Journal of Applied And Basic Machine Intelligence Research (ABMIR)

پژوهش‌های نظری و کاربردی هوش ماشینی

سال سوم، شماره اول، بهار و تابستان ۱٤۰٤، صفحات ۲۱۱ - ۱۹۳

سیستم توصیه‌گر مبتنی بر جلسه با استفاده از روش ترکیبی شبکه‌های عصبی بازگشتی و گراف

میثم میرزائی[۱*]، حسین محمدی[۲] سجاد احمدیان[۳]

[۱] دانشجوی دکتری، گروه کامپیوتر، دانشکده مهندسی، دانشگاه زنجان، ایران

[۲] استادیار گروه کامپیوتر، دانشکده مهندسی، دانشگاه زنجان، ایران

[۳] استادیار، دانشکده فناوری اطلاعات، دانشگاه صنعتی کرمانشاه، کرمانشاه، ایران

مقاله پژوهشی

تاریخ دریافت:

۱٤۰٤/۷/۵

تاریخ پذیرش:

۱٤۰٤/۸/۲٤

کلیدواژه‌ها:

سیستم توصیه‌گر مبتنی بر جلسه، شبکه‌های عصبی گراف، شبکه‌های عصبی بازگشتی، مدل ترکیبی

نویسنده‌مسئول:

m.mirzaei@znu.ac.ir

چکیده

سیستم‌های توصیه‌گر مبتنی بر جلسه نوعی از سیستم‌های توصیه‌گر هستند که با مدل‌سازی وابستگی‌های بین تعاملات کاربر و آیتم‌ها، به دنبال معرفی مناسب‌ترین آیتم یا مجموعه‌ای از آیتم‌ها به کاربر هستند. موضوع مهم در عملکرد این سیستم‌ها، توانایی کشف وابستگی‌های مختلف در یک جلسه است. درحالی‌که مدل‌های موجود اغلب تنها بر یکی از این دو نوع وابستگی ترتیبی یا پیچیده تمرکز دارند، در این پژوهش یک مدل جدید ارائه‌شده است که از شبکه عصبی گراف برای شناسایی وابستگی‌های پیچیده و غیرترتیبی بین آیتم‌ها، و از شبکه عصبی بازگشتی برای مدل‌سازی وابستگی‌های ترتیبی درون جلسه استفاده می‌کند. این ترکیب به مدل امکان می‌دهد تا درکی غنی‌تر و دقیق‌تر از هدف کاربر به دست آورده و در نتیجه، نسبت به روش‌های متعارف که تنها به یکی از این جنبه‌ها می‌پردازند، دقت پیش‌بینی بالاتری را ارائه دهد. ارزیابی جامع بر روی دو مجموعه داده پرکاربرد **Yoochoose** و **Diginetica**، برتری روش معرفی‌شده را در مقایسه با جدیدترین روش‌های موجود به اثبات می‌رساند. نتایج نشان می‌دهد که **RGNF-SRec** به ترتیب به بهبود ۰/٤۰٪ در معیار **P@20** و ۱/۰۷٪ در معیار **MRR@20** نسبت به قوی‌ترین مدل پایه بر روی این مجموعه داده‌ها دست‌یافته و توانایی خود را در ارائه توصیه‌های دقیق‌تر برای آیتم بعدی به نمایش می‌گذارد.

: 10.22034/ABMIR.2025.23730.1172

E-ISSN: 2821-2037

/The Author 2025. Published by Yazd University This is an open access article under the CC BY 4.0 License (https://creativecommons.org/licenses/by/4.0/).

۱- مقدمه

رشد قابل توجه اطلاعات تولید شده توسط کاربران به‌واسطه استفاده از وب در دو دهه اخیر، باعث بروز مشکل گرانباری اطلاعات[1] شده است که در برخی موارد باعث دورکردن کاربر از اطلاعات مورد نیاز شده است. به همین دلیل نیاز به یک ابزار پالایش‌گر برای کار با این حجم اطلاعات و دسترسی بهتر افراد به موارد مرتبط با نیازشان به وجود آمده است. سیستم‌های توصیه‌گر[2] ابزاری برای حل این مشکل هستند که به صورت گسترده برای معرفی موارد[3] مختلف به کاربران ازجمله لباس، فیلم، موزیک، کتاب، بازی، اخبار و غیره مورد استفاده قرار می‌گیرند.

به صورت کلی، سیستم‌های توصیه‌گر به دنبال فهم علایق کاربر از طریق تحلیل رفتار و تعاملات قبلی کاربر با آیتم‌ها هستند تا با استخراج علاقه‌مندی کاربر، در انتخاب مرتبط‌ترین آیتم‌ها به وی کمک کنند. با پیشنهاد آیتم‌های مفید به کاربر، نیاز کاربر برآورده شده و رضایت وی افزایش می‌یابد. تعامل کاربر می‌تواند به‌صورت صریح (امتیاز داده شده توسط کاربر به آیتم) و یا ضمنی (بازدید، گوش‌کردن، هدایت کردن و غیره) باشد. توصیه‌گرها امروزه نقشی حیاتی در سکوهای مختلف اینترنتی ایفا می‌کنند؛ برای مثال، بخش عمده‌ای از محتوای مصرفی در سرویس‌های پخشی[4]، محصولات پیشنهادی در پلتفرم‌های تجارت الکترونیک، و محتوای خبری در شبکه‌های اجتماعی توسط این سیستم‌ها پالایش و پیشنهاد می‌شوند [۱] [۲] [۳].

سیستم‌های توصیه‌گر کلاسیک از روش‌های پالایش مشارکتی یا فاکتورسازی ماتریس[5] استفاده کرده‌اند. این سیستم‌ها متکی بر امتیازدهی صریح کاربران و یا اطلاعات ضمنی یعنی تعاملات کاربران بر روی مجموعه آیتم‌ها بوده و با محاسبه شباهت بین کاربران، علایق آن‌ها را پیش‌بینی می‌کنند. این روش‌ها به حجم زیادی از فعالیت‌های گذشته کاربران نیاز دارند و برای کاربران جدید یا آیتم‌های جدید با مشکل شروع سرد[6] مواجه می‌شوند. همچنین به دلیل کمبود داده‌ها و خلوت‌بودن تعاملات کاربر- آیتم قادر به ارائه توصیه‌های دقیق و مطمئنی به کاربران نیستند. همچنین اکثر روش‌های موجود متعهد به استفاده از تمام اطلاعات تاریخی کاربران و ثبت علاقه‌مندی بلندمدت برای توصیه هستند [٤]. این روش ممکن است در سناریوهای واقعی عملی نباشد زیرا (۱) ترجیح کاربر با گذشت زمان تغییر می‌کند و تکامل می‌یابد، و (۲) این روش‌ها در مدل‌سازی پویایی رفتارهای کاربر و تحقق توصیه در لحظه دچار مشکل هستند. در حقیقت این روش‌ها برای پیش‌بینی علایق کاربران از ویژگی‌های کم عمق استفاده کرده‌اند و اطلاعات مفید توالی جلسات را که در بسیاری از کاربردها وجود داشته و دارای ویژگی‌های مفید بوده و می‌تواند به پیش‌بینی آیتم هدف کمک کند، را نادیده گرفته‌اند[۵]. برای حل این مشکل، سیستم‌های توصیه‌گر مبتنی بر جلسه معرفی شدند. یک جلسه دنباله‌ای از تعاملات کاربر از قبیل کلیک‌های کاربر در یک بازه زمانی مشخص است که می‌تواند تغییرات علایق و ترجیحات کاربر را در طول زمان منعکس کند. سیستم‌های توصیه‌گر مبتنی بر جلسه با مدل‌سازی وابستگی‌های بین تعاملات کاربر و آیتم‌ها، به دنبال معرفی مناسب‌ترین آیتم یا مجموعه‌ای از آیتم‌ها به کاربر هستند. این توصیه‌گرها درواقع بر پویایی تعاملات متمرکز هستند و از آن برای توصیه استفاده می‌کنند [٦]. کاربردهای عملی این سیستم‌ها بسیار گسترده است؛ ازجمله پیشنهاد محصول بعدی در یک خرید آنلاین، توصیه مقاله یا خبر بعدی بر اساس چند مقاله اخیر خوانده‌شده در پرتال‌های خبری، یا پیشنهاد آهنگ بعدی در سرویس‌های پخش موسیقی. این توصیه‌گرها با کشف وابستگی‌های پیچیده بین آیتم‌های یک یا چند توالی یا جلسه کاربر و ارتباط بین جلسات مختلف، آیتم یا آیتم‌ها را به کاربر پیشنهاد می‌کنند. این توصیه‌گرها تعاملات کاربر و آیتم‌ها را دقیق‌تر و منعطف‌تر بررسی می‌کنند [۷].

با توجه به پتانسیل بالای سیستم‌های توصیه‌گر مبتنی بر جلسه، روش‌های متعددی برای انجام توصیه‌های مبتنی بر جلسه توسعه یافته‌اند. به‌طور خاص، برخی روش‌های مبتنی بر زنجیره مارکوف[۸، ۹] پیشنهاد شده‌اند تا آیتم بعدی را در یک جلسه بر

[1] Information Overload
[2] Recommender systems
[3] Items
[4] Streaming
[5] Matrix Factorization
[6] Cold Start

اساس آیتم قبلی در همان جلسه پیش‌بینی کنند، که تنها وابستگی‌های مرتبه اول بین دو آیتم مجاور را در نظر می‌گیرند. با این حال، فرضیه قوی زنجیره مارکوف ممکن است با داده‌های دنیای واقعی سازگار نباشد و باعث کاهش عملکرد پیش‌بینی شود. علاوه بر این، روش‌های مبتنی بر زنجیره مارکوف وابستگی‌های مرتبه بالاتر بین چندین آیتم در یک جلسه را نادیده می‌گیرند. در سال‌های اخیر، شبکه‌های عصبی بازگشتی[1] به‌طور گسترده‌ای در توصیه‌های مبتنی بر جلسه برای درک وابستگی‌های ترتیبی مرتبه بالا بین دنباله‌ای از آیتم‌ها درون جلسات به کار گرفته شده‌اند[۱۰–۱۳]. اگرچه این روش‌های مبتنی بر شبکه‌های عصبی بازگشتی عملکرد خوبی داشته‌اند، اما با مشکلات نیز مواجه هستند. این روش‌ها فرض می‌کنند که تمام آیتم‌های قبلی به آیتم فعلی مرتبط هستند و سپس آیتم بعدی را بر اساس وابستگی‌های ترتیبی دقیق فرض‌شده بر روی تمام آیتم‌های موجود در جلسه پیش‌بینی می‌کنند. اما درواقعیت، معمولاً همه آیتم‌های موجود در جلسه به آیتم فعلی مرتبط نیستند و وجود آیتم‌های غیرمرتبط می‌تواند بر توصیه‌های آیتم بعدی تأثیر منفی بگذارد[۱٤]. همچنین، این روش‌ها تنها امکان کشف وابستگی‌های ترتیبی ساده و مستقیم یک‌طرفه بر روی دنباله‌ای از آیتم‌ها را دارند در حالی‌که معمولاً انتقال‌های پیچیده‌تری مانند وابستگی‌های بین آیتم‌های دور از هم نیز در یک جلسه وجود دارند [۱۵]. برای درک انتقال‌های پیچیده آیتم‌ها در داده‌های جلسه، شبکه‌های عصبی گراف[2] موردتوجه قرار گرفته‌اند. این سیستم‌ها مانند SR-GNN [۳۱] از شبکه‌های عصبی گراف برای درک انتقال‌های پیچیده و غیرترتیبی آیتم‌ها (مانند آیتم ۱ به ۵) بین آیتم‌های درون جلسات استفاده می‌کنند تا آیتم بعدی را به کاربران توصیه کنند. با توجه به انتقال‌های پیچیده بین آیتم‌ها در جلسات، روش‌های مبتنی بر گراف قادر به تولید نمایش دقیق‌تری برای هر آیتم هستند. هرچه نمایش آیتم دقیق‌تر باشد، توصیه بهتر است. با این حال، محدودیت‌هایی نیز وجود دارد. اول، روش‌های مبتنی بر گراف نمی‌توانند اطلاعات توالی جلسه، مانند گره‌های تکراری در یک جلسه و گره شروع گراف جهت‌دار را به‌طور کامل بیان کنند. به‌طور خاص، گراف دو جلسه متفاوت S1: (A→B→C→A و C→A→B→C :S2) یکسان خواهد بود که نشان می‌دهد برخی از اطلاعات توالی از قبیل آیتم شروع جلسه را نمی‌توان به درستی توسط گراف ثبت کرد. دوم، آیتم‌های نویزی به ناچار زمانی که کاربر از بین محصولات متنوع انتخاب می‌کند در جلسات وجود دارند. با این حال، نویز در سیستم‌های توصیه مبتنی بر جلسه نادیده گرفته می‌شود. به‌طور خلاصه، این مشکلات، روش‌های توصیه مبتنی بر گراف را محدود کرده و آن‌ها را قادر به دستیابی به نتایج بهتر نمی‌کند.

برای رفع این محدودیت‌ها، در این پژوهش یک مدل جدید ترکیبی برای توصیه‌های مبتنی بر جلسه پیشنهاد شده است. این مدل از گراف عصبی برای ثبت انتقال‌های پیچیده بین آیتم‌ها در یک جلسه و از شبکه عصبی بازگشتی برای به دست آوردن اطلاعات مربوط به وابستگی‌های ترتیبی برای رفع کاستی‌های مدل‌سازی گراف جلسه استفاده می‌کند. مهم‌ترین دستاوردهای روش پیشنهادی به شرح زیر خلاصه می‌شود:

• گراف عصبی و شبکه بازگشتی به ترتیب جلسه را مدل‌سازی می‌کنند تا اطلاعات انتقال آیتم و اطلاعات توالی را به دست آورند، که نقص هریک را در مدل‌سازی جلسه کاهش می‌دهد.

• دو ماژول به صورت مستقل تحلیل خود را انجام می‌دهند و خطای موجود در هرکدام بر تحلیل دیگری اثر منفی نمی‌گذارد.

• آزمایش‌های انجام شده روی دو مجموعه داده واقعی انجام شده است. نتایج این آزمایش‌ها نشان می‌دهد که مدل پیشنهادی از سایر روش‌های بهتر عمل می‌کند و نشان می‌دهد که اجزای مدل مؤثر هستند.

در ادامه مقاله و در بخش ۲ به تشریح و بررسی تحقیقات پیشین در زمینه سیستم‌های توصیه‌گر مبتنی بر جلسه پرداخته شده است. در بخش ۳ روش پیشنهادی توضیح داده شده و نتایج پیاده‌سازی و همچنین مقایسه عملکرد روش پیشنهادی با سایر روش‌های مشابه در بخش ٤ ارائه گردیده است. نهایتاً در بخش ۵ نتیجه‌گیری و مسیر آتی تحقیق بیان شده است.

[1] Recurrent Neural Networks (RNNs)

[2] Graph Neural Networks (GNNs)

۲- کارهای مرتبط

سیستم‌های توصیه‌گر مبتنی بر جلسه را می‌توان به دو دسته اصلی روش‌های سنتی و روش‌های مبتنی بر یادگیری عمیق تقسیم‌بندی کرد. در این بخش به بیان و مرور برخی از پژوهش‌های انجام‌گرفته در این حوزه پرداخته شده است.

۲-۱ روش‌های سنتی

روش‌های سنتی مبتنی بر فرض وابستگی و ارتباط بین آیتم‌های یک جلسه توسعه داده نشده بودند و بنابراین فاقد مدل‌سازی وابستگی‌های درون جلسه هستند. این روش‌ها ترجیحات کلی کاربران را مبتنی بر کل تاریخچه تعاملات کاربران و بااهمیت یکسان در نظر می‌گیرند و اطلاعات توالی تعاملات کاربر را نادیده می‌گیرند. هدف این روش‌ها یافتن آیتم‌های مشابه با آیتم‌های تعامل شده توسط کاربر برای توصیه است. روش‌های مبتنی بر شباهت، فراوانی، و ماتریس عامل‌بندی نمونه‌هایی از روش‌های سنتی هستند. این روش‌ها به دلیل ماهیت ایستا، قابلیت کشف پویایی و تغییر در علاقه‌مندی‌های کاربر را ندارند. هم‌چنین به دلیل نیاز به مقایسه یک توالی با سایر توالی‌ها، مشکل شروع سرد یک چالش اساسی در آن‌ها است. برخلاف روش‌های فاقد مدل، زنجیره مارکف اصلی‌ترین رویکرد شناخته شده سنتی است که از مدل‌سازی تعاملات کاربر در یک جلسه برای کشف وابستگی بین آیتم‌ها و ترجیحات کاربران استفاده می‌کند [۱۷]. فرض وابستگی قوی وضعیت فعلی به وضعیت‌های قبلی، زنجیره مارکف را به ابزاری برای تحلیل رفتار در توالی‌ها تبدیل کرده است [۱۸]. زیرا در زنجیره مارکف امکان انتخاب یک آیتم در یک زمان، به آیتم‌های انتخاب شده در زمان‌های قبلی وابسته است. مدل‌های مرتبه اول مارکف در کشف وابستگی‌های ساده [۱۹] و مدل‌های مرتبه بالاتر [۲۰] در کشف وابستگی‌های پیچیده کاربرد دارند. از ترکیب زنجیره مارکف و سایر روش‌ها نیز برای بهبود توصیه استفاده شده است. FPMC [۲۱] روشی است که از ترکیب زنجیره مارکف و ماتریس عامل‌بندی برای ساخت یک ماتریس سفارشی به ازای هر کاربر بجای همه کاربران و درک بهتر علاقه‌مندی کاربر استفاده می‌کند. روش معرفی‌شده در [۲۲] نیز روشی ترکیبی است که از زنجیره مارکف برای مدل‌سازی رفتار کاربر و تحلیل محتوای صفحات بازدیدشده توسط کاربر برای بهبود توصیه در وب استفاده می‌کند. یکی از مزایای اصلی زنجیره مارکف در مقایسه با سایر روش‌های سنتی، درنظرگرفتن تمام آیتم‌ها و عدم حذف آیتم‌های کم تکرار در توصیه است. با این حال، روش‌های مبتنی بر زنجیره مارکوف اغلب متکی بر آخرین تعامل کاربر و یا چند تعامل اخیر کاربر هستند و فقط می‌توانند وابستگی‌های کوتاه‌مدت و محلی بین تعاملات را مدل کنند [۲۳]. درواقع زنجیره مارکف ارتباط بین آیتم‌های متصل را استخراج می‌کند و امکان استخراج هدف اصلی یک توالی را ندارد. مدل‌های مرتبه بالا نیز با جود درک بهتر توالی‌های طولانی به دلیل افزایش وضعیت، پیچیده هستند. روش‌های سنتی بیشتر به اطلاعات محلی موجود در جلسه تکیه می‌کنند، یعنی تنها تعاملات کوتاه‌مدت و بلافاصله نزدیک در یک جلسه را در نظر می‌گیرند. این امر باعث می‌شود که از نظر محاسباتی کارآمد باشند و پیچیدگی کمی داشته باشند، اما به دلیل محدودیت به اطلاعات محلی، اطلاعات کلی جلسه را نادیده می‌گیرند. این محدودیت‌ها ممکن است دقت توصیه‌ها را در شناسایی الگوهای بلندمدت و پیچیده کاهش دهد.

۲-۲ روش‌های مبتنی بر یادگیری عمیق

در سال‌های اخیر، یادگیری عمیق با قابلیت استخراج ویژگی‌های عمیق، عملکرد سیستم‌های توصیه‌گر مبتنی بر جلسه را به‌طور گسترده‌ای بهبود بخشیده است [۵]. شبکه‌های عصبی بازگشتی توسط بسیاری از سیستم‌های توصیه‌گر برای مدل‌کردن رفتار متوالی کاربران استفاده شده‌اند. GRU4REC [۲۴] اولین کاری است که از شبکه‌های عصبی بازگشتی برای توصیه‌های مبتنی بر جلسه استفاده می‌کند و از یک واحد بازگشتی دروازه‌دار چندلایه برای مدل‌سازی وابستگی‌های ترتیبی بین آیتم‌ها استفاده می‌کند. در ادامه این کار، تان و همکاران با استفاده از تکنیک‌های افزایش داده و در نظر گرفتن تغییرات رفتار کاربران در طول زمان، مدل‌های مبتنی بر شبکه‌های عصبی بازگشتی را بهبود بخشیدند [۲۵]. NARM یک مکانیسم توجه را به یک رمزگذار ترکیبی اضافه می‌کند تا ترجیحات کاربر را در جلسه فعلی درک کند [۲۶]. روش STAMP [۲۷] از

یک شبکه عصبی پرسپترون چندلایه[1] و یک شبکه توجه استفاده می‌کند که علایق کلی کاربر را از حافظه بلندمدت یک جلسه و علایق فعلی کاربر را از حافظه کوتاه‌مدت آخرین کلیک کاربر به دست می‌آورد. Tag2Seq [۲۸] یک سیستم توصیه‌گر جلسه‌محور مبتنی بر LSTM است که با بهره‌گیری از بازنمایی‌های پیش‌آموزش‌دیده تگ‌ها اطلاعات معنایی و زمینه‌ای تگ‌ها را در کنار توالی کلیک‌ها مدل‌سازی کرده و دقت پیش‌بینی آیتم بعدی را افزایش می‌دهد. [۲۹] یک مدل جلسه محور مبتنی بر LSTM با لایه توجه چندمعیاره است که با استفاده از طرح ضریب همبستگی و انحراف معیار وزن هر معیار را در توالی رتبه‌بندی‌های چندمعیاره محاسبه می‌کند. این روش در دو حالت کار می‌کند: فقط با رتبه‌بندی‌های صریح چندمعیاره یا با متادیتای مجاز کاربر و دقت پیش‌بینی را در کنار حفظ حریم خصوصی بهبود می‌بخشد.

روش‌های مبتنی بر شبکه‌های عصبی گراف به دلیل توانایی قدرتمند آن‌ها در مدل‌سازی انتقال‌های پیچیده بین آیتم‌ها، توجه گسترده‌ای را به خود جلب کرده‌اند[۳۰]. وو و همکاران یک مدل شبکه عصبی گراف دروازه‌ای به نام سیستم توصیه‌های مبتنی بر جلسه با شبکه‌های عصبی گراف پیشنهاد کردند که ابتدا انتقال‌های پیچیده بین آیتم‌ها درون جلسات را درک می‌کند و سپس ترجیحات بلندمدت و فعلی کاربر را استخراج و ادغام می‌کند تا آیتم بعدی کاربر را پیش‌بینی کند[۳۱]. در روش دیگری از شبکه خودتوجه زمینه‌ای گراف برای توصیه‌های مبتنی بر جلسه استفاده شده است[۳۲]. شبکه عصبی کانولوشنی گراف چندرفتاری[2]، بر اساس رفتارهای چندگانه کاربران در طول خرید، با ساخت یک گراف یکپارچه برای نمایش داده‌های چندرفتاری، توصیه‌های مبتنی بر جلسه را انجام می‌دهد[۳۳]. شبکه‌های عصبی گراف شخصی‌سازی‌شده با مکانیزم توجه، شبکه عصبی گراف شخصی‌سازی شده و مکانیسم توجه ضرب داخلی را ترکیب می‌کند تا انتقال‌ها را در سطح آیتم و سطح جلسه برای توصیه‌های بهتر آیتم بعدی یاد بگیرد[۳۴]. توصیه‌گر مبتنی بر جلسه با گراف دینامیک[3] یک مدل شبکه عصبی گراف مبتنی بر توجه پویا برای توصیه‌های ترتیبی است. این مدل دنباله‌های کاربران مختلف را با یک گراف پویا ادغام می‌کند تا تعاملات بین کاربران و بین آیتم‌ها را بررسی کند. LESSR یک لایه تجمیع حفظ‌کننده ترتیب یال‌ها و یک لایه توجه گراف میان‌بر طراحی می‌کند تا مشکل از دست رفتن اطلاعات در روش‌های مبتنی بر شبکه‌های عصبی گراف را کاهش دهد[۳۵]. SASRec [۳۶]و SGRec [۳۷] از گراف برای ساخت بازنمایی آیتم استفاده می‌کنند. روش اول با درنظرگرفتن آیتم بعدی در سایر جلسات و روش دوم با درنظرگرفتن آیتم قبلی و بعدی در سایر جلسه‌ها بازنمایی مرتبط با آیتم را به دست می‌آورند. DHCN [۳۸] با ساخت یک هایپرگراف از کل جلسات برای مدل‌سازی بهتر روابط مرتبه بالا استفاده می‌کند. یو و همکاران [۳۹] یک مدل شبکه عصبی گراف را برای توصیه مبتنی بر جلسه ارائه کردند. آن‌ها علایق کاربر را با در نظر گرفتن آیتم هدف و همچنین انتقالات پیچیده بین آیتم‌ها در جلسه مدل می‌کنند. درنهایت از ترکیب ترجیحات فعلی و ترجیحات کلی هر جلسه برای نمایش جلسه و ایجاد توصیه استفاده می‌نمایند. TSGNN [۴۰] یک شبکه گرافی حساس به زمان برای توصیه آیتم‌های جدید در جلسه است که با ساخت گراف جلسه با وزن زمانی، تقویت گراف و مکانیزم توجه زمان‌آگاه برای تعادل تأثیر همه آیتم‌ها نه فقط آخرین آیتم عمل می‌کند. این روش با تطبیق علاقه‌مندی کاربر و بازنمایی آیتم‌های جدید، دقت توصیه را افزایش می‌دهد. مدل پیشنهادی در [۴۱] با افزودن مکانیزم وزن‌دهی تطبیقی به بردارهای خروجی گراف در مدل SR-GNN، اهمیت متغیر آیتم‌ها را در هر جلسه بر اساس اطلاعات جانبی تنظیم می‌کند. این روش با تعدیل پویای وزن‌ها، دقت پیش‌بینی را افزایش داده و به‌ویژه مشکل شروع سرد در سیستم‌های توصیه‌گر مبتنی بر جلسه را بهبود می‌بخشد.

در سال‌های اخیر، مدل‌های مبتنی بر ترنسفورمر[4] به دلیل قابلیت بالای خود در مدل‌سازی وابستگی‌های بلندمدت از طریق مکانیزم خودتوجه، برای توصیه آیتم بعدی در جلسه مورد استفاده قرار گرفته‌اند. Bert4Rec[۴۲] اولین روش مبتنی بر ترنسفورمر است که مشابه مدل زبانی Bert از یک رمزگذار دوطرفه[5] برای مدل‌سازی

[1] Multi-Layer Perceptron
[2] Multi-Behavior Graph Convolutional Network
[3] Dynamic Graph Session-based Recommendation
[4] Transformers
[5] Bidirectional encoder

توالی تعاملات کاربر با یادگیری دوطرفه استفاده می‌کند و با معرفی پیش‌بینی آیتم ماسک‌گذاری شده برای توصیه به‌جای پیش‌بینی گام بعدی، از زمینه کامل توالی استفاده می‌کند. BERT4Sess [۴۳] روشی مبتنی بر BERT4Rec است که برای داده‌های مبتنی بر جلسات کوتاه و بدون شناسه کاربر معرفی‌شده است که با ماسک‌گذاری سطح جلسه و یادگیری بازنمایی زمینه جلسه، بدون نیاز به اطلاعات مربوط به جلسات قبلی کاربر، هدف کاربر از جلسه فعلی را استخراج و برای پیش‌بینی استفاده می‌کند. روش معرفی‌شده در [۴۴] با ترکیب ترنسفورمر و شبکه‌های کانولوشنی به دنبال مدل‌سازی و یادگیری بهتر وابستگی‌های کوتاه مدت و بلندمدت و افزایش دقت توصیه است. این روش از ترتسفورمر برای کشف وابستگی‌های بلندمدت یا سراسری بین آیتم‌ها و از شبکه کانولوشنی برای کشف وابستگی‌های محلی استفاده می‌کند. روش TSESRec [۴۵] با هدف بهبود توصیه‌گر مبتنی بر جلسه، از معماری رمزگذار ترنسفور برای یادگیری روابط درون جلسه بین آیتم‌ها استفاده می‌کند، اما برخلاف مدل‌های ترتیبی، جلسه را به‌صورت یک مجموعه از آیتم‌ها در نظر می‌گیرد نه یک توالی خطی. این مدل مجموعه تلاش می‌کند آیتم‌هایی را که به مجموعه فعلی کاربر تعلق دارند ولی هنوز مشاهده نشده‌اند پیش‌بینی کند. برای این منظور، از مکانیزم خودتوجه جهت استخراج وابستگی‌های پیچیده و غیرترتیبی میان آیتم‌ها بهره می‌گیرد. در نتیجه می‌تواند نیت کاربر را بدون وابستگی به ترتیب زمانی دقیق تعاملات، با دقت و تعمیم‌پذیری بالاتری مدل کند.

جدول (۱): خلاصه‌ای از نقاط قوت و ضعف روش‌های توصیه مبتنی بر جلسه.

دسته کلی	روش	مراجع	نقاط قوت	نقاط ضعف
روش‌های سنتی	ماتریس عامل‌بندی	[۲۱] [٤]	سادگی، تفسیرپذیری بالا، امکان افزودن اطلاعات جانبی، کشف مؤثر هم‌رخدادی آیتم‌ها و تعاملات پنهان	عدم توجه به ترتیب آیتم‌ها، نیاز به حجم بالای اطلاعات، حساس به شروع سرد و پراکندگی داده
	نزدیک‌ترین همسایه	[٤١]	سرعت بالا در پیش‌محاسبه شباهت‌ها، تفسیرپذیری بالای توصیه‌ها، به‌روزرسانی ساده (اضافه/حذف آیتم‌ها بدون بازآموزی کامل)	دریافت اطلاعات محلی همسایگان مستقیم و ناتوانی در کشف الگوهای پیچیده یا چندمرحله‌ای، مقیاس‌پذیری ضعیف با افزایش تعداد آیتم‌ها($O(n^2)$) ، حساسیت بالا به نویز و مقیاس ناپذیری در داده‌های بزرگ، عدم تطبیق با تغییرات علاقه کاربر درون یک جلسه
	زنجیره مارکف	[۱۸] [۱۷] [۱۹] [۱۹]	مدل‌سازی صریح احتمالات انتقال بین آیتم‌ها، توانایی جذب وابستگی‌های کوتاه‌مدت، سادگی آموزش و استنتاج در جلسات منظم	ناتوانی در مدل‌سازی وابستگی‌های بلندمدت، انفجارحالت با افزایش مرتبه، حساسیت شدید به نویز و جلسات پراکنده، ضعف در مدل‌سازی جلسه‌های طولانی به دلیل نقض استقلال شرطی
روش‌های مبتنی بر یادگیری عمیق	شبکه‌های بازگشتی	[۲٦] [۲۵] [۲٤] [۲۹] [۲۸] [۲۷]	طراحی ذاتی برای تحلیل دنباله‌ها، مدل‌سازی مؤثر وابستگی‌های ترتیبی با طول متغیر، یادگیری وابستگی‌های بلندمدت، امکان گنجاندن ویژگی	مشکل محوشدن گرادیان در توالی‌های بسیار طولانی، تمرکز بیش از حد روی آیتم‌های اخیر، ناتوانی در مدل‌سازی روابط غیرخطی و چندمسیره بین آیتم‌های غیرمجاور
	شبکه خود توجه و ترنسفورمرها	[٤٢] [۳۵] [۳۲] [٤٥] [٤٤] [٤۳]	توجه پویا به تمام آیتم‌های جلسه بدون محدودیت طول، مدل‌سازی روابط غیرخطی و چندمرحله‌ای، یادگیری موازی و سرعت بالا در استنتاج، مقاومت در برابر نویز و جلسات نامنظم	نیاز به داده زیاد برای آموزش، پیچیدگی بالای محاسباتی، دشواری تفسیر، عملکرد ضعیف در جلسات بسیار کوتاه (کمتراز ۵ آیتم)
	گراف	[۳۳] [۳۱] [۳۸] [۳۵] [٤۱] [٤۰]	مدل‌سازی دقیق انتقال‌های پیچیده و چندمرحله‌ای، توانایی جذب الگوهای غیرخطی و حلقه‌ای در جلسه	پیچیدگی ساخت گراف، هزینه محاسباتی بالا به دلیل پیام‌رسانی گرافی، خطر بیش‌برازش در گراف‌ها (ازبین رفتن تمایز بین گره‌ها)

توانایی بالای ترنسفورمرها در مدل‌سازی وابستگی‌های بلندمدت میان آیتم‌ها، باعث توجه به آن‌ها در این حوزه شده است. زیرا برخلاف شبکه‌های عصبی بازگشتی برای پردازش دنباله‌ها به‌صورت ترتیبی عمل نمی‌کنند و در نتیجه از مشکل وابستگی طولانی‌مدت رنج نمی‌برند. ترنسفورمرها با استفاده از مکانیزم توجه می‌توانند تمام آیتم‌های موجود در جلسه را به‌صورت هم‌زمان در نظر بگیرند. این ویژگی باعث می‌شود که مدل بتواند روابط پیچیده‌تری را میان آیتم‌ها شناسایی کرده و توصیه‌های دقیق‌تری ارائه دهد. با این حال، باوجود عملکرد بهتر در داده‌های حجیم و پیچیده، قابلیت ترنسفورمرها در جلسات کوتاه یا زمانی که داده‌ها محدودند کاهش می‌یابد. مورد بعدی آن است که مدل‌های ترنسفورمری هم فقط توالی را می‌بینند و ساختار ارتباطی بین آیتم‌ها و روابط ضمنی بین آن‌ها را در نظر نمی‌گیرد.

همان‌گونه که تحلیل‌های این بخش بیان شد و در جدول ۱ نشان داده شده است، اکثر روش‌های موجود یا بر مدل‌سازی ترتیبی با شبکه‌های بازگشتی یا بر مدل‌سازی وابستگی‌های ساختاری و پیچیده با شبکه‌های گراف متمرکز هستند. ضعف اساسی این رویکردهای تک‌بعدی، نادیده گرفتن ماهیت دوگانه رفتار کاربر است که هم شامل روندهای ترتیبی و هم شامل پرش‌های غیرترتیبی و پیچیده است. روش پیشنهادی برای رفع همین شکاف طراحی شده است و با الهام از نقاط قوت هر دو خانواده، از یک رویکرد ترکیبی استفاده می‌کند که در آن وابستگی‌های ترتیبی و پیچیده توسط ماژول‌های مجزا تحلیل بازنمایی نهایی از ترکیب آن‌ها حاصل می‌شود.

۳- مدل پیشنهادی

در این بخش، یک توصیه‌گر مبتنی بر جلسه به نام RGNF-SRec[1] که به تحلیل وابستگی‌های ترتیبی و پیچیده با استفاده از شبکه عصبی بازگشتی و گراف می‌پردازد، تشریح می‌شود. هدف این سیستم، پیشنهاد آیتم بعدی با ترکیب شبکه‌های عصبی بازگشتی و گراف برای تحلیل جلسه و کشف وابستگی‌های ترتیبی و پیچیده در جلسه است. شکل ۱ معماری RGNF-SREC را نشان می‌دهد که از چهار مؤلفه اصلی تشکیل شده است: یادگیری وابستگی ترتیبی، یادگیری انتقال‌های پیچیده، تولید بازنمایی مرتبط با جلسه و محاسبه احتمال توصیه آیتم‌ها. در ادامه ابتدا مسئله توصیه مبتنی بر جلسه بیان می‌گردد و سپس معماری سیستم پیشنهادی با جزئیات توضیح داده می‌شود.

۳-۱ توصیه‌گر مبتنی بر جلسه

در یک سیستم توصیه‌گر مبتنی بر جلسه، $V= \{v1, v2, v3,....vm\}$ مجموعه آیتم‌های منحصر به فرد در تمام جلسات است. هر جلسه نیز با $S = [v_{s,1}, v_{s,2}, ...,v_{s,m}]$ که $v_{s,m} \in V$ نشان داده می‌شود و شامل کلیک‌های کاربر در یک جلسه است که به ترتیب زمانی مرتب شده‌اند. هدف سیستم توصیه‌گر مبتنی بر جلسه، پیش‌بینی آیتم بعدی یعنی $v_{s,m+1}$ برای این جلسه است. بر این اساس، یک سیستم توصیه‌گر مبتنی بر جلسه یک توزیع احتمال $\hat{y}$ برای تمام آیتم‌های کاندید تولید می‌کند که نشان‌دهنده احتمال این است که آیتم مربوطه v_j آیتم بعدی باشد. درنهایت، آیتم‌های کاندید با بالاترین K احتمال انتخاب می‌شوند تا لیست توصیه‌ها را تشکیل دهند و به کاربر ارائه شوند.

۳-۲ یادگیری وابستگی ترتیبی

این ماژول وابستگی‌های ترتیبی بین آیتم‌ها در جلسه فعلی را با استفاده از یک روش مبتنی بر شبکه‌های عصبی بازگشتی، یعنی واحد بازگشتی دروازه‌دار یاد می‌گیرد. این شبکه یک شبکه عصبی بازگشتی پیچیده‌تر است که هدف آن مقابله با مشکل گرادیان ناپدیدشونده است. ورودی‌های این شبکه آیتم‌های کلیک شده توسط کاربر و خروجی آن بازنمایی متناظر با جلسه است که هر دو بردارهای با ابعاد ۱۰۰ هستند. در این شبکه بازنمایی هر مرحله از درون‌یابی خطی بین فعال‌سازی مرحله قبل و فعال‌سازی کاندید طبق رابطه (۱) به دست می‌آید.

$$s_t = (1-z_t)\, s_{t-1} + z_t\, \hat{s}_t \qquad (۱)$$

که در آن z_t، r_t و $\hat{s}_t$ به ترتیب گیت به‌روزرسانی، گیت بازنشانی و تابع فعال‌سازی کاندید بوده و از روابط زیر محاسبه می‌شوند. W و U نیز ماتریس‌های وزن هستند.

[1] Recurrent and Graph Network Fusion for Session-Based Recommendation

$$(W_z\, x_t + U_z\, s_{t-1}) = z_t \quad (۲)$$

$$r_t = (W_r\, x_t + U_r\, s_{t-1}) \quad (۳)$$

$$\hat{s}_t = \tanh[W x_t + U(r_t \odot s_{t-1})] \quad (۴)$$

وضعیت پنهان نهایی شبکه به‌عنوان بازنمایی ترتیبی جلسه یا همان sf در نظر گرفته شده و در ایجاد بازنمایی نهایی شبکه استفاده می‌شود.

۳-۳ یادگیری انتقال آیتم

در این ماژول مشابه روش [۳۱] ابتدا گراف مرتبط با جلسه ساخته شده و با به دست آوردن بازنمایی‌های آیتم‌ها، بازنمایی جلسه به دست می‌آید. هر جلسه می‌تواند به‌صورت یک گراف جهت‌دار $G_s = (V_s, E_s)$ مدل‌سازی شود. در این گراف جلسه، هر گره نشان‌دهنده یک آیتم و هر یال به این معنی است که کاربر پس از $V_{s,i-1}$ روی آیتم $V_{s,i}$ در جلسه کلیک کرده است. از آنجایی که ممکن است چندین آیتم به‌طور مکرر در دنباله ظاهر شوند، به هر یال یک وزن نرمال‌شده اختصاص داده می‌شود که با تقسیم تعداد وقوع یال بر درجه خروجی گره شروع یال محاسبه می‌شود. سپس بازنمایی‌های متناظر با آیتم‌های جلسه از طریق انتشار اطلاعات در گراف به دست می‌آید. شبکه‌های عصبی گراف برای توصیه مبتنی بر جلسه بسیار مناسب هستند، زیرا می‌توانند به‌طور خودکار ویژگی‌های گراف‌های جلسه را با در نظر گرفتن اتصالات غنی گره‌ها استخراج کنند.

فرآیند یادگیری بازنمایی گره $v_{s,i}$ از گراف $\mathcal{G}_s$، با استفاده از توابع به‌روزرسانی به شرح زیر انجام می‌شود:

$$a_{s,i}^t = A_{s,i:}[v_1^{t-1}, \dots, v_n^{t-1}]^T H + b \quad (۵)$$

$$z_{s,i}^t = \sigma(W_z a_{s,i}^t + U_z v_i^{t-1}) \quad (۶)$$

$$r_{s,i}^t = \sigma(W_r a_{s,i}^t + U_r v_i^{t-1}) \quad (۷)$$

$$\tilde{v}_i^t = tanh(W_o a_{s,i}^t + U_o(r_{s,i}^t \odot v_i^{t-1})) \quad (۸)$$

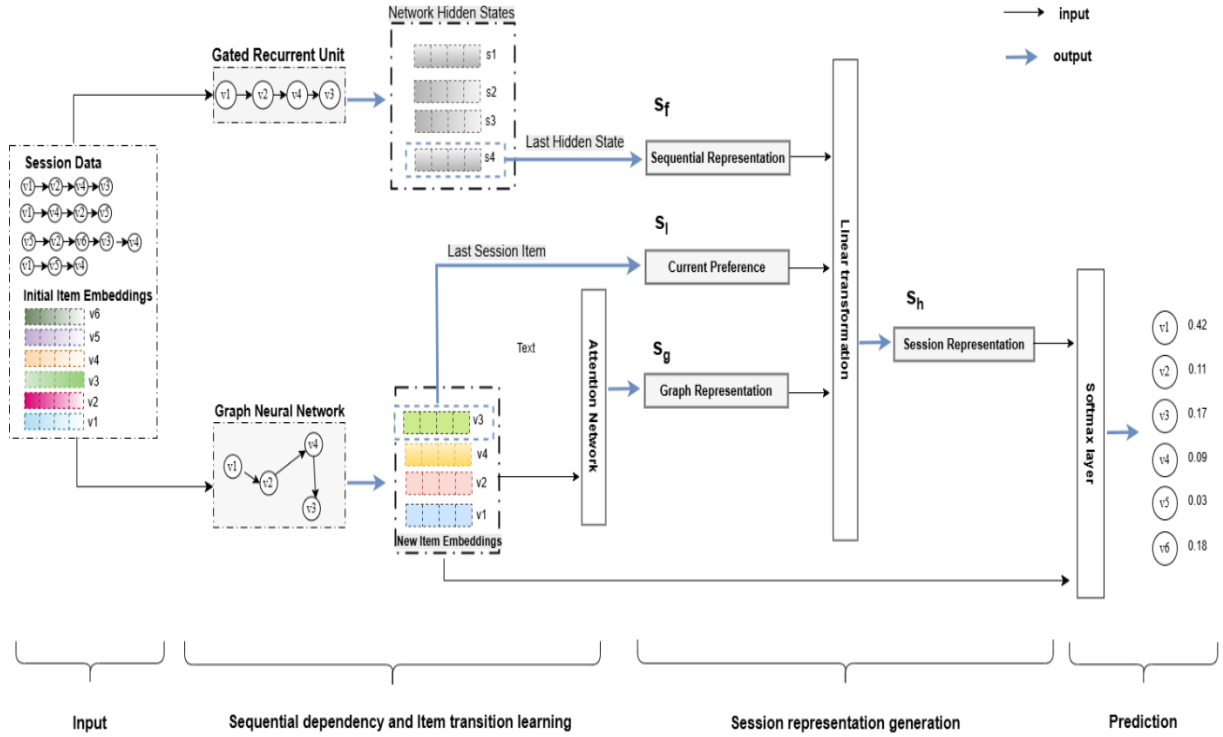
$$v_i^t = (1 - z_{s,i}^t) \odot v_i^{t-1} + z_{s,i}^t \odot \tilde{v}_i^t \quad (۹)$$

جایی که $H \in \mathbb{R}^{d \times 2d}$ وزن را کنترل می‌کند، $z_{s,i}$ و $r_{s,i}$ به ترتیب گیت‌های بازنشانی و به‌روزرسانی هستند، $[v_1^{t-1}, \dots, v_n^{t-1}]$ لیست بردارهای گره در جلسه s است، $\sigma(.)$ تابع سیگموئید و $\odot$ عملگر ضرب عنصر به عنصر است. $v_i \in \mathbb{R}^d$ نشان‌دهنده بردار پنهان گره $v_{s,i}$ است. ماتریس اتصال $A_s \in \mathbb{R}^{n \times 2n}$ تعیین می‌کند که چگونه گره‌ها در گراف با یکدیگر ارتباط برقرار می‌کنند و $A_{s,i:} \in \mathbb{R}^{1 \times 2n}$ دو ستون از بلوک‌های A_s مربوط به گره $v_{s,i}$ هستند.

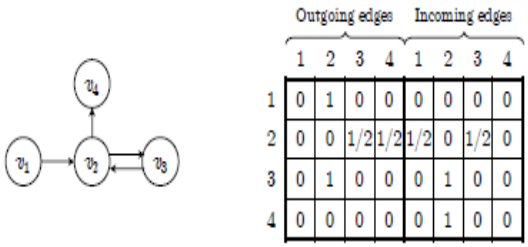
A_s به‌عنوان الحاق دو ماتریس مجاورت $A_s^{(out)}$ و $A_s^{(in)}$ تعریف می‌شود که به ترتیب نشان‌دهنده اتصالات وزن‌دار یال‌های خروجی و ورودی در گراف جلسه هستند. به‌عنوان مثال برای جلسه $s = [v_1, v_2, v_3, v_2, v_4]$، گراف $\mathcal{G}_s$ و ماتریس A_s مربوطه در شکل ۲ نشان داده شده‌اند.

برای هر گراف جلسه، شبکه عصبی گراف گیت‌دار گره‌ها را به‌صورت هم‌زمان پردازش می‌کند. در این فرآیند، اطلاعات از طریق همسایگی گره‌ها منتشر می‌شود، به‌طوری‌که بردارهای پنهان گره‌های مجاور استخراج و به‌عنوان ورودی به شبکه مورد استفاده قرار می‌گیرند. در ادامه، دو گیت به‌روزرسانی $(z_{s,i}^t)$ و گیت بازنشانی $(r_{s,i}^t)$ تعیین می‌کنند که کدام بخش از اطلاعات باید حفظ شود و کدام بخش باید کنار گذاشته شود. سپس، حالت کاندید $(\tilde{v}_i^t)$ با استفاده از ترکیبی از وضعیت قبلی، وضعیت فعلی و خروجی گیت‌ها محاسبه می‌گردد. درنهایت، حالت نهایی (v_i^t) هر گره با تلفیق مناسب حالت پنهان قبلی و حالت کاندید به‌دست می‌آید. پس از به‌روزرسانی گره‌ها در گراف جلسه و رسیدن به همگرایی، بردارهای نهایی گره‌ها قابل استخراج خواهند بود.

یک جلسه توسط گره‌های درگیر در آن نمایش داده می‌شود. پس از پردازش گراف جلسه توسط شبکه عصبی گراف گیت‌دار، بازنمایی تمام گره‌ها به دست می‌آید. با توجه به اهمیت آخرین آیتم کلیک شده توسط کاربر در جلسه، برای ایجاد بهتر بازنمایی جلسه، ابتدا بازنمایی آخرین آیتم جلسه را به عنوان بازنمایی محلی (s_l) در نظر می‌گیریم.



شکل (۱): معماری کلی مدل پیشنهادی RGNF-SRec. (هرجلسه و بازنمایی‌های اولیه به صورت موازی به دو ماژول تغذیه می‌شود: (الف) ماژول بازگشتی برای یادگیری وابستگی ترتیبی(s_f)، و (ب) ماژول گراف برای ساخت گراف جلسه و یادگیری انتقال‌های پیچیده(s_g). بازنمایی آخرین آیتم جلسه (s_l) با این دو بازنمایی ترکیب شده و بازنمایی‌های نهایی جلسه (s_h) ایجاد شده و برای پیش‌بینی آیتم بعدی استفاده می‌شود.)



شکل (۲): مثالی از گراف جلسه و ماتریس اتصال آن

برای مثال در جلسه $s = [v_{s,1}, v_{s,2}, \dots, v_{s,n}]$ بازنمایی محلی برابر است با: $s_l = v_n$.

سپس، بازنمایی کلی (s_g) گراف جلسه با تجمیع تمام بردارهای گره به دست می‌آید. با توجه به اینکه اطلاعات در این بازنمایی‌ها ممکن است سطوح اولویت متفاوتی داشته باشند، از مکانیزم توجه نرم برای نمایش بهتر بازنمایی کلی جلسه استفاده می‌کنیم. ورودی شبکه توجه، بازنمایی آیتم آخر به عنوان کوئری و بازنمایی سایر آیتم‌ها به عنوان کلید هستند، وزن توجه بین هر جفت از این بردارها و تجمیع آن‌ها به عنوان بازنمایی کلی توسط روابط ۱۰ و ۱۱ به دست می‌آید.

$$\alpha_i = q^T \sigma(W_1 v_n + W_2 v_i + c) \qquad (10)$$

$$s_g = \sum_{i=1}^{n} \alpha_i v_i \qquad (11)$$

جایی که پارامترهای $q \in \mathbb{R}^d$ و $W_1, W_2 \in \mathbb{R}^{d \times d}$ وزن بردارهای آیتم را کنترل می‌کنند.

۴-۳ تولید بازنمایی جلسه

پس از پردازش جلسه توسط شبکه بازگشتی دروازه‌ای و گراف، دو بازنمایی s_f و s_g تاکنون ایجاد شده‌اند. این دو بازنمایی به همراه بازنمایی مرتبط با آخرین آیتم کلیک شده توسط کاربر (s_l)، با یکدیگر ترکیب شده و بازنمایی نهایی جلسه را ایجاد می‌کنند.

$$s_h = W_3[s_l; s_g; s_f] \quad (۱۲)$$

که در آن ماتریس $W_3 \in \mathbb{R}^{d\times 2d}$ سه بردار را به فضای پنهان $\mathbb{R}^d$ فشرده می‌کند.

۳-۵ محاسبه احتمال توصیه آیتم‌ها

پس از به دست آوردن بازنمایی نهایی جلسه، احتمال کلیک شدن آیتم کاندید $v_i \in V$ پیش‌بینی می‌شود. برای این کار، بازنمایی آیتم v_i را در بازنمایی جلسه s_h ضرب می‌کنیم و سپس از تابع بیشینه هموار[1] برای به دست آوردن خروجی $\hat{y}_i$ استفاده می‌کنیم که برابر است با:

$$\hat{y}_i = softmax(s_h^T v_i) \quad (۱۳)$$

که در آن $\hat{y}_i \in \hat{y}$ نشان‌دهنده احتمال کلیک شدن آیتم v_i یا همان امتیاز توصیه است.

تابع زیان به عنوان آنتروپی متقاطع[2] نتایج پیش‌بینی $\hat{y}$ تعریف می‌شود:

$$\mathcal{L}(\hat{y}) = -\sum_{i=1}^{m} y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i) \quad (۱۴)$$

که y بردار کدگذاری one-hot متناظر با مقدار واقعی[3] است.

الگوریتم کامل در الگوریتم ۱ بیان شده است.

الگوریتم ۱: فرآیند آموزش و پیش‌بینی مدل RGNF-Srec

Input: S: ordered session lists, s ∈ S is given by s = [$v_{s,1}$,$v_{s,2}$,...,$v_{s,t-1}$]. Where $v_{s,i}$ ∈ V and V = v_1,v_2,...,v_m is the set of distinct items in all sessions and item_embeddings = {v_1 , v_2 , ..., v□}
Output: Recommendation list for users
for each session s do
feed session to GRU and learn sequential representation (S_f) using (1) - (4)
construct an incoming A_{in} and outgoing A_{out} adjacency matrices
build a directed session graph G_s from unique items and consecutive edges; A_{in} and A_{out}
compute node representation after T propagation steps using (5) - (9)
compute graph representation (S_g) by attention over node embeddings using (10) - (11)
compute current preference (S_l) from the last-clicked item in session (last node embedding).
compute the final session representation S_h using S_f, S_g and S_l as in (12)
compute the probability scores of each item$\hat{y}_i$ using softmax as in (13)
select the top-K as the recommended list of items for the session
compute loss and do backpropagation using (14)
end for

۳-۶ تحلیل پیچیدگی محاسباتی

پیچیدگی محاسباتی مدل RGNF-SRec از دو بخش اصلی تشکیل شده است: ماژول مبتنی بر شبکه بازگشتی و ماژول گراف. در ماژول اول پردازش هر جلسه شامل n گام برابر با تعداد آیتم‌های جلسه است. در هر گام پیچیدگی محاسبات گیت‌های بازنشانی و به‌روزرسانی و همچنین محاسبه حالت کاندید برابر با $O(d^2)$ است که d ابعاد بازنمایی است، بنابراین پیچیدگی این ماژول در هر جلسه $O(nd^2)$ خواهد بود. در ماژول گراف، ۳ مرحله یعنی ساخت گراف، انتشار اطلاعات بین گره‌ها و ساخت بازنمایی جلسه وجود دارد که پیچیدگی آن‌ها به ترتیب برابر با $O(E)$، $O(n.d)$ و $O(T.n.d^2)$ است که E تعداد یال‌های بین همه آیتم‌ها و T تعداد مراحل انتشار اطلاعات در گراف است. پیچیدگی محاسباتی مرحله پیش‌بینی نیز $O(v.d)$ است که v تعداد آیتم‌های منحصر به فرد در مجموعه داده است.

[1] Softmax
[2] Cross-entropy
[3] Ground Truth

٤- پیاده‌سازی و ارزیابی

در این بخش ابتدا مجموعه داده استفاده شده در پژوهش توصیف و سپس به بیان روش‌های مقایسه‌شده، معیارهای ارزیابی و تنظیم پارامترها پرداخته می‌شود. درنهایت نتایج مقایسه مدل پیشنهادی با روش‌های دیگر تجزیه و تحلیل می‌شود.

٤-١ مجموعه داده

در این پژوهش از دو مجموعه داده Yoochoose و Diginetica استفاده شده است. Yoochoose مربوط به کنفرانس RecSys سال ۲۰۱۵ است و شامل جریانی از کلیک‌های کاربران بر روی یک وب‌سایت تجارت الکترونیک در طی شش ماه است. Diginetica نیز از مسابقات CIKM Cup سال ۲۰۱٦ تهیه شده و شامل داده‌های تراکنشی است. پیش‌پردازش انجام شده روی مجموعه داده مشابه کارهای [۳۱] و [۳۲] انجام شده است. برای این کار، جلساتی با طول ۱ و آیتم‌هایی که کمتر از ۵ بار در مجموعه داده‌ها ظاهرشده‌اند، حذف می‌شوند. بر اساس روش پیش‌پردازش داده‌ها[۲٦]، برای یک جلسه $s = [v_{s,1}, v_{s,2}, \dots, v_{s,n}]$ در مجموعه داده‌ها، آن را به چندین دنباله تقسیم می‌کنیم و برچسب‌های متناظر را اختصاص می‌دهیم، یعنی:

$$([v_{s,1}], v_{s,2}), ([v_{s,1}, v_{s,2}], v_{s,3}), \dots, ([v_{s,1}, v_{s,2}, \dots, v_{s,n-1}], v_{s,n})$$

که در آن $[v_{s,1}, v_{s,2}, \dots, v_{s,n-1}]$ دنباله تولیدشده و $v_{s,n}$ برچسب متناظر است. پس از انجام پیش‌پردازش، مجموعه داده Yoochoose شامل ۷،۹۸۱،۵۸۰ جلسه و ۳۷،٤۸۳ آیتم منحصربه‌فرد، و مجموعه داده Diginetica شامل ۲۰٤،۷۷۱ جلسه و ٤۳،۰۹۷ آیتم منحصربه‌فرد است.

برای تقسیم‌بندی داده‌ها به مجموعه‌های آموزشی و آزمایشی، از رویکردهای استاندارد مورد استفاده در این حوزه، مشابه کارهای [۳۱، ٤۱] پیروی شد. در مجموعه داده Diginetica، جلسات مربوط به چند روز آخر به عنوان مجموعه آزمون و جلسات باقی‌مانده به عنوان مجموعه آموزش در نظر گرفته شدند. برای مجموعه داده Yoochoose، مطابق با چالش RecSys 2015، از تقسیم‌بندی رایج (Yoochoose 1/64) استفاده شد که در آن بخش کوچکی از داده‌های انتهایی (۱/٦٤) به عنوان مجموعه آزمون و مابقی به عنوان مجموعه آموزش استفاده می‌شوند. آمار دقیق این تقسیم‌بندی در جدول ۲ ارائه‌شده است.

جدول (۲): خلاصه اطلاعات مجموعه داده‌های استفاده شده

آمار	Diginetica	Yoochoose
تعداد کلیک	۹۸۲۹٦۱	۵۵۷۲٤۸
تعداد جلسه آموزش	۷۱۹٤۷۰	۳٦۹۸۵۹
تعداد جلسه تست	٦۰۸۵۸	۵۵۸۹۸
تعداد آیتم	٤۳۰۹۷	۱٦۷٦٦
میانگین طول جلسه	۵/۱۲	٦/۱٦

٤-۲ معیارهای ارزیابی

دو معیار ارزیابی زیر در بسیاری از سیستم‌های توصیه‌گر مبتنی بر جلسه استفاده شده است. برای مقایسه عملکرد مدل پیشنهادی با روش‌های مرتبط از این معیارها استفاده می‌کنیم:

۱) P@K (دقت): این معیار به عنوان یک معیار ارزیابی برای دقت پیش‌بینی به‌طور گسترده استفاده می‌شود. P@K به نسبت آیتم‌های صحیح در بین K آیتم توصیه‌شده برتر اشاره دارد.

$$P@K = \frac{n_{hit}}{N} \quad (۱۵)$$

که N تعداد جلسات مجموعه آزمون و nhit تعداد جلساتی را که شامل آیتم هدف در میان K تا آیتم بالا در لیست رتبه‌بندی هستند، نشان می‌دهد.

۲) MRR@K (میانگین رتبه متقابل): این معیار میانگین رتبه‌های متقابل آیتم‌هایی است که به درستی توصیه شده‌اند. اگر رتبه از K بیشتر شود، رتبه معکوس ۰ در نظر گرفته می‌شود. معیار MRR ترتیب رتبه‌بندی توصیه‌ها را در نظر می‌گیرد، به‌طوری‌که مقدار MRR بزرگ‌تر نشان‌دهنده این است که توصیه‌های صحیح در بالای لیست رتبه‌بندی قرار دارند.

$$MRR@K = \frac{1}{N} + \sum_{i=1}^{n} \frac{1}{rank\ i} \quad (۱٦)$$

که N تعداد جلسات مجموعه آزمون و $rank_i$ رتبه آیتم هدف را در میان K تا آیتم بالا در لیست رتبه‌بندی مربوط به i امین جلسه نشان می‌دهد.

۴-۳ جزئیات پیاده‌سازی و تنظیم پارامترها

مدل پیشنهادی با استفاده از کتابخانه PyTorch پیاده‌سازی شد. تنظیم پارامترها بر نحوه عملکرد سیستم و دقت آن تأثیر می‌گذارد. در این مقاله ابعاد بازنمایی آیتم‌ها و همچنین وضعیت‌های پنهان هر دو ماژول بازگشتی و گراف ۱۰۰ بعدی در نظر گرفته شد. تعداد لایه‌های انتشار ماژول گراف نیز برابر با ۱ تنظیم گردید. تمامی پارامترها با استفاده از توزیع گاوسی با میانگین ۰ و انحراف معیار ۰/۱ مقداردهی اولیه می‌شوند. بهینه‌ساز Adam با اندازه دسته ۱۰۰ برای بهینه‌سازی این پارامترها به کار گرفته می‌شود، که در آن نرخ یادگیری اولیه بر روی ۰/۰۰۱ تنظیم‌شده و پس از هر ۳ دوره آموزشی، به میزان ۰/۱ کاهش می‌یابد. علاوه بر این، اندازه دسته و جریمه L2 به ترتیب بر روی ۱۰۰ و ۱۰ به توان ۵- تنظیم‌شده‌اند. تمام آزمایش‌ها بر روی مشخصات سخت‌افزار GPU Tesla V100 اجرا گردید.

۴-۴ مدل‌های پایه

روش پیشنهادی با هشت مدل سنتی و پیشرفته مقایسه شده که به شرح زیر هستند:

۱) POP: این روش k آیتمی که بیشترین فراوانی را دارند به کاربران توصیه می‌کند

۲) Item-KNN [۴۶]: این روش شباهت بین آیتم‌های دیگر و آیتم فعلی را محاسبه می‌کند و سپس k آیتمی که بیشترین شباهت را به آیتم فعلی دارند به کاربران توصیه می‌کند

۳) FPMC [۲۱]: این روش بر ماتریس‌های انتقال شخصی‌سازی شده تکیه دارد و مزایای روش‌های فاکتورسازی ماتریس و روش‌های زنجیره مارکوف را ادغام می‌کند.

۴) GRU4Rec [۲۴]: اولین کار مبتنی بر شبکه‌های عصبی بازگشتی برای توصیه‌های مبتنی بر جلسه است و از یک واحد بازگشتی دروازه‌دار چندلایه برای مدل‌سازی وابستگی‌های ترتیبی آیتم‌ها استفاده می‌کند.

۵) NARM [۲۶]: این روش GRU4Rec را با استفاده از یک مکانیسم توجه بهبود می‌بخشد تا هدف اصلی کاربر و رفتار ترتیبی او را درک کند.

۶) STAMP [۲۷]: این روش ترجیح کلی کاربر در جلسه و ترجیح فعلی کاربر از آخرین آیتم کلیک‌شده را در نظر می‌گیرد تا توصیه نهایی را انجام دهد.

۷) SR-GNN [۳۱]: این روش از یک شبکه عصبی گراف دروازه‌ای برای درک انتقال‌های پیچیده آیتم‌ها استفاده می‌کند و سپس ترجیح بلندمدت و ترجیح فعلی کاربر را استخراج و ادغام می‌کند تا آیتم مورد علاقه بعدی کاربر را بهتر پیش‌بینی کند.

۸) SimGNN [۴۷]: از یک شبکه عصبی گراف ساده‌شده و یک مدل گراف غیرترتیبی و خطی برای نمایش جلسه استفاده می‌کند و از نمایش تعامل یادگیری شده توسط مدل غیرترتیبی و خطی پیشنهادی، ترجیحات محلی و ترجیحات کلی را مدل‌سازی و از یک مکانیسم دروازه‌بندی برای تجمیع آن‌ها استفاده می‌کند.

۹) Transformer – CNN [۴۴]: یک روش ترکیبی است که از ترنسفورمر برای کشف وابستگی‌های سراسری و از شبکه کانولوشنی برای کشف وابستگی‌های محلی استفاده می‌کند.

۱۰) TSESRec [۴۵]: یک روش مبتنی بر ترنسفورمرها است که از رمزگذار دوطرفه جهت یادگیری هدف کاربر از جلسه بدون درنظرگرفتن ترتیب آیتم‌ها استفاده می‌کند.

۴-۵ نتایج آزمایش‌ها

جدول ۳ نتایج آزمایشی مدل‌های پایه و مدل پیشنهادی را در دو مجموعه داده نشان می‌دهد، که بهترین نتیجه در هر ستون به صورت پررنگ مشخص‌شده است. طبق این نتایج، مدل پیشنهادی RGNF-SRec عملکرد بهتری نسبت به تمام مدل‌های پایه مقایسه‌شده دارد. با توجه به جدول ۳، مشاهدات زیر را داریم:

روش‌های سنتی شامل POP، Item-KNN و FPMC عملکرد ضعیف‌تری نسبت به سایر روش‌ها دارند. روش POP تنها با در نظر گرفتن ویژگی ساده محبوبیت محصول و بدون توجه به تعاملات کاربر خاص با آیتم‌ها، توصیه‌ها را ایجاد می‌کند و در بین روش‌ها بدترین عملکرد را داشته است. در بین روش‌های سنتی، Item-KNN بهترین عملکرد را در هر دو مجموعه داده نشان می‌دهد اما با توجه به اینکه ترتیب تعاملات جلسه را نادیده

می‌گیرد، عملکرد رضایت‌بخشی ندارد.FPMC با در نظر گرفتن اطلاعات جلسه با استفاده از زنجیره مارکف نسبت به دو روش قبلی نتایج بهتری به دست آورده اما به‌دلیل فرض مستقل‌بودن تعاملات متوالی کاربر در جلسه، دقت پایینی در ایجاد توصیه دارد. مدل‌های توصیه‌گر مبتنی بر یادگیری عمیق نسبت به روش‌های سنتی عملکرد بهتری دارند. این روش‌ها قادر به استخراج بهتر ویژگی از توالی‌های جلسه هستند. GRU4Rec با استفاده از شبکه‌های عصبی بازگشتی، نتیجه نسبتاً خوبی به دست آورده است. این روش متکی بر یادگیری اطلاعات متوالی جلسه است و علایق کلی کاربر در جلسه را نادیده می‌گیرد. NARM و STAMP مدل‌های مبتنی بر توجه هستند. NARM از یک مکانیسم توجه روی یک رمزگذار ترکیبی استفاده می‌کند که آخرین گام زمانی را به عنوان ترجیح اصلی کاربر در نظر می‌گیرد. STAMP از یک مکانیسم توجه برای استخراج ترجیح کوتاه‌مدت کاربر استفاده می‌کند. این دو روش برتری قابل‌توجهی نسبت به GRU4Rec نشان می‌دهند. دلیل این موضوع آن است که این روش‌ها از مکانیسم توجه برای محاسبه تأثیر آیتم‌های مختلف بر علاقه کاربر استفاده می‌کنند. SRGNN با نمایش توالی‌های جلسه به صورت داده ساختاری گراف، انتقالات پیچیده بین آیتم‌های جلسه را مدل می‌کند و عملکرد بهتری نسبت به روش‌های قبلی دارد. SimGNN نیز که یک رویکرد مبتنی بر گراف است توانسته با انجام تغییراتی در ساختار گراف، نتایج بهتری نسبت به SRGNN به دست آورد. در مورد مدل‌های مبتنی بر ترنسفورمر، مدل TSESRec نتایج ضعیف‌تری نسبت به مدل Transformer - CNN دارد که می‌تواند ناشی از ضعف در یادگیری وابستگی‌های کوتاه مدت یا مرتبط با طول جلسات کاربر باشد.

جدول (۳): مقایسه روش پیشنهادی با سایر روش‌ها

روش	YOOCHOOSE 1/64		DIGINETICA	
	MRR@20	P@20	MRR@20	P@20
POP	۱/٦٥	٦/٧١	۰/۲۰	۰/۸۹
Item-KNN	۲۱/۸۱	٥۱/٦۰	۱۱/٥۷	۳٥/۷٥
FPMC	۱٥/۰۱	٤٥/٦۲	٦/۹٥	۲٦/٥۳
GRU4Rec	۲۲/۸۹	٦۰/٦٤	۸/۳۳	۲۹/٤٥
NARM	۲۸/٦۳	٦۸/۳۲	۱٦/۱۷	٤۹/۷۰
STAMP	۲۹/٦۷	٦۸/۷٤	۱٤/۳۲	٤٥/٦٤
SR-GNN	۳۰/۹٤	۷۰/٥۷	۱۷/٥۹	٥۰/۷۳
SimGNN	۳۱/۳۷	۷۱/۲۲	۱۸/۲۳	٥۳/۱۳
Transformer - CNN	۳۰/٥٤	۷۱/۱۳	۱۷/۷۷	٥۲/٥۸
TSESRec	۳٤/٥۳	٦٥/۰۹	۱٤/۳۸	۳۸/۰۸
RGNF-SRec	۳۱/۸۲	۷۱/٦۲	۱۹/۳	٥۲/۸٥

طبق نتایج، مدل پیشنهادی در دو معیار ارزیابی بهترین عملکرد و در دو معیار ارزیابی دومین عملکرد را دارد و در مجموع از بهترین عملکرد را در بین مدل‌ها برخوردار است. دلیل این موضوع می‌تواند ساختار ترکیبی مدل باشد که مزایای مدل‌های مبتنی بر

شبکه‌های عصبی بازگشتی و شبکه‌های عصبی گراف را ترکیب و وابستگی‌های ترتیبی و انتقال‌های آیتم‌ها را به‌طور هم‌زمان یاد می‌گیرد. ضمن آن‌که با وجود نتایج بهتر نسبت به ترنسفورمرها، با محدودیت‌های مربوط به منابع محاسباتی و میزان داده‌های آموزشی نیز مواجه نخواهد بود.

٤-٦ تحلیل حساسیت و مطالعه حذف اجرا

به منظور درک عمیق‌تر پایداری[1] مدل پیشنهادی و ارزیابی سهم نسبی هر یک از مؤلفه‌های کلیدی آن، دو دسته آزمایش جامع طراحی و اجرا گردید. در بخش اول، حساسیت مدل به تغییرات ابرپارامترِ ابعاد بردار پنهان[2] سنجیده می‌شود. در بخش دوم، یک مطالعه حذف اجزا[3] برای بررسی نقش و اهمیت هر یک از سه مؤلفه اصلی معماری یعنی شبکه بازگشتی، گراف و آخرین آیتم کلیک شده انجام شده است.

٤-٦-١ تحلیل بر روی مجموعه داده Yoochoose

تحلیل حساسیت به ابعاد بردار پنهان

ابعاد بردار پنهان (d) یکی از حیاتی‌ترین ابرپارامترها است که مستقیماً بر مدل تأثیر می‌گذارد. ابعاد بسیار کوچک ممکن است منجر به کم‌برازش[4] شود، زیرا مدل قادر به یادگیری الگوهای پیچیده در داده نخواهد بود. در مقابل، ابعاد بسیار بزرگ می‌تواند منجر به بیش‌برازش[5]، افزایش شدید هزینه‌های محاسباتی و یادگیری نویز در داده‌های آموزشی شود. ما مدل RGNF-SRec را با ابعاد $d \in \{32, 64, 100, 128\}$ بر روی مجموعه داده Yoochoose (1/64) اجرا کردیم. (مقدار d=100 همان مقداری است که در بخش ٤-٣ به عنوان مقدار بهینه انتخاب شد). نتایج (میانگین ± انحراف معیار در ٥ اجرا) در جدول ٤ گزارش‌شده است.

نتایج جدول ٤ نشان می‌دهد که عملکرد مدل تحت تأثیر ابعاد پنهان بوده و مقدار بهینه انتخاب شده است.

جدول (٤): نتایج تحلیل حساسیت به ابعاد d بر روی Yoochoose

D	MRR@20 (MEAN ± STD)	P@20 (MEAN ± STD)
٣٢	٠/٣٨ ± ٢٩/٥٤	٠/٤٥ ± ٦٨/١١
٦٤	٠/٢٥ ± ٣١/١٢	٠/٣١ ± ٧٠/٩٣
١٠٠	٠/٢١ ± ٣١/٨٢	٠/٢٨ ± ٧١/٦٢
١٢٨	٠/٢٣ ± ٣١/٦٩	٠/٣٠ ± ٧١/٤٠

مطالعه حذف اجزا بر روی مجموعه داده Yoochoose

برای سنجش سهم دقیق هر یک از سه مؤلفه اصلی مدل (GNN, GRU, Last-Item)، چهار مدل را مقایسه کردیم:

١) **RGNF-SRec (Full)**: مدل کامل پیشنهادی (GNN + GRU + Last-Item).

٢) **Model (w/o GNN)**: مدل کامل بدون مؤلفه GNN (شبکه عصبی گراف). در این حالت، بازنمایی جلسه فقط از خروجی GRU و Last-Item (s_f و s_l) تشکیل می‌شود.

٣) **Model (w/o GRU)**: مدل کامل بدون مؤلفه GRU (شبکه عصبی بازگشتی). بازنمایی جلسه از خروجی GNN و Last-Item (s_g و s_l) تشکیل می‌شود.

٤) **Model (w/o Last)**: مدل کامل بدون استفاده از بازنمایی آخرین آیتم (s_l). بازنمایی نهایی فقط ترکیبی از s_f و s_g است.

جدول (٥): نتایج مطالعه حذف اجزا بر روی Yoochoose

مدل	اجزا	MRR@20 (MEAN ± STD)	P@20 (MEAN ± STD)
FULL	GNN + GRU + Last	٠/٢١ ± ٣١/٨٢	٠/٢٨ ± ٧١/٦٢
W/O GNN	GRU + Last	٠/٣٥ ± ٢٩/١٥	٠/٤٠ ± ٦٨/٥١
W/O GRU	GNN + Last	٠/٢٩ ± ٣٠/٢٢	٠/٣٤ ± ٦٩/٨٠
W/O LAST	GNN + GRU	٠/٢٤ ± ٣٠/٩٤	٠/٣١ ± ٧٠/٧٣

نتایج جدول ٥ نشان می‌دهد که:

١) **اهمیت گراف**: حذف مؤلفه گراف منجر به بیشترین افت عملکرد در هر دو معیار (افت نسبی ٨,٤٪ در MRR@20) می‌شود. این یافته به‌شدت اهمیت مدل‌سازی انتقال‌های

[1] Robustness
[2] Latent Dimension
[3] Ablation Study
[4] Underfitting
[5] Overfitting

پیچیده و غیرترتیبی بین آیتم‌ها را تأیید می‌کند. جلسات کاربران صرفاً یک دنباله خطی نیستند، بلکه شامل پرش‌های ساختاری (مثلاً بازگشت به آیتمی مشابه آیتم اول) هستند که شبکه بازگشتی به تنهایی قادر به درک آن‌ها نیست.

۲) **اهمیت شبکه بازگشتی**: حذف مؤلفه بازگشتی در رتبه دوم از نظر اهمیت قرار دارد (افت نسبی ۵٫۰٪ در MRR@20). این نشان می‌دهد که وابستگی‌های ترتیبی نیز نقش مهمی در تحلیل جلسه دارند. مدل گراف به تنهایی (w/o GRU) نمی‌تواند ترتیب دقیق کلیک‌ها و پویایی زمانی جلسه را به‌خوبی مدل کند.

۳) **اهمیت آخرین آیتم**: حذف بازنمایی آخرین آیتم که برای درک قصد کوتاه‌مدت استفاده می‌شود با وجود اینکه کمترین افت را ایجاد می‌کند (افت نسبی ۲٫۸٪ در MRR@20)، اما بر بهبود عملکرد مدل تأثیر دارد.

٤-٦-٢ تحلیل بر روی مجموعه داده Diginetica

تحلیل حساسیت به ابعاد بردار پنهان

همانند Yoochoose، آزمایش‌ها برای ابعاد $d \in \{32, 64, 100, 128\}$ بر روی مجموعه داده Diginetica تکرار شد. جدول ٦ شامل نتایج این تحلیل است و نشان می‌دهد در این مجموعه داده نیز بهترین ابعاد d انتخاب شده است.

جدول (٦): نتایج تحلیل حساسیت به ابعادd بر روی Diginetica

D	MRR@20 (MEAN ± STD)	P@20 (MEAN ± STD)
۳۲	۰/٤٥ ± ۱۷/۲۲	۰/۵۱ ± ۵۰/۱٤
٦٤	۰/۳۰ ± ۱۸/۹۱	۰/۳۹ ± ۵۲/۲۳
۱۰۰	۰/۲٤ ± ۱۹/۳۰	۰/۳۳ ± ۵۲/۸۵
۱۲۸	۰/۲۹ ± ۱۹/۱٤	۰/۳۷ ± ۵۲/٦۰

مطالعه حذف اجزا بر روی Diginetica

چهار مدل (Full, w/o GNN, w/o GRU, w/o Last)مشابه مجموعه داده قبلی، بر روی مجموعه دادهDiginetica نیز ارزیابی‌شده است. نتایج مربوط به حذف و نیز افت نسبی هریک از مدل‌ها به ترتیب در جدول ۷ آمده است. نتایج روی این مجموعه داده نیز تائید کننده اهمیت بخش‌های مدل پیشنهادی است که با حذف هرکدام از آن‌ها، مدل دچار کاهش در معیارهای ارزیابی‌شده است.

جدول (۷): نتایج مطالعه حذف اجزا بر روی Diginetica

مدل	اجزا	MRR@20 (MEAN ± STD)	P@20 (MEAN ± STD)
FULL	GNN + GRU + Last	۰/۲٤ ± ۱۹/۳۰	۰/۳۳ ± ۵۲/۸۵
W/O GNN	GRU + Last	۰/٤۱ ± ۱۷/۰۸	۰/٤۹ ± ٤۹/۹۲
W/O GRU	GNN + Last	۰/۳۵ ± ۱۷/۹۳	۰/٤۲ ± ۵۱/۱۰
W/O LAST	GNN + GRU	۰/۲۹ ± ۱۸/۵۵	۰/۳۸ ± ۵۲/۰٤

این جدول نتایج یافته‌های قبلی را تأیید و حتی تقویت می‌کند که:

۱) **اهمیت گراف**: حذف مؤلفه **گراف** منجر به افت عملکرد **بیشتری** (افت نسبی ۱۱٫۵٪ درMRR@20) نسبت به مجموعه داده قبلی شده است. با توجه ماهیت جلسات در Diginetica (که شامل داده‌های تراکنشی است) احتمالاً روابط غیرخطی و پیچیده‌تری وجود دارد که مدل‌های ترتیبی قادر به درک آن‌ها نیستند.

۲) **اهمیت شبکه بازگشتی و آخرین کلیک**: همانند قبل، GRU (افت ۷٫۱٪) و Last-Item (افت ۳٫۹٪) به ترتیب در رتبه‌های بعدی اهمیت قرار دارند.

۵- نتیجه‌گیری و کارهای آتی

در این مقاله، مدل RGNF-SRec برای توصیه‌های مبتنی بر جلسه معرفی گردید که با مدل‌سازی هم‌زمان وابستگی‌های ترتیبی با شبکه بازگشتی و وابستگی‌های پیچیده آیتم‌ها با گراف، به درک عمیق‌تری از رفتار کاربر دست می‌یابد. استفاده از این رویکرد ترکیبی، توانایی یادگیری بازنمایی جلسه را بهبود بخشیده و نتایج آزمایش‌ها، برتری آن را نسبت به روش‌هایی که تنها بر یک جنبه متمرکز هستند، نشان داده است. از منظر کاربردهای عملی، مدل RGNF-SRec می‌تواند به‌طور مستقیم در سیستم‌های تجارت الکترونیک و سرویس‌های محتوایی برای بهبود دقت توصیه‌های «آیتم بعدی» در طول یک جلسه فعال کاربر، مورد استفاده قرار گیرد.

با وجود نتایج مطلوب، این پژوهش دارای محدودیت‌هایی نیز است. اولاً، مدل پیشنهادی به دلیل ماهیت ترکیبی و نیاز به پردازش

موازی، ممکن است در مقایسه با مدل‌های مبتنی بر یک جزء، با پیچیدگی محاسباتی بیشتری در فاز آموزش مواجه باشد. ثانیاً، عملکرد مدل می‌تواند تحت تأثیر نویز موجود در جلسات (مانند کلیک‌های تصادفی یا غیرمرتبط کاربر) قرار گیرد

برای کارهای آتی، چند مسیر مشخص پیشنهاد می‌شود:

توسعه مکانیزم‌های حذف نویز پویا: طراحی مکانیزم‌های توجه یا فیلترینگ برای شناسایی و کاهش وزن تعاملات غیرمرتبط در جلسه.

ادغام اطلاعات جانبی: غنی‌سازی بازنمایی آیتم‌ها با استفاده از ویژگی‌های محتوایی (مانند دسته، قیمت، یا توضیحات آیتم).

بهینه‌سازی محاسباتی: بررسی روش‌های کاهش پیچیدگی مدل ترکیبی (مانند اشتراک وزن) برای تسریع آموزش.

References

[1] P.N. Vicente and C.D. Burnay, "Recommender Systems and Over-the-Top Services: A Systematic Review Study (2010–2022)," Journal. Media, vol. 5, no. 3, 2024.

[2] Y. Li, K. Liu, R. Satapathy, S. Wang and E. Cambria, "Recent Developments in Recommender Systems: A Survey," in arxiv preprint, arXiv:2306.12680, 2023.

[3] A. Valencia-Arias, H. Uribe-Bedoya, J. González-Ruiz, G. Santos, E. Ramírez and E. Rojas, "Artificial intelligence and recommender systems in e-commerce. Trends and research agenda," Intelligent Systems with Applications, vol. 24, 2024.

[4] E. Elahi et al.,"Knowledge graph enhanced contextualized attention-based network for responsible user-specific recommendation," ACM Trans. Intell. Syst. Technol, vol. 15, no. 4, 2024.

[5] A. Ramazani and A. Zareh, "Provide a Personalized Session-Based Recommender System with Self-Attention Networks," (in Fa), Iranian Journal of Electrical and Computer Engineering, vol. 20, no. 3, pp. 236-244, 2022.

[6] A. Peintner, “Sequential Recommendation Models: A Graph-based Perspective,” in Proceedings of the 17th ACM Conference on Recommender Systems (RecSys '23), Sept. 2023, pp. 1295–1299.

[7] J. H. Yoon and B. Jang, “Evolution of Deep Learning-Based Sequential Recommender Systems: From Current Trends to New Perspectives,” IEEE Access, vol. 11, pp. 54265–54279, Jan. 2023.

[8] G. Shani, D. Heckerman, and R. I. Brafman, "An MDP-based recommender system," Journal of machine Learning research, vol. 6, no. Sep, pp. 1265-1295, 2005.

[9] S. Rendle, C. Freudenthaler and L. Schmidt-Th, “Factorizing Personalized Markov Chains for Next-Basket Recommendation,” in Proceedings of the 19th international conference on World wide web (WWW '10), Apr. 2010, pp 811 – 820.

[10] M. Ruocco and H. Langseth, “Inter-Session Modeling for Session-Based Recommendation,” in Proceedings of the 2nd Workshop on Deep Learning for Recommender Systems (DLRS), Aug. 2017, pp. 24-31.

[11] M. Quadrana, A. Karatzoglou, B. Hidasi, and P. Cremonesi, “Personalizing Session-based Recommendations with Hierarchical Recurrent Neural Networks,” in Proceedings of the 11th ACM Conference on Recommender Systems (RecSys 17), Aug. 2017, pp. 130-137.

[12] T. Donkers, B. Loepp and A. Ziegler, “A Variational Recurrent Neural Network for Session-Based Recommendations using Bayesian Personalized Ranking,” in Proceedings of the Eleventh ACM Conference on Recommender Systems (RecSys 17), Aug. 2017, pp. 152-160.

[13] S. Wang, L. Hu, Y. Wang, Q. Z. Sheng, M. Orgun, and L. Cao, "Intention2basket: A neural intention-driven approach for dynamic next-basket planning," in 29th International Joint Conference on Artificial Intelligence and Seventeenth Pacific Rim International Conference on Artificial Intelligence (IJCAI-PRICAI-20), 2020.

[14] Z. Wang, W. Wei, G. Cong, X. Li, X.-L. Mao, and M. Qiu, “Global Context Enhanced Graph Neural Networks for Session-based Recommendation,” in Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '20), Jul. 2020, pp. 169-178.

[15] S. Wang, L. Hu, Y. Wang, L. Cao, Q. Z. Sheng, and M. Orgun, "Sequential recommender systems: challenges, progress and prospects," arXiv preprint arXiv:2001.04830, 2019.

[16] S. Wang, L. Cao, Y. Wang, Q. Z. Sheng, M. A. Orgun, and D. Lian, “A Survey on Session-based Recommender Systems,” ACM Computing Surveys, vol. 54, no. 7, pp. 1–38, July. 2021.

[17] S. Yakhchi, “Learning Complex Users’ Preferences for Recommender Systems,” arXiv preprint arXiv:2107.01529, Jul. 2021.
[18] L. Cao, Y. Ou and P. S. Yu, “Coupled Behavior Analysis with Applications,” IEEE Transactions on Knowledge and Data Engineering, vol. 24, no. 8, pp. 1378-1392, Aug. 2012.
[19] I. V. Cadez, D. Heckerman, C. Meek and P. Smyth, “Visualization of Navigation Patterns on a Web Site using Model-Based Clustering,” in Proceedings of the 6th ACM SIGKDD international conference on Knowledge discovery and data mining, Aug. 2000, pp. 280-284.
[20] M. Eirinaki, M. Vazirgiannis and D. Kapogiann, “Web Path Recommendations Based on Page Ranking and Markov Models,” in Proceedings of the 7th ACM international workshop on Web information and data management, Nov. 2005, pp. 2-95.
[21] S. Rendle, C. Freudenthaler and L. Schmidt-Th, “Factorizing Personalized Markov Chains for Next-Basket Recommendation,” in Proceedings of the 19th international conference on World wide web (WWW '10), Apr. 2010, pp 811 – 820.
[22] Z. Zhang and O. Nasraoui, “Efficient Hybrid Web Recommendations Based on Markov Clickstream Models and Implicit Search,” in IEEE/WIC/ACM International Conference on Web Intelligence (WI'07), Nov. 2007, pp. 621-627.
[23] M. Nasir and C. I. Ezeife, “A Survey and Taxonomy of Sequential Recommender Systems for E-commerce Product Recommendation,” SN computer science, vol. 4, no. 6, Sept. 2023.
[24] B. Hidas, A. Karatzoglou, L. Baltrunas, and D. Tikk, "Session-based recommendations with recurrent neural networks," in International Confer-ence on Learning Representations, San Juan, 2016.
[25] Y. Tan, X. Xu and Y. Liu, “Improved Recurrent Neural Networks for Session-based Recommendations,” in 1st Workshop on Deep Learning for Recommender Systems, Sept. 2016.
[26] J. Li, P. Ren, Z. Chen, Z. Ren, T. Lian, and J. Ma, "Neural attentive session-based recommendation," in ACM on Conference on Information and Knowledge Management, 2017.
[27] Q. Liu, Y. Zeng, R. Mokhosi and H. Zhang, “Stamp: Short-Term Attention/Memory Priority Model for Session-Based Recommendaion,” in Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '18), Jul. 2018, pp. 1831-1839.
[28] Y. Bougteb, E. Akachar, B. Ouhbi and B. Frikh, " Tag2Seq: Enhancing Session-Based Recommender Systems with Tag-Based LSTM" in Proceedings of International Conference on Information Integration and Web Intelligence, 2023.
[29] Y. Bougteb, B. Ouhbi, B. Frikh and E. M. Zemmouri,"A multi-criteria attention-LSTM approach for enhancing privacy and accuracy in recommender systems", Social Network Analysis and Mining, vol. 38, 2025.
[30] S. Wang et al., “Graph Learning based Recommender Systems: A Review,” in Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (IJCAI-210), Aug. 2021, pp. 4644-4652.
[31] S. Wu, Y. Tang, Y. Zhu, L. Wang, X. Xie, and T. Tan, “Session-Based Recommendation with Graph Neural Networks,” in AAAI Conference on Artificial Intelligence, Jul. 2019, pp. 346–353.
[32] C. Xu et al., “Graph Contextualized Self-Attention Network for Session-based Recommendation,” in Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI 19), Aug. 2019, pp. 3940-3946.
[33] B. Jin, C. Gao, X. He, D. Jin, and Y. Li, "Multi-behavior recommendation with graph convolutional networks," in Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval, 2020, pp. 659-668.
[34] M. Zhang, S. Wu, M. Gao, X. Jiang, K. Xu, and L. Wang, "Personalized graph neural networks with attention mechanism for session-aware recommendation," IEEE Transactions on Knowledge and Data Engineering, vol. 34, no. 8, pp. 3946-3957, 2020.
[35] T. Chen and R. C.-W. Wong, "Handling information loss of graph neural networks for session-based recommendation," in Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining, 2020, pp. 1172-1180.
[36] T. Zhu, L. Sun and G. Chen, “Graph-based Embedding Smoothing for Sequential Recommendation,” IEEE Transactions on Knowledge and Data Engineering, vol. 35, no. 1, pp. 496-508, Apr. 2021.
[37] Y. Li, T. Chen, Y. Luo, H. Yin and Z. Huang, “Discovering Collaborative Signals for Next POI Recommendation with Iterative Seq2Graph Augmentation,” in 13th International Joint

Conference on Artificial Intelligence (IJCAI-21), Aug. 2021.
[38] X. Xia, H. Yin, J. Yu, Q. Wang, L. Cui and X. Zhang, “Self-Supervised Hypergraph Convolutional Networks for Session-based Recommendation,” in AAAI Conference on Artificial Intelligence, Feb. 2021, pp. 4503-4511.
[39] F. Yu, Y. Zhu, Q. Liu, S. Wu, L. Wang and T. Tan, “TAGNN: Target Attentive Graph Neural Networks for Session-Based Recommendation,” in Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '20), Jul. 2020, pp. 1921-1924.
[40] L. Wang and D. Jin, "A Time-Sensitive Graph Neural Network for Session-Based New Item Recommendation," Electronics, vol. 13, 2024.
[41] B. Ozbay, R. Tugay and S. G. Oguducu, " A GNN Model with Adaptive Weights for Session-Based Recommendation Systems", in Proceedings of the 9th International Conference on Machine Learning Technologies,2024.
[42] F. Sun etal., "BERT4Rec: Sequential Recommendation with Bidirectional Encoder Representations from Transformer", in Proceedings of the 28th ACM International Conference on Information and Knowledge Management, 2019, pp 1441 - 1450.
[43] J. Seol, Y. Ko and S. Lee," Exploiting Session Information in BERT-based Session-aware Sequential Recommendation," in Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2022. pp. 2639-2644.
[44] J. Wang, H. Xi, F. L. Wang and L. K. Lee, "A transformer–convolution model for enhanced session-based recommendation," Neurocomputing, vol. 531, pp. 21-33, 2023.
[45] C. Liu, T. Yu, X. Zhou, L. Zhou and X. Gong, "TSESRec: A transformer-facilitated set extension model for session-based recommendation," Supercomputing, vol. 81, no.1, 2024.
[46] G. Sarwar, J. Konstan and J. Riedl, "Item-based collaborative filtering recommendation algorithms," in Proceedings of WWW, 2001, pp. 285-295.
[47] T.R. Gwadabe, M.A.M. Al-hababi, and Y. Liu, "SimGNN: simplified graph neural networks for session-based recommendation," Applied Intelligence, vol. 53, no. 19, pp. 22789-22802, 2023.

Journal of Applied And Basic Machine Intelligence Research (ABMIR)

Applied and Basic Machine Intelligence Research

Vol. 3, No. 1, Spring & Summer 2025, pp. 193 - 211

Session-based recommender system integrating analysis of sequential and complex dependencies using recurrent neural networks and graphs

Meysam Mirzaee[1*], Hossein Mohammadi[2], Sajad Ahmadian [3]

[1] PhD Student, Faculty of Computer, Zanjan University, Zanjan, Iran

[2] Assistant Professor, Faculty of Computer, Zanjan University, Zanjan, Iran

[3] Assistant Professor, Faculty of Information Technology, Kermanshah University of Technology, Kermanshah, Iran

Article Information

Original Research Paper

Received:
2025 September 27

Accepted:
2025 November 15

Keywords:
Session-based recommender systems, Graph, Recurrent Neural Networks

Corresponding Author*:
m.mirzaei@znu.ac.ir

Abstract

Session-based recommender systems are a type of recommender systems that model the dependencies between items in user interactions to find the most appropriate item or set of items to the user. An important issue in the performance of these systems is the ability to discover different dependencies in a session. While existing models often focus on only one of the two dependency types—either sequential or complex—this study proposes a novel model that integrates both. Specifically, a Graph Neural Network (GNN) is employed to identify complex and non-sequential dependencies among items, while a Recurrent Neural Network (RNN) is utilized to model the sequential dependencies within sessions. This hybrid architecture enables the model to obtain a more accurate representation of user intent, thereby achieving higher prediction accuracy compared to conventional approaches that address only one dependency type. Comprehensive evaluations conducted on two widely used benchmark datasets, Yoochoose and Diginetica, demonstrate the superiority of the proposed method over state-of-the-art baselines. The results show that RGNF-SRec achieves improvements of 0.40% in P@20 and 1.07% in MRR@20 compared to the strongest baseline model on these datasets, highlighting its enhanced capability in providing more accurate next-item recommendations.

: 10.22034/Abmir.2025.23730.1172

E-ISSN: 2821-2037 /The Author 2025. Published by Yazd University This is an open access article under the CC BY 4.0 License (https://creativecommons.org/licenses/by/4.0/).