای مدل‌های فرآیندها بدست آمده است. طبق معیار واریانس ناهماهنگ معماری رزنت توانسته است برای یکی از مدل‌های پیچیده به نام مدل رانش انتشار فروپاشی به مقدار ۰٫۸۷ نسبت به معماری استنتاج-عمیق بهتر عمل کند. اگر به سراغ مدل‌های ساده تر برویم این پیشرفت به مقدار بالاتری هم می‌رسد.

کلمات کلیدی

تقریب پارامتر، مدل‌های سری زمانی، یادگیری عمیق، شبکه پیچشی.

۱- مقدمه

به طور کلی در طبیعت مدل‌های مختلفی وجود دارند که هرکدام رفتار یک پدیده طبیعی را توجیه می‌کنند. هدف سازگار کردن این مدل‌ها بر روی داده‌ها است. از آن جایی که مدل‌ها اکثرا دارای فرایند به نسبت پیچیده‌ای هستند، پارامترهای ناشناخته‌ای را نیز دارا هستند و نیاز است تا از روش‌هایی برای تخمین پارامترها استفاده شود. به عنوان مثال روند پیشرفت انتشار یک بیماری مانند کوید-۱۹ [12,22]، صعود و تنزل در عرصه بورس، تحلیل مدل‌های تصمیم‌گیری در حوزه‌ی علوم شناختی [4,15] و علم ستاره‌شناسی برای تقریب پارامترهای امواج گرانشی در فضا [7,9] که هر کدام دارای پارامترهای خاص خود هستند را می‌توان نام برد. از این رو تقریب پارامتر توجه محققان را به خود جلب کرده است تا که بتوانند در زمینه‌های مختلف، پارامترهای مدل‌های مختلفی را تقریب بزنند. در آزمایش‌ها و تحقیق‌های علوم شناختی برای آنکه روند تصمیم‌گیری یک آزمودنی را مورد بررسی قرار دهند از مدل‌های مختلفی استفاده می‌کنند. یکی از این مدل‌ها رانش-انتشار می‌باشد که با در دست داشتن زمان‌های پاسخ آزمودنی پارامترهای بخصوصی را بدست می‌آورد که با آن می‌توان فرایند تصمیم‌گیری فرد را مورد بررسی قرار داد [18]. به این دلیل استفاده از مدل‌های شبکه عصبی برای تقریب این پارامترها با محاسبه همزمان-با-وقوع امری حیاتی است که بتوان در حین اجرای تکلیف شناختی، پاسخ مناسب در کمترین زمان به آزمودنی داده شود. لذا شبکه‌های عصبی طراحی شده باید تا حد امکان ساده و از لحاظ بار محاسباتی قابل قبول باشند. روشی که در این تحقیق استفاده شده است، یادگیری عمیق است. ابتدا با استفاده از پیاده‌سازی معماری رزنت [10] برای مثال ساده‌ای از دو مدل توزیع نرمال چند متغیره و رگرسیون چندگانه نتایج را به دست آورده و نتایج با مدل دیگری مقایسه می‌شود. مدل‌های زیادی در این زمینه تحقیقاتی معرفی شده‌اند، از جمله : استنتاج-عمیق [16] و BAYESFLOW [17] و Likelihood Approximation networks (LANs) [6] که هرکدام از شبکه‌های مطرح شده، دارای مشکلات و محدودیت‌هایی می‌باشند که به آن ها اشاره خواهد شد. عملکرد معماری پیاده‌سازی شده در این تحقیق بر روی مدل های پیچیده تر مانند مدل‌های سری زمانی، مدل تصمیم‌گیری انباشته کننده رقابتی نشتی [1] و مدل پیش‌بینی روند انتشار بیماری بررسی شده است. نتیجه‌ای که از این مرحله به دست آمد این است که مدل‌هایی که با زمان مرتبط هستند نیاز به یک روشی کاربردی‌تر دارند، که ما را به استفاده از شبکه‌های عصبی مبتنی بر معادله دیفرانسیل معمولی [3] که با مدل‌های زمانی سازگار هستند و ارتباط نزدیکی با معماری رزنت دارند، سوق می‌دهد. ما به دنبال روشی هستیم که بتوان محدودیت‌های ذکر شده را دور بزند و به مدلی کاربردی‌تر دست پیدا کرد.

۲- تقریب پارامتر

یکی از اهداف مهم در مدل‌سازی آماری آن است که فرایند تولید داده را از طریق مجموعه‌ای از پارامترها توصیف کرد. بدلیل وجود منابع مختلف عدم قطعیت، این توصیف به صورت ذاتی تصادفی است. یک روش برای نشان دادن این عدم قطعیت قرار دادن توزیع احتمال بر روی پارامترها است. در واقع، این‌ایده اساسی در آمار بیزی گنجانده شده است [8,16].

(*) نویسنده مسئول

$$p(\theta, x) = \frac{p(x, \theta)\pi(x)}{\int p(x, \theta)\pi(x) d_\theta}. \quad (۱)$$

در فرمول (۱)، $p(\theta.x)$ بیانگر توزیع پســین پارامترهای مدل، $p(x.\theta)$ بیانگر تابع درســت‌نمایی، $\pi(x)$ بیانگر توزیع احتمال قبلی و عبارت ثابت در مخرج توزیع حاشیه‌ای داده است که به دلیل بار محاسباتی بالا معمولا درنظر گرفته نمی‌شود. بنابراین می‌توان فرمول (۱) را به صورت زیر بازنویسی کرد

$$p(\theta, x) \propto p(x, \theta) \times \pi(x). \quad (۲)$$

بسیاری از مدل‌های امروزی بسیار پیچیده هستند و یا از فرایندهای شبیه‌سازی برای تولید داده استفاده می‌کنند، درنتیجه تابع درست‌نمایی از نظر محاسباتی غیرقابل حل خواهد شد و یا نمی‌توان آن را به صورت جبری مشخص کرد. بنابراین برای انجام هر نوع استنتاج بیزی به روش‌های تقریبی نیاز است. در این بخش به بررسی روش‌های مختلف برای تقریب مقادیر پارامترهای مدل می‌پردازیم.

۲-۱- روش های تقریب پارامتر

روش محاسبه تقریبی بیزی [5] برای اولین بار در تحقیقات ژنتیکی با موفقیت پیاده‌سازی شد [21] و اهمیت و سودمندی این روش در سایر حوزه‌های تحقیقاتی گسترش یافت. ایده اصلی این روش آن است که تابع درست‌نمایی فرمول (۲) را از طریق نمونه‌گیری‌های مکرر از توزیع های قبلی تخمین بزند و به تولید داده‌های مصنوعی مشروط به پارامترهای نمونه‌گیری شده بپردازد [14,19].

به دست آوردن تابع درست‌نمایی با دقت بالا از لحاظ محاسباتی امکان پذیر نیست؛ به همین علت به فرم $q(.\theta)$ مشخص می‌شود. نقطه شروع تمامی الگوریتم های محاسبه تقریبی بیزی آن است که جدول مرجع ایجاد کنند. این جدول یک ساختمان داده مخصوص است که برای ذخیره سازی تعداد زیادی مجموعه داده‌های شبیه سازی شده است [14,11,19,20].

جنگل‌های تصمیم تصادفی [2] یک روش یادگیری با نظارت برای دسته‌بندی رگراسیون است. به این الگوریتم دسته‌ای از جفت‌های متناظر ورودی – خروجی داده می‌شود و سعی می‌کند که یک نگاشت غیر خطی که ورودی را به خروجی تبدیل می‌کند بر روی درخت‌های تصمیم‌گیری یاد بگیرد که در حوزه محاسبه تقریبی بیزی، داده‌های شبیه‌سازی شده از جدول مرجع هستند و خروجی‌ها پارامترهای متناظر دیتاهای شبیه‌سازی شده است.

یکی از روش‌های پرکاربرد برای محاسبه تقریبی بیزی با بهره‌گیری از یادگیری ماشین، استفاده از شبکه های عصبی عمیق به منظور یادگیری ویژگی ها از جدول مرجع است [11,20]. در حالت کلی یک شبکه عصبی با تعریف یک نگاشت $y = f_W(x)$ از ورودی x به خروجی y و یادگیری پارامترهای مدل، ما را به بهترین تقریب هدایت می‌کند. از مزیت شبکه‌های عصبی عمیق می‌توان به آن اشاره کرد که قادر به تقریب هر تابع پیوسته به یک درجه دقت دلخواه هستند .

۲-۲- استنتاج-عمیق

این روشی است که از مزایای دو روش محاسبه تقریبی بیزی با استفاده از جنگل‌های تصمیم تصادفی [19] و یادگیری عمیق [11] استفاده می‌کند. به طور کلی این الگوریتم به تعلیم شبکه‌های عصبی تمام پیچشی یا همگشتی [13] بر روی داده‌های شبیه‌سازی شده، با هدف تقریب میانگین پارامترهای مدل و همزمان تقریب عدم قطعیت و یا واریانس از طریق کمینه کردن تابع ضرر می‌پردازد.

معماری پیاده‌سازی شده شامل سه لایه مخفی می‌باشد که آخرین لایه میانگین مقادیر خروجی هر هسته را حساب می‌کند. اگر m تعداد پارامترهای مدل تولید شده باشد، لایه خروجی شبکه‌ی عصبی تمام پیچشی شامل m واحد فعال سازی خطی برای پیش‌بینی هر یک از اجزاء پارامتر θ و سایر m واحد بعدی برای پیش بینی واریانس $\sigma^2(\dot{x})$ است. آخرین واحدها نیز از یک تابع سافت پلاس (فرمول (۳)) عبور کرده‌اند. علت استفاده از این تابع در واحدهای آخر برای آن است که از مثبت بودن تقریب واریانس σ^2 (x ̇) اطمینان حاصل شود. برای تمامی مثال‌ها از تعداد بیست epoch با الگوریتم بهینه سازی Adam و نرخ یادگیری ۰.۰۰۱ تعلیم یافته اند [16]

$$softp(z) = 1 + \log(1 + e^z). \quad (۳)$$

مدل استنتاج-عمیق دارای معماری بسیار ساده‌ای است که با وجود این سادگی از دو مدل محاسبه تقریبی بیزی با استفاده از جنگل‌های تصمیم تصادفی و یادگیری عمیق بهتر عمل می‌کند. اما سادگی این مدل در همه‌ی جهات مناسب نیست. به طور مثال مدل هایی همچون تصمیم‌گیری انباشته کننده رقابتی نشتی و روند انتشار بیماری که از مدل‌های سری‌زمانی می‌باشند. پارامترهای این نوع مدل‌های سری زمانی به خوبی توسط مدلی ساده مثل استنتاج-عمیق بازیابی نمی‌شوند. همچنین با پیچیدگی مدل یا افزایش پارامترها، نمی‌تواند به خوبی عمل کند.

۲-۳- BAYESFLOW, LANs

مدل BAYESFLOW دارای شبکه بسیار پیچیده‌ای است و از لحاظ زمانی و هزینه محاسباتی هزینه بر است. زیرا که معماری بسیار پیچیده‌ای دارد، که شامل یک شبکه عصبی معکوس‌پذیر عمیق برای بازیابی پارامترها و یک شبکه انکودر برای خلاصه کردن داده‌های ورودی می‌باشد، چنین شبکه‌ی عصبی دارای پارامترهای بسیار زیادی می‌باشد. همچنین مدل LANs نیز کاربرد عام ندارد و فقط مخصوص مدل‌های تصمیم‌گیری می‌باشد.

۳- معرفی معماری رزنت

قبل از رزنت، محققها با عمیق کردن شبکه‌های عصبی برای بدست آوردن خروجی‌های چشم‌انگیزتر، به مشکلات مختلفی از جمله محوشدگی گرادیان و انفجار گرادیان و در نهایت کاهش دقت عملکرد شبکه بر می‌خوردند. شِیوکینگ رن و همکارانش [10] با ارائه معماری رزنت کمک بزرگی به علم یادگیری ماشین کرده اند و باعث تحول این حوزه شده اند. رن و همکارانش با معرفی اتصالات میانبر راه حل مشکل‌های شبکه‌های بسیار عمیق را پیدا کرده اند. اتصال میانبر، همان طور که از اسمش پیداست، از یک یا چند لایه عبور می‌کند

چهارمین کنفرانس ملی انفورماتیک ایران
پژوهشگاه دانش‌های بنیادی، پردیس فرمانیه، تهران
۱۴ و ۱۵ دی ماه ۱۴۰۱

و آن‌ها را در نظر نمی‌گیرد. با این کار (اضافه کردن خروجی لایه‌های پیشین به خروجی لایه فعلی)، ثابت می‌شود با وجود شبکه‌های بسیار عمیق، وزن‌های مدل مقادیر غیر صفر خواهند داشت و به دنبال آن مشکل محو شدن گرادیان حل می‌شود.

در این بخش معماری ساده رزنت تعریف شده‌است. دلیل انتخاب ساده بودن مدل این است که سعی داشتیم بتوانیم تا حد امکان پیچیدگی معماری را کاهش دهیم و تعداد پارامترهای حاصل از شبکه عصبی قابل مقایسه با سایر مدل‌ها همچون استنتاج-عمیق باشد. پس از تعریف معماری، نتایج آن را در مقایسه با نتایج معماری [16] در تخمین پارامتر دو مثال ساده نرمال چند متغیره و رگرسیون چندگانه و همچنین بر روی دو مدل تصمیم‌گیری رانش-انتشار و رانش-انتشار فروپاشی گزارش خواهد شد.

معماری رزنت که در شکل (۱) قابل مشاهده است. در ابتدا بر روی ورودی یک لایه کانولوشن با مقدار فیلتر ۶۴ و اندازه کرنل یک و سپس تابع فعال ساز اعمال شده است. در ادامه معماری شامل دو بخش است که در هر بخش سه لایه وجود دارد که به ترتیب در لایه‌ها از لایه پیچشی، لایه نرمال‌سازی دسته‌ای و تابع فعال ساز یکسوساز استفاده شده است. در تابع میانبر که برای مرحله اول می‌باشد به ترتیب از لایه پیچشی و لایه نرمال‌سازی دسته‌ای استفاده شده است و در مرحله دوم، تابع میانبر به صورت ساده است.

این تابع میانبر در انتهای هر بخش با خروجی آن بخش جمع می‌شود. تمامی مدل‌های شبکه‌های پیچشی از طریق واسط برنامه‌نویسی کاربردی مخصوص شبکه‌های عصبی، Tensorflow [23]، با زبان پایتون پیاده سازی شده است.

۳-۱- مقایسه نتایج استنتاج-عمیق و رزنت بر روی توزیع نرمال چند متغیره و رگرسیون چندگانه

با تعریف مدل توزیع نرمال چند متغیره با میانگین و انحراف معیار ناشناخته، خروجی معماری را بر روی آن می‌سنجیم. توزیع مزدوج برای میانگین μ و ماتریس کوواریانس Σ به صورت فرمول (۴) است

$$\Sigma \sim \mathrm{InvWisHart}(\Lambda_0, v_0)\,, \mu \mid \Sigma \sim \mathrm{N}_d(\mu_0, \frac{\Sigma}{k_0})$$
$$x_i \mid \mu . \Sigma \sim \mathrm{N}_d(\mu_0, \Sigma) \rightarrow i\ 1 \ldots m . \qquad (۴)$$

که در این مدل $InvWisHart(\Lambda_0, v_0)$ توزیع معکوس Wishart پارامتر شده توسط یک ماتریس مقیاس Λ_0 به ابعاد d × d و درجه آزادی v_0، همچنین $\mathrm{N}_d(\mu_0, \frac{\Sigma}{k_0})$ توزیع نرمال چند متغیره با بردار میانگین d بعدی μ و ماتریس کوواریانس Σ به ابعاد d × d است. با تعریف پارامتر های زیر، صدهزار داده از فرمول فوق تولید شد.

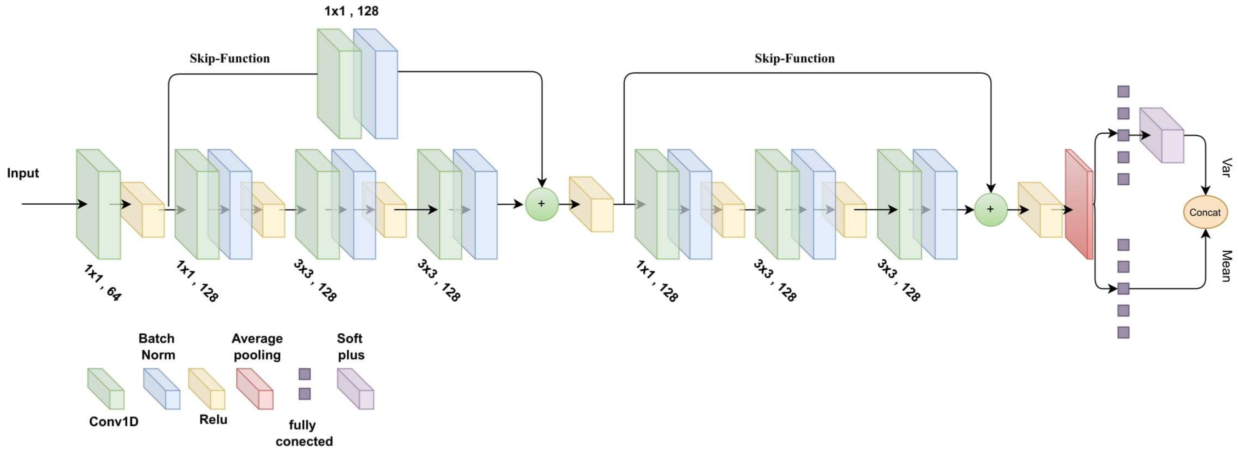
$$d = 2\ , \quad m = 100\,, \quad \Lambda_0 = I_2\,, \quad v_0 = 5\,,$$
$$k_0 = 3\ , \quad k_0 = 3$$

اگر مثال را کمی پیچیده‌تر کنیم ، با تعریف داده های شبیه‌سازی شده ماتریس کوواریانس $m \times d$، $X = (x_1 \ldots x_d)$ می‌توان مدل را به صـورت زیر تعریف کرد:

$$\sigma^2 \sim IG(a_0 . b_0)\,,$$
$$\beta | \sigma^2 \sim N_d(0 . n\sigma^2 (X^T X)^{-1}),$$
$$y | \beta . \sigma^2 \sim N_m(x\beta . \sigma^2 I),$$

که $IG(a_0 . b_0)$ نشان دهنده توزیع گاما معکوس با بعد a_0 و مقیاس b_0 و وکتور $\beta = (\beta_1 \ldots . \beta_d)$ وزن‌های رگراسیون با توزیع نرمال d بعدی هستند و تابع درست‌نمایی توزیع نرمال m بعدی با ماتریس کواریانس ایزوتروپیک است.

با تعریف پارامترهای $d = 2\,, a_0 = 4\,,\ b_0 = 3\ , m = 100$ و تولید یک جدول مرجع شامل صد هزار ردیف از نمونه‌های یکسان توزیع شده $\{\tilde{x}_k^{(i)}\}_{k=1}^{100}$ که هر کدام شامل یک سه تایی از دو متغیر و یک خروجی حاصل از آن دو $(x_1^i . x_2^i . y^i)$ است، تولید می‌کنیم و سپس به کمک شبکه عصبی پارامترهای داده‌های تولید شده متناظر را بدست می‌آوریم $(\beta_1^i . \beta_2^i . \sigma^{2^{(i)}})$.



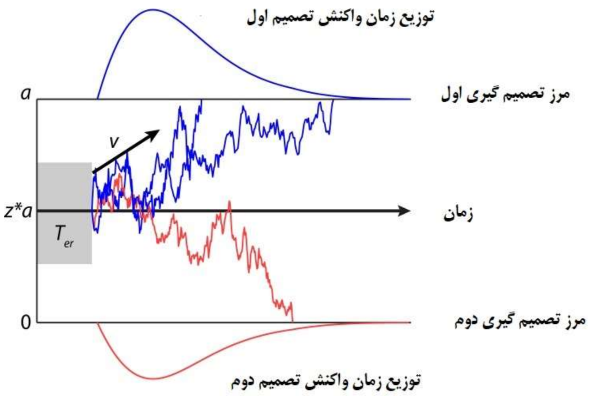
شکل (۱) : معماری طراحی شده که شامل دو بلوک میانبر و اتصال میانبر اول دارای لایه‌های پیچشی و نرمال‌سازی دسته‌ای است.

۳-۲- مدل تصمیم‌گیری رانش-انتشار و رانش-انتشار فروپاشی

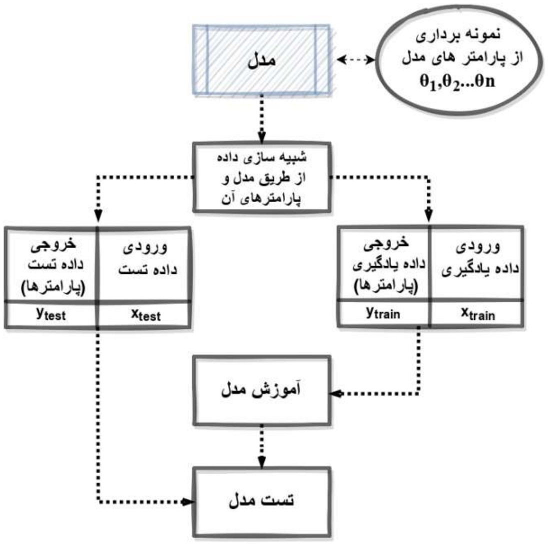
مدل رانش–انتشار ، یکی از مدل های مهم در زمینه مسائل تصمیم‌گیری است. این مدل رفتار دو انتخاب و یک زمان واکنش را توصیف می‌کند. این مدل فرض بر آن دارد که فرد برای گرفتن یک تصمیم برای خود دو مرز در نشر می‌گیرد؛ یک مرز برای گرفتن یک تصمیم و مرز دیگری را برای رد این تصمیم. برای رسیدن به این دو مرز با یک نرخ رانش از طریق قدم‌زنی تصادفی استفاده می‌شود. به این صورت که هر قدمی که به نفع تصمیم‌گیرنده باشد انتخاب می‌شود و به مرز مربوطه نزدیک می‌شود. این مدل شامل چهار پارامتر a ، v ، z و t_0 است. پارامتر a ، جدا کننده آستانه است و نشان‌دهنده فاصله بین دو مرز بالا و پایین یا همان میزان اطلاعات لازم برای گرفتن تصمیم است. پارامتر v برای نمایش نرخ رانش و نشان‌دهنده میانگین شیب فرایند تصمیم‌گیری از نقطه شروع تا نقطه پایان است. در نهایت پارامتر z، نقطه شروع فرآیند تصمیم‌گیری و نشان دهنده میزان تمایل به یکی از تصمیم ها و پارامتر t_0 برای نمایش زمان عدم تصمیم گیری است. عدم تصمیم‌گیری بیانگر آن است که فرد تصمیم گیرنده از همان ابتدا شروع به تصمیم‌گیری نمی‌کند بلکه زمانی را صرف پردازش اطلاعات می‌کند. علاوه بر آن زمان تاخیر آنکه تصمیم در مغز گرفته شود و دستور از آن به سایر اندام محرک انسان برسد نیز وجود دارد، به مجموع این دو زمان، پردازش اطلاعات و تاخیر صورت گرفته، زمان عدم تصمیم‌گیری گفته می‌شود. نمای کلی این مدل در شکل (۲) آورده شده است. به عنوان مثال نقطه شروع z آغاز تصمیم‌گیری نام گذاری می‌شود که باید تصمیمی گرفته شود (سقف a) یا نشود (کف ۰) ، این مدل رفتار روند تصمیم‌گیری را در طی زمان به نمایش می‌گذارد

$$x(t+t') = x(t) + v \times t' + N(1.\alpha) \times t'. \qquad (۵)$$

از آنجایی که ما رفتار مدل را با مدل ریاضی، فرمول (۵)، می‌دانیم و به داده‌ای دسترسی نداریم، داده ها را تولید می‌کنیم. به این صورت که ابتدا مقادیر تصادفی و قابل قبولی از پارامترها را تولید می‌کنیم. سپس با مقادیر داده‌های مختلفی مقادیر y یا خروجی را به دست آورده و ذخیره می‌کنیم. شمای کلی روند تولید داده و استفاده از آن را می‌توان در شکل (۳) مشاهده کرد.



شکل (۲) : نمایی از مدل رانش انت شار، T_{er} زمان عدم ت صمیم‌گیری را نشان می‌دهد که تصمیم گیرنده از مرز تصمیم‌گیری خنثی با نرخ رانش v به سوی یکی از مرزها حرکت می‌کند.



شکل (۳) : روند کلی تولید داده برای مدل و استفاده از آن. یک فرایند طبیعی با استفاده از معادلات دیفرانسیل معمولی مدل می‌شود که شامل تعدادی پارامتر است. با مقداردهی مناسب به این پارامترها و جایگذاری آن‌ها در مدل تعدادی داده شـبیه‌سـازی می‌شـود که از آن‌ها برای یادگیری اسـتفاده می‌شـود. پس از یادگیری برای ارزیابی مدل از داده های شبیه‌سازی شده‌ی متفاوتی استفاده می‌شود.

با استفاده از روش کار فوق چهل هزار داده تولید شد و سپس داده‌ها را بر روی مدل پیشنهادی خود رزنت و مدل استنتاج–عمیق اعمال کرده و نتایج مناسبی به دست آورده‌شد.

برخلاف مدل رانش–انتشار، مرزهای تصمیم‌گیری در رانش–انتشار فروپاشی ثابت نیستند و وابسته به زمان هستند و به صورت فرمول (۶) تعریف می‌شوند

$$l(t) = \frac{at}{2 * T_{\infty}} \qquad , \qquad u(t) = a - l(t). \qquad (۶)$$

که در آن a مرز تصمیم‌گیری است و وابسته به زمان است و T_{∞} پارامتری است که نشان می‌دهد مرزهای بالا و پایین چه زمانی به یکدیگر الحاق می‌شوند. مشابه قبل تعداد چهل هزار داده شبیه‌سازی شد و سپس داده‌ها بر روی مدل پیاده شد. از آنجایی که تعداد پارامترهای مدل رانش–انتشار فروپاشی بیشتر است و همچنین برای اثبات آنکه با افزایش لایه‌های رزنت خروجی بهتر می‌شود، معماری رزنت پیشین کمی عمیق‌تر درنظر گرفته شد.

۴- مقایسه نتایج

نتایجی که به دست آوردیم نشان می‌دهد که در اکثر موارد رزنت یا مشابه معماری استنتاج–عمیق عمل کرده و یا نتایج بهتری به دست آورده است. در جدول (۱) نتایج به دست آمده توسط معماری رزنت بر روی مثال اول و در جدول (۲) نتایج بدست آمده توسط معماری رزنت برروی مدل مثال دوم قابل مشاهده می باشد.

جدول (۱) : نتایج دو معماری بر توزیع نرمال چند متغیره با معیار خطای جذر میانگین مربعات

	مدل رزنت	استنتاج-عمیق
$E[\mu_1\|x]$	**۰.۰۱۱**	۰.۰۲۲
$E[\mu_2\|x]$	**۰.۰۰۹**	۰.۰۱۷
$E[\sigma_1^2\|x]$	۰.۰۵۳	**۰.۰۴۲**
$E[\sigma_2^2\|x]$	**۰.۰۳۰**	۰.۰۳۸
$Var[\mu_1\|x]$	**۰.۰۰۲**	۰.۰۰۲
$Var[\mu_2\|x]$	**۰.۰۰۰**	۰.۰۰۱
$Var[\sigma_1^2\|x]$	۰.۰۶۳	**۰.۰۴۶**
$Var[\sigma_2^2\|x]$	**۰.۰۰۳**	۰.۰۱۳

جدول (۲) : نتایج دو معماری بر رگرسیون چندگانه با معیار خطای جذر میانگین مربعات

	مدل رزنت	استنتاج-عمیق
$E[\beta_1\|x]$	**۰.۰۵۹**	۰.۱۰۳
$E[\beta_2\|x]$	**۰.۰۶۹**	۰.۰۷۸
$E[\sigma^2\|x]$	**۰.۱۴۸**	۰.۲۳۴
$Var[\beta_1\|x]$	**۰.۰۱۸**	۰.۰۲۵
$Var[\beta_2\|x]$	**۰.۰۲۴**	۰.۰۳۲
$Var[\sigma^2\|x]$	۰.۲۴۵	**۰.۲۳۵**

همچنین نتایج حاصل بر روی دو مدل تصمیم‌گیری رانش-انتشار و رانش-انتشار فروپاشی نیز نشان‌دهنده آن است که مدل پیشنهادی می‌تواند به خوبی پارامترهای این دو مدل را تخمین بزند.

نتایج حاصل از شبکه بر روی این دو مدل در جدول (۳ و ۴) به ترتیب آورده شده است که نشان می‌دهد از رزنت برای مدل‌های پیچیده‌تر همچون این مدل می‌توان استفاده کرد و نتایج بهتری دریافت کرد.

جدول (۳) : نتایج دو معماری بر روی مدل رانش-انتشار با معیار سنجش ضریب همبستگی پیرسون

	مدل رزنت	استنتاج-عمیق
a	**۰.۹۱**	۰.۸۴
v	**۰.۹۸**	۰.۸۹
zr	**۰.۸۶**	۰.۸۱
Ndt	**۰.۹۵**	۰.۸۷

جدول (٤) : نتایج دو معماری بر روی مدل رانش-انتشار فروپاشی با معیار سنجش ضریب همبستگی پیرسون

	مدل رزنت	استنتاج-عمیق
a	**۰.۷۳**	۰.۶۹
v	**۰.۷۹**	۰.۶۴
zr	**۰.۷۲**	۰.۷
ndt	**۰.۸۸**	۰.۸۳
T_{inf}	**۰.۶۹**	۰.۶۴

۵- کارهای آینده

مسأله‌ای که وجود دارد این است که ممکن است در مواقعی به سراغ مدل‌هایی برویم که به زمان وابستگی دارند مانند مدل‌هایی که برای پیش‌بینی وضعیت شیوع یک بیماری در یک جامعه در طی زمان به کار می‌روند و یا اگر لازم باشد که وضعیت بورس کشور در طی زمان را مدل کنیم. در این شرایط به دنبال خانواده‌ی خاصی از شبکه‌های عصبی به نام شبکه‌های عصبی دیفرانسیل معمولی [3] می‌رویم و آن را مورد بررسی قرار می‌دهیم.

در شبکه عصبی میانبر تعدادی لایه مخفی وجود دارند که به صورت گسسته و مجزاء می باشند. اگر گام‌های شبکه را به سمت بی‌نهایت سوق دهیم و مدل از حالت گسسته به پیوسته تبدیل شود می‌توان امید داشت که از این شبکه برای مدل‌های وابسته به زمان استفاده کرد. از طرف دیگر برای مسائلی که با داشتن یک وضعیت ابتدایی بتواند یک روند در آینده را پیش‌بینی کند که برای حل شدن نیاز به یک حل‌کننده دارند (فرمول (۷)). به دلیل شباهتی که بین رابطه روش اویلر که یک حل‌کننده است با بلوک میانبر وجود دارد می‌توان نتیجه گرفت که از آن می‌توان به عنوان یک حل‌کننده ساده استفاده کرد

$$z(t_1) = z(t_0) + \int_{t_1}^{t_0} f(z(t_0).t.\theta)dt. \qquad (۷)$$

شبکه میانبر شامل چندین بلوک میانبر است و طبق مطالب گفته شده هر بلوک به عنوان یک حل‌کننده عمل می‌کند. حال اگر از یک حل‌کننده معادلات دیفرانسیل معمولی به جای تمامی بلوک‌ها استفاده کنیم تعداد ارزیابی‌های بین لایه‌ها افزایش می‌یابند. در این شبکه هرچه زمان بیشتری صرف آموزش شبکه شود در بخش تست دقت بالاتری خواهیم داشت و به دلیل اینکه شبکه سنگین و با تعداد لایه‌های بسیار زیادی است برای بهینه‌سازی پارامترها نمی‌توان از پس انتشار استفاده کرد و از روش حالت الحاقی برای این کار استفاده می‌شود که این کار باعث می‌شود میزان مصرف از حافظه به شدت کاهش پیدا کند. با جای‌گذاری تمامی بلوک‌های میانبر با یک حل کننده معادله دیفرانسیل معمولی به یک شبکه‌ی عصبی دیفرانسیل معمولی می‌رسیم.

۶- نتیجه گیری

با توجه به مفاهیمی که تا به الان گفته شد، مدل‌هایی وجود دارند که پارامترهای آن‌ها برای ما مهم است و اگر این مدل‌ها پیچیده شوند تقریب این پارامترها کار آسانی نیست پس باید به سوی روش‌های تقریب پارامتر رفت. یادگیری عمیق و استفاده از شبکه‌های عصبی یکی از این روش‌ها است. در این روش به بررسی معماری رزنت پرداختیم که خروجی مطلوبی را در مقایسه با موارد اشاره شده داشت. همچنین طبق ویژگی‌های اصلی شبکه‌های عصبی میانبر، هر مقدار این شبکه‌ها را عمیق‌تر کنیم، نتایج بهتری را به دست خواهیم آورد که این مورد در استنتاج–عمیق قابل مشاهده نبود. با این حال مدل‌های زیاد و مهمی وجود دارد که وابسته به زمان هستند. مدل ساده مطرح شده نمی‌تواند پارامترهای این مدل را به خوبی بازیابی کند، مگر این که از شبکه‌های بسیار عمیق رزنت استفاده شود، که بدلیل بار محاسباتی بسیار بالا و طولانی بودن فرآیند یادگیری توجه مارا جلب نمی‌کند. توضیح داده شد که معماری جدیدی به نام شبکه‌های عصبی دیفرانسیل معمولی وجود دارد که به دلیل نوع طراحی این مدل بسیار با مدل‌های زمانی سازگار است و همچنین جایگزین معماری رزنت است. روشی با نام روش حالت الحاقی وجود دارد که نسبت تغییرات تابع خسارت به پارامتر را می‌سنجد. روش حالت الحاقی یک روش عددی برای محاسبه بهینه گرادیان یک تابع یا عملگر در یک مسئله بهینه سازی عددی است. در ادامه قرار است که نسبت تابع خسارتی را که در رزنت استفاده کردیم را به پارامترهایی که می‌خواهیم تقریب بزنیم، بسنجیم و معماری را با تبدیل به شبکه‌های عصبی دیفرانسیل معمولی برای سازگاری با مدل‌های زمانی تغییر دهیم.

مراجع

[1] Bird, S.M., Sir David, C., Farewell, V.T., Harvey, G., Tim, H. and Peter C, S., 2005. *Performance indicators: good, bad, and ugly. Journal of the Royal Statistical Society: Series A (Statistics in Society),* 168(1), pp.1-27.

[2] Breiman, L., 2001. *Random forests*. Machine learning, 45(1), pp.5-32.

[3] Chen, R.T., Rubanova, Y., Bettencourt, J. and Duvenaud, D.K., 2018. *Neural ordinary differential equations. Advances in neural information processing systems*, 31.

[4] Chen, Y. and Zhou, Y., 2020. *Machine learning based decision making for time varying systems: Parameter estimation and performance optimization*. Knowledge-Based Systems, 190, p.105479.

[5] Csilléry, K., Blum, M.G., Gaggiotti, O.E. and François, O., 2010. *Approximate Bayesian computation (ABC) in practice.* Trends in ecology & evolution, 25(7), pp.410-418.

[6] Fengler, A., Govindarajan, L.N., Chen, T. and Frank, M.J., 2021. *Likelihood approximation networks (LANs) for fast inference of simulation models in cognitive neuroscience*. Elife, 10, p.e65074.

[7] Gabbard, H., Messenger, C., Heng, I.S., Tonolini, F. and Murray-Smith, R., 2022. *Bayesian parameter estimation using conditional variational autoencoders for gravitational-wave astronomy*. Nature Physics, 18(1), pp.112-117.

[8] Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B., 1995. *Bayesian data analysis*. Chapman and Hall/CRC.

[9] Green, S.R., Simpson, C. and Gair, J., 2020. *Gravitational-wave parameter estimation with autoregressive neural network flows*. Physical Review D, 102(10), p.104057.

[10] He, K., Zhang, X., Ren, S. and Sun, J., 2016. *Deep residual learning for image recognition*. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[11] Jiang B, Wu TY, Zheng C, Wong WH. *Learning summary statistic for approximate Bayesian computation via deep neural network*. Statistica Sinica. 2017 Oct 1:1595-618.

[12] Liu, F., Li, X. and Zhu, G., 2020. *Using the contact network model and Metropolis-Hastings sampling to reconstruct the COVID-19 spread on the "Diamond Princess".* Science Bulletin, 65(15), pp.1297-1305.

[13] Long, J., Shelhamer, E. and Darrell, T., 2015. *Fully convolutional networks for semantic segmentation*. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3431-3440).

[14] Mertens UK, Voss A, Radev S. ABrox—A user-friendly *Python module for approximate Bayesian computation with a focus on model comparison*. PloS one. 2018 Mar 8;13(3):e0193981.

[15] Palestro JJ, Sederberg PB, Osth AF, Van Zandt T, Turner BM. *Likelihood-free methods for cognitive science*. Springer International Publishing; 2018 Feb 7.

[16] Radev, S.T., Mertens, U.K., Voss, A. and Köthe, U., 2020. *Towards end-to-end likelihood-free inference with convolutional neural networks*. British Journal of Mathematical and Statistical Psychology, 73(1), pp.23-43.

[17] Radev, S.T., Mertens, U.K., Voss, A., Ardizzone, L. and Köthe, U., 2020. BayesFlow: *Learning complex stochastic models with invertible neural networks*. IEEE transactions on neural networks and learning systems.

[18] Ratcliff, R., 1978. *A theory of memory retrieval. Psychological review*, 85(2), p.59.

[19] Raynal L, Marin JM, Pudlo P, Ribatet M, Robert CP, Estoup A. *ABC random forests for Bayesian parameter inference.* Bioinformatics. 2019 May 15;35(10):1720-8.

[20] Sheehan, S. and Song, Y.S., 2016. *Deep learning for population genetic inference*. PLoS computational biology, 12(3), p.e1004845

[21] Tavaré, S., Balding, D.J., Griffiths, R.C. and Donnelly, P., 1997. *Inferring coalescence times from DNA sequence data.* Genetics, 145(2), pp.505-518.

[22] Yin, M.Z., Zhu, Q.W. and Lü, X., 2021. *Parameter estimation of the incubation period of COVID-19 based on the doubly interval-censored data model*. Nonlinear Dynamics, 106(2), pp.1347-1358.

[23] Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M. and Kudlur, M., 2016. {TensorFlow}: *a system for {Large-Scale} machine learning*. In 12th USENIX symposium on operating systems design and implementation (OSDI 16) (pp. 265-283).

چهارمین کنفرانس ملی انفورماتیک ایران
پژوهشگاه دانش‌های بنیادی، پردیس فرمانیه، تهران
۱۴ و ۱۵ دی ماه ۱۴۰۱

نگاشت عمق تصاویر به تصاویر دید از بالا: رویکرد شبکه‌های عصبی عمیق

آبتین ماه‌یار[۱]، حسین معتمد نیا[۲]، دارا رحمتی[۳و۴] (*)

[۱] گروه علوم کامپیوتر و علوم داده‌ها، دانشگاه شهید بهشتی ، تهران،
a.mahyar@mail.sbu.ac.ir
[۲] پژوهشگر، مرکز پردازش سریع، پژوهشگاه دانش‌های بنیادی ، تهران،
h.motamednia@ipm.ir
[3] استادیار، دانشکده مهندسی و علوم کامپیوتر، دانشگاه شهید بهشتی، تهران
[4] پژوهشکده علوم کامپیوتر، پژوهشگاه دانش‌های بنیادی، تهران
d_rahmati@sbu.ac.ir

چکیده

دوربین‌ها در راهبری ربات‌ها، ماشـین‌های خودران و پرنده‌های کنترل از راه دور نقش اساسی دارند. الگوریتم‌های پردازش تصویر می‌توانند با داده‌هایی که دوربین‌ها در اختیار آن‌ها قرار می‌دهند اطلاعاتی پیرامون محیط اطراف برای هسته پردازشی دستگاه‌های یاد شده را فراهم کنند که می‌توانند آن‌ها را برای تصـمیم‌گیری راهنمایی نماید. در این بین چالش واسـنجی[۱] موضـوعی بسـیار اساسـی برای این دسـته از الگوریتم‌ها است. واسنجی صـحیح دوربین می‌تواند موردی باشد که توسط آن الگوریتم‌ها به درک خطی از فاصله اشیا از همدیگر برسند. موردی که این مساله را دشوارتر می‌نماید آن است که تغییر در محل قرارگیری دوربین می‌تواند نتایجی که به کمک اطلاعات واسنجی بدست آمده است را دستخوش تغییر کند.

یک مورد پرکاربرد از موارد اسـتفاده واسـنجی دید از بالا اسـت. در این کاربرد تصـویری که از جلو تهیه شـده به یک تصـویر از نگاه یک عامل که از بالا اشیا صحنه را ثبت می‌کند، نگاشت می‌شود. به تازگی از شبکه‌های عصبی عمیق برای حل این مساله استفاده شده است. قابلیتی که این شبکه‌ها در حل این چالش ارائه می‌کنند مورد تغییر مکان دوربین است. شبکه‌های عصـبی در حل این چالش می‌توانند تا حد خوبی همچنان نتایج نزدیک به واقعیتی پس از تغییر مکان دوربین ارائه دهند. در این پژوهش مجموعه داده‌ای مرتبط با مساله حاضر تولید شده است و یک مدل بر پایه شبکه‌های پیچشی ارائه شده است که تا حد مناسبی توانسته است بر روی داده‌ها آموزش ببیند.

کلمات کلیدی

پردازش تصویر، دید از بالا، واسنجی، شبکه‌های عصبی عمیق

۱- مقدمه

در راهبری ربات‌ها و دستگاه‌های خودران از حسگرهای متنوعی استفاده می‌شود. بطور معمول دوربین تصویربرداری رنگی یکی از حسگرهای پرکاربرد در چنین کاربردهایی هست. دوربین تصویربرداری با توجه به قیمت مناسب خود، اطلاعاتی را در اختیار ما قرار می‌دهد که می‌تواند بسیار متنوع و باارزش باشد. در این راستا الگوریتم‌های پردازش تصویر توانسته‌اند به روش‌های مختلف این اطلاعات را پردازش کنند و تحلیل‌های گوناگونی را ارائه کنند.

یک نمونه از اطلاعات استخراج شده پیرامون اطلاعات مکانی اشیا است. محل قرارگیری اشیا و فاصله آن‌ها از یکدیگر به طریقی که با یک ضریب می‌توان فاصله‌ها را به متر یا معیار دیگری تبدیل کرد، از جمله این اطلاعات است. تصویر از بالا نمونه تصویر پردازش شده‌ای است که به نحو مطلوبی می‌تواند این اطلاعات را در اختیار ما قرار دهد.

رویکردی که برای محاسبه تصویر از بالا در بسیاری از کاربردها به کارگرفته می‌شود، استفاده از روش‌های انتقال با استفاده از ماتریس‌های هوموگرافی است. در این روش با تنظیم‌کردن دوربین، مقادیری برای ماتریس انتقال محاسبه می‌شود که هر پیکسل از تصویر را به فضای تصویر از بالا انتقال می‌دهد. در این روش چالشی که ایجاد می‌شود، این است که با تغییر مکان دوربین مقادیر ماتریس انتقال نیز باید بروز رسانی شود تا انتقال به نحو مناسبی انجام شود.

روش‌های مبتنی بر یادگیری عمیق در موضوعات گوناگونی به کارگرفته شده‌اند و نتایج بسیار دقیقی را تولید کرده‌اند. برای حل چالش فوق نیز روش-هایی پیشنهاد شده است که می‌توان با بکارگیری آن‌ها به نتایج دقیقی رسید. قابلیتی که شبکه‌های عصبی می‌توانند در اختیار ما قرار دهند زمانی است که مکان دوربین تغییر کرده است. در این شرایط نیز می‌توان نتایج نسبتا دقیقی را تولید نمود.

در رویکردهای قبل قسمتی از تصویر به همراه حدود اشیا در تصویر از جلو در اختیار شبکه قرار می‌گرفت تا مکان قرارگیری اشیا در تصویر آرمانی از بالا محاسبه شود. در اینجا روشی بررسی شده است که کل تصویر به همراه مکان قرار گیری اشیا توسط شبکه پردازش شده است تا مکان قرار گیری شی در تصویر آرمانی محاسبه شود. این روش برای بار نخست به کار گرفته شده است و معماری برای آن پیشنهاد شده است.

جهت آموزش چنین شبکه‌هایی تولید دادگان مناسب بر مبنای دادگان واقعی بسیار امر سخت و هزینه‌بری است. همچنین دادگان در دسترس و کاملی که بتوان چنین شبکه‌هایی را آموزش داد در حال حاضر موجود نیست. برای همین

[۱] Calibration

(*) نویسنده مسئول

موضوع در این پژوهش رویکردی برای تولید دادگانی مناسب آموزش شبکه اتخاذ شده است. با کمک این رویکرد مقدار مناسبی داده به کمک شبیه سازهای موجود تولید شده است که در دیگر کارها نیز می‌توان به کارگرفت.

در ادامه مروری برکارهای گذشته انجام شده است و سپس الگوریتم معرفی شده، تشریح شده است. رویکرد تولید دادگان قسمت بعدی است که نحوه تولید دادگان در آن مشخص شده است. در نهایت ارزیابی از عملکرد مدل روی دادگان گزارش شده است.

۲- کارهای پیشین

جهت افزایش کارکرد و بهبود الگوریتم‌ها و روش‌های جهت‌یابی اتومبیل خودران، نقشه‌های از بالا به پایین که از تصاویر به دست آمده از دوربین‌هایی که در اطراف ماشین و از جمله سمت جلویی ماشین بدست آمده باشد میتواند کمک‌کننده باشد. بیشتر تحقیق‌های انجام شده در این زمینه با استفاده از تصاویر رنگی سعی در ایجاد تصویری به سبک بالا به پایین که در آن شی‌های مختلف مشخص شده باشند داشتند. برای این کار، نتیجه بدست آمده به این صورت بود که روش یادگیری عمیق نسبت به بقیه روش‌های کلاسیک، هندسی و پردازش تصویری عملکرد بهتری هم در زمینه دقت و هم در زمینه کارایی داشتند.

۲-۱- ایجاد نمایش بالا به پایین از تصاویر

برای تبدیل تصاویر گرفته شده به نقشه بالا به پایین، کارهای محدودی در گذشته انجام شده که به طور کلی میتوان آن‌ها را به سه بخش تقسیم کرد. در بخش اول از آنجایی که این کار به نوعی جزو دسته کارهای ترجمه تصویر به تصویر[2] قرار میگیرد و فضای تصاویر ورودی با فضای تصاویر خروجی کاملا متفاوت است، از مدل‌های مولد مقابله ای[۳] استفاده شده است. در [۵] با استفاده از این مدل تصاویر و فیلم‌های رنگی که از جهت جلوی ماشین گرفته شده اند به تصاویر و فیلم‌های از بالا به پایین تبدیل می شوند. در [۳] کار مشابهی انجام شده است و مدل علاوه بر تصویر از جلو، مختصات جعبه‌های مرزبندی[۴] های ماشین‌های داخل تصویر را به عنوان پارامتر کمکی شبکه عصبی دریافت می‌کند. در [۹] به مدل مولد استاندارد یک تمیز دهنده[۵] اضافه شده جهت تطبیق یکتایی خروجی تولید شده توسط تولیدکننده که باعث افزایش کارایی مدل شده است. در بعضی از کارها نیز از روش هندسی هموگرافی در کنار مدل تولیدکننده استفاده شده است تا ابتدا یک تصویر از بالا به پایین توسط این روش تولید شود، در [۶] با معرفی مدل BridgeGAN که از چند تولیدکننده تشکیل شده است[۶]، تصویر اصلی به همراه تصویر تولید شده توسط روش هموگرافی وارد مدل می شوند.

در تقسیم بندی بعدی، از مدل‌های خودرمزگذار[۷] استفاده شده است. در [۷] با استفاده از یک شبکه Variational Encoder-Decoder تصویر از سمت جلو ماشین به نقشه از بالا به پایین شبکه اشغال معنایی تبدیل می‌شود و هر پیکسل نشان دهنده یکی از ۴ کلاس پیش‌بینی شده توسط مدل می‌باشد که نشان‌دهنده اشغال و یا آزاد بودن آن است. در [۴] با استفاده از تصاویر بدست آمده از سنجنده‌های لیدار[۸]، رادار[۹]، و تصویر از جلو و با استفاده از چندین مدل خودرمزگذار که به صورت موازی بر روی این تصاویر اجرا می‌شوند، تصاویر به صورت فریم‌های پشت هم وارد مدل شده و تصویرهای نقشه از بالابه‌پایین فریم‌های بعدی پیش بینی می شود. در [۱۱] کار مشابهی و فقط با استفاده از تصاویر قطعه بندی معنایی[۱۰] به عنوان تصویر ورودی پیاده سازی شده است.

در تقسیم بندی آخر، از تبدیل‌کننده‌ها[۱۱] در کنار شبکه‌های عصبی پیچشی[۱۲] استفاده شده تا از ویژگی‌های استخراج شده توسط شبکه عصبی بهره ببرند و آن‌ها را به فضای نقشه از بالا به پایین تبدیل کنند [۱۲, ۱۰, ۱]. در این روش عموما مدل شامل یک قسمت پشتی[۱۳] است که از یک مدل آموزش دیده استفاده می شود تا ویژگی‌های داخل تصویر استخراج شود. سپس، این ویژگی‌ها تجمیع‌شده و وارد تبدیل‌کننده می شوند تا از فضای تصویر ورودی به فضای ۲ بعدی نقشه از بالا به پایین تبدیل شوند. سپس با استفاده از یک شبکه عصبی دیگر نقشه تولید شده قطعه‌بندی معنایی میشود.

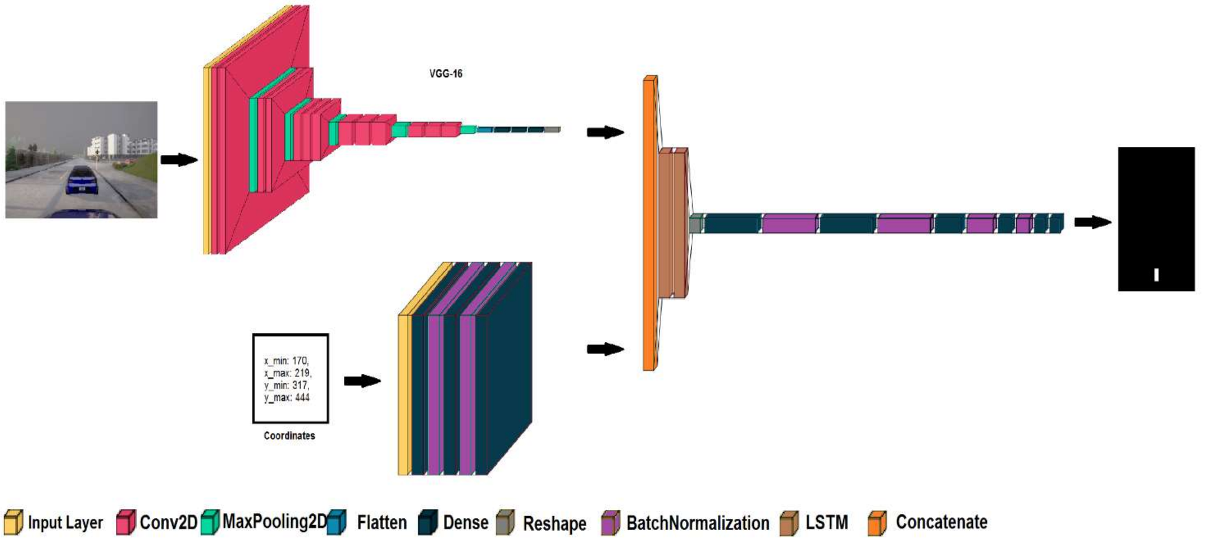
۲-۲- مجموعه داده های مرتبط

در سال های گذشته چندین مجموعه داده از فضاهای خیابان، عابران پیاده و ماشین‌ها عموما به خاطر سادگی در گرفتن عکس‌های رنگی منتشر شده است که در بعضی از آن‌ها تفسیرهای ۲ بعدی مانند مرزبندی برای تشخیص شی و نقشه‌های معنایی برای قطعه قطعه سازی در شرایط میحطی و آب و هوایی گوناگون وجود دارد. در منبع [۱۳] مجموعه داده‌ای مشهوری در این زمینه ارائه شده است که علاوه‌بر تصاویر رنگی که از دوربین های نصب شده در ۶ جهت ماشین می‌گیرد، از حسگرهای لیدار و رادار نیز بهره می‌برد و یک نمای ۳۶۰ درجه را به عنوان خروجی تولید می کند. این مجموعه داده از ۱۰۰۰ سکانس ۲۰ ثانیه‌ای تشکیل شده و شامل جعبه‌های مرزبندی ۳ بعدی برای ۲۳ کلاس مختلف نیز است که نسبت به مجموعه داده‌های قبلی داده های بیشتر و با کیفیت‌تری را دارا است. در منبع [۸] نیز مجموعه داده مشابهی ارائه شد که علاوه‌بر عکس‌های رنگی از حسگر لیدار نیز بهره می‌برد ولی برخلاف مجموعه داده قبلی، تنها شامل نمای جلو ماشین بوده و نمای ۳۶۰ درجه را پشتیبانی نمی‌کرد.

مجموعه داده مشابه‌ای که همراه با عکس‌های رنگی از جلو و جعبه‌های مرزبندی، نقشه‌های از بالا به پایین که شی‌های مختلف را به صورت قطعه بندی معنایی نمایان کند وجود ندارد و مجموعه داده معرفی شده در این مقاله اولین کار در این زمینه است که الگوریتم‌های اتومبیل خودران میتوانند با استفاده از نقشه‌های از بالا به پایین تولید شده، فاصله و جهت حرکت دیگر ماشین‌ها را بدست بیاورند و کارکردشان را بهبود ببخشند.

---

[2] Image-to-image translation
[۳] GAN
[۴] Bounding box
[۵] Discriminator
[۶] Multi-GAN
[۷] Autoencoder
[۸] Lidar
[۹] Radar
[۱۰] Semantic segmentation
[۱۱] Transformers
[۱۲] Convolution neural networks
[۱۳] Backbone



شکل ۱: شمای معماری مدل ارایه شده.

۳- الگوریتم پیشنهادی

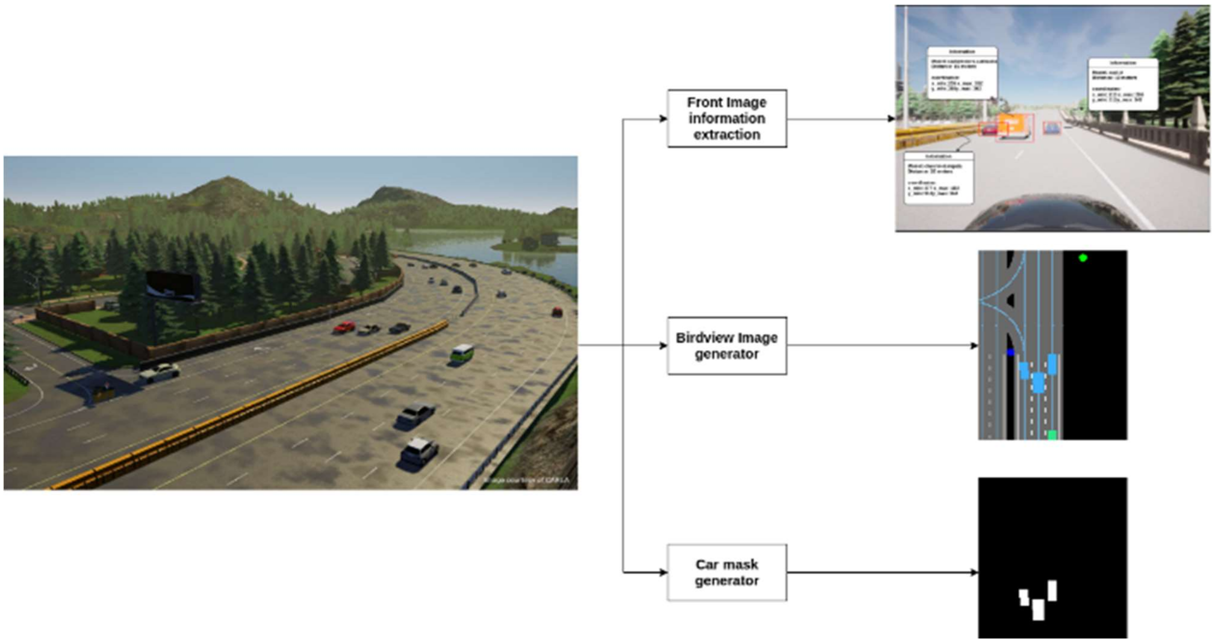
در این پژوهش، هدف تولید نمای بالا به پایین از تصاویر رنگی است. به طور معمول تولید تصاویر نمای از بالا نیازمند قرارگیری دوربین در مکانی است که در کاربرد های عملی، ایجاد چنین شرایطی غیرممکن است. تصاویر در کاربردهای معمول از نمای جلو ثبت می‌شوند. تصاویر نمای جلو ماشین توسط دوربینی که بر روی داشبورد خودرو به سمت جلو قرار داده شده است، تولید می‌شوند. تبدیل تصاویر نمای جلو به تصویر از بالا خود چالش‌هایی دارد که یکی از آن‌ها عدم ثبت اطلاعاتی است که در تصویر آرمانی از بالا باید بازسازی شوند. به عنوان نمونه در تصویر از جلو دوربین خودرو ممکن است، ماشین‌های دیگر به صورت کامل قابل مشاهده نباشند. بنابراین ممکن است در تخمین اندازه و موقعیت مکانی آن‌ها خطایی صورت گیرد، لذا از الگوریتمهای کلاسیک و هندسی که مقدار این خطا را افزایش می دهند نمی‌توان استفاده کرد زیرا فضای تصاویر ورودی با نقشه‌های خروجی به صورت کامل متفاوت است. به عنوان جایگزین برای حل این مساله مدل شبکه عصبی‌ای را ارایه خواهیم کرد که با استفاده از تصاویر ورودی و مختصات جعبه‌های مرزبندی در تصویر ورودی ارتباط بین آن‌ها را یاد میگیرد و سپس مختصات جعبه مرزبندی ماشین ورودی را در تصویر مقصد تخمین میزند.

الگوریتم پیشنهادی ما همانطور که در تصویر ۱ مشخص است از ۲ شاخه اصلی تشکیل شده است. در شاخه اول تصاویر بدست آمده از نمای جلویی ماشین به عنوان ورودی استفاده می‌شوند. در این شاخه از مدل یادگیری عمیق VGG۱۶ [۱۵] از قبل آموزش دیده بر روی دادگان ImageNet [۱۶] به دلیل داشتن پارامترهای کمتر و عملکرد خوب نسبت به بقیه مدل‌های از قبل آموزش دیده استفاده می‌شود تا بازنمایی عمیق و ویژگی‌های تصویر ورودی از آن استخراج شود. استفاده از این مدل به ما این قابلیت را می دهد که در حین کم بودن تعداد پارامترها و در نتیجه ساده‌تر و سریعتر شدن محاسبات، بتوانیم ویژگی‌های معنایی داخل تصاویر ورودی را حتی بدون داشتن مختصات جعبه-های مرزبندی ماشین‌ها استخراج کنیم. لازم به ذکر است که آخرین لایه متراکم[۱۴] این مدل که به منظور حل مساله‌ی طبقه‌بندی[۱۵] طراحی شده، حذف شده است. به صورت موازی در شاخه دوم، از مختصات جعبه مرزبندی یک ماشین در تصویر ورودی به صورت ۴ عدد مجزا استفاده می‌شود. این شاخه از یک شبکه چندلایه پرسپترون[۱۶] تشکیل شده است، که از ۴ لایه متراکم کاملا متصل بهم[۱۷] استفاده شده تا مقادیر ورودی را به یک بردار ۲۵۶ بعدی تبدیل و رمزگذاری کند. این قسمت از مدل به این خاطر که صرفا مختصات جعبه‌های مرزبندی را دریافت می کند نوع و معنای شی‌ها و ویژگی‌های داخل تصاویر را یاد نمی‌گیرد و هدف اصلی آن یادگیری نگاشت مکانی[۱۸] بین مختصات ورودی و مختصات در تصویر خروجی و زاویه قرار گیری دیگر ماشین‌ها نسبت به ماشین آرمانی است.

ویژگی‌های استخراج شده از این دو لایه در شاخه نهایی به یکدیگر می‌پیوندند. در ابتدا این ویژگی‌ها وارد دو لایه کاملا بهم متصل شده دو سویه حافظه کوتاه-مدت طولانی[۱۹] می‌شوند. این دو لایه ارتباط بین ویژگی‌های ستونی استخراج شده در تصویر اصلی را با یکدیگر استخراج می‌کنند. سپس این خروجی وارد یک شبکه کاملا به هم متصل رمزگشا می‌شود که در آخر مختصات جعبه مرزبندی ماشین ورودی در نقشه از بالا به پایین را محاسبه می‌کند.

---

[۱۴] Dense layer
[۱۵] Classification
[۱۶] MLP
[۱۷] Fully-connected
[۱۸] Spatial transformation
[۱۹] Bidirectional LSTM



شکل ۲: نمای روش بکارگرفته شده برای استخراج دادگان.

۳-۱- جزییات آموزش

در ابتدا نرمال سازی بر روی مقادیر پیکسل‌های تصاویر ورودی انجام گرفت و مختصات‌های جعبه‌های مرزبندی در ورودی و خروجی به نسبت اندازه تصاویر به ترتیب ورودی و خروجی مقیاس‌گذاری شده‌اند به طوری که تمامی مختصات‌ها به صورت نسبی بین اعداد صفر و یک می باشند. سپس تصاویر به اندازه ۲۲۴ * ۲۲۴ پیکسل که اندازه مقدار ورودی مدل VGG۱۶ است تبدیل شدند. پارامترهای این مدل در طول مدت فرایند یادگیری ثابت هستند و تغییر نمی‌کنند. از لایه‌های نرمال سازی دسته‌ای[۲۰] به منظور حفظ پایداری[۲۱] مدل و مقابله با بیش‌برازش[۲۲] قبل از هر لایه متراکم استفاده شده است. مدل با استفاده از تابع هزینه[۲۳] خطای میانگین مربعات[۲۴] بین مختصات پیش‌بینی‌شده و مختصات اصلی آموزش داده شده و سعی دارد این مقدار را به کمترین مقدار خود برساند. از بهینه‌ساز[۲۵] Adam [۱۷] برای این منظور استفاده شده است.

۴- رویکرد ساخت دادگان

برای ساخت دادگان از محیط شبیه سازی متن باز CARLA [14] استفاده شده است. در این محیط ابزارهای شبیه سازی اتومبیل های خودران و امکان توسعه طرح‌های متفاوت خودرانی وجود دارد. در این شبیه‌ساز امکان تعریف طرح حرکتی برای هر یک از المان‌های جهان شبیه‌سازی شده موجود است. همچنین نقشه‌هایی با محیط‌های متفاوت طراحی شده است که امکان تعریف حالت‌های مختلف را فراهم کرده است. شرایط آب و هوایی مختلف، چیدمان شهری مختلف و همچنین انواع مختلف وسایل نقلیه از جمله امکانات این شبیه ساز است.

جهت توسعه دادگان با استفاده از شبیه ساز ابتدا یک ماشین آرمانی[۲۶] تولید شده است، المان‌هایی همچون وسایل نقلیه و افراد که با ماشین آرمانی در تعامل هستند نیز ایجاد شده است. برای هر یک از این المان‌ها یک طرح تصادفی در نظر گرفته شده است که مشخص کننده نحوه حرکت آن در جهان شبیه سازی شده و تعامل آن با ماشین آرمانی است.

جهت آموزش مدل نگاشت دهنده لازم است تصویر از جلو، اطلاعات مکانی اشیا در تصویر از جلو و اطلاعات مکانی اشیا در تصویر آرمانی که از بالا اطلاعات اشیا را نشان می‌دهد مشخص باشد. در طراحی و ساخت مجموعه داده علاوه بر موارد اشاره شده اطلاعات دیگری نیز در نظر گرفته شده است که مجموعه داده حاضر را قابل استفاده در دیگر کاربردها نیز تبدیل نموده است. مواردی همچون مکان قرار گیری اشیا ثابت مانند ساختمان‌ها، حدود خیابان‌ها در تصویر قطعه‌بندی شده‌ی معنایی از بالا، فاصله واقعی ماشین ها از یکدیگر و مشخص نمودن زمان برخورد از جمله این موارد است.

برای ثبت و نمونه برداری از ماشین آرمانی از چند نوع حسگر مختلف استفاده شده است. یک نمونه از حسگرهای استفاده شده یک دوربین تصویربرداری رنگی است که در جلوی ماشین تنظیم شده است. از اطلاعات مکانی ماشین‌ها استفاده شده است تا تصویر آرمانی از بالا ساخته شود. در شبیه ساز حدود فضایی که هر شی در جهان شبیه‌سازی شده اشغال کرده است با یک مکعب فرضی مشخص شده است. با استفاده از مختصات قرارگیری این مکعب حدود قرارگیری هر شی در تصویر دو بعدی نیز مشخص شده است. برای

---

[۲۰] Batch normalization
[۲۱] Regularization
[۲۲] Overfitting
[۲۳] Loss function
[۲۴] MSE
[۲۵] Optimizer
[۲۶] Ego agent

محاسبه فاصله هر دو اتومبیل از یکدیگر از مقادیر مکانی که در شبیه‌ساز موجود است، استفاده شده تا این مقدار فاصله محاسبه شود. اطلاعات جاده و مکان قرار گیری اشیا دیگر از جمله اطلاعاتی است که از تحلیل برخط اطلاعات شبیه ساز در هر فریم حساب شده و به مجموعه داده اضافه می شود.

در تصویر برداری با استفاده از حسگر دوربین در یک زمان مشخص یکسان بطور متناوب از محیط یک نمونه ثبت می شود. بطور همزمان جعبه‌های مرزبندی[۲۷] برای ماشین‌هایی که در محدوده‌ی ۵۰ متری باشد محاسبه شده و برای ماشین‌ها به عنوان اطلاعات جانبی در نظر گرفته می‌شود. کلیه اطلاعاتی که از روبرو جمع‌آوری شده در قالب تصویر و اعداد ساختار یافته ذخیره شده است.

بطور موازی اطلاعات هر یک از ماشین‌ها و اشیایی که در محدوده ماشین آرمانی هستند پردازش می‌شوند. پس از پردازش این اطلاعات بصورت یک تصویر قطعه بندی شده از بالا ، اطلاعات ماشین‌ها در تصویر از بالا بصورت دادگان ساختار یافته و یکسری از اطلاعات جانبی ذخیره می‌شود. در تصویر قطعه بندی شده از بالا هر مقدار نمایانگر نوع دادگانی است که اطلاعات آن ذخیره شده است. هشت مقدار متفاوت برای هشت نوع دادگانی که ذخیره شده است در نظر گرفته شده است.

در این دادگان نزدیک به 16 مدل خودرو و ۶ نقشه مختلف که هر کدام وضعیت آب و هوایی متفاوتی دارند استفاده شده است. اگر خودرو آرمانی در ۸ فریم زمانی به طور متوالی ثابت بماند به منظور کاهش داده ی تکراری مکان خودرو در نقشه عوض می شود. از هر نقشه نزدیک به ۲۰۰۰ داده به همراه ۵۰ خودرو دیگر تولید شده است. نمایی از دادگان تولید شده در جدول ۱ قابل مشاهده است. در تصویر شماره ۲ شمای کلی تولید دادگان با استفاده از شبیه ساز و رابط‌های توسعه داده شده نشان داده شده است. همچنین اطلاعات سیستم مورد استفاده در مراحل آموزش مدل و ساخت دادگان در جدول ۲ ذکر شده است.

جدول ۱: اطلاعات دادگان تولید شده

پارامتر	تعداد
تصاویر تولیدشده	۱۰،۰۰۰
جعبه‌های مرزبندی یکتا	۴۰۰۰
نرخ فریم	۲۰FPS
نرخ نمونه‌برداری	۱ عکس در هر ۲ ثانیه
اندازه تصاویر نمای جلویی ماشین	۶۰۰ * ۸۰۰ پیکسل
اندازه تصاویر نقشه از بالا به پایین	۲۰۰ * ۱۵۰ پیکسل
سایز تصاویر نمای جلوی ماشین به طور میانگین	۶۶۵ کیلوبایت برای هر تصویر
سایز تصاویر نمای نقشه از بالا به پایین به طور میانگین	۴ کیلوبایت برای هر تصویر
محدوده فاصله اندازه‌گیری شده	۵۰ متر از ماشین آرمانی
حجم دادگان	۱۰ گیگابایت

جدول ۲: اطلاعات سیستم استفاده شده برای آموزش مدل و ساخت دادگان

قطعه	مدل
CPU	Ryzen ۱ ۱۷۰۰
GPU	Nvidia GeForce RTX ۲۰۸۰ Super
RAM	۱۶ گیگابایت

۵- نتایج و ارزیابی

در این قسمت نتایج مدل معرفی شده بر روی دادگان تولید شده را بررسی می‌کنیم. از آنجایی که این کار برای اولین بار با دادگان شبیه‌سازی شده و با استفاده از تصاویر نمای جلویی ماشین و مختصات جعبه‌های مرزبندی صورت گرفته است مدل‌های معرفی شده قبلی کارآمد نخواهند بود و نتایج را بر روی مدل معرفی شده در این مقاله بررسی خواهیم کرد.

مدل معرفی شده در هنگام یادگیری نگاشت بین تصاویر ورودی و مختصات خروجی سعی می کند که مختصات خروجی را به مختصات مرکز تصویر خروجی نزدیک کند زیرا این کار باعث می شود تا مقادیر تابع هدف به یک مقدار کمینه محلی[۲۸] برسند و بعد از مدتی در هنگام آموزش این مقدار در تابع هدف دیگر کاهش نمی‌یابد. مقدار تابع هدف برای دادگان آموزش بعد از همگرایی مدل برابر با ۰.۲۱ و برای دادگان ارزیابی که حدود ۵۰ داده بودند برابر با ۰.۳۰ گزارش شده است.

این مدل بعد از طی کردن ۱۰ دوره به همگرایی می رسد. همچنین از ۱۰ درصد مجموعه دادگان به عنوان داده‌ی ارزیابی استفاده شده است به طوری که مدل در هنگام آموزش این نوع داده را مشاهده نمی‌کند و نتایج را برای آن تخمین میزند که باعث می شود رفتار مدل در برابر مثال هایی که از قبل مشاهده نکرده است افزایش یابد. همچنین پارامتر سرعت یادگیری[۲۹] با استفاده از یک زمان‌بند تنظیم شده است تا به مقدار کمینه عمومی نزدیک‌تر شود. اطلاعات مدل معرفی شده در جدول زیر مشخص می‌باشند.

جدول ۳: اطلاعات مدل

المان	تعداد
تعداد پارامترها	۱۵۴،۸۳۳،۸۳۶
تعداد پارامترهای یادگیرنده	۱۶،۴۶۹،۶۳۶
تابع هدف	MSE
بهینه ساز	Adam with momentum
زمان آموزش برای هر دوره	۵ ثانیه
تعداد داده‌ها در هر دسته	۳۲

[۲۷] bounding box
[۲۸] Local minimum
[۲۹] Learning rate

۶- نتیجه‌گیری

در این پژوهش روشی مبتنی بر یادگیری عمیق ارائه شده است که توانایی تعیین محدوده اشیا تشخیص داده شده در تصویر از جلو، در تصویر آرمانی از بالا را دارد. برای آموزش این مدل از دادگانی استفاده شده است که برای توسعه چنین مدل‌هایی در این پژوهش تولید شده است. دادگان به گونه‌ای ساخته شده‌اند که قابلیت بکارگیری در دیگر کاربردهای مشابه را نیز دارند. بنا بر بررسی های ما این رویکرد برای نخستین بار به صورتی که در این مقاله ارایه شد در این حوزه انجام شده است. در کارهای آینده می‌توان با ارایه یک معماری بزرگتر با تعداد پارامترهای بیشتری مدلی ارایه کرد که دقت بسیار مطلوب‌تری را برای این کاربرد تولید کند. همچنین مجموعه دادگان می‌تواند در تعداد تصاویر و مقدار اطلاعات بیشتری تولید شود.

مراجع

[1] N. Gosala and A. Valada, "Bird's-Eye-View Panoptic Segmentation Using Monocular Frontal View Images," *IEEE Robotics and Automation Letters,* vol. 7, no. 2, 2022.

[2] K. Regmi and A. Borji, "Cross-view image synthesis using geometry-guided conditional GANs," *Computer Vision and Image Understanding,* vol. 187, 2019.

[3] H. Fraser and S. Wang, "DeepBev: A conditional adversarial network for bird's eye view generation," *Proceedings - International Conference on Pattern Recognition,* 2020.

[4] N. Hendy, C. Sloan, F. Tian, P. Duan, N. Charchut, Y. Xie, C. Wang and J. Philbin, "FISHING Net: Future Inference of Semantic Heatmaps In Grids," arXiv:2006.09917v1, 2020.

[5] V. Jain, Q. Wu, S. Grover, K. Sidana, G. Chaudhary, S. H. Myint and Q. Hua, "Generating Bird's Eye View from Egocentric RGB Videos," *Wireless Communications and Mobile Computing,* 2021.

[6] X. Zhu, Z. Yin, J. Shi, H. Li and D. Lin, "Generative adversarial frontal view to bird view synthesis," International Conference on 3D Vision (3DV), 2018.

[7] C. Lu, M. J. G. Van De Molengraft and G. Dubbelman, "Monocular Semantic Occupancy Grid Mapping With Convolutional Variational Encoder-Decoder Networks," *IEEE Robotics and Automation Letters,* vol. 4, no. 2, 2019.

[8] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan and O. Beijbom, "Nuscenes: A multimodal dataset for autonomous driving," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition,* 2020.

[9] T. Zhou, D. He and C. H. Lee, "Pixel-level bird view image generation from front view by using a generative adversarial network," *2020 6th International Conference on Control, Automation and Robotics, ICCAR 2020,* 2020.

[10] T. Roddick and R. Cipolla, "Predicting semantic map representations from images using pyramid occupancy networks," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition,* 2020.

[11] L. Hoyer, P. Kesper, A. Khoreva and V. Fischer, "Short-term prediction and multi-camera fusion on semantic grids," *Proceedings - 2019 International Conference on Computer Vision Workshop, ICCVW 2019,* 2019.

[12] A. Saha, O. M. Maldonado, C. Russell and R. Bowden, "Translating Images into Maps," International Conference on Robotics and Automation (ICRA), 9200-9206, 2022.

[13] A. Geiger, P. Lenz, C. Stiller and R. Urtasun, "Vision meets robotics: The KITTI dataset," *International Journal of Robotics Research,* vol. 32, no. 11, 2013.

[14] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez and V. Koltun, "CARLA: An Open Urban Driving Simulator," *Proceedings of the 1st Annual Conference on Robot Learning,* pp. 1-16, 2017.

[15] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," CoRR journal, vol. 1409.1556, 2014.

[16] O. Russakovsky, J. a. S. H. Deng, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," International Journal of Computer Vision 115(3), 2014.

[17] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," CoRR journal, vol. 1412.6980, 2014.

ANN-EAM: یک چارچوب شبکه عصبی مصنوعی – انباشت شواهد برای مدل سازی تصمیم‌گیری واژگان انسانی

آرش دادرس[۱]، ماه وش سیاوش پور[۲]، جمال امانی‌راد[۳]

[۱] دپارتمان مدل‌سازی شناختی، پژوهشکده علوم شناختی و مغز، دانشگاه شهید بهشتی، تهران،
ar.dadras@mail.sbu.ac.ir

[۲] دانشکده مهندسی کامپیوتر، دانشگاه امیرکبیر، تهران،
m.siavashpour@aut.ac.ir

[۱۳] دپارتمان مدل‌سازی شناختی، پژوهشکده علوم شناختی و مغز، دانشگاه شهید بهشتی، تهران،
J_amanirad@sbu.ac.ir

چکیده

بازنمایی دانش مشکل اصلی مدل‌های تصمیم‌گیری مانند مدل‌های انباشت شواهد می‌باشد. پیشرفت‌های اخیر حوزه هوش مصنوعی نویدبخش حل این مشکل با استفاده از چارچوب شبکه عصبی مصنوعی – انباشت شواهد یا به اختصار ANN-EAM می‌باشد. در این چارچوب محرک‌های ورودی توسط یک شبکه عصبی مصنوعی بازنمایی می‌شوند. سپس از بازنمایی‌های تولید شده پارامترهای مدل انباشت شواهد به وسیله توابع نگاشت تعیین و تخمین زده می‌شوند. در مطالعه پیش‌رو توانایی پردازش کلمه در انسان با ارائه دو مدل از چارچوب شبکه عصبی مصنوعی – انباشت شواهد مدل‌سازی شده است. مدل اول از ترکیب مدل FastText با مدل انتشار مسابقه‌ای و مدل دوم از ترکیب مدل BERT با مدل انتشار مسابقه‌ای ساخته شده‌اند. برای آزمایش عملکرد این مدل‌ها، آن‌ها بر روی داده های تکلیف تصمیم‌گیری واژگانی که از مجموعه‌داده English Lexicon Project بدست آمده‌اند برازش می‌شوند. نتایج به‌دست‌آمده نشان می‌دهد که مدل اول ضمن ارائه بازنمایی معنایی برای کلمات وغیرکلمات، زمان واکنش و دقت افراد را در مقایسه با مدل دوم بهتر پیش‌بینی می‌کند. همچنین مدل اول عملکردی قابل قبول و مشابه با مدل انتشار مسابقه‌ای معیار که به تنهایی و بر داده‌ها برازش شده بود را دارد. به طور کل می‌توان نتیجه گرفت که ANN-EAM چارچوبی مناسب برای مدل‌سازی تکلیف تصمیم‌گیری واژگانی است.

کلمات کلیدی

مدل‌سازی شناختی، شبکه عصبی مصنوعی، تصمیم‌گیری واژگانی، مدل انباشت شواهد، تصمیم‌گیری

۱- مقدمه

حوزه‌های علوم شناختی و هوش مصنوعی (AI) هر دو در تلاش هستند که چگونگی کسب، بازنمایی و تغییر دانش توسط یک عامل هوشمند را توضیح دهند [۲۰]. تمرکز علوم شناختی و روانشناسی بر بررسی و توضیح مکانیسم‌هایی است که در این فرآیندها دخیل می‌باشند. این در حالی است که تمرکز اصلی هوش مصنوعی برساخت سیستم‌هایی است که توانایی انجام برخی پیش‌بینی‌ها را براساس دانش کسب شده داشته باشند [۱]. درسال‌های اخیر این دو حوزه از نظریات و کشفیات یکدیگر به منظور رسیدن به اهداف خود استفاده می‌کنند. نظریات مطرح در علوم شناختی سبب بهبود عملکرد شبکه‌های عصبی شده‌اند برای مثال، استفاده از مکانیسم توجه در شبکه‌های عصبی [۹,۳۳]. از طرف دیگر در علوم شناختی نیز از روش‌های هوش مصنوعی، در مدل‌سازی و توضیح مکانیسم‌های زیربنایی ذهن و رفتار انسان استفاده شده است [۴,۵,۱۷].

نظریه انباشت شواهد قوی‌ترین و محبوب‌ترین چارچوب برای مدل‌سازی تصمیم‌گیری است. بر اساس این چارچوب، هر زمان که مقدار اطلاعات به نفع یک گزینه به یک حد آستانه خاص برسد، تصمیم گیری انجام می‌شود. بنابراین، یک انباشتگر شواهد، فرآیند انباشتن شواهد (اطلاعات) را از یک نقطه شروع کرد و شواهد را به صورت نویزی، با یک نرخ ثابت (نرخ رانش) تا رسیدن به یک مقدار مشخص (حد آستانه) انباشته می کند [۲۴].

ایراد اصلی مدل‌های انباشت شواهد عدم درنظرگرفتن مکانیزمی برای بازنمایی ویژگی‌های محرک می‌باشد. علاوه بر این، مکانسیم بازنمایی دانش نیز در این مدل‌ها درنظر گرفته نشده است [۲۸]. برای حل این مشکل مدل‌های ترکیبی از شبکه‌های عصبی مصنوعی و مدل‌های انباشت شواهد پیشنهاد شده‌اند. نقش شبکه‌های عصبی در این مدل‌های ایجاد بازنمایی از محرک برای مدل انباشت شواهد می‌باشد. این خانواده از مدل‌ها ANN-EAM نامیده می‌شوند.

در این مطالعه، از چارچوب ANN-EAM برای مدل‌سازی عملکرد انسان در تکلیف تصمیم‌گیری واژگانی استفاده شده است. این تکلیف محبوب ترین چارچوب آزمایشی روانشناختی برای بررسی توانایی تشخیص کلمات، می‌باشد. در این تکلیف، به شرکت‌کنندگان مجموعه‌ای از رشته‌های حروف ارائه می‌شود و از آنها خواسته می‌شود که تصمیم بگیرند که آیا رشته حروف ارائه شده یک کلمه در زبان آن‌ها می‌باشد یا خیر [۳]. به منظور مدل سازی رفتار انسان و کسب دانش در مورد مکانیسم‌های اساسی تصمیم‌گیری واژگانی، بسیاری از مطالعات از مدل‌های انباشت شواهد استفاده کرده اند [۲۳,۳۰,۳۱,۳۲]. با وجود پیش‌بینی‌های دقیق این مدل‌ها از رفتار انسان، این مدل‌ها از هیچ مکانیزمی برای نمایش ویژگی‌های محرک استفاده نمی‌کنند.

ادامه این مقاله به شرح زیر تنظیم شده است: در بخش بعدی، مروری بر سابقه مدل‌های ANN-EAM ارائه شده است. پس از آن، در بخش «روش‌ها»، ANN-EAM های پیشنهادی برای تکلیف تصمیم‌گیری واژگانی مورد بحث قرار می‌گیرند و مجموعه داده[۱]هایی که در این تحقیق

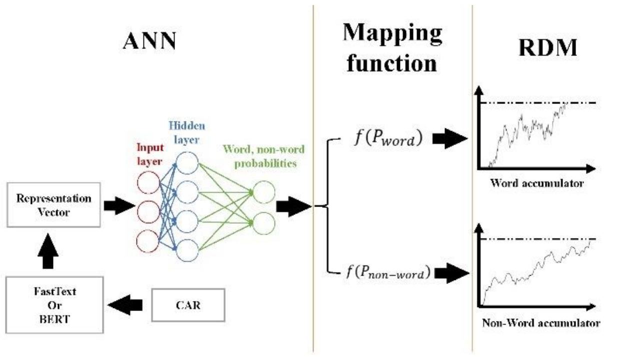
برای ارزیابی مدل‌ها استفاده می‌شوند، ارائه می‌شوند. سپس در بخش «نتایج»، نتایج به‌دست‌آمده از برازش مدل‌ها بر روی داده‌های رفتاری ارائه می‌شود. در نتیجه‌گیری و نکات پایانی در بخش پایانی ارائه شده است.

۲- سابقه مدل‌های ANN-EAM

این دسته از مدل ها سابقه‌ای طولانی ندارند. اولین لیده از این گونه مدل ترکیبی، در سال ۲۰۱۵ ارائه شد که در آن یک مدل یادگیری بیزی همراه با یک مدل انباشت شواهد برای مدل‌سازی فرآیند تصمیم‌گیری انسانی در یک تکلیف یادگیری ابزاری استفاده شد [۱۲]. پس از آن، چندین مدل انباشت شواهد بر مبنای یادگیری تقویتی توسط محققان با هدف مدل‌سازی رفتار انسان در تکلیف multi-armed bandit توسعه داده شد [۱۱,۲۲]. اولین ترکیب یک مدل پیوندگرا با یک مدل انباشت شواهد در سال ۲۰۱۹ برای یک تکلیف یادگیری طبقه‌بندی احتمالی ارائه شد [۲۵]. این مدل از چارچوب نمونه محور برای بازنمایی هر دسته استفاده کرده و سپس به وسیله تابع نگاشت، خروجی مدل نمونه محور را به پارامترهای مدل انباشت شواهد متصل می‌کند. اولین مدل رسمی ANN-EAM از مدل GoogLeNet از پیش آموزش دیده برای ساختن بازنمایی از محرک، به همراه با مدل رانش–انتشار[۲] استفاده کرده است. به طور دقیق تر، این مدل برای تکلیف دسته بندی تصاویر استفاده شده است [۱۸]. علاوه بر این، حیدری–گرجی [۱۶] و همکاران در سال ۲۰۲۱، یک شبکه عصبی اسپایکینگ را توسعه دادند که به دو انباشتگر شواهد برای تکلیف دسته‌بندی چهره/خانه متصل شده بود.

۳- روش‌ها

هدف این بخش ارایه دو مدل محاسباتی برای تصمیم گیری واژگانی بر اساس چارچوب ANN-EAM می‌باشد. به طور خاص، مدل‌های مورد بحث در اینجا براساس دو خوشه اصلی مدل‌های شناختی در حافظه معنایی ارائه شده‌اند. مدل های توزیعی و پیوندگرا. مدل‌های توزیعی مدل‌هایی هستند که نحوه یادگیری مفاهیم و کلمات را با استفاده از تجربیات آماری بیان می‌کنند [۵,۸]. مدل‌های پیوندگرا، مدل‌هایی را توصیف می‌کنند که از بازنمایی‌های مفهومی در فرآیندهای شناختی استفاده می‌کنند [۵,۸]. شماتیک کلی مدل‌های ارائه شده در شکل۱ قابل مشاهده است. فرآیند مدل با ساخت یک بردار بازنمایی از کلمات آغاز میشود. مدل اول که مدلی توزیعی است از FastText برای ساخت بازنمایی استفاده می‌کند و مدل دوم از معماری BERT برای این منظور بهره می‌برد. پس از به دست آوردن بازنمایی از رشته حروف ورودی، با استفاده از یک شبکه عصبی تماما متصل دسته‌بندی رشته‌های ورودی به دو دسته کلمه و غیرکلمه انجام می‌پذیرد. خروجی شبکه بدین صورت می‌باشد که به هر رشته ورودی دو مقدار بین ۰ و ۱ به عنوان احتمال کلمه و غیرکلمه بودن اختصاص می‌یابد. این احتمالات به همراه معیاری از فراوانی کلمات به عنوان ورودی وارد قسمت انباشت شواهد می‌گردند. پارامترهای مدل انباشت شواهد با استفاده از توابع نگاشت ارائه شده تعیین و تخمین زده می‌شوند و درنهایت مدل زمان واکنش شرکت‌کنندگان را پیش‌بینی می‌کند. مدل انتشار مسابقه‌ای در هر دو مدل به عنوان بخش انباشت شواهد در نظر گرفته شده است وبه همین علت از این پس از مدل اول به عنوان FT-RDM و از مدل دوم به عنوان BERT-RDM نام می‌بریم.



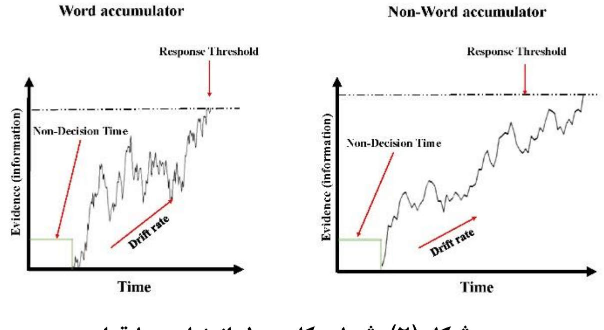
شکل (۱): شمای کلی مدل‌های پیشنهادی ANN-EAM

۱-۳- مدل انتشار مسابقه‌ای

همانطور که در شکل۲ قابل مشاهده است، مدل انتشار مسابقه‌ای[۳] نوعی مدل انباشت شواهد است که که برای هر گزینه یک انباشتگر متفاوت دارد. هر انباشتگر میتواند مجموعه‌ای مشترک یا مجزا از پارامترها با مقادیر متفاوت برای هر گزینه یا شرایط داشته باشد. فرآیند انباشت در این مدل به صورت زیر انجام می‌گیرد:

$$\dot{X}_i(t) = v_i + \xi(t) \qquad (۱)$$

که در آن $\dot{X}_i(t)$ بیانگر انباشتگر i ام در زمان t و v_i بیانگر نرخ رانش (نرخ انباشت شواهد) برای انباشتگر i ام است. همچنین $\xi(t)$ توزیع نرمال نویز سفید می‌باشد. در این مطالعه فرض می‌شود که تمامی نقاط شروع برای انباشتگرها برابر با صفر است $(\dot{X}_i(\cdot) = ۰)$. همچنین، هر زمان که یک انباشتگر به حد آستانه خود a_i برسد، این فرآیند متوقف خواهد شد یعنی، $\dot{X}_i(t) > a_i$. علاوه بر این پارامترها، یک پارامتر دیگر برای زمان غیر تصمیم‌گیری نیز وجود دارد که زمان رمزگردانی محرک (زمان ایجاد بازنمایی از محرک در مغز) و زمان انجام فرمان حرکتی(زمان ثبت پاسخ روی صفحه کلید پس از تصمیم‌گیری) را ثبت می‌کند [۲۶].



شکل (۲): شمای کلی مدل انتشار مسابقه‌ای

یک مدل انتشار مسابقه‌ای به صورت تنها نیز به عنوان یک مدل معیار بر داده‌های رفتاری برازش می‌شود و نتایج آن با مدل‌های ANN-RDM مقایسه می‌شود. در این مدل معیار، آستانه پاسخ و زمان غیر تصمیم‌گیری متناسب با نوع محرک متفاوت است (پارامترهای آستانه پاسخ و زمان غیر تصمیم‌گیری جداگانه برای محرک‌های کلمه و غیرکلمه). پارامتر نرخ رانش

متناسب با نوع محرک (کلمه یا غیر کلمه) و فراوانی کلمه (فراوانی زیاد، فراوانی کم، فراوانی صفر مربوط به غیرکلمه ها) متفاوت است.

۲-۳- اتصال شبکه عصبی به مدل انتشار مسابقه‌ای

به منظور اتصال خروجی های شبکه عصبی به پارامترهای مدل انتشار مسابقه‌ای، اثرات گزارش شده در تکلیف تصمیم‌گیری واژگانی مورد بررسی قرار گرفتند. در بسیاری از مطالعات، ویژگی های آماری برای به دست اوردن بینشی در مورد چگونگی پردازش واژگانی و توسعه مدل‌های محاسباتی پیچیده مورد آزمایش و بررسی قرار گرفته‌اند [۳,۱۷,۳۰]. یکی از مهم‌ترین ویژگی‌های آماری مورد مطالعه، فراوانی کلمات است [۱۷,۱۹]. مطابق با مطالعات صورت گرفته، فراوانی کلمات بر عملکرد شرکت‌کنندگان تأثیر زیادی دارد. شرکت‌کنندگان به واژگانی که فراوانی بیشتری در زبان دارند سریع‌تر و دقیق‌تر پاسخ می‌دهند [۲,۱۷,۳۰,۳۱] . همچنین زمان غیرتصمیم‌گیری نیز در واژگانی که فراوانی بیشتری دارند کمتر است [۱۰]. بنابراین، ما مکانیسم‌هایی مبتنی بر تغییر نرخ رانش و تغییر زمان غیرتصمیم‌گیری براساس فراوانی کلمات را به طور همزمان در مدل‌های مورد مطالعه در نظر می گیریم.

همانگونه که در شکل۱ نشان داده شده است، پس از ساختن یک بردار بازنمایی از یک محرک ورودی با استفاده از FastText یا BERT، یک شبکه عصبی تماما متصل با دو نورون خروجی، بردار بازنمایی را به احتمال کلمه یا غیرکلمه بودن محرک تبدیل می کند. سپس این احتمالات به توابع نگاشت داده می‌شوند و به مقادیر نرخ رانش تبدیل می‌شوند. لازم به ذکر است که مدل رانش مسابقه‌ای دارای دو انباشتگر با نرخ رانش متفاوت است؛ یکی برای کلمه‌ها و دیگری برای غیرکلمه‌ها. بنابراین برای بدست آوردن مقادیر نرخ رانش باید توابع مختلفی بر هر احتمال اعمال شود. در اینجا ما یک تابع نگاشت سیگموئیدی را برای کلمه‌ها و غیرکلمه‌ها با مجانب‌های متفاوت در نظر می گیریم. از نقطه نظر ریاضی، نرخ رانش برای هر کلمه به صورت زیر تعریف شده‌است:

$$v_w = \frac{k_۱ + \beta * F}{۱ + e^{-\alpha*(p_w-۰.۵)}} \qquad (۲)$$

در اینجا $K_۱$ یک متغیر آزاد است که به آن مقیاس رانش می‌گویند، F مقدار Zipf است که معیاری از فراوانی کلمات می‌باشد. به این معیار در بخش مجموعه‌داده ها بیشتر پرداخته می‌شود. افراد مختلف به صورت یکسان در معرض کلمات قرار ندارند و ممکن است با فراوانی متفاوتی از کلمات برخورد داشته باشند. برای همین نیاز است که برای افراد مختلف مقدار Zipf تفاوت داشته باشد. به همین دلیل متغیر آزاد β در فرمول در نظر گرفته شده است که در تعیین مجانب فرمول نرخ رانش نقش دارد. P_w احتمال کلمه بودن محرک ورودی است و α یک متغیر آزاد است که برای محاسبه تفاوت میان پیش‌بینی‌های شبکه عصبی و افراد در نظر گرفته شده است. از سوی دیگر، نرخ رانش برای انباشتگر غیرکلمه به صورت زیر تعریف می‌شود:

$$v_{nw} = \frac{k_۲}{۱ + e^{-\alpha*(p_{nw}-۰.۵)}} \qquad (۳)$$

که در آن p_{nw} احتمال غیرکلمه بودن ورودی و $k_۲$ پارامتر مقیاس است. همچنین α نیز یک متغیر آزاد است که برای محاسبه تفاوت میان پیش‌بینی‌های شبکه عصبی و افراد در نظر گرفته شده است.

مکانسیم دیگر در مدل‌های پیشنهادی، تغییر زمان غیر تصمیم‌گیری بر اساس فراوانی کلمات است. در مقایسه با داشتن تنها یک زمان غیر تصمیم‌گیری برای همه کلمات، اضافه کردن این مکانیسم، عملکرد مدل انباشت شواهد را بهبود می‌بخشد [۱۰]. مکانیسم تغییر زمان غیر تصمیم‌گیری بر اساس فراوانی کلمات را به صورت زیر فرموله می‌شود:

$$T_{er} = m + \gamma * e^{-F} \qquad (۴)$$

که در آن F مقدار Zipf یا همان فراوانی واژه را نشان می‌دهد، m پارامتر آزادی است که زمان انجام فرمان حرکتی را نمایش می‌دهد و γ پارامتر آزاد دیگری است که بیانگر تفاوت افراد در میزان برخورد با کلمات است. در تابع مربوطه، زمان انجام فرمان حرکتی و زمان رمزگردانی محرک قابل تفکیک می‌باشند. به بیان دیگر، $\gamma * e^{-F}$ برای محاسبه زمان مزگردانی محرک می‌باشد و عبارت e^{-F} تضمین می‌کند کلمات با فراوانی بالاتر سرعت رمزگردانی بیشتری خواهند داشت. همچنین فراوانی غیرکلمات، صفر در نظر گرفته می‌شود و در نتیجه زمان مزگردانی برای غیرکلمات در مقایسه با کلمات کندتر است.

۳-۳- مدل FT-RDM

در این مدل، برای بازنمایی کلمات از مدل FastText استفاده می‌شود. FastText یک چارچوب بازنمایی کلمات است که در سال‌های گذشته برای حل تکالیف و مسائل مختلفی استفاده شده است و نتایج بسیار خوبی داشته است [۶,۱۵]. این مدل کلمات را به صورت کیسه‌هایی از کاراکترهای n-gram نشان می‌دهد و بازنمایی‌های سطح کلمه‌ای را از بردارهای زیرکلمه‌ای ایجاد می‌کند. این مدل بر این اساس طراحی شده است که کلمات مشابه در زمینه‌های مشابه ظاهر می شوند. به همین دلیل دو تکلیف به خصوص مبتنی بر زمینه CBOW و Skip-gram در این مدل برای تسهیل یادگیری بازنمایی‌ها، مورد استفاده قرارگرفته‌اند [۶]. از آن جایی که FastText برخلاف بسیاری از مدل‌های دیگر به اطلاعات زیرکلمه‌ای متکی می‌باشد، می‌تواند برای غیرکلمات نیز، بازنمایی برداری تولید کند [۱۶].

در مدل FT-RDM بازنمایی‌های به دست آمده از FastText برای کلمات و غیرکلمات به عنوان ورودی به شبکه عصبی داده می‌شود. مدل FastText استفاده شده در این پروژه دارای ۲ میلیون بردار کلمه با طول ۳۰۰ است که با اطلاعات زیرکلمات مجموعه‌داده Common Crawl آموزش داده شده است (در مجموع، ۶۰۰ میلیارد توکن) [۱۵]. به منظور تبدیل رشته‌های ورودی به نرخ رانش در مدل انتشار مسابقه‌ای، شبکه عصبی بر روی مجموعه‌ای از بردارهای کلمات و غیرکلمات موجود در مجموعه‌داده English Lexcion Project برای دسته‌بندی آن‌ها به دو کلاس مجزا آموزش داده شده است. شبکه بر روی ۸۰٪ از کل داده‌ها که به داده‌های آموزشی اختصاص یافته است با روش اعتبارسنجی متقاطع ، با پنج تقسیم آموزش داده می‌شود. پس از آموزش، میانگین دقت ۹۹.۶٪ به دست می آید. پیش‌بینی‌های مدل آموزش‌دیده برای همه کلمات و غیرکلمات مجموعه‌داده استخراج می‌شوند و توابع نگاشت برای دستیابی به مقدار عددی برای نرخ‌های رانش بر آن‌ها اعمال می‌شوند.

۳-۴- مدل BERT-RDM

در سال‌های اخیر، ما شاهد پیشرفت‌هایی در زبان‌شناسی محاسباتی با نوع جدیدی از شبکه‌های عصبی عمیق مبتنی بر مکانیزم توجه به نام ترنسفورمر بوده‌ایم [۲۹]. برت یک مدل زبانی مبتنی بر زمینه است که یکی از انواع مدل‌هایی است که از معماری ترنسفورمر استفاده می‌کند [۹]. این مدل یک شبکه مبتنی بر ترنسفورمر چندلایه است که هر لایه، ورودی را به وسیله یک شبکه پیش‌رو و یک مکانیسم خودتوجه تغییر می‌دهد. میان انواع گوناگونی از مدل‌های برت منتشر شده، مطالعه پیش رو از نسخه از پیش آموزش داده شده BERT-base uncased که بر روی مجموعه‌های عظیم داده انگلیسی آموزش داده شده است [۹]، استفاده می‌کند. در واقع، در مطالعه حاضر، ما بر مکانیسم یادگیری معنایی مبتنی بر زمینه در مدل برت تکیه می‌کنیم که می‌تواند به ما در یادگیری نحوه ساخت و بازنمایی دانش بر اساس تجربه قبلی کمک کند. برت مدل پیوندگرا را برای تبدیل ورودی به دانش از طریق واحدهای شبه نورون که با اتصالات وزنی به هم می‌پیوندند، به کار می‌گیرد. از سوی دیگر، با استفاده از یک معماری دو سویه، برت می‌تواند در حین آموزش زمینه‌های چپ و راست را با هم ترکیب کند که منجر به بازنمایی معنایی بهتر کلمات می‌شود.

در مدل BERT-RDM، از یک معماری مبتنی بر برت و به دنبال آن مجموعه‌ای از لایه‌های شبکه عصبی تماما متصل به همراه با یک لایه طبقه‌بندی softmax استفاده می‌شود. مدل برت از پیش آموزش داده شده رشته ورودی را رمزگردانی می کند و یک بازنمایی در فضای برداری به دست می دهد که به عنوان ورودی لایه های بعدی عمل می کند. بخش بعدی شبکه با استفاده از بازنمایی‌های تولید شده و دو مقدار عددی را پیش بینی می کند. این دو مقدار، احتمال کلمه و غیرکلمه بودن است که شبکه به هر رشته حروف ورودی اختصاص می‌دهد.

این مدل بر روی داده‌های English Lexcion Project که به دو دسته آموزشی با اندازه ۷۵٪ کل داده‌ها و دسته تست با اندازه ۲۵٪ کل داده تقسیم شده است، آموزش داده شد. ورودی مدل یک رشته حروف است و خروجی یک بردار دو بعدی است که هر عنصر احتمال اختصاص داده شده برای هر دسته را نشان می‌دهد. در طول آموزش، وزن‌های لایه‌های مدل برت ثابت درنظر گرفته شده‌اند و آموزش در این بخش تنها بر روی لایه‌های شبکه عصبی دسته‌بندی صورت می‌گیرد .پس از آموزش دقت ۸۸٪ به دست آمده است. مانند مدل قبل، پیش‌بینی‌های مدل آموزش‌دیده‌شده برای همه کلمات و غیرکلمات در مجموعه داده استخراج می‌شوند و توابع نگاشت برای دستیابی به مقدار عددی برای نرخ‌های رانش به آنها اعمال می‌شوند

۳-۵- مجموعه‌داده ها

برای آزمایش مدل‌های ارائه شده در این مطالعه از دو مجموعه‌داده English Lexicon Project (ELP) و SUBTLEXus استفاده شده است. ELP مجموعه‌داده‌ای شامل معیارهای رفتاری و آماری از کلمات و غیرکلمات انگلیسی می‌باشد. [۳]. برای این مطالعه، مجموعه‌ای شامل ۴۰۴۸۳ کلمه و ۴۰۳۴۱ غیرکلمه به همراه داده‌های آزمایشی افراد در تکلیف تصمیم گیری واژگانی از این مجموعه داده استخراج شده است.

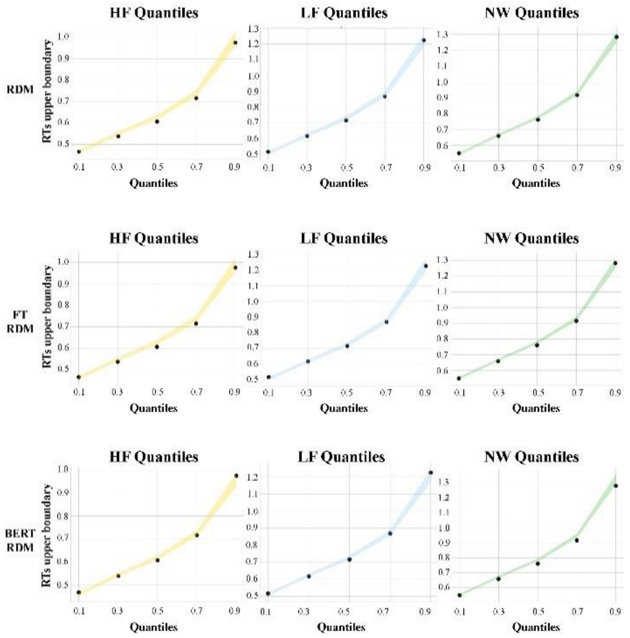
برای بدست آوردن فراوانی هر واژه از مجموعه داده SUBTLEXus استفاده شده است. این مجموعه داده دارای معیارهای دقیق فراوانی کلمات است که از زیرنویس‌های برنامه‌های تلویزیونی و فیلم‌های پخش شده در آمریکا برای ۵۱ میلیون کلمه ایجاد بدست آمده‌اند [۷]. از میان انواع گوناگون معیارهای فراوانی ارائه شده در این مجموعه داده، برای این مطالعه از معیار Zipf استفاده شده است. مقیاس Zipf از کمی کمتر از ۱ تا کمی بیشتر از ۷ متغیر است. با استفاده از این معیار، دسته بندی کلمات بر اساس فراوانی آنها به سادگی امکان پذیر است؛ در این مقیاس کلمات با فراوانی کم مقدار ۳ یا کمتر و کلمات با فراوانی بالا مقدار ۴ یا بیشتر دارند [۲۷]. در این مطالعه مقدار Zipf برای هر یک از کلمات ELP استخراج شده و کلماتی که هیچ مقدار Zipf ای در SUBTLEXus ندارند حذف می شوند. پس از استخراج مقادیر Zipf، کلمات به دو دسته با فراوانی زیاد و کم دسته‌بندی می شوند. دسته‌بندی کلمات به با فراوانی زیاد و کم برای آموزش بخش شبکه عصبی مدل ها کاربردی ندارد بلکه شبکه‌عصبی برای دسته بندی رشته های ورودی به عنوان کلمه یا غیرکلمه، مشابه آنچه شرکت کنندگان در تکلیف تصمیم‌گیری واژگانی انجام می دهند، آموزش دیده است. دسته‌بندی براساس فراوانی برای تحلیل الگوهای مورد انتظار از پیش‌بینی‌های انجام گرفته توسط مدل‌ها پس از آموزش شبکه عصبی و تخمین پارامترهای مربوط به مدل انتشار مسابقه‌ای است.

۴- نتایج

به منظور برازش مدل های پیشنهادی بر روی داده های رفتاری، از رویکرد بیزی سلسله مراتبی استفاده شده است. تخمین پارامترها با استفاده از الگوریتم (Markov Chain Monte Carlo MCMC) برای ۴ زنجیره با ۷۰۰۰ تکرار صورت گرفت. نیمی ازین تکرارها ،یعنی ۳۵۰۰ تا، برای warmup استفاده شده اند. با توجه به اینکه ضریب کاهش مقیاس بالقوه ، $\hat{r}$ ، برای تمامی پارامترها کمتر از ۱.۰۱ بود می‌توان گفت تمامی زنجیرها همگرا شدند [۱۳].

از آنجایی که در علوم شناختی پیش‌بینی‌های کیفی از اهمیت زیادی برخوردار هستند، در شکل ۳ پیش‌بینی پستریور[۴] همه مدل‌ها برای چندک‌های زمان پاسخ نمایش داده شده است. باتوجه به نتایج بدست آمده، تمامی مدل‌ها پیش‌بینی‌های دقیق و نسبتا مشابهی برای چندک‌های زمان پاسخ ارائه می‌دهند.

برای ارزیابی کمی دقت مدل‌ها مقادیر چگالی پیش‌بینی نقطه ای[۵] (lppd)، معیار اطلاعات قابل اعمال گسترده[۶] (WAIC)[۱۴]، خطای استاندارد WAIC یا ($WAIC_{SE}$) و P_{waic} مورد بررسی قرار گرفته اند و نتایج در جدول ۱ گزارش شده‌لند. معیارهای lppd و WAIC هر دو دقت پیش‌بینی مدل را روی داده‌های برازش شده نشان می‌دهند، با این تفاوت که WAIC پیچیدگی مدل را نیز به عنوان یک جنبه منفی در نظر می‌گیرد. مقادیر بالاتر lppd و مقادیر پایین تر برای WAIC با در نظر گرفتن پیچیدگی مدل، برازش بهتر مدل بر داده ها را نشان می دهند. P_{waic} معیاری است که تعداد موثر پارامترهای مدل را بیان می‌کند. مطابق با معیارهای موجود بالاترین مقدار lppd و کمترین WAIC متعلق به مدل FT-RDM است.



شکل (۳): پیش‌بینی‌های پستریور مدل‌ها از چندک‌های زمان واکنش افراد. نقاط نشانگر چندک‌ها و نواحی سایه زده شده نشانگر بازه‌ی ٪۹۵ معتبر بیزی[۷] می‌باشند. کلمات با فراوانی زیاد : HF، کلمات با فراوانی کم: LF، غیرکلمات:NW

BERT-RDM با داشـــتن کمترین مقدار P_{waic} دارای کمترین پیچیدگی اســـت، اما بدترین عملکرد را با توجه به کمترین مقدار lppd و بالاترین WAIC دارد. پیچیدگی بخش ANN مدل‌ها با معیار P_waic نشـــان داده نمی‌شود، و از منظر پیچیدگی شبکه عصبی مدل BERT-RDM پیچیدگی بیشتری دارد. از منظر تخمین پارامتر، مدل رانش مسابقه‌ای تنها، پیچیده ترین مدل برای تخمین می‌باشد.

جدول (۱) : مقایسه مدل‌ها: طبق معیارهای WAIC و lppd مدل FT-RDM با داشتن بالاتری lppd و پایین‌ترین WAIC بهترین دقت را دارد.

$WAIC_{SE}$	WAIC	P_{waic}	lppd	نام مدل
۲۵۳	۶۵۱	۱۹۵	-۱۳۱	FT RDM
۲۴۹	۱۳۰۴	۱۸۴	-۴۶۸	BERT RDM
۲۵۳	۹۱۴	۲۳۵	-۲۲۲	RDM

۵- نتیجه‌گیری

در این مطـالعـه، دو مـدل ANN-EAM متفـاوت برای تکلیف تصـمیم‌گیری واژگانی پیشـنهاد شـده‌اند که قادر به ارائه بازنمایی از کلمات و غیرکلمات ورودی مدل انتشـار مسـابقه‌ای می‌باشـند. اگرچه از منظر پیش‌بینی پسـتریور برای زمان پاسخ، مدل‌ها عملکردی مشـابه داشـتند ولی مدل FT-RDM از نظر WAIC و lppd بهترین عملکرد را داشـــت. همچنین نتایج به‌دســت‌آمده نشــان می‌دهد که FT-RDM عملکرد بهتری نسـبت به مدل BERT-RDM دارد. شـــایان ذکر اســـت که ویژگی‌های مختلف محرک ورودی مـانند ویژگی‌های معنایی، املایی و واجی تـأثیر بســـزایی در فرآیند تصـمیم‌گیری واژگانی دارد. با این حال، مدل‌های پیشـنهادی تنها ویژگی‌های معنایی و فراوانی محرک ارائه شـــده را در نظر گرفتند. به همین دلیل به نظر بتوان با افزودن مکانیسـم‌های اضافی دیگر به مدل‌ها مانند طول رشته حروف و یا همســـایه های املایی عملکرد مدل‌ها را بهبود بخشـید. سـهم دیگر این مطالعه تلاش برای تجزیه زمان غیر تصمیم‌گیری به اجزای تشکیل دهنده آن یعنی زمان انجام فرمان حرکتی و زمان رمزگردانی محرک می‌باشـــد که یک مشکلی سـنتی در زمینه مدل‌های انباشت شواهد است [۲۱]. مکانیسـم تغییر زمان غیر تصـمیم گیری براسـاس فراوانی واژگان می تواند تقریبی برای زمان انجام فرمان جرکتی با تخمین پارامتر m و زمان رمزگردانی محرک با تخمین پارامتر γ ارائه دهد.

در پایان، به عنوان یک نکته کلی در مورد چارچوب ANN-EAM، نقطه ضـعف اصـلی همه ANN-EAM های پیشـنهادی عدم سـازگاری با زیسـت شـناسـی تصـمیم گیری اسـت. بنابراین، یکی از جهت‌گیری‌های آینده برای توسعه مدل‌های ANN-EAM جدید برای تکلیف‌های مختلف تصمیم‌گیری، پیشـنهاد شبکه‌هایی است که با زیست‌شناسی تصمیم‌گیری سازگار باشند. با این وجود چارچوب ANN-EAM می‌تواند نویدبخش حل مشـــکل بازنمایی محرک و دانش در مدل‌های انباشـــت شـــواهد باشـــد و به درک بهتر ما از فرآیندهای زیربنایی تصمیم‌گیری کمک کند.

مراجع

[۱] Agrawal, M.; Peterson, J. C.; and Griffiths, T. L., "*Using* machine learning to guide cognitive modeling*: A case study in moral reasoning,"* arXiv preprint arXiv:۱۹۰۲,۰۶۷۴۴., ۲۰۱۹

[۲] Balota, D. A.; Yap, M. J.; and Cortese, M. J., Chapter ۹ - Visual Word Recognition: The Journey from Features to Meaning (A Travel Update), In Traxler, M. J.; and Gernsbacher, M. A., eds., Handbook of Psycholinguistics (Second Edition), ۲۸۵–۳۷۵. London: Academic Press, second edition. ISBN ۹۷۸-۰-۱۲-۳۶۹۳۷۴-۷, ۲۰۰۶

[۳] Balota, D. A.; Yap, M. J.; Cortese, M. J.; Hutchison, K. A.; Kessler, B.; Loftis, B.; Neely, J. H.; Nelson, D. L.; Simpson, G. B.; and Treiman, R., "*The English lexicon project,"* Behavior Research Methods, ۳۹(۳): ٤٤٥–٤٥۹, ۲۰۰۷

[۴] Bhatia, S., "Associative judgment and vector space semantics," Psychological Review, ۱۲٤(۱):۱-۲۰, ۲۰۱۷

[۵] Bhatia, S.; and Richie, R., "Transformer networks of human conceptual knowledge," Psychological Review, ۲۰۲۱

[۶] Bojanowski, P.; Grave, E.; Joulin, A.; and Mikolov, T., *"Enriching word vectors with subword information,"* Transactions of the association for computational linguistics, ٥:۱۳٥–۱٤٦, ۲۰۱۷

[۷] Brysbaert, M.; and New, B., *"Moving beyond Kuˇcera* and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English," Behavior Research Methods, ٤۱(٤): ۹۷۷–۹۹۰, ۲۰۰۹

[۸] Busemeyer, J. R.; Wang, Z.; Townsend, J. T.; and Eidels, A., *The Oxford handbook of computational and mathematical psychology*, Oxford University Press, ۲۰۱٥

[۹] Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K, *"Bert: Pre-training of deep bidirectional transformers for language Understanding,"* arXiv preprint arXiv:۱۸۱۰,۰٤۸۰٥, ۲۰۱۸

[۱۰] Donkin, C.; Heathcote, A.; Brown, S.; and Andrews, S., *"Non-decision time effects in the lexical decision task,"* Proceedings of the ۳۱st annual conference of the cognitive science society, Austin: Cognitive Science Society, ۲۰۰۹

[۱۱] Fontanesi, L.; Gluth, S.; Spektor, M. S.; and Rieskamp, J., *"A reinforcement learning diffusion decision model for value-based decisions,"* Psychonomic Bulletin and Review, ۲٦(٤): ۱۰۹۹–۱۱۲۱, ۲۰۱۹

[۱۲] Frank, M. J.; Gagne, C.; Nyhus, E.; Masters, S.; Wiecki, T. V.; Cavanagh, J. F.; and Badre, D., *"fMRI and EEG predictors of dynamic decision parameters during human reinforcement learning,"* Journal of Neuroscience, ۳٥(۲): ٤۸٥–٤۹٤, ۲۰۱٥

[۱۳] Gelman, A.; and Rubin, D. B., *"Inference from iterative simulation using multiple sequences,"* Statistical science, ۴۵۷□۴۷۲, ۱۹۹۲

[۱۴] Gelman, A.; Hwang, J.; and Vehtari, A., *"Understanding predictive information criteria for Bayesian models,"* Statistics and computing, ۲٤(٦): ۹۹۷–۱۰۱٦, ۲۰۱٤

[۱۵] Grave, E.; Bojanowski, P.; Gupta, P.; Joulin, A.; and Mikolov, T., *"Learning word vectors for 157 languages"*, arXiv preprint arXiv:۱۸۰۲,۰٦۸۹۳., ۲۰۱۸

[۱۶] Heidari-Gorji, H.; Ebrahimpour, R.; and Zabbah, S., "A temporal hierarchical feedforward model explains both the time and the accuracy of object recognition," Scientific Reports, ۱۱(۱): ۱–۱۲, ۲۰۲۱

[۱۷] Hendrix, P.; and Sun, C. C., "A word or two about nonwords: Frequency, semantic neighborhood density, and orthography-to-semantics consistency effects for nonwords in the lexical decision task," Journal of Experimental Psychology: Learning Memory and Cognition, ٤۷(۱): ۱٥۷–۱۸۳, ۲۰۲۱

[۱۸] Holmes, W. R.; O'Daniels, P.; and Trueblood, J. S., "A Joint Deep Neural Network and Evidence Accumulation Modeling Approach to Human Decision-Making with Naturalistic Images," Computational Brain and Behavior, ۳(۱), ۲۰۲۰

[۱۹] Juhasz, B. J.; Yap, M. J.; Raoul, A.; and Kaye, M., "A further examination of word frequency and age-ofacquisition effects in English lexical decision task performance: The role of frequency trajectory," Journal of Experimental Psychology: Learning, Memory, and Cognition, ٤٥(۱): ۸۲–۹٦, ۲۰۱۹

[۲۰] Kaadoud, I. C.; Bennetot, A.; Mawhin, B.; Charisi, V.; and D´ıaz-Rodr´ıguez, N., "Explaining Aha! moments in artificial agents through IKE-XAI: Implicit Knowledge Extraction for eXplainable AI", Neural Networks, ۲۰۲۲

[۲۱] Lin, Y.; and Tadin, D., *"Estimating decision time in perceptual decision making,"* Journal of Vision, ۲۱(۹): ۲٦۹٤–۲٦۹٤, ۲۰۲۱

[۲۲] Pedersen, M. L.; Frank, M. J.; and Biele, G., *"The drift diffusion model as the choice rule in reinforcement learning,"* Psychonomic bulletin & review, ۲٤(٤): ۱۲۳٤–۱۲٥۱, ۲۰۱۷

[۲۳] Ratcliff, R.; McKoon, G.; and Gomez, P., *"A Diffusion Model Account of the Lexical Decision Task,"* Psychological Review, ۱۱۱(۱): ۱٥۹–۱۸۲, ۲۰۰٤

[۲۴] Ratcliff, R.; Smith, P. L.; Brown, S. D.; and McKoon, G., *"Diffusion decision model: Current issues and history,"* Trends in cognitive sciences, ۲۰(٤): ۲٦۰–۲۸۱, ۲۰۱٦

[۲۵] Sewell, D. K.; Jach, H. K.; Boag, R. J.; and Van Heer, C. A., *"Combining error-driven models of associative learning with evidence accumulation models of decision-making,"* Psychonomic Bulletin & Review, ۲٦(۳): ۸٦۸–۸۹۳, ۲۰۱۹

[۲۶] Tillman, G.; Van Zandt, T.; and Logan, G. D., "Sequential sampling models without random between-trial variability: The racing diffusion model of speeded decision making," Psychonomic Bulletin & Review, ۲۷(٥): ۹۱۱–۹۳٦, ۲۰۲۰

[۲۷] Van Heuven,W. J.; Mandera, P.; Keuleers, E.; and Brysbaert, M., *"SUBTLEX-UK: A new and improved word frequency database for British English,"* Quarterly journal of experimental psychology, ٦۷(٦): ۱۱۷٦–۱۱۹۰, ۲۰۱٤

[۲۸] van Ravenzwaaij, D.; Brown, S. D.; Marley, A.; and Heathcote, A., *"Accumulating advantages: A new conceptualization of rapid multiple choice,"* Psychological review, ۱۲۷(۲): ۱۸٦–۲۱٥, ۲۰۲۰

[۲۹] Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I., *"Attention is all you need,"* Advances in neural information processing systems, ۳۰, ۲۰۱۷

[۳۰] Wagenmakers, E. J.; Ratcliff, R.; Gomez, P.; and McKoon, G., *"A diffusion model account of criterion shifts in the lexical decision task,"* Journal of Memory and Language, ۵۸(۱): ۱۴۰□۱۵۹, ۲۰۰۸

[۳۱] Yap, M. J.; Balota, D. A.; Sibley, D. E.; and Ratcliff, R., *"Individual differences in visual word recognition: Insights from the English Lexicon Project,"* Journal of Experimental Psychology: Human Perception and Performance, ۳۸(۱): ٥۳–۷۹, ۲۰۱۲

[۳۲] Yap, M. J.; Sibley, D. E.; Balota, D. A.; Ratcliff, R.; and Rueckl, J., *"Responding to nonwords in the lexical decision task: Insights from the english Lexicon project,"* Journal of Experimental Psychology: Learning Memory and Cognition, ٤۱(۳): ٥۹۷–٦۱۳, ۲۰۱٥

[۳۳] Yin,W.; Sch¨utze, H.; Xiang, B.; and Zhou, B., *"Abcnn: Attention-based convolutional neural network for modeling sentence pairs"*, Transactions of the Association for Computational Linguistics, ٤: ۲٥۹–۲۷۲, ۲۰۱٦

پانویس ها

Dataset [۱]
Drift Diffusion model [۲]
Race Diffusion model [۳]
Posterior [٤]
log pointwise predictive density [٥]
widely applicable information criterion [٦]
Bayesian credible interval (BCI) [۷]

ساخت و جاسازی گراف ناهمگن برای خوشه‌بندی مؤثر محتوای متنی شبکه اجتماعی توییتر

ملیکا حیدری دستجردی[۱]، وحید معتقد [۲]، فاطمه بهاری‌فرد [۳]

[۱] دانشکده مهندسی برق و کامپیوتر، دانشگاه تهران
melikaheidari@ut.ac.ir

[۲] پژوهشکده علوم کامپیوتر، پژوهشگاه دانش‌های بنیادی (IPM)
vahid.motaghed.2020@gmail.com

[۳] پژوهشکده علوم کامپیوتر، پژوهشگاه دانش‌های بنیادی (IPM)
f.baharifard@ipm.ir

چکیده

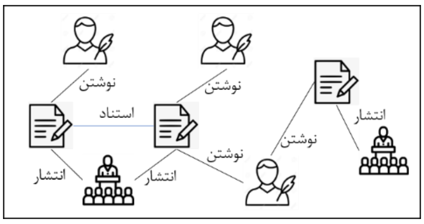
با رشد روزافزون شبکه‌های اجتماعی، فعالیت‌های پژوهشی متنوعی برای درک محتوای منتشرشده در آن‌ها انجام شده است. اکثر تحقیقات به صورت یکسان به انواع داده‌های موجود در این شبکه‌ها نگاه می‌کنند و آن‌ها را با گراف همگن مدل می‌کنند. این درحالی است که این مجموعه داده‌ها شامل بیش از یک نوع موجودیت با ارتباطات متنوعی هستند و بهتر است برای مدل‌کردن آن‌ها از شبکه‌های ناهمگن استفاده کرد. در این مقاله چارچوبی برای ساخت گرافی ناهمگن و چند لایه از داده‌های شبکه اجتماعی توییتر ارائه می‌شود که می‌تواند علاوه بر روابط ساختاری موجود بین انواع موجودیت‌ها، ارتباط محتوایی متن پیام را هم در گراف نمایش دهد. در این روش برای جاسازی رئوس گراف ناهمگن، با در نظر گرفتن فرامسیرهای احتمالی به عنوان الگویی برای حرکت بین انواع رئوس، یک الگوریتم قدم‌زنی تصادفی-احتمالی اجرا می‌شود که می‌تواند از ویژگی رئوس برای افزایش دقت فرایند خوشه‌بندی استفاده کند. چارچوب و الگوریتم ارائه‌شده روی مجموعه‌ای با بیش از صد هزار داده واقعی توییتر که از طریق رابط کاربری توییتر جمع‌آوری شده، مورد ارزیابی قرار گرفته است که نتایج حاصل، نشانگر دقت بیش‌تر روش پیشنهادی نسبت به روش‌های قبلی بدون ناظر کلاسیک و مبتنی بر فرامسیرها برای گروه‌بندی توییت‌های منتشرشده در شبکه اجتماعی توییتر است.

کلمات کلیدی

گراف ناهمگن، یادگیری بدون ناظر، پردازش زبان طبیعی، الگوریتم خوشه‌بندی، شبکه‌های اجتماعی، توییتر

۱- مقدمه

با توجه به نقش بسیار مهم و حیاتی داده‌ها در زندگی امروزه، یکی از حوزه‌های اصلی پژوهش در این زمینه نمایش کارآمد داده‌ها و در نتیجه تحلیل آن‌ها است. برای نمایش اکثر مجموعه داده‌ها می‌توان از ساختار قدرتمند گراف استفاده کرده و موجودیت‌ها و روابط موجود در داده‌ها را به ترتیب با رئوس و یال‌های مابین آن‌ها نشان داد. به عنوان مثال، مجموعه داده‌های مربوط به شبکه‌های اجتماعی [۱]، شبکه‌های ترافیکی [۲]، شبکه‌های استنادی [۳] از نمونه‌هایی هستند که امکان نمایش آن‌ها با گراف‌ها وجود دارد. اگر گراف فقط یک نوع رأس و یک نوع یال داشته باشد، *گراف همگن*[۱] نامیده می‌شود. اما، در بسیاری از مجموعه داده‌های دنیای واقعی، جنبه‌های مختلف اطلاعات با هم ترکیب می‌شوند و داده‌ها تمایل دارند به‌عنوان یک *گراف ناهمگن*[۲] [۴] ارائه شوند که در این نوع گراف‌ها، رئوس و یال‌ها انواع مختلفی دارند. *شبکه‌های اطلاعاتی ناهمگن*[۳] (HIN) به دلیل داشتن اطلاعات جامع و معناشناسی غنی در بسیاری از وظایف داده کاوی استفاده می‌شوند. مثالی از این شبکه‌ها در شکل (۱) آورده شده است که مربوط به مجموعه داده یک شبکه استنادی است که شامل سه نوع رأس (مقاله، نویسنده و محل برگزاری) و سه نوع یال (نوشتن، استناد و انتشار) می‌باشد. بر اساس هدف مطالعه، این شبکه‌ها می‌توانند به صورت جهت‌دار یا بی‌جهت و هم‌چنین وزن‌دار یا بی‌وزن نمایش داده شوند.



شکل (۱) : نمونه شبکه استنادی ناهمگن

وابسته به مساله مورد پژوهش، در مطالعه برخی از مجموعه داده‌ها، علاوه بر روابط ساختاری موجود، می‌توان اطلاعات غنی کمکی مانند برچسب، صفت و ویژگی رئوس را به توپولوژی گراف اضافه کرد تا بتوان به صورت ترکیبی از مجموع اطلاعات ساختاریافته و بدون ساختار در تحلیل و انجام تسک مورد نظر استفاده نمود [۵،۶]. در واقع با تجزیه و تحلیل کارآمد شبکه‌های ناهمگن می‌توان اطلاعات پنهان موجود در آن‌ها را کشف کرده و در کاربردهای بسیاری مانند خوشه‌بندی رئوس[۴]، طبقه‌بندی رئوس[۵]، پیش‌بینی پیوندها[۶] و غیره مورد استفاده قرار داد. یکی از چالش‌های موجود در این حوزه، چگونگی ترکیب این منابع اطلاعاتی ساختاریافته و بدون ساختار است به گونه‌ای که مفهوم شباهت رئوس حفظ شده و در نتیجه در کاربرد مورد نظر تاثیرگذار و مفید باشد.

در اکثر مطالعات قبلی، شبکه‌های ناهمگن صرفاً بر اساس ساختارهای از پیش تعریف شده مورد تحلیل قرار گرفته‌اند و بسیاری از مزایای سایر اطلاعات موجود در مجموعه داده نادیده گرفته شده است [۹-۷]. هم‌چنین برخی از مطالعات پیشین که از این اطلاعات در فرایند جاسازی[۷] شبکه استفاده کرده‌اند، دانش معنایی و ارتباطات نهفته‌ای که مابین رئوس وجود دارد را مورد بررسی قرار نداده‌اند [۵،۶،۱۰،۱۱].

طی سالیان اخیر استفاده از شبکه‌های اجتماعی در بین عموم جامعه به شدت مورد استقبال قرار گرفته است. در این شبکه‌ها حجم عظیمی از داده‌ها با سرعت

و تنوع زیاد (به عنوان مثال تصاویر، متن و فیلم) تولید و منتشر می‌شود. کاربرد این شبکه‌ها از برقراری ارتباط بین افراد تا کسب اطلاعات و اخبار در حوزه‌های وسیعی از مسائل سیاسی، اجتماعی، فرهنگی و غیره گسترده شده است. با این حال بسیاری از داده‌های تولیدشده در این بسترها، ساختارمند نیستند. پس نمی‌توان به سادگی در مورد چگونگی انتشار موضوعات مختلف و هم‌چنین تعامل کاربران با یکدیگر اطلاعاتی به دست آورد.

در این مقاله برای غلبه بر محدودیت‌های ذکر شده، از ساختار گراف ناهمگن برای نمایش داده‌های موجود در *شبکه اجتماعی توییتر* که یکی از بسترهای تعاملی پرکاربرد است استفاده و یک چارچوب جدید برای خوشه‌بندی مجموعه داده‌های توییتر ارائه خواهیم کرد. این چارچوب با استفاده از دانش ساختاری و معنایی، سازگاری کلی بین انواع مختلف اشیاء در مجموعه داده‌ها را تضمین می‌کند. برای این منظور در این چارچوب، ابتدا یک شبکه ناهمگن سه لایه (لایه‌های هشتگ، توییت و کاربر) ساخته شده و سپس از روش‌های *پردازش زبان طبیعی*[8] (NLP) برای کشف ارتباطات معنایی ویژگی‌های متنی داده‌ها بهره خواهیم گرفت. هم‌چنین از داده‌های اضافی موجود در مجموعه داده مانند تعداد لایک، تعداد ریتوییت و غیره برای تعیین ویژگی[9] رئوس گراف ناهمگن سه لایه استفاده خواهیم کرد. در نهایت با استفاده از یک مکانیزم جدید احتمالاتی برای تعبیه گراف ناهمگن، گراف تولیدشده را جاسازی کرده و عملیات خوشه‌بندی را روی آن اعمال می‌کنیم تا بتوانیم موضوعات مطرح‌شده در مجموعه داده را به صورت بدون ناظر[10] دسته‌بندی کنیم.

به طور خلاصه، ایده‌ها و موارد زیر در این مقاله مطرح شده است:

- چارچوبی برای خوشه‌بندی توییت‌های مشابه در یک مجموعه داده با روشی بدون ناظر معرفی می‌شود که در آن یک گراف ناهمگن سه لایه، بر اساس ترکیب اطلاعات ساختاری، ویژگی‌ها و هم‌چنین دانش معنایی ساخته خواهد شد.
- از داده‌های اضافی موجود در مجموعه داده مانند تعداد لایک، تعداد ریتوییت، تعداد نقل‌قول و غیره به عنوان ویژگی‌های رئوس استفاده می‌شود.
- با استفاده از روش جدید فرامسیر[11] احتمالی، الگوهای احتمالاتی برای شناسایی مسیرهای متفاوت بین رئوس گراف را در نظر گرفته و با استفاده از مکانیزم skip-gram رئوس نزدیک به هم از نظر فاصله‌ای و معنایی مدل و جاسازی می‌شوند.
- برای ارزیابی دقت، خوشه‌بندی روی مجموعه داده‌های واقعی که از شبکه اجتماعی توییتر جمع‌آوری شده است، اعمال می‌شود. آزمایش‌ها روی این مجموعه داده‌ها نشان می‌دهد که چارچوب ارائه شده برای این مجموعه داده‌ها، نسبت به روش‌های بدون ناظر قبلی کلاسیک و مبتنی بر فرامسیرها بهتر عمل می‌کند.

۲- پژوهش‌های پیشین

خوشه‌بندی شامل تقسیم تعدادی نمونه داده بدون برچسب به مجموعه‌ای از دسته‌ها است به طوری که اعضای هر دسته نسبت به سایر دسته‌ها به یکدیگر شباهت بیش‌تری داشته باشند [۱۲]. تحقیقات بسیاری برای خوشه‌بندی داده‌های متنی ارائه شده است [۱۳]. برای خوشه‌بندی داده‌های متنی، روش‌های سنتی مانند نمایش متن وزن‌دار [۱۴] و مدل‌های از پیش آموزش‌دیده جاسازی کلمه [۱۵] به‌طور گسترده مورد استفاده قرار گرفته‌اند. اخیراً نیز بازنمایی‌های رمزگذار دوطرفه از مدل تعبیه‌سازی متن ترانسفورماتورها (BERT) [۱۶]، توجه زیادی را به خود جلب کرده است. این روش نمایش‌هایی از یک کلمه را به دست می‌آورد که مطابقت بهتری با معنای اصلی کلمه در جملات مختلف دارد که این نمایش موثر می‌تواند عملکرد خوشه‌بندی متن را بهبود بخشد.

علاوه بر خوشه‌بندی متن، رویکرد خوشه‌بندی را می‌توان برای انواع داده‌ها و هم‌چنین ویژگی‌های مختلف یک مجموعه داده اعمال کرد. بسیاری از روش‌های کلاسیک که در چند دهه اخیر مورد مطالعه قرار گرفته‌اند، از جمله k-means [۱۷] و خوشه‌بندی طیفی [۱۸]، تنها به ویژگی‌های نقاط داده وابسته هستند و نمی‌توانند از روابط بین نقاط استفاده کنند. برخی روش‌های دیگر، با استفاده از روابط بین داده‌ها، ابتدا عملیات جاسازی گراف را انجام می‌دهند و سپس خوشه‌بندی را اعمال می‌کنند. جاسازی گراف چارچوبی برای نمایش یک گراف در فضای با بعد پایین است. برخی از روش‌های موفق جاسازی مانند node2vec [۱۹] و DeepWalk [۲۰] در حوزه پردازش زبان طبیعی نیز مورد توجه قرار گرفته‌اند. در روش DeepWalk از قدم‌زنی تصادفی کوتاه و skip-gram [۲۱] برای جاسازی رئوس استفاده می‌شود.

هم‌چنین برای خوشه‌بندی گراف‌ها می‌توان از الگوریتم‌های شناسایی تشکل[12] نیز استفاده نمود که از روش‌های محبوب به شمار رفته و از روابط بین موجودیت‌ها در گراف استفاده می‌کنند. این روش‌ها معمولاً گراف اصلی را به دنباله‌ای از زیرگراف‌ها (با همپوشانی یا غیرهمپوشانی) تقسیم می‌کنند و از معیارهایی مانند پیمانه‌ای بودن [۲۲]، الگوهای موتیف [۲۳] و ساختار محلی اگو-نت [۲۴] در این راستا بهره می‌برند. روش‌های مبتنی بر یادگیری عمیق نیز برای حل مساله خوشه‌بندی گراف پیشنهاد شده‌اند که از میان آن‌ها می‌توان به CDDTA [۲۵] و CommunityGAN [۲۶] اشاره نمود که به‌ترتیب مبتنی بر خودرمزگذار عمیق و شبکه‌های مولد متخاصم هستند.

تمام روش‌های ذکرشده برای گراف‌های همگن طراحی شده‌اند، اما برای شبکه‌های اطلاعاتی ناهمگن (HIN) خیلی کارآمد نیستند. بنابراین، مطالعه HINها در سال‌های اخیر به دلیل توانایی آن‌ها در مدل‌سازی مجموعه داده‌ها با دو نوع اطلاعات، یعنی ویژگی‌های رئوس و ساختار شبکه، توجه فزاینده‌ای را به خود جلب کرده است [۲۷،۲۸]. یکی از روش‌های اصلی برای خوشه‌بندی این نوع گراف‌ها، RankClus [۲۹] است که در آن یک خوشه از شبکه با استفاده از یال‌های میان چند نوع رأس ساخته می‌شود و در یک روش تکراری، یال‌ها برای دستیابی به خوشه‌بندی مؤثرتر بر اساس رتبه‌بندی بهبود می‌یابند. این روش فقط ویژگی‌های ساختاری را در نظر می‌گیرد و ویژگی‌های رئوس را مورد بررسی قرار نمی‌دهد. بسیاری از روش‌های ارائه‌شده برای گراف‌های ناهمگن مبتنی بر *یادگیری نمایش گراف*[13] هستند که به چند نمونه از آن‌ها اشاره می‌کنیم. فرامسیرها یک مفهوم تعریف شده در اکثر روش‌های جاسازی گراف ناهمگن است که در روش‌هایی مانند metapath2vec [۷] و HIN2vec [۸] از این مفهوم استفاده شده است. *فرامسیر* الگویی است که از دنباله‌ای از انواع رئوس تشکیل شده است که رابطه ترکیبی بین انواع مختلف رئوس در شبکه ناهمگن با این الگو نشان داده می‌شود. به طور مثال اگر در شبکه استنادی شکل (۱) نوع رئوس مقالات و نویسندگان مقالات را با P و A نمایش دهیم دنباله (APA) یک فرامسیر خواهد بود. روش metapath2vec به دنبال بیشینه‌کردن احتمال حفظ ساختار و معنا در یک گراف ناهمگن است. در این روش ابتدا به کمک فرامسیرها و قدم زنی تصادفی روی گراف،

نمونه‌هایی از همسایگی‌های رئوس ساخته شده و سپس از skip-gram برای جاسازی رئوس استفاده می‌شود.

روش‌های فوق قابلیت استفاده از ویژگی‌های رئوس را ندارند، اما اخیرا روش‌های نیمه–نظارتی[۱۴] مانند MAGNN [۱۰] و MeGNN [۳۰] که مبتنی بر یادگیری عمیق گرافی هستند برای استفاده از ویژگی رئوس پیشنهاد شده‌اند. در MAGNN ویژگی‌های رئوس از اطلاعات معنایی تولید شده و نتایج شناسایی تشکل با در نظر گرفتن هم‌زمان این اطلاعات و توپولوژی شبکه بهبود می‌یابد. اخیراً، یک روش جاسازی برای شبکه‌های ناهمگن پیچیده نیز پیشنهاد شده است [۱۱]، که با در نظر گرفتن مفهوم گرافلت برای شبکه‌های ناهمگن، خوشه‌بندی طیفی مرتبه بالاتر را تعمیم می‌دهد. با این حال، بسیاری از روش‌های مورد استفاده برای مطالعه خوشه‌بندی در گراف‌های ناهمگن، نیمه–نظارت‌شده بوده یا فقط شباهت ساختاری را ارزیابی می‌کنند. به همین دلیل، هدف این مقاله ارائه یک روش بدون ناظر برای خوشه‌بندی HINها است که در آن اطلاعات توپولوژیکی، اطلاعات متنی و ویژگی‌های موجود در مجموعه داده هم‌زمان درنظر گرفته شود تا شباهت‌یابی دقیق‌تری انجام گیرد.

۳- الگوریتم پیشنهادی

در این بخش چارچوب اصلی الگوریتم پیشنهادی برای دسته‌بندی داده‌های متنی توییتر ارائه می‌شود. می‌دانیم هر توییت توسط یک کاربر در یک زمان خاص ارسال شده است که می‌تواند توسط سایر کاربران ریتوییت هم شود. مجموعه $T = [t_1,\ldots,t_n]$ که شامل متن توییت‌های n پیام منتشرشده/ریتوییت‌شده در شبکه اجتماعی توییتر است را در نظر بگیرید. $U = [u_1,\ldots,u_f]$ مجموعه‌ای از کاربرانی است که یا پیامی را منتشر/ریتوییت کرده‌اند و یا در پیامی مورد خطاب قرار گرفته‌اند. $H = [h_1,\ldots,h_g]$ مجموعه‌ای از هشتگ‌هایی است که در این پیام‌ها استفاده شده‌اند. هدف گروه‌بندی مجموعه T به مجموعه‌های $C = [C_1,\ldots,C_m]$ است. الگوریتم پیشنهادی برای این هدف را می‌توان به چهار بخش اصلی تقسیم کرد:

۱) پردازش و جاسازی متن پیام‌ها: برخی از عملیات پیش‌پردازش روی محتوای توییت‌ها اعمال شده و سپس متون پردازش‌شده به صورت بردارهایی جاسازی و تعبیه می‌شوند.

۲) ساخت گراف ناهمگن: یک گراف ناهمگن سه لایه‌ای با استفاده از رابطه بین توییت‌ها، هشتگ‌ها و کاربران ساخته می‌شود. در این گراف بر اساس فراداده‌های[۱۵] هر رأس، ویژگی آن رأس مشخص می‌شود.

۳) الگوریتم فرامسیر احتمالی: مبتنی بر فرامسیرهای تعیین‌شده و با مکانیزم قدم‌زنی تصادفی–احتمالی، مجموعه‌ای از نمونه‌های فرامسیرهای احتمالی ساخته می‌شود.

۴) جاسازی گراف ناهمگن و خوشه‌بندی: عملیات جاسازی گراف ناهمگن با اجرای روش skip-gram روی مجموعه نمونه‌های فرامسیرهای احتمالی انجام شده و توسط یک الگوریتم خوشه‌بندی، توییت‌ها به صورت موضوعی گروه‌بندی می‌شوند.

- **پردازش و جاسازی متن پیام‌ها**

مجموعه متن $T = [t_1,\ldots,t_n]$ از پست‌های توییت را به عنوان مجموعه داده ورودی چارچوب در نظر می‌گیریم. در این مقاله فقط متون انگلیسی بررسی شده و توییت‌های سایر زبان‌ها از مجموعه داده‌های ورودی حذف شده‌اند و برای این منظور از سرویس گوگل برای شناسایی زبان متون استفاده شده است. پیام‌هایی که در شبکه اجتماعی توییتر ارسال می‌شوند معمولا کوتاه و حاوی کلمات عامیانه و غلط املایی هستند. بنابراین، ابتدا باید همه متون پیش‌پردازش شوند. در این مرحله، هر توییت کلمه‌بندی شده و عملیات ریشه‌یابی روی آن اعمال می‌شود و سپس علائم نقطه‌گذاری، کلمات توقف و لینک‌ها (url) از آن حذف می‌شوند. بنابراین برای توییت t_k، بردار کلمات $\vec{t}_k$ به صورت زیر به دست می‌آید

$$\vec{t}_k = [T_1^k,\ldots,T_{m_k}^k], \quad k \in \{1,\ldots,n\},$$

که در آن m_k تعداد کلماتی است که از پیش‌پردازش توییت t_k حاصل شده است. پس از پیش‌پردازش، برای محاسبه شباهت مفهومی بین متون مختلف، یک بردار ویژگی متمایز برای هر متن ایجاد می‌شود. برای این منظور، از BERT [۱۶] استفاده و هر کلمه از مجموعه داده به یک فضای ۷۶۸ بعدی نگاشت می‌شود. بنابراین، برای i–امین کلمه از بردار $\vec{t}_k$، بردار زیر محاسبه می‌شود که در آن مقدار d، ۷۶۸ است.

$$BERT(T_i^k) = [x_{i_1}^k,\ldots,x_{i_d}^k], \quad i \in \{1,\ldots,m_k\}$$

در نهایت، برای توییت t_k، میانگین بردارهای d بعدی کلماتی که متعلق به بردار $\vec{t}_k$ هستند، به عنوان بردار ویژگی آن توییت در نظر گرفته می‌شود که آن را $\overrightarrow{E(t_k)}$ می‌نامیم.

- **ساخت گراف ناهمگن**

گراف ناهمگن پیشنهادی شامل سه لایه از رئوس توییت، رئوس هشتگ و رئوس کاربر است. در واقع برای هر یک از اعضای مجموعه های T، H و U، یک رأس در گراف در نظر می‌گیریم و بنابراین به ترتیب مجموعه‌هایی از رئوس V_t، V_h و V_u خواهیم داشت. در ادامه، یک تعریف رسمی برای گراف ناهمگن برای شبکه اجتماعی توییتر ارائه می‌دهیم.

تعریف ۱. (گراف ناهمگن توییتر) یک گراف ناهمگن شبکه اجتماعی توییتر یک گراف بدون جهت $G = (V,E,F)$ است که در آن مجموعه رئوس $V = \{V_t \cup V_h \cup V_u\}$ است. هر یال متعلق به یکی از انواع یال‌های $E \subseteq V_i \times V_j$ است که در آن $i,j \in \{t,h,u\}$ و حضور یال‌ها توسط *قوانین اتصال* تعریف می‌شوند. هم‌چنین بر اساس فراداده‌های موجود (MD)، برای هر رأس یک تابع ویژگی $F: MD \to \mathbb{R}$ مطابق با *قوانین ویژگی* نسبت داده می‌شود. قوانین اتصال و قوانین ویژگی به شرح زیر است:

- قوانین اتصال: در قوانین زیر، فرض می‌کنیم $v_t \in V_t$، $v_h \in V_h$ و $v_u \in V_u$.

- ✓ قانون ۱: رابطه توییت–هشتگ (v_h،v_t). هنگامی که هشتگ h در متن توییت t استفاده شده باشد، یک یال بین رأس هشتگ v_h و رأس توییت v_t ایجاد می‌گردد.
- ✓ قانون ۲: رابطه توییت–کاربر (v_u،v_t). یک یال اتصال بین رأس کاربر v_u و رأس توییت v_t ساخته می‌شود، اگر کاربر u توییت t را پست یا ریتوییت کرده باشد یا در آن توییت مورد خطاب قرار گرفته باشد.
- ✓ قانون ۳: رابطه توییت–توییت (v_{t_i}،v_{t_j}). اگر فرض کنیم شباهت دو توییت t_i و t_j را با فرمول $w_{ij} = \left\|\overrightarrow{E(t_i)} - \overrightarrow{E(t_j)}\right\|$ محاسبه کنیم، بین دو رأس توییت v_{t_i} و v_{t_j} یال برقرار می‌شود اگر $w_{ij} \le \alpha$

که $\alpha = 1/4 \sum_{\{i,j\}=1}^{n} w_{ij}$ و $\overrightarrow{E(t_i)}$ بردار تعبیه‌شده توییت t_i در فضای d بعدی است.

- ✓ قانون ۴: رابطه هشتگ–هشتگ (v_{h_i}, v_{h_j}). اگر هشتگ h_i و h_j حداقل در یک توییت به صورت مشترک استفاده شده باشند، بین آن‌ها یال برقرار می‌شود.
- ✓ قانون ۵: رابطه کاربر–کاربر (v_{u_i}, v_{u_j}). اگر کاربر u_i و u_j حداقل یک پست مشترک (توییتی را پست یا ریتوییت کرده باشند و یا در توییتی مورد خطاب قرار گرفته باشند) داشته باشند، بین آن‌ها یال برقرار می‌شود.

- ➢ <u>قوانین ویژگی</u>: در قوانین زیر با نرمال‌سازی هر مقدار همه ویژگی‌ها در بازه صفر تا یک قرار خواهند گرفت.
- ✓ قانون ۱: ویژگی رئوس هشتگ (v_h). تعداد دفعات حضور هشتگ h در مجموعه داده به عنوان ویژگی رأس v_h در نظر گرفته می‌شود.
- ✓ قانون ۲: ویژگی رئوس کاربر (v_u). مجموع تعداد فالوور و فالوینگ کاربر u به عنوان ویژگی رأس v_u در نظر گرفته می‌شود.
- ✓ قانون ۳: ویژگی رئوس توییت (v_t). مجموع تعداد لایک، تعداد ریتوییت و تعداد نقل‌قول‌های پیام t به عنوان ویژگی رأس v_t در نظر گرفته می‌شود.

- **الگوریتم فرامسیر احتمالی**

پس از ساخت گراف ناهمگن ابتدا الگوهایی از ارتباط بین رئوس به عنوان فرامسیر در نظر گرفته شده و مبتنی بر آن‌ها و با استفاده از قدم‌زنی تصادفی–احتمالی، نمونه‌هایی از این الگوها به دست می‌آید. این نمونه‌ها به عنوان ورودی skip-gram که در بخش بعد توضیح داده می‌شود استفاده خواهد شد.

یک فرامسیر عبارت است از مسیری که به صورت

$V_1 \xrightarrow{R_1} V_2 \xrightarrow{R_2} \dots V_s \xrightarrow{R_s} V_{s+1} \dots \xrightarrow{R_{l-1}} V_l$

تعریف می‌شود که در آن $R = R_1 \circ R_2 \circ \dots \circ R_{l-1}$ رابطه ترکیبی بین رأس‌های از نوع V_1 و V_l است. به عنوان مثال در شکل (۱) فرامسیر APA نشان‌دهنده نویسنده مشترک در یک مقاله و فرامسیر APVPA بیانگر آن است که دو نویسنده مقالات خود را در یک کنفرانس به چاپ رسانده‌اند. در مجموعه داده توییتر نیز می‌توان فرضا الگوهای HTH که نشان‌دهنده هشتگ مشترک در یک توییت و UTHTU که نشانگر استفاده دو کاربر از یک هشتگ یکسان است را در نظر گرفت. فرامسیرها معمولا به صورت متقارن تعریف می‌شوند.

در این قسمت یک روش احتمالی–تصادفی برای نشان دادن چگونگی استفاده از فرامسیرها برای قدم‌زنی تصادفی روی گراف ناهمگن معرفی می‌کنیم. فرض کنید گراف $G = (V, E, F)$ و الگوی فرامسیر $\mathcal{P}: V_1 \xrightarrow{R_1} V_2 \xrightarrow{R_2} \dots V_s \xrightarrow{R_s} V_{s+1} \dots \xrightarrow{R_{l-1}} V_l$ را در اختیار داریم. احتمال انتقال در مرحله i–ام از قدم‌زنی تصادفی را به صورت زیر تعریف می‌کنیم:

$$p(v^{i+1} | v_s^i, \mathcal{P}) = \begin{cases} \dfrac{1-\beta}{|N_{s+1}(v_s^i)| - 1}, & (v^{i+1}, v_s^i) \in E, \delta(v^{i+1}) = s+1 \\ \beta, & (v^{i+1}, v_s^i) \in E, \delta(v^{i+1}) = s+1, \\ & F(v^{i+1}) = \max_{j \in s} F(v_j) \\ 0, & O.w \end{cases}$$

در این‌جا $v_s^i \in V_s$ و $N_{s+1}(v_s^i)$ نشان‌دهنده همسایه‌های v_s^i است که از نوع V_{s+1} می‌باشند. هم‌چنین تابع δ نوع هر رأس را مشخص می‌کند. در این فرمول پارامتر احتمالی $0 < \beta \leq 1$ تعریف شده است که بر اساس این پارامتر در قدم‌زنی تصادفی به احتمال β به همسایه‌ای می‌رویم که ویژگی آن بیشینه باشد و بقیه همسایه‌ها با احتمال $(1 - \beta)$ به صورت مساوی انتخاب خواهند شد. بنابراین تأثیر ویژگی‌های رئوس در الگوریتم پیشنهادی در نظر گرفته می‌شود.

- **جاسازی گراف ناهمگن و خوشه‌بندی**

در این مرحله با استفاده از skip-gram [۲۱] و حداکثرسازی احتمال داشتن زمینه ناهمگن $N_s(v)$ بر اساس فرمول زیر، یک جاسازی موثر برای رئوس گراف ناهمگن به دست می‌آوریم:

$$\arg\max_{\theta} \sum_{v \in V} \sum_{s \in T_v} \sum_{c_s \in N_s(v)} \log p(c_s | v; \theta).$$

در این فرمول، T_v مجموعه‌ای از نوع رئوس و $N_s(v)$ همسایه‌های رأس v که از نوع s–امین نوع رئوس هستند را نشان می‌دهند. $p(c_s|v;\theta)$ که یک تابع Soft-max [۳۱] به شمار می‌رود برابر است با $p(c_s|v;\theta) = \frac{e^{X_{c_s} \cdot X_v}}{\sum_{u \in V} e^{X_u \cdot X_v}}$ که در آن X_v، v–امین سطر از ماتریس X است که این سطر جاسازی رأس v را نشان می‌دهد.

برای آموزش skip-gram از نمونه‌های تصادفی فرامسیرهای به دست آمده از قسمت قبل استفاده می‌شود. در واقع از طریق مسیرهای حاصل‌شده، ساختار یک شبکه ناهمگن را به صورت کارا به skip-gram تبدیل می‌کنیم. حال با اعمال روش k-means روی بردارهای تعبیه‌شده مربوط به رئوس توییت، گروه‌بندی موضوعی (خوشه‌بندی) روی آن‌ها انجام می‌شود.

۴- پیاده‌سازی

توییتر یک سرویس شبکه اجتماعی است که در آن کاربران پیام‌کوتاه خود را توییت می‌کنند یا با مکانیزم ریتوییت پیام‌های دیگر کاربران را باز نشر می‌کنند. در این مقاله برای ارزیابی چارچوب ارائه‌شده، به جمع‌آوری متن توییت‌ها با استفاده از API توییتر پرداختیم. برای این منظور بازه زمانی ۱ آوریل تا ۲ آوریل ۲۰۲۲ را در نظر گرفتیم و پست‌های توییت و ریتوییت را که حاوی کلمه کلیدی «بایدن» بودند را جمع‌آوری کردیم. در این مجموعه به دلیل گستردگی موضوع، با حذف محتواهایی با زبان غیر از انگلیسی تلاش شده است اندکی از پراکندگی موضوعات کاسته شود. در این مجموعه تعداد صد و بیست هزار توییت و ریتوییت انتخاب شد. اطلاعات این مجموعه داده را می‌توان به یک گراف ناهمگن سه لایه‌ای تبدیل کرد که آمار آن در جدول (۱) ارائه شده است.

جدول (۱) : اطلاعات مجموعه داده توییتر

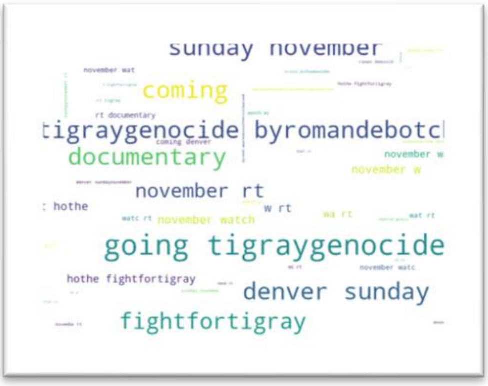
نوع و تعداد رئوس	# توییت (T) : ۱۲۰۰۰۰ # هشتگ (H) : ۱۴۰۷ # کاربر (U) : ۳۹۴۰۴
نوع یال‌ها	یال‌های میان لایه‌ای: T–H، T–U یال‌های درون لایه‌ای: T–T، U–U، H–H

از آن‌جایی که موضوعات واقعی برای این مجموعه داده در دسترس نیست، ما از مجموعه‌ای از کلمات کلیدی با تعداد استفاده بالا به عنوان برچسب‌های هر موضوع استفاده کرده‌ایم. این کلمات کلیدی که توسط نیروی انسانی استخراج

شده‌اند، در جدول (۲) نشان داده شده است. به عنوان نمونه ابرکلمات دسته چهارم این مجموعه داده را می‌توان در شکل (۲) مشاهده نمود.

جدول (۲) : عناوین و کلمات کلیدی مرتبط با هر موضوع

شماره	کلمات نماینده
۱	leaks, dhs, disinformation
۲	oil, gas, price
۳	vote, trump, election
۴	tigray, tigraygenocide, fightfortigray
۵	halloween, bidenhalloween, candy



شکل (۲) : ابرکلمات دسته چهارم مجموعه داده توییتر

۵- ارزیابی و مقایسه

برای ارزیابی دقت روش ارائه‌شده از دو معیار استاندارد *اطلاعات متقابل نرمال‌شده*[۱۶] (NMI) و *شاخص تصادفی تنظیم‌شده*[۱۷] (ARI) استفاده می‌کنیم. هم‌چنین الگوریتم‌های زیر را نیز روی مجموعه داده ورودی اجرا کرده و نتایج به دست آمده را با نتیجه حاصل از الگوریتم پیشنهادی مقایسه می‌کنیم.

- k-means [۱۷]: یک روش سنتی کمی‌سازی برداری است که هدف آن تقسیم داده‌ها به خوشه‌ها بر اساس نزدیک‌ترین میانگین برای هر داده است. برای اجرای این روش، متن توییت‌ها با روش BERT جاسازی شده و این بردارها به عنوان ورودی به الگوریتم k-means داده شده است.
- node2vec [۱۹]: روشی برای جاسازی گراف همگن که به دنبال حفظ همسایگی‌های رئوس موجود در گراف، نمایش کم بعدی را برای رئوس گراف به دست می‌آورد. با نادیده گرفتن ناهمگنی ساختار گراف و با در نظر گرفتن ارتباط معنایی موجود بین متن توییت ها (BERT) برای اتصال یال بین توییت‌ها روی گراف توییتر این روش اعمال شده است.
- metapath2vec [۷]: این روش، روشی برای جاسازی گراف ناهمگن با استفاده از مفهوم فرامسیر و skip-gram است. مانند روش node2vec این الگوریتم با بهره‌گیری از روش BERT برای اتصال یال بین توییت‌ها و در نظر گرفتن ارتباطات ساختاری برای سه نوع رأس، اجرا شده است.

در الگوریتم پیشنهادی فرامسیرهای UTU، HTUTH و HTH به عنوان الگوهایی برای ارتباط بین رئوس گراف ناهمگن توییتر در نظر گرفته شده است. هم‌چنین در محاسبه احتمال انتخاب همسایه‌ها در قدم‌زنی تصادفی-احتمالی $\beta = 0.2$ نسبت داده شده است. در الگوریتم k-means مقدار k را متناسب با تعداد موضوعات یعنی پنج قرار داده و برای کم‌کردن تاثیر انتخاب مراکز در این الگوریتم، نتایج را بر اساس ۵۰ بار تکرار الگوریتم k-means و انتخاب بهترین نتیجه گزارش داده‌ایم.

در جدول (۳) مقادیر NMI و ARI به دست آمده برای الگوریتم پیشنهادی و سه الگوریتم پایه‌ای ذکرشده، به ازای مجموعه داده مربوط به متن توییت‌های ارسال‌شده با کلمات کلیدی «بایدن» گزارش شده است.

جدول (۳) : مقایسه نتایج خوشه‌بندی (٪) در مجموعه داده توییتر

روش	NMI	ARI
k-means(+BERT)	۰.۸۶	۰.۴۳
node2vec(+BERT)	۰.۶۰	۰.۳۰
Metapath2vec(+BERT)	۰.۸۴	۰.۶۵
روش پیشنهادی	**۰.۹۲**	**۰.۷۸**

مطابق جدول (۳)، الگوریتم ارائه‌شده، نسبت به دیگر الگوریتم‌ها از دقت بیش‌تری برخوردار است. دلیل این عملکرد را می‌توان بهره‌گیری مناسب از کلیه اطلاعات موجود در مجموعه داده و هم‌چنین در نظر گرفتن فراداده‌ها به عنوان ویژگی رئوس دانست. دقت الگوریتم k-means با وجود فقدان استفاده از ساختار گراف، نسبت به الگوریتم node2vec برتری دارد و علت آن عدم توانایی کافی الگوریتم node2vec برای حل این مساله پیچیده است. الگوریتم ارائه‌شده در مقایسه با الگوریتم پایه‌ای تحلیل ناهمگن گراف (در این حالت الگوریتم پایه metapath2vec مستقیما روی ساختار گراف اعمال شده است) از برتری چشم‌گیری برخوردار است که دلیل آن وجود ویژگی برای هر رأس و ارائه مکانیزم قدم‌زنی تصادفی-احتمالی می‌باشد. مکانیزم قدم‌زنی تصادفی-احتمالی تلاش می‌کند کاربران، توییت‌ها و هشتگ‌ها را مبتنی بر محتوای مشترک و هم‌چنین میزان تأثیر آن‌ها در جامعه، گروه‌بندی کند.

۶- نتیجه‌گیری

علیرغم اینکه اخیرا خوشه‌بندی در شبکه‌های اطلاعاتی ناهمگن مورد مطالعه قرار گرفته است، یافتن یک روش بدون ناظر برای این منظور کم‌تر مورد بررسی قرار گرفته است. در این مقاله یک روش خوشه‌بندی مبتنی بر گراف ناهمگن برای دسته‌بندی موضوعی پیام‌های موجود در شبکه اجتماعی توییتر ارائه شده است که هم‌زمان از اطلاعات ساختاری، اطلاعات متنی و ویژگی‌های موجود در فراداده‌ها استفاده می‌کند تا شباهت‌یابی دقیق‌تری انجام دهد. در الگوریتم ارائه‌شده فرآیند تعیین ویژگی رأس‌ها به صورت خبرگانی بوده و تلاش می‌شود در آینده دقت این مساله با رویکرد یادگیری عمیق گرافی بهبود یابد و علاوه بر افزودن قابلیت استفاده از سایر ویژگی‌های رئوس (مانند زمان ارسال، تعداد پاسخ و ...) و افزودن لایه مدیا، برای یال‌ها نیز وزن تعیین کنیم تا بتوان در هر لایه، میزان شباهت هر دو نود را به صورت صریح در نظر گرفت.

مراجع

[1] R. Churchill and L. Singh, “Topic-noise models: Modeling topic and noise distributions in social media post collections,” *in 2021 IEEE International Conference on Data Mining (ICDM). IEEE*, 2021, pp .71–80.

[2] Y. Li, R. Yu, C. Shahabi, and Y. Liu, “Diffusion convolutional recurrent neural network: Data-driven traffic forecasting,” *arXiv preprint arXiv:1707.01926*, 2017.

[3] J. Atwood and D. Towsley, “Diffusion-convolutional neural networks,” *Advances in neural information processing systems*, vol. 29, 2016.

[4] C. Shi, Y. Li, J. Zhang, Y. Sun, and S. Y. Philip, “A survey of heterogeneous information network analysis,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 1, pp. 17–37, 2016.

[5] V. Moscato and G. Sperli, “A survey about community detection over on-line social and heterogeneous information networks,” *Knowledge-Based Systems*, vol. 224, p. 107112, 2021.

[6] X. Wang, D. Bo, C. Shi, S. Fan, Y. Ye, and S. Y. Philip, “A survey on heterogeneous graph embedding: methods, techniques, applications and sources,” *IEEE Transactions on Big Data*, 2022.

[7] Y. Dong, N. V. Chawla, and A. Swami, “metapath2vec: Scalable representation learning for heterogeneous networks,” *in Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, 2017, pp. 135–144.

[8] T.-y. Fu, W.-C. Lee, and Z. Lei, “HIN2vec: Explore meta-paths in heterogeneous information networks for representation learning,” *in Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 2017, pp. 1797–1806.

[9] X. Li, Y. Wu, M. Ester, B. Kao, X. Wang, and Y. Zheng, “Semi-supervised clustering in attributed heterogeneous information networks,” *in Proceedings of the 26th international conference on world wide web*, 2017, pp. 1621–1629.

[10] X. Fu, J. Zhang, Z. Meng, and I. King, “MAGNN: Metapath aggregated graph neural network for heterogeneous graph embedding,” *in Proceedings of The Web Conference*, 2020, pp. 2331–2341.

[11] A. G. Carranza, R. A. Rossi, A. Rao, and E. Koh, “Higher-order clustering in complex heterogeneous networks,” *in Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2020, pp. 25–35.

[12] F. D. Malliaros and M. Vazirgiannis, “Clustering and community detection in directed networks: A survey,” *Physics reports*, vol. 533, no. 4, pp. 95–142, 2013.

[13] C. C. Aggarwal and C. Zhai, “A survey of text clustering algorithms”, *in Mining text data.* Springer, 2012, pp. 77–128.

[14] J. Leskovec, A. Rajaraman, and J. D. Ullman, Mining of massive data sets. *Cambridge university press*, 2020.

[15] J. Pennington, R. Socher, and C. D. Manning, “GloVe: Global vectors for word representation,” *in Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014, pp. 1532–1543.

[16] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” *arXiv preprint arXiv:1810.04805*, 2018.

[17] K. Wagstaff, C. Cardie, S. Rogers, S. Schrödl et al., “Constrained k-means clustering with background knowledge,” *in ICML*, vol. 1, 2001, pp. 577–584.

[18] A. Ng, M. Jordan, and Y. Weiss, “On spectral clustering: Analysis and an algorithm,” *Advances in neural information processing systems*, vol. 14, 2001.

[19] A. Grover and J. Leskovec, “node2vec: Scalable feature learning for networks,” *in Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, 2016, pp. 855–864.

[20] B. Perozzi, R. Al-Rfou, and S. Skiena, “DeepWalk: Online learning of social representations,” *in Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2014 ,pp. 701–710.

[21] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, “Distributed representations of words and phrases and their compositionality”, *In Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems*, 2013, pp. 3111–3119.

[22] M. E. Newman and M. Girvan, “Finding and evaluating community structure in networks,” *Physical review E*, vol. 69, no. 2, p. 026113, 2004.

[23] P.-Z. Li, L. Huang, C.-D. Wang, and J.-H. Lai, “EdMot: An edge enhancement approach for motif-aware community detection,” *in Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019, pp. 479–487.

[24] A. Epasto, S. Lattanzi, and R. Paes Leme, “Ego-splitting framework: From non-overlapping to overlapping clusters,” *in Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2017, pp. 145–154.

[25] Y. Xie, X. Wang, D. Jiang, and R. Xu, “High-performance community detection in social networks using a deep transitive autoencoder”, *Information Sciences*, vol. 493, pp. 75–90, 2019.

[26] Y. Jia, Q. Zhang, W. Zhang, and X. Wang, “Communitygan: Community detection with generative adversarial nets,” *in The World Wide Web Conference*, 2019, pp. 784–794.

[27] C. Zhang, D. Song, C. Huang, A. Swami, and N. V. Chawla, “Heterogeneous graph neural network,” *in Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019, pp. 793–803.

[28] X. Wang, H. Ji, C. Shi, B. Wang, Y. Ye, P. Cui, and P. S. Yu, “Heterogeneous graph attention network,” *in The world wide web conference*, 2019, pp. 2022–2032.

[29] Y. Sun, Y. Yu, and J. Han, “Ranking-based clustering of heterogeneous information networks with star network schema,” *in Proceedings of the15th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2009, pp. 797–806.

[30] Y. Chang, C. Chen, W. Hu, Z. Zheng, X. Zhou, and S. Chen, “MEGNN: Meta-path extracted graph neural network for heterogeneous graph representation learning,” *Knowledge-Based Systems*, vol. 235, p. 107611, 2022.

[31] Y. Bengio, A. Courville, and P. Vincent, “Representation learning: A review and new perspectives.” *IEEE TPAMI*, vol. 8, 2013, pp. 1798–1828.

پانویس‌ها

1 Homogeneous Graph
2 Heterogeneous Graph
3 Heterogeneous Information Network (HIN)
4 Node Clustering
5 Node Classification
6 Link Prediction
7 Embedding
8 Natural Language Processing (NLP)
9 Feature
10 Unsupervised
11 Metapath
12 Community Detection
13 Graph Representation Learning
14 Semi-Supervised
15 Metadata
16 Normalized Mutual Information (NMI)
17 Adjusted rand index (ARI)

چهارمین کنفرانس ملی انفورماتیک ایران
پژوهشگاه دانش‌های بنیادی، پردیس فرمانیه، تهران
۱۴ و ۱۵ دی ماه ۱۴۰۱

احتمال تعمیم‌پذیر نبودن روش‌های ارزیابی کیفیت مبتنی بر بردار پشتیبان

آرمین احمدزاده *[۱]، پوریا چراغی †[۲]

[۱،۲]مرکز پردازش سریع، پژوهشکده علوم کامپیوتر، پژوهشگاه دانش‌های بنیادی، تهران، ایران

چکیده

یک مدل محاسباتی که بتواند نظرِ انسان در مورد کیفیت تصاویر را پیش‌بینی نماید، کاربردهای فراوانی در مسائل مربوط به پردازش تصویر دارد. به دلیلِ پیچیدگی دستگاه بینایی بشر، مدل‌های یادگیری ماشین برای شبیه‌سازیِ نحوه‌ی قضاوت انسان‌ها استفاده می‌شوند. یکی از این مدل‌ها، وایازش بردار پشتیبان است. در این مقاله، نشان می‌دهیم که روش رایج برای بکارگیریِ بردارِ پشتیبان در ارزیابی کیفیت تصویر، لزوماً به مدل‌های قابلِ تعمیم نمی‌انجامد.

واژه‌های کلیدی: هوش مصنوعی، پردازش تصویر، دستگاه بینایی انسان، یادگیری ماشین، ارزیابی کیفیت تصویر، نمامفاد، ماشین بردار پشتیبان، وایازش

۱ مقدمه

تصاویر دیجیتال بخش قابل توجهی از مصرف رسانه‌ای بشر را تشکیل می‌دهند [۱]. بدیهی است که کیفیّتِ کَمِ این رسانه، منجر به نارضایتی کاربران و ناکامی ما در دریافت اطّلاعات مدّنظرمان خواهد شد. لذا، مطلوب است که وضعیت تصاویر ارسالی، از منظر کیفیت‌شان، تحت نظارت باشد.

مطمئن‌ترین راه برای ارزیابی کیفیت تصویر (که به اختصار «اَکْت» می‌خوانیمش)، پرسش از انسان‌هاست [۲]. به این ترتیب که تصویری را به جمعی از ناظرین[۱] نشان داده و نظر آن‌ها در مورد کیفیت تصویر را در قالب نمره‌ای از یک بازه‌ی مشخّص (، مثلاً $[0, 100]$،) جویا می‌شویم. میانگین نظرات[۲]، یک کمّیّت قابل اطمینان از کیفیّت تصویر است [۳]. چنین آزمایشی نیز اَکْتِ «انسانی[۳]» نام دارد.

بدیهی است که اَکْتِ انسانی برای کاربردهای بر-خط[۴] و حجیم امکان‌پذیر نیست. در این موارد، نرم‌افزاری مطلوب است که بتواند قضاوت دستگاه بینایی بشر در مورد کیفیت یک تصویر را پیش‌بینی نماید. دقت و سرعت این پیش‌بینی، معیارهای عملکرد یک مدل محاسباتیِ اَکْت[۵] هستند. ساخت چنین مدلی، موضوع مورد مطالعه‌ی بخش قابل توجهی از تحقیقاتِ پردازش تصویر است[۴]. برای ارزیابی و آموزش روش‌های محاسباتی، از نتایج اَکْتِ انسانی استفاده می‌شود.

ساده‌ترین راه برای اَکْتِ محاسباتی، محاسبه‌ی اختلاف دو تصویر است. موارد زیادی وجود دارند که یک تصویر سالم (با کیفیت احتمالاً مطلوب)، دچار تخریب شده و یک نسخه‌ی تخریب‌شده از آن بوجود می‌آید. وقتی هردوی این دو تصویر در دسترس باشند، می‌توان برای ارزیابی کیفیتِ تصویرِ تخریب‌شده، به اطلاعات موجود در تصویر سالم رجوع کرد. به این کار اَکْتِ «مرجع کامل» گفته می‌شود [۵]. یک مثال از کاربرد اَکْتِ مرجع کامل، فشرده‌سازی تصاویر است. الگوریتم فشرده‌کننده با کاهشِ اطلاعاتِ نشانک[۶]، حجم آن را کاهش داده و از طرفی کیفیّت آن را نیز خدشه‌دار می‌نماید. الگوریتم اَکْتِ مرجع کامل می‌تواند در هر لحظه تصویر فشرده‌شده را با تصویر اصلی مقایسه کرده و در صورت تخریب بیش از حدِّ کیفیت، این مورد را به الگوریتم فشرده‌کننده بازخورد دهد [۶].

اگر $x(i, j)$ نشانک مرجع و $y(i, j)$، نشانک تخریب‌شده باشد، میانگین مربّعات خطا، «MSE[۷]»، اختلاف مقادیر آن دو را خواهد سنجید:

$$MSE(x, y) = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} (x(i, j) - y(i, j))^2 \quad (۱)$$

M و N در (۱)، ابعاد تصاویر هستند. نشان داده می‌شود که «MSE» یا بیشینه‌ی نسبت نشانک به نوفه[۸]، «PSNR[۹]»، نمی‌توانند درک انسان از کیفیت تصویر را شبیه‌سازی نمایند [۷]. وانْگْ[۱۰] و همکارانش نشان دادند که تخریب ساختارهای تصویر در کاهش کیفیت آن مؤثر است[۷]. در سال ۲۰۰۴، روشی به نام «SSIM[۱۱]» برای کمّی‌سازی شباهت ساختاری دو تصویر ارائه کردند [۸] که با اختلاف از روش‌های مبتنی بر-«MSE» بهتر بود [۹].

با موفقیت SSIM، محققین سعی کردند ساختارهای تصویر را با ویژگی‌های دیگری، مثل لبه‌ها، مدل کنند. آماره‌های تصاویر طبیعی[۱۲] [۱۰] و پایداری اطلاعاتی[۱۳] [۱۱]، دیگر معیارهای پیشنهاد شده برای کیفیت تصویر هستند. روش‌های یادگیری ماشین سر-تا-سری[۱۴] نیز، برای اَکْتِ محاسباتی استفاده شده‌اند [۱۲، ۱۳، ۱۴].

یک سازَ و کار رایج برای اَکْتْ، استفاده از ماشین بردار پشتیبان، «SVR[۱۵]» [۱۵] است [۱۶]. کلّیات بکارگیری «SVR» در شکل ۱

a.ahmadzadeh@ipm.ir*
cheraaqee@ipm.ir†
azad@sharif.edu‡
azad@ipm.ir§

[۱] همان «subject»ها در آزمایش‌های علمی

[۲] که به اختصار «MOS» نامیده می‌شود و مخفّف عبارت «Mean Opinion Score» است. لازم به توضیح است که «MOS» نسبت مستقیمی با کیفیت ادراکی و «Difference-MOS» یا «DMOS» نسبتی عکس با کیفیت تصویر دارد.

[3] subjective image quality assessment

[۴] معادل فارسی «on-line»

[5] objective image quality assessment

[۶] معادل فارسی «signal»

[7] Mean Squared Error

[۸] معادل فارسی «noise»

[9] Peak Signal-to-Noise Ratio

[10] Zhou Wang

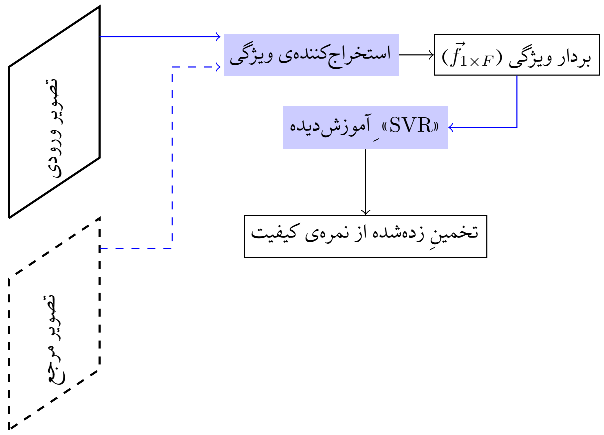
[11] Structural Similarity

[12] Natural Scene Statistics- NSS

[13] Information Fidelity

[14] end-to-end

[15] Support Vector Regression



شکل ۱: بکارگیری «SVR» برای اَکْتْ

دیده می‌شود. در روش‌های مبتنی بر بردار پشتیبان، نوآوری اصلی، در طراحی ویژگی‌ها صورت می‌گیرد. چنین روشی، ابتدا از تصویر ورودی یک بردار ویژگی استخراج می‌کند. اگر روش مرجع کامل باشد، می‌تواند برای استخراج ویژگی به نسخه‌ی سالم نیز رجوع کند. بردار ویژگی، $\vec{f}$، آرایه‌ای از اعداد است. اگر F ویژگی استخراج شوند، ابعاد این بردار، $F \times 1$ خواهد بود. یک مدل مبتنی بر وایازش[۱۶] بردار پشتیبان، آموخته است که این بردار ویژگی را به نمره‌ی کیفیت نگاشت کند. برای آموزش چنین مدلی، از تصاویر و نمرات ارزیابی‌های انسانی استفاده می‌شود. در این مقاله نشان می‌دهیم که مدل حاصل از روش رایج برای آموزش بردار پشتیبان که در اَکْتْ‌های محاسباتی استفاده می‌شود، لزوماً قابل تعمیم[۱۷] به تصاویر مختلف نیست.

مبانی مورد نیاز و برخی از کارهای مرتبط در قسمت۲ مرور می‌شوند. قسمت ۳، آزمایش‌های انجام شده را تشریح خواهد کرد و مقاله در قسمت ۴ جمع‌بندی می‌شود.

۲ مبانی و مرور ادبیات

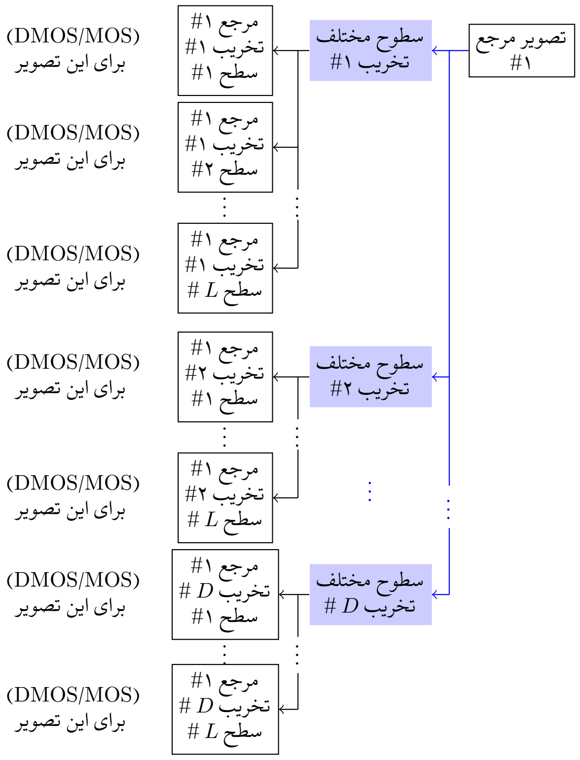
در این قسمت برخی مفاهیم و قراردادهای رایج در تحقیقات اَکْتِ محاسباتی تشریح می‌شوند. همچنین، یک دسته‌بندی از حوزه‌های ارزیابیِ کیفیت ارائه می‌گردد.

۱.۲ مجموعه‌داده‌ها

در اکثر کاربردها، انسان ناظرِ نهایی تصاویر است. لذا، نظر انسان بهترین معیارِ اَکْتْ برای این کاربردها خواهد بود. روش‌های محاسباتی سعی می‌کنند که نمراتی مثل نتایج اَکْتِ انسانی تولید کنند. مجموعه‌داده‌هایی وجود دارد که نظرات انسان به همراه تصاویر مربوطه را در اختیار محققین قرار می‌دهند، تا روش‌های اَکْتِ محاسباتی خود را محک بزنند [۳]. به این مجموعه‌داده‌ها، مجموعه‌داده‌های اَکْتْ گفته می‌شود.

یک مجموعه‌داده‌ی اَکْتْ، شامل یک سری تصویر مرجع، نسخ تخریب‌شده و نمرات انسانی است (شکل ۲). برخی اتفاقات برای تصاویر رایج هستند؛ مثلاً تار شدن به علت فاصله‌ی کانونی نامناسب، یا تَصَنُّعاتی[۱۸] که به دلیل فشرده‌سازی ایجاد می‌شود. به دلیل رایج بودن

[۱۶] معادل فارسی «regression»
[17] generalization
[18] artifacts



شکل ۲: قسمتی از آنچه که در یک مجموعه‌داده‌ی اَکْتْ وجود دارد. فرض می‌کنیم که R تصویر مرجع و D نوع تخریب وجود دارد که هر تخریب با L سطح از شدّت اعمال می‌شود. شکل، موارد را برای مرجع #۱ نشان می‌دهد. همین موارد برای مراجع #۲ تا R# نیز تکرار می‌شوند.

این تخریب‌ها، مهم است که توانایی الگوریتم‌های اَکْتْ در سنجش آن‌ها بررسی شود. به همین دلیل، یک سری تصویر سالم، با کیفیت قابل قبول، انتخاب کرده و این تخریب‌ها را به صورت مصنوعی روی آن‌ها اعمال می‌کنند. هر تخریب هم می‌تواند با شدّت متفاوتی اعمال شود. یک تصویر می‌تواند مقداری تار، یا خیلی تار باشد. بدیهی است که نمره‌ی انسانی برای شدّت‌های مختلف، متفاوت خواهد بود. انتظار می‌رود که نمره‌ی حاصل‌شده از روش محاسباتی نیز با نمره‌ی انسانی مطابق باشد [۱۷].

بنابراین، هر مجموعه‌داده‌ی اَکْتْ، شاملِ مجموعه‌ای از «نمونه»ها است. هر نمونه، یک زوج مرتب به شکل (نمره‌ی انسانی,تصویر) است. نمره‌ی انسانی می‌تواند به صورت MOS یا DMOS ذخیره شده باشد. به این ترتیب که، «تصویر»، ورودیِ یک روش محاسباتی، و «نمره‌ی انسانی»، پاسخِ صحیحِ موردِ انتظار[۱۹] است. حال اگر روش مرجع کامل باشد، تصویرِ مرجعِ متناظر نیز به عنوان ورودی به الگوریتم داده می‌شود. برای ارزیابی یک الگوریتم، نمرات آن برای تصاویر موجود در یک مجموعه‌داده محاسبه شده و سپس همبستگیِ نمرات الگوریتم با نمرات انسانی اندازه‌گیری می‌شود. معیار همبستگی، همان شاخص‌های آماری (پیِرسُنْ[۲۰]، اِسْپیِرْمَنْ[۲۱] و غیره) هستند. هرچه این همبستگی بیشتر باشد، الگوریتم در پیش‌بینی نظر انسان، دقیق‌تر عمل کرده است [۹].

[۱۹] همان «ground truth» در ادبیات یادگیری ماشین
[20] Pearson Correlation Coefficient- PLCC
[21] Spearman Rank Order Correlation Coefficient- SROCC

چهارمین کنفرانس ملی انفورماتیک ایران
پژوهشگاه دانش‌های بنیادی، پردیس فرمانیه، تهران
۱۴ و ۱۵ دی ماه ۱۴۰۱

۲.۲ آموزش و آزمون روش‌های مبتنی بر SVR

طبق روش رایج برای بکارگیری SVR در اَکْتْ [۱۸]، ابتدا باید صحنه‌های یک مجموعه‌داده را افراز[22] کنیم. منظور از «صحنه[23]» در یک مجموعه‌داده، مجموعه‌ی تمامیِ تصاویرِ تخریب‌شده‌ای است، که متعلق به **یک** مرجع هستند. (یعنی یک منظره‌ی یکسان را نشان می‌دهند، منتها با تخریب‌های متفاوت و شدّت‌های متفاوت.) اگر ٪۸۰ **صحنه‌ها** را، به طور تصادفی، برای آموزش، و ٪۲۰ را برای آزمون کنار گذاشته باشیم، یک «**افراز ۸۰–۲۰**» انجام داده‌ایم.

روشِ مبتنی بر SVR، ابتدا از تصویر یک بردار ویژگی استخراج می‌کند. اگر در مجموعه‌داده، S نمونه، به شکل DS_{image} در (۲) داشته باشیم، استخراج‌کننده‌ی ویژگی، S بردار ویژگی محاسبه می‌کند که مجموعه‌ی $DS_{\text{feature vector}}$ در (۳) را تشکیل می‌دهند.

$$DS_{image} = \{(\text{تصویر}_1, \text{نمره‌ی انسانی}_1), \dots, (\text{تصویر}_S, \text{نمره‌ی انسانی}_S)\} \quad (۲)$$

$$DS_{\text{vector feature}} = \{(\vec{f}_1, \text{نمره‌ی انسانی}_1), \dots, (\vec{f}_S, \text{نمره‌ی انسانی}_S)\} \quad (۳)$$

وقتی افرازِ تصادفیِ ۸۰–۲۰ را انجام دهیم، برخی اعضای $DS_{\text{feature vector}}$ مربوط به صحنه‌هایی هستند که در مجموعه‌ی آموزش[24] قرار گرفته‌اند و سایرِ اعضای آن مربوط به صحنه‌هایی هستند که در مجموعه‌ی آزمون[25] قرار گرفته‌اند. می‌توانیم عامل‌های[26] یک بردار پشتیبان را با استفاده از بردارهای ویژگی و نمرات انسانیِ مجموعه‌ی آموزش بهینه کرده و عملکرد آن را روی مجموعه‌ی آزمون ارزیابی کنیم.

از آنجایی که انتخاب صحنه‌ها برای افراز به مجموعه‌های آموزش و آزمون به صورت تصادفی صورت می‌گیرد، می‌توانیم افرازِ تصادفیِ ۸۰–۲۰ را چند بار انجام دهیم و میانه‌ی[27] عملکرد وایازنده[28] در این چند بار را به عنوان دقت الگوریتم در نظر بگیریم. اگر این کار را n دفعه انجام دهیم، اصطلاحاً می‌گویند که «**اعتبار سنجیِ متقابلِ n–لایه[29]**» انجام داده‌ایم. مقدار پیشنهاد شده برای n، در اَکْتْ، ۱۰۰۰ است [۱۸]. یعنی هزار بار، افرازِ ۸۰–۲۰ را به صورت تصادفی انجام دهیم، آموزش و آزمون را اجرا نمائیم و میانه‌ی عملکرد مدل در این هزار دفعه را به عنوان دقت الگوریتم گزارش کنیم.

البته باید این نکته را هم در نظر بگیریم که وایازنده‌ی مبتنی بر بردار پشتیبان، دو فراعامل[30]، به نام‌های c(ost) و γ، دارد. می‌توانیم نحوه‌ی کار وایازنده را به صورت رابطه‌ی (۴) بیان نماییم.

$$\text{پیش‌بینی} = \text{SVR}_{c,\gamma}(\vec{W} \cdot \vec{f}^T + \vec{b}) \quad (۴)$$

که $\vec{W}$ و $\vec{b}$ وزن‌های بهینه شده با استفاده از داده‌های آموزشی و $\vec{f}$ بردار ویژگیِ ورودی است. مقادیر c و γ قابل یادگیری نبوده و به صورت دستی تعیین می‌شوند. می‌توان بازه‌ای از مقادیر برای این دو فراعامل در نظر گرفت و اعتبار سنجی متقابل ۱۰۰–لایه را روی یک مجموعه‌داده برای تمامی این مقادیر انجام داد. زوج (c, γ)ای که بهترین عملکرد را داشته باشند، به عنوان فراعامل‌های انتخابی در نظر گرفته می‌شوند [۱۸].

۱.۲.۲ ارزیابی جداگانه‌ی عملکرد الگوریتم روی هر یک از تخریب‌ها

همانطور که در قسمت ۱.۲ گفتیم، برخی تخریب‌ها رایج هستند و مجموعه‌داده‌ها نمونه‌هایی با این تخریب‌ها دارند. برای اندازه‌گیری دقت یک الگوریتم در ارزیابی این تخریب‌ها، همان اعتبار سنجیِ ۱۰۰۰–لایه را انجام می‌دهیم. تفاوت این است که هنگام سنجش الگوریتم روی مجموعه‌ی آزمون، تنها نمونه‌هایی را در نظر می‌گیریم که دارای تخریب مورد نظر هستند [۱۸].

۳.۲ حوزه‌های مختلف اَکْتْ

مشخص است که دقت الگوریتمِ مبتنی بر بردار پشتیبان، به طور اصلی، به ویژگی‌هایی که استخراج می‌شوند بستگی دارد. هر چه این ویژگی‌ها گویاتر[31] باشند، SVR هم کار راحت‌تری برای یادگیریِ نحوه‌ی نگاشت ویژگی‌ها به نمره‌ی کیفیت خواهد داشت [۱۷]. ویژگی‌های گویا برای هر تصویری متفاوت خواهند بود و طراحی روشی که بتواند برای هر تصویری ویژگی‌های مربوط به کیفیت را محاسبه کند، کار دشواری است [۱۹].

غالب تحقیقاتِ اَکْتِ محاسباتی، برای تصاویر طبیعی انجام گرفته است [۲۰]. تصویرِ طبیعی، تصویری است که با دوربین‌های حساس به طیفِ مرئیِ موج الکترومغناطیس و از جهان فیزیکیِ پیرامون گرفته شده باشد [۲۱].

با شیوع تصویربرداری دیجیتال، تصاویر زیادی از اسناد و نوشته‌ها بوجود آمده است. ارزیابی کیفیت این تصاویر به روش‌هایی نیاز دارد که با روش‌های مربوط به تصاویر طبیعی متفاوت هستند [۲۲]. دلیل این تفاوت، وجود مقادیر زیاد نویسه‌ها و نواحی سیاه–سفید است.

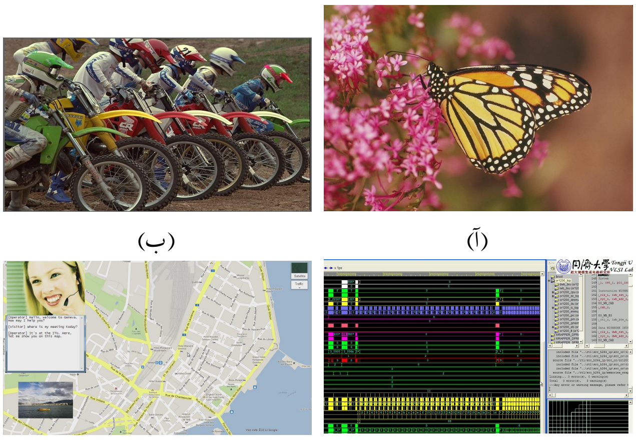
ترکیبِ تصاویر طبیعی و تصاویر اسناد، «**نَمامُفادها[32]**» را می‌سازد (شکل ۳). محاسبات از راه دور[33] و آموزشِ مجازی، شرایطی را بوجود می‌آورند که نیاز است مُفادِّ صفحه‌نمایشِ یک رایانه به عنوان تصویر مخابره شود. در [۲۰] نشان داده می‌شود که نمامفادها هم به روش‌های مختص به خود نیازمند هستند. مجموعه‌داده‌های کیفیت متخصص نمامفادها نیز برای تحقیقات اَکْتْ ارائه شده است [۲۰، ۲۳].

در این قسمت، نحوه‌ی آموزش و آزمون روش‌های اَکْتِ مبتنی بر بردار پشتیبان را مرور کردیم. همچنین، دیدیم که اَکْتِ محاسباتی، کاربردهای مختلفی برای انواع تصاویر دیجیتال دارد. در قسمتِ بعد، برخی جنبه‌های مربوط به یادگیری ماشین در آموزش SVRها را تحلیل می‌نماییم.

۳ آزمایش‌های پیشنهادی و نتایج

در این قسمت می‌بینیم که SVRهای آموزش‌دیده برای اَکْتْ، ممکن است مشکلاتی از نظر تعمیم‌پذیری داشته باشند. ابتدا یک آزمایش تعریف کرده و سپس به تحلیلِ عملکردِ SVR می‌پردازیم.

[22] افراز یک مجموعه به معنی تقسیم آن به زیرمجموعه‌هایی است که با هم اشتراک ندارند. یعنی اگر یک مجموعه‌داده را به زیرمجموعه‌هایی برای آموزش و آزمون افراز کنیم، آن زیرمجموعه‌ها عضو مشترکی ندارند.
[23] scene
[24] training set
[25] test set
[26] معادل فارسی «parameter»
[27] median
[28] regressor
[29] n-fold cross validation
[30] hyper-parameter
[31] expressive
[32] معادل فارسی «screen content image». دقت شود که «نماگرفت»ها، «screen-shots»، یک حالت خاصِّ نمامفادها هستند.
[33] remote computing



(ب) (آ)

(د) (ج)

شکل ۳: دو نمونه تصویر طبیعی (در (آ) و (ب)) و دو نمونه نمامفاد (در (ج) و (د))

۱.۳ آزمایش پیشنهادی

گفتیم که نمامفادها از نوشته‌ها، تصاویر طبیعی و نگاشتارها[34] تشکیل شده‌اند (قسمت ۳.۲). یک راه برای ارزیابی کیفیت نمامفادها، تحلیل جداگانه‌ی این نواحی است [۲۴]. اگر یک روش، مختصّ ارزیابی کیفیت تصاویر اسناد داشته باشیم، به نظر می‌رسد که ترکیبِ آن با روشی مختصِّ تصاویر طبیعی، برای ارزیابی کیفیت نمامفاد، شانس داشته باشد.

$LBPSI(.,.)$ [۲۵]، یک روش مرجع کامل، برای اَکثَرِ اسناد و نوشته‌هاست. به این ترتیب که اگر x و y مثل (۱) تعریف شده باشند، خواهیم داشت:

$$\text{LBPSI} = LBPSI(x, y) \quad \text{(نمره‌ی کیفیت تخمینی به وسیله‌ی LBPSI)} \qquad (۵)$$

همینطور، $SQMS(.,.)$ [۲۶]، یک روش مرجع کامل برای نمامفاد است. برای ارزیابی نواحی طبیعی، از $HaarPSI(.,.)$ [۲۷] و $GMSD(.,.)$ [۲۸] استفاده می‌کنیم که جنبه‌های مختلف ساختاری را می‌سنجند. از آنجایی که ارزیابی تباین[35] کار دشواری است [۲۹]، می‌توانیم از $VIF(.,.)$ [۱۱] هم استفاده کنیم که در این زمینه دقیق است.

با استفاده از این روش‌ها، می‌توانیم یک روش ترکیبیِ مرجع کامل [۳۰]، برای ارزیابی نمامفادها بسازیم. SQMS، LBPSI، HaarPSI، GMSD و VIF را برای x و y محاسبه می‌کنیم. در این صورت، ۵ عدد خواهیم داشت که می‌توانیم با آن‌ها یک بردار ویژگی تشکیل دهیم:

$$\vec{f} = [\text{SQMS}, \text{LBPSI}, \text{HaarPSI}, \text{GMSD}, \text{VIF}] \qquad (۶)$$

با داشتن این بردار ویژگی، می‌توانیم یک SVR را برای نگاشت آن به نمره‌ی کیفیت آموزش دهیم. برای داده‌ی آموزشی، از مجموعه‌داده‌های کیفیت نمامفاد، به نام‌های SIQAD [۲۰] و SCID [۲۳] استفاده می‌کنیم.

SIQAD شامل ۲۰ تصویر مرجع است، که هر کدام با ۷ نوع تخریب، تغییر یافته‌اند. هر تخریب در ۷ سطح اعمال شده است. بنابراین، $20 \times 7 \times 7 = 980$ تصویر تخریب‌شده در SIQAD وجود دارد. علاوه بر تخریب‌های SIQAD، SCID دو تخریب دیگر را نیز به ۴۰ تصویرِ مرجعِ خود اعمال کرده است. هر تخریب در ۵ سطحِ شدّت شبیه‌سازی

[34] graphics
[35] contrast

جدول ۱: ترکیب‌های درنظر گرفته‌شده برای تشکیل بردار ویژگی. وجودِ ✓ به معنی استفاده شدنِ نمره‌ی روشِ آن ستون است.

بردار	SQMS	HaarPSI	GMSD	VIF	LBPSI
$\vec{f_1}$	✓	✓			
$\vec{f_2}$	✓		✓		
$\vec{f_3}$	✓			✓	
$\vec{f_4}$	✓				✓
$\vec{f_5}$	✓	✓	✓		
$\vec{f_6}$	✓	✓		✓	
$\vec{f_7}$	✓	✓			✓
$\vec{f_8}$	✓		✓	✓	
$\vec{f_9}$	✓		✓		✓
$\vec{f_{10}}$	✓			✓	✓
$\vec{f_{11}}$	✓	✓	✓	✓	
$\vec{f_{12}}$	✓	✓	✓		✓
$\vec{f_{13}}$	✓	✓		✓	✓
$\vec{f_{14}}$	✓		✓	✓	✓

شده است، لذا در SCID، ۱۸۰۰ تصویر تخریب‌شده موجود است. برای تصاویر تخریب‌شده در این دو مجموعه‌داده، نمرات کیفیت انسانی تهیه شده است.

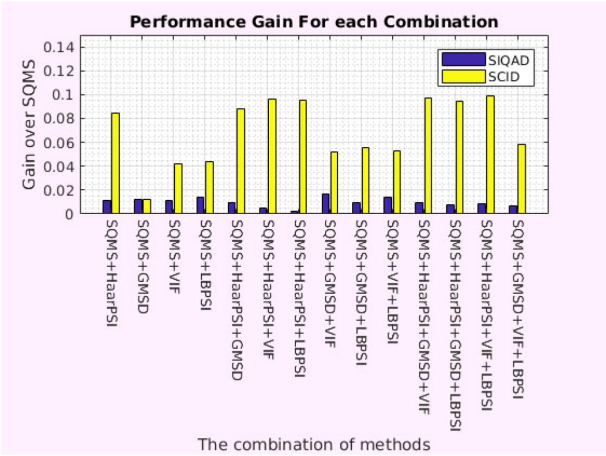
علاوه بر (۶)، می‌توانیم ترکیب‌های دیگری، مثل $\vec{f'} = [SQMS, LBPSI, GMSD]$ را در نظر بگیریم. از آنجایی که $SQMS(.,.)$، خود یک روش مختصّ نمامفاد است، تأثیر ترکیب سایر روش‌ها با آن را بررسی می‌کنیم. به این ترتیب، ترکیباتی که می‌توان امتحان کرد را در جدول ۱ خلاصه نموده‌ایم. برای مثال، $\vec{f_1}$، حاصلِ ترکیبِ $SQMS(.,.)$ و $HaarPSI(.,.)$ است. یعنی: $\vec{f_1} = [\text{SQMS}, \text{HaarPSI}]$. یا $\vec{f_6} = [\text{SQMS}, \text{HaarPSI}, \text{VIF}]$. بدیهی است که برای هر یک از بردارهای ویژگیِ $\vec{f_1}$ تا $\vec{f_{14}}$، باید SVR مجزّایی آموزش داده شود. در ادامه، عملکرد هر یک از این مدل‌ها بررسی می‌شوند.

۲.۳ موفقیت ترکیبِ روش‌ها، در هر یک از مجموعه‌داده‌ها

بردارهای $\vec{f_1}$ تا $\vec{f_{14}}$ را برای نمونه‌های SIQAD و SCID محاسبه کرده، و اعتبارسنجیِ متقابلِ ۱۰۰۰ـ‌لایه را برای یک یکِ آن‌ها انجام می‌دهیم. شکل ۴، بهره‌وریِ عملکردِ مدل‌های ترکیبی، نسبت به استفاده‌ی مستقیم از SQMS را نشان می‌دهد. منظور از بهره‌وری، محاسبه‌ی مقدار زیر است:

$$\text{بهره‌وری}_{f_i} = |\text{SROCC}_{f_i}| - |\text{SROCC}_{\text{SQMS}}| \qquad (۷)$$

که $i \in \{1, \dots, 14\}$ و SROCC_{f_i}، ضریب همبستگی اِسْپیرمَنِ محاسبه‌شده برای SVRی است که روی f_i آموزش دیده باشد. همانطور که در قسمت ۱.۲ گفته شد، این ضریب همبستگی، بین نمرات الگوریتم و نمرات انسانی محاسبه می‌شود. $\text{SROCC}_{\text{SQMS}}$ هم دقتِ $SQMS(.,.)$ نشان می‌دهد. می‌بینیم که تمام مدل‌ها عملکرد بهتری نسبت به $SQMS(.,.)$ داشته‌اند.



شکل ۴: بهره‌وری هر یک از روش‌های ترکیبی نسبت به $SQMS(.,.)$. برای هر ترکیب، بهره‌وریِ عملکرد بر روی دو محموعه‌داده، با دو رنگ متفاوت نشان داده شده است.

۳.۳ عدم تعمیم‌پذیری به تخریب‌های مختلف

با افزایش دقتِ $SQMS(.,.)$، روی کُلِّ مجموعه‌داده، انتظار می‌رود که ترکیب‌ها بهبودی مشابه را روی تک تکِ تخریب‌ها نیز داشته باشند. میزان دقت مدل‌ها روی تخریب‌ها طبق ۱.۲.۲ محاسبه گردیده و بهره‌وری آن نسبت به $SQMS(.,.)$ در شکل ۵ نشان داده می‌شود. می‌بینیم که نه تنها همه‌ی مدل‌ها عملکرد مشابهی نداشتند، بلکه در بسیاری از موارد، بهره‌وری منفی بوده است. کاملاً منطقی است که یکی از علّت‌ها را ضعف بردارهای ویژگی بدانیم. فارغ از این نظریه، مسئله‌ی دیگری وجود دارد که باید در نظر گرفته شود و آن *فراعامل‌ها* هستند.

همانطور که در انتهای ۲.۲ گفته شد، بهینه‌سازیِ c و γ با اعتبارسنجیِ متقابلِ ۱۰۰ـ لایه صورت می‌گیرد:

$$(c^*, \gamma^*) = \text{opt}(r_c, r_\gamma, \text{dataset}_{80,\text{all}}, \text{dataset}_{20,\text{all}}, 100) \quad (۸)$$

رابطه‌ی ۸ مواردی که باید در نظر گرفته شوند را خلاصه می‌کند. c^* و γ^* فراعامل‌های بهینه هستند. تابع opt نتایج بهینه‌سازی را برمی‌گرداند. r_c و r_γ، به ترتیب، مجموعه‌ی مقادیری هستند که برای c و γ امتحان می‌شوند. مثلاً اگر $r_c = \{1, 10, 10^2, 10^3, 10^4, 10^5, 10^6\}$ باشد، بهترین مقدار برای c از این مجموعه پیدا می‌شود. مقادیر مورد بررسی به صورت دستی انتخاب شده و تعدادشان به توان پردازشی در دسترس بستگی دارد. در هر بار آموزش و ارزیابی، نیاز به یک مجموعه‌ی آموزش و یک مجموعه‌ی آزمون داریم.

طبق رابطه‌ی۸، مجموعه‌ی آموزش، $\text{dataset}_{80,\text{all}}$ است. این عبارت یک زیرمجموعه از نمونه‌های مجموعه‌داده‌ی dataset را مشخص می‌نماید. این زیرمجموعه، عبارت است از **همه‌ی** (all) تخریب‌های موجود در زیرمجموعه‌ی آموزشِ افرازِ ۸۰-۲۰. برخی نمونه‌ها از این عبارت، می‌توانند $\text{SIQAD}_{80,\text{JPEG}}$ و $\text{SCID}_{20,\text{GB}}$ باشند. که اوّلی، یعنی تمامیِ نمونه‌های SIQAD؛ که در افرازِ ۸۰-۲۰، در مجموعه‌ی آموزش قرار گرفته‌اند؛ و تخریبِ آن‌ها از نوعِ JPEG است. دومی، نمونه‌هایی از SCID را مشخص می‌کند که در مجموعه‌ی آزمون قرار گرفته‌اند و تخریبِ آن‌ها از نوعِ GB[۳۶] است. آخرین ورودیِ تابعِ opt هم، تعداد لایه‌های (دفعاتِ) اعتبارسنجیِ متقابل است.

[۳۶]نوعی تار شدگی تصویر. برای شرح سایر تخریب‌ها به مقاله‌ی خودِ مجموعه‌داده‌ها رجوع شود.

اگر به جای $\text{opt}(r_c, r_\gamma, \text{SIQAD}_{80,all}, \text{SIQAD}_{20,all}, 100)$، از $\text{opt}(r_c, r_\gamma, \text{SIQAD}_{80,all}, \text{SIQAD}_{20,\text{JPEG}}, 100)$ استفاده کنیم، نتایج متفاوت خواهند بود (شکل ۶). قبل از بررسیِ نتایج، یک رابطه مانند ۸ برای اعتبارسنجیِ متقابل قرارداد می‌کنیم، که نقشِ فراعامل‌ها را نیز در نظر بگیرد:

$$\text{دقّت الگوریتم} = cv(\text{dataset}_{80,\text{all}}, \text{opt}(r_c, r_\gamma, \text{dataset}_{80,\text{all}}, \text{dataset}_{20,\text{all}}, 100), \text{dataset}_{20,\text{all}}, 1000) \quad (۹)$$

تابعِ $cv(.,.,.,.)$، چهار ورودی دارد که عبارت‌اند از: مجموعه‌ی آموزش، فراعامل‌های بهینه، مجموعه‌ی آزمون و تعداد لایه‌های اعتبارسنجیِ متقابل. این تابع میانه‌ی SROCCهای محاسبه‌شده در ۱۰۰۰ مرتبه را بر می‌گرداند.

شکل ۴ نتایج را برای آزمایش زیر نشان می‌دهد:

$$cv(\text{dataset}_{80,\text{all}}, \text{opt}(r_c, r_\gamma, \text{dataset}_{80,\text{all}}, \text{dataset}_{20,\text{all}}, 100), \text{dataset}_{20,\text{all}}, 1000) \quad (۱۰)$$

که dataset در هر بار، یا SIQAD و یا SCID است. برای ارزیابی روی هر یک از تخریب‌ها (شکل ۵)، رابطه (۱۰) به شکل زیر تغییر می‌کند:

$$cv(\text{dataset}_{80,\text{all}}, \text{opt}(r_c, r_\gamma, \text{dataset}_{80,\text{all}}, \text{dataset}_{20,\text{all}}, 100), \text{dataset}_{20,\text{dst}}, 1000) \quad (۱۱)$$

که dataset مطابقِ (۱۰) و $\text{dst} \in \{\text{dataset}$ تخریب‌های موجود در$\}$.

اگر در آزمایشی که رابطه (۱۱) نشان می‌دهد، فراعامل‌ها را روی تخریب‌های متناظر بهینه کنیم، نتایجِ شکل ۶ حاصل می‌شوند. به طور رسمی، نتایج شکل ۶، حاصلِ آزمایشِ زیر هستند:

$$cv(\text{dataset}_{80,\text{all}}, \text{opt}(r_c, r_\gamma, \text{dataset}_{80,\text{all}}, \text{dataset}_{20,\text{dst}}, 100), \text{dataset}_{20,\text{dst}}, 1000) \quad (۱۲)$$

پس می‌بینیم که علاوه بر تأثیرِ احتمالیِ بردارهای ویژگی، فراعامل‌ها نیز نقش مهمی در عملکرد SVRها دارند.

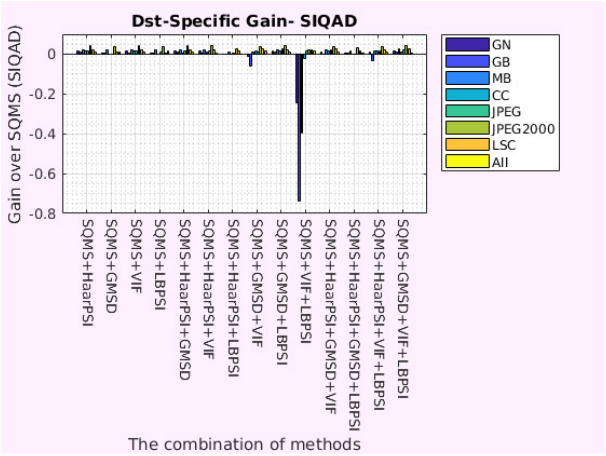
۴.۳ عدم تعمیم‌پذیری به سایر مجموعه‌داده‌ها

برای بررسی عدمِ وابستگیِ یک مدل به صحنه‌های یک مجموعه‌داده، آن را روی یک مجموعه‌داده آموزش داده و روی مجموعه‌داده‌ی دیگر می‌آزمایند:

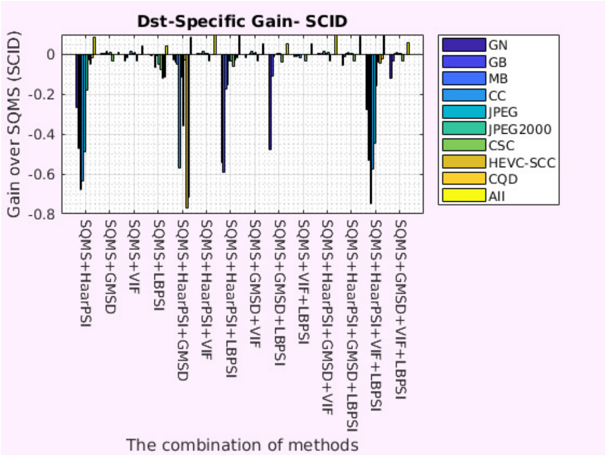
$$cv(\text{SIQAD}_{\text{all},\text{all}}, \text{opt}(r_c, r_\gamma, \text{SIQAD}_{80,\text{all}}, \text{SIQAD}_{20,\text{all}}, 100), \text{SCID}_{\text{all},\text{all}}, 1) \quad (۱۳)$$

$$cv(\text{SCID}_{\text{all},\text{all}}, \text{opt}(r_c, r_\gamma, \text{SCID}_{80,\text{all}}, \text{SCID}_{20,\text{all}}, 100), \text{SIQAD}_{\text{all},\text{all}}, 1) \quad (۱۴)$$

شکل ۷ نتایج را برای آزمایش‌های (۱۴) و (۱۳) نشان می‌دهد.

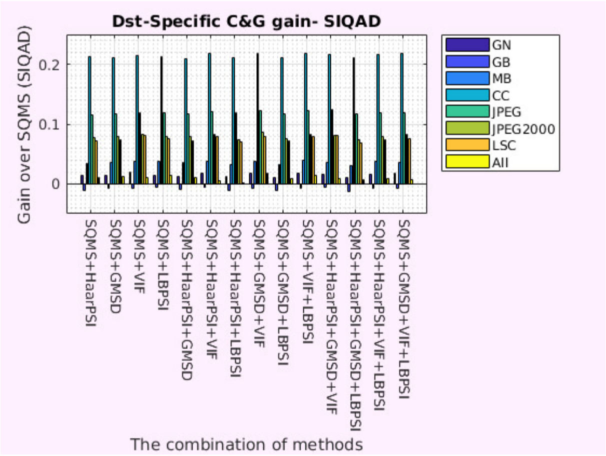


(آ)

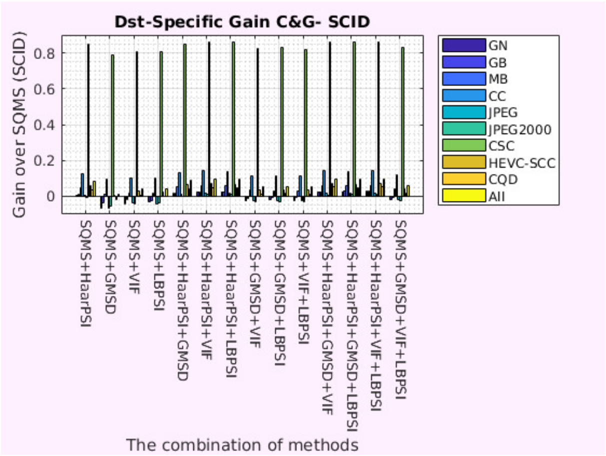


(ب)

شکل ۵: نتیجه‌ی ارزیابی به ازای هر تخریب. نتایج SIQAD در (آ) و SCID در (ب) قابل ملاحظه هستند. بهره‌وری به ازای هر تخریب، با رنگ متفاوتی مشخص شده است.



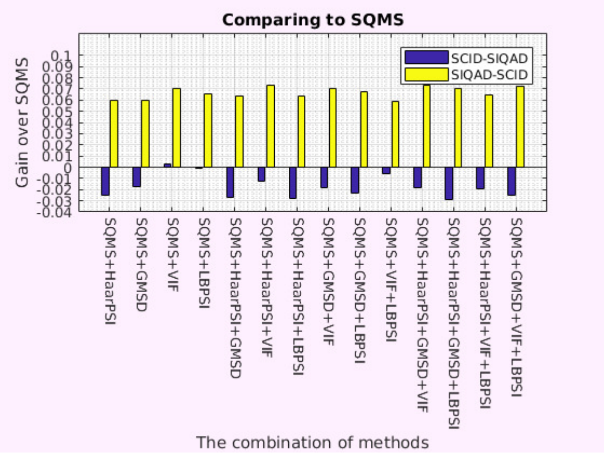
(آ)



(ب)

شکل ۶: نتیجه‌ی ارزیابی به ازای هر تخریب بعد از بهینه‌سازی فراعامل‌ها روی هر یک از تخریب‌ها. نتایج SIQAD در (آ) و SCID در (ب) قابل ملاحظه هستند.

چهارمین کنفرانس ملی انفورماتیک ایران
پژوهشگاه دانش‌های بنیادی، پردیس فرمانیه، تهران
۱۴ و ۱۵ دی ماه ۱۴۰۱



شکل ۷: نتایج آزمونِ تعمیم‌پذیری. ستون‌های زرد مربوط به آموزش روی SIQAD و آزمون روی SCID هستند. ستون‌های سُرمه‌ای مربوط به آموزش و آزمونِ عکس هستند.

می‌بینیم که بهبودِ مشاهده شده در آزمایش (۱۰) (با نتایجِ قابلِ مشاهده در شکل ۴) برای SIQAD اتفاق نمی‌افتد. به غیر از این مسئله، اگر مدلِ آموزش‌دیده را برای تک تکِ تخریب‌ها بیازمائیم، باز هم با اُفتِ عملکردی مشابهِ قسمتِ ۳.۳ مشابه می‌شویم. بیان رسمیِ این آزمایش‌ها در روابط (۱۵) و (۱۶) و نتایج آن‌ها در شکل ۸ ارائه شده‌اند.

$$\begin{aligned} &cv(\mathrm{SCID}_{\mathrm{all,dst}},\\ &\mathrm{opt}(r_c, r_\gamma, \mathrm{SCID}_{80,\mathrm{all}}, \mathrm{SCID}_{20,\mathrm{all}}, 100),\\ &\mathrm{SIQAD}_{\mathrm{all,dst}}, 1) \end{aligned} \qquad (۱۵)$$

$$\begin{aligned} &cv(\mathrm{SIQAD}_{\mathrm{all,dst}},\\ &\mathrm{opt}(r_c, r_\gamma, \mathrm{SIQAD}_{80,\mathrm{all}}, \mathrm{SIQAD}_{20,\mathrm{all}}, 100),\\ &\mathrm{SCID}_{\mathrm{all,dst}}, 1) \end{aligned} \qquad (۱۶)$$

dst در روابط بالا، عضوِ مجموعه‌ی تخریب‌هایی است که در SCID و SIQAD مشترک هستند. می‌بینیم که نتایج بهترِ آزمون تعمیم‌پذیری برای مجموعه‌ی SCID هم به تک تکِ تخریب‌های آن قابل تعمیم نیست.

۴ جمع‌بندی

در این مقاله بردارهای پشتیبان را، طبق روش رایج، برای اَکْتْ آموزش دادیم. دیدیم که وقتی گزارش عملکرد به روش رایج انجام می‌شود، نقش فراعامل‌ها هم باید بازتاب گردد. یک فرمول‌بندی ارائه شد که این موارد را لحاظ نماید. تأثیرِ بهینه‌سازیِ فراعامل‌ها را مشاهده کردیم و دیدیم که به این مسئله در آزمایش‌های ارزیابی کیفیت پرداخته نمی‌شود. محدود بودنِ تصاویر مجموعه‌داده‌ها، نگرانی در موردِ تعمیم‌پذیری مدل‌ها به تصاویر دنیای واقعی را بیشتر می‌کند. مخصوصاً وقتی که تخریب‌های مجموعه‌داده‌ها مصنوعی بوده و ممکن است با تخریب‌های طبیعی متفاوت باشند. شاید بهتر باشد که نحوه‌ی بهینه‌سازیِ فراعامل‌های مدل‌های یادگیری نیز در گزارش‌های عملکرد مقالات تشریح گردد. آزمایش‌های انجام‌شده با کُدهای موجود در https://github.com/cheraaqee/fusion_iqa قابلِ شبیه‌سازی هستند.

مراجع

[1] CISCO, “Cisco annual internet report,” `https://www.cisco.com/c/en/us/solutions/executive-perspectives/annual-internet-report/index.html`, 2022. Accessed: 2022-July-15.

[2] E. Allen and S. Triantaphillidou. *The manual of photography*. CRC Press, 2012.

[3] P. Mohammadi, A. Ebrahimi-Moghadam, and S. Shirani, “Subjective and objective quality assessment of image: A survey,” *arXiv preprint arXiv:1406.7799*, 2014.

[4] G. Zhai and X. Min, “Perceptual image quality assessment: a survey,” *Science China Information Sciences*, vol.63, no.11, pp.1–52, 2020.

[5] E. C. Larson and D. M. Chandler, “Most apparent distortion: full-reference image quality assessment and the role of strategy,” *Journal of electronic imaging*, vol.19, no.1, p.011006, 2010.

[6] M. P. Eckert and A. P. Bradley, “Perceptual quality metrics applied to still image compression,” *Signal processing*, vol.70, no.3, pp.177–200, 1998.

[7] Z. Wang and A. C. Bovik, “Mean squared error: Love it or leave it? a new look at signal fidelity measures,” *IEEE signal processing magazine*, vol.26, no.1, pp.98–117, 2009.

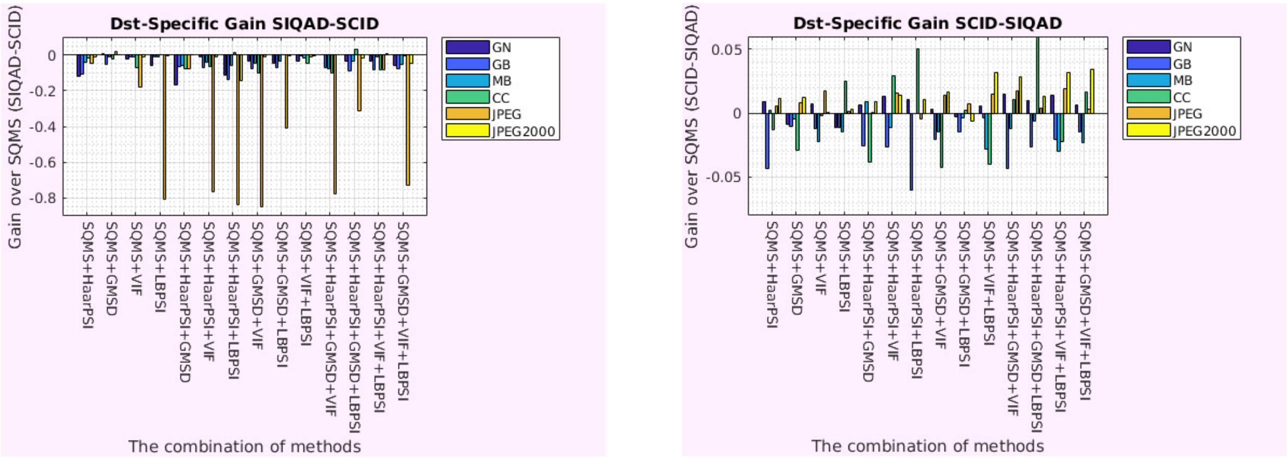
[8] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *IEEE transactions on image processing*, vol.13, no.4, pp.600–612, 2004.

[9] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, “A statistical evaluation of recent full reference image quality assessment algorithms,” *IEEE Transactions on image processing*, vol.15, no.11, pp.3440–3451, 2006.

[10] D. L. Ruderman and W. Bialek, “Statistics of natural images: Scaling in the woods,” in *Advances in neural information processing systems*, pp.551–558, 1994.

[11] H. R. Sheikh and A. C. Bovik, “Image information and visual quality,” *IEEE Transactions on image processing*, vol.15, no.2, pp.430–444, 2006.

[12] J. Kim, H. Zeng, D. Ghadiyaram, S. Lee, L. Zhang, and A. C. Bovik, “Deep convolutional neural models for picture-quality prediction: Challenges and solutions to data-driven image quality assessment,” *IEEE Signal processing magazine*, vol.34, no.6, pp.130–141, 2017.



(آ) (ب)

شکل ۸: نتایجِ ارزیابی هر تخریب، وقتی مجموعه‌داده‌های آموزش و آزمون متفاوت‌اند. (آ): برای SIQAD → SCID و (ب) برای SCID → SIQAD

[13] X. Yang, F. Li, and H. Liu, "A survey of dnn methods for blind image quality assessment," *IEEE Access*, vol.7, pp.123788–123806, 2019.

[14] P. Ye, J. Kumar, L. Kang, and D. Doermann, "Unsupervised feature learning framework for no-reference image quality assessment," in *2012 IEEE conference on computer vision and pattern recognition*, pp.1098–1105, IEEE, 2012.

[15] V. Vapnik. *The nature of statistical learning theory.* Springer science & business media, 1999.

[۱۶] پوریا چراغی، "ارائه‌ی یک روش بدون مرجع برای ارزیابی کیفیت تصاویر با تخریب چندگانه،" پایان‌نامه کارشناسی‌ارشد، گروه مهندسی برق و کامپیوتر دانشگاه خوارزمی، شهریور ۱۳۹۸.

[17] D. M. Chandler, "Seven challenges in image quality assessment: past, present, and future research," *International Scholarly Research Notices*, vol.2013, 2013.

[18] W. Xue, X. Mou, L. Zhang, A. C. Bovik, and X. Feng, "Blind image quality assessment using joint statistics of gradient magnitude and laplacian features," *IEEE Transactions on Image Processing*, vol.23, no.11, pp.4850–4862, 2014.

[19] L. Kang, P. Ye, Y. Li, and D. Doermann, "Convolutional neural networks for no-reference image quality assessment," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp.1733–1740, 2014.

[20] H. Yang, Y. Fang, and W. Lin, "Perceptual Quality Assessment of Screen Content Images," *IEEE Transactions on Image Processing*, vol.24, no.11, pp.4408–4421, 2015.

[21] A. Mittal, *Natural scene statistics-based blind visual quality assessment in the spatial domain.* Phd thesis, The University of Texas at Austin, 2013.

[22] P. Ye and D. Doermann, "Document image quality assessment: A brief survey," in *2013 12th International Conference on Document Analysis and Recognition*, pp.723–727, IEEE, 2013.

[23] Z. Ni, L. Ma, H. Zeng, J. Chen, C. Cai, and K. K. Ma, "ESIM: Edge Similarity for Screen Content Image Quality Assessment," *IEEE Transactions on Image Processing*, vol.26, no.10, pp.4818–4831, 2017.

[24] Y. Zhang, D. M. Chandler, and X. Mou, "Quality Assessment of Screen Content Images via Convolutional-Neural-Network-Based Synthetic/Natural Segmentation," *IEEE Transactions on Image Processing*, vol.27, no.10, pp.5113–5128, 2018.

[25] A. Alaei, D. Conte, M. Blumenstein, and R. Raveaux, "Document image quality assessment based on texture similarity index," in *2016 12th IAPR Workshop on Document Analysis Systems (DAS)*, pp.132–137, IEEE, 2016.

[26] K. Gu, S. Wang, H. Yang, W. Lin, G. Zhai, X. Yang, and W. Zhang, "Saliency-guided quality assessment of screen content images," *IEEE Transactions on Multimedia*, vol.18, no.6, pp.1098–1110, 2016.

[27] R. Reisenhofer, S. Bosse, G. Kutyniok, and T. Wiegand, "A haar wavelet-based perceptual similarity index for image quality assessment," *Signal Processing: Image Communication*, vol.61, pp.33–43, 2018.

[28] W. Xue, L. Zhang, X. Mou, and A. C. Bovik, "Gradient magnitude similarity deviation: A highly efficient perceptual image quality index," *IEEE transactions on image processing*, vol.23, no.2, pp.684–695, 2013.

[29] A. Shokrollahi, B. M.-N. Maybodi, and A. Mahmoudi-Aznaveh, "Histogram modification based enhancement along with contrast-changed image quality assessment," *Multimedia Tools and Applications*, pp.1–22, 2020.

[30] K. Okarma, P. Lech, and V. V. Lukin, "Combined full-reference image quality metrics for objective assessment of multiply distorted images," *Electronics*, vol.10, no.18, p.2256, 2021.