*مقاله پژوهشی*

# ارائه سیستم توصیه‌گر مبتنی بر جلسه شخصی‌سازی شده با استفاده از شبکه‌های خودتوجه

اعظم رمضانی و علی محمد زارع بیدکی

**چکیده: سیستم‌های توصیه‌گر مبتنی بر جلسه بر اساس رفتار و تعاملات کاربر در یک جلسه، رفتار بعدی یا علاقه کاربر را پیش‌بینی کرده و بر این اساس، آیتم‌های مناسب را به کاربر پیشنهاد می‌دهند. مطالعات اخیر برای ایجاد توصیه‌ها عمدتاً روی اطلاعات جلسه فعلی متمرکز شده‌اند و اطلاعات جلسات قبلی کاربر را نادیده می‌گیرند. در این مقاله، یک مدل توصیه‌گر مبتنی بر جلسه شخصی‌سازی شده با شبکه‌های خودتوجه پیشنهاد می‌شود که علاوه بر جلسه فعلی از جلسات قبلی اخیر کاربر هم استفاده می‌کند. مدل پیشنهادی به منظور یادگیری وابستگی کلی بین همه آیتم‌های جلسه، از شبکه‌های خودتوجه (SAN) استفاده می‌کند. ابتدا SAN مبتنی بر جلسات ناشناس آموزش داده می‌شود و سپس برای هر کاربر، توالی‌های جلسه فعلی و جلسات قبلی به صورت جداگانه به شبکه داده می‌شود و با ترکیب وزنی نتایج رتبه‌بندی حاصل از هر جلسه، آیتم‌های توصیه‌شده نهایی به دست می‌آید. مدل پیشنهادی بر روی مجموعه داده واقعی Reddit در دو معیار دقت و میانگین رتبه متقابل، تست و ارزیابی شده است. مقایسه نتایج حاصل از مدل پیشنهادی با رویکردهای قبلی، توانایی و اثربخشی مدل پیشنهادی را در ارائه توصیه‌های دقیق‌تر نشان می‌دهد.**

***کلیدواژه:* توصیه‌گر شخصی‌سازی شده، توصیه‌گر مبتنی بر جلسه، شبکه‌های خودتوجه، یادگیری عمیق.**

## ۱- مقدمه

با توجه به افزایش روزافزون حجم اطلاعات در اینترنت و فضای وب، جستجو و پیداکردن اطلاعات و یا اقلام مورد نظر برای کاربران به یک چالش عمده تبدیل شده است. امروزه در اکثر سرویس‌های برخط وب و سایت‌های تجارت الکترونیک، سیستم‌های توصیه‌گر بر اساس رفتار کنونی و تاریخچه کاربران، خواسته‌ها و نیازهای کاربران را پیش‌بینی کرده و با فراهم‌کردن پیشنهاد و توصیه‌های متناسب با ترجیحات و علایق کاربران، نقش بسیار مهمی در کمک به کاربران ایفا می‌کنند. در نتیجه چنین سیستم‌هایی موجب صرفه‌جویی زمان جستجوی کاربران و افزایش رضایتمندی و وفاداری آنها می‌شوند.

سیستم‌های توصیه‌گر کلاسیک از روش‌های پالایش مشارکتی[1] [۱] و یا ماتریس فاکتوربندی (MF)[2] استفاده کرده‌اند. این سیستم‌ها متکی بر رتبه‌بندی 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[3] مواجه می‌شوند. همچنین به دلیل کمبود داده‌ها و خلوت‌بودن[4] تعاملات کاربر- آیتم قادر به ارائه توصیه‌های دقیق و مطمئنی به کاربران نیستند. در حقیقت این روش‌ها برای پیش‌بینی علایق کاربران از ویژگی‌های کم‌عمق[5] استفاده کرده‌اند و اطلاعات مفید توالی جلسات را که می‌تواند به پیش‌بینی آیتم هدف کمک کند، نادیده گرفته‌اند. به منظور بررسی و حل این چالش، سیستم‌های توصیه‌گر مبتنی بر جلسه[6] شدند. یک جلسه دنباله‌ای از تعاملات کاربر از قبیل کلیک‌های کاربر در یک بازه زمانی مشخص می‌باشد که می‌تواند تغییرات علایق و ترجیحات کاربر را در طول زمان منعکس کند. وظیفه یک سیستم توصیه‌گر مبتنی بر جلسه این است که بر اساس رفتارهای قبلی کاربر در یک جلسه، رفتار یا کلیک بعدی کاربر را پیش‌بینی کند [۲]. مطالعات گسترده‌ای روی سیستم‌های توصیه‌گر مبتنی بر جلسه صورت گرفته است. تعدادی از پژوهش‌ها از زنجیره مارکوف (MC)[7] [۳] و [۴] برای پیش‌بینی کلیک بعدی کاربر بر اساس رفتار قبلی کاربر استفاده کرده‌اند. اما این روش به دلیل فرض مستقل‌بودن تعاملات یا کلیک‌های متوالی کاربر در یک جلسه، دقت پایینی در ایجاد توصیه‌ها دارد.

تحقیقات اخیر از تکنیک‌های قدرتمند یادگیری عمیق استفاده کرده‌اند که آنها در مقایسه با رویکردهای کلاسیک، کیفیت پیش‌بینی و توصیه‌ها را به طور چشم‌گیری بهبود داده‌اند. در میان این تکنیک‌ها، شبکه‌های عصبی مکرر (RNN)[8] [۵] تا [۹] با یادگیری ویژگی‌ها و اطلاعات متوالی در جلسات، نتایج نسبتاً خوبی ارائه کرده‌اند. این رویکردها مبتنی بر ترتیب ورودی آیتم‌ها در یک جلسه آموزش داده می‌شوند و وابستگی کلی[9] بین آیتم‌های یک جلسه را در نظر نمی‌گیرند. به طوری که دو جلسه با ترتیب آیتم متفاوت اما با مجموعه آیتم یکسان را متفاوت در نظر می‌گیرند، در حالی که آیتم هدف هر دو جلسه یکسان است که این می‌تواند موجب یادگیری نادرست و کاهش عملکرد مدل شبکه شود [۱۰]. روش‌های

این مقاله در تاریخ ۶ مهر ماه ۱۴۰۰ دریافت و در تاریخ ۲۵ اردیبهشت ماه ۱۴۰۱ بازنگری شد.

اعظم رمضانی، دانشکده مهندسی کامپیوتر، دانشگاه یزد، یزد، ایران، (email: ramazani.azam@stu.yazd.ac.ir).

علی محمد زارع بیدکی (نویسنده مسئول)، دانشکده مهندسی کامپیوتر، دانشگاه یزد، یزد، ایران، (email: alizareh@yazd.ac.ir).

1. Collaborative Filtering
2. Matrix Factorization
3. Cold Start
4. Sparsity
5. Shallow
6. Session-Based Recommender Systems
7. Markov Chain
8. Recurrent Neural Network
9. Global

توصیه‌گر مبتنی بر شبکه‌های عصبی گراف (GNN)[1] [۱۱] تا [۱۳] با مدل‌کردن انتقالات پیچیده بین آیتم‌ها نسبت به روش‌های مبتنی بر RNN دقت بالاتری به دست آوردند. این روش‌ها، بردارهای نهفته[2] آیتم‌ها را با در نظر گرفتن وابستگی فقط بین آیتم‌های مجاور یا متصل به یک آیتم در یک گراف جلسه به دست می‌آورند و اطلاعات آیتم‌های دورتر را نادیده می‌گیرند. اخیراً شبکه‌های خودتوجه (SAN)[3]، برگرفته از معماری شبکه ترنسفورمرها[4] [۱۴]، با به دست آوردن وابستگی کلی بین همه آیتم‌های یک جلسه و بدون در نظر گرفتن فاصله آنها، توانایی قدرتمندی را در توصیه مبتنی بر جلسه و تولید بردارهای نهفته دقیق برای آیتم‌ها از خود نشان داده‌اند [۱۵] و [۱۶]. اما اغلب این پژوهش‌ها تنها مبتنی بر جلسات فعلی کاربر هستند و رفتار و فعالیت‌های گذشته کاربر را در جلسات قبلی‌اش که ممکن است دارای اطلاعات ارزشمندی باشد از دست می‌دهند.

در این مقاله به منظور یادگیری وابستگی کلی میان همه آیتم‌های یک جلسه، یک مدل توصیه‌گر مبتنی بر جلسه با استفاده از شبکه‌های خودتوجه برای ایجاد توصیه‌های شخصی‌سازی شده به کاربران ارائه گردیده است. مدل پیشنهادی از توالی جلسه فعلی و توالی‌های جلسات قبلی کاربر که به ترتیب بیانگر علاقه و ترجیحات کوتاه‌مدت و بلندمدت کاربر هستند، استفاده می‌کند. ابتدا مدل مبتنی بر مکانیزم SAN با استفاده از جلسات کاربران ناشناس آموزش داده می‌شود و بردارهای نهفته آیتم‌ها تولید می‌گردند. در SAN بردار نهفته آخرین آیتم به طور مشترک هم علاقه فعلی و هم علاقه کلی کاربر در آن جلسه را منعکس می‌کند. بنابراین به عنوان نماینده هر جلسه، به جای تجمیع بردارهای نهفته همه آیتم‌های جلسه، فقط از بردار نهفته آخرین آیتم استفاده شده است. سرانجام برای ایجاد توصیه به کاربر، توالی جلسه فعلی و تعدادی از توالی‌های جلسات قبلی اخیر کاربر به صورت جداگانه از شبکه عبور کرده و با ترکیب وزنی امتیازهای رتبه‌بندی آیتم‌ها حاصل از هر جلسه، آیتم کلیک‌شده بعدی کاربر محاسبه و پیش‌بینی می‌شود. مدل پیشنهادی بر روی مجموعه داده واقعی Reddit ارزیابی و با سایر روش‌های مشابه مقایسه شده است. نتایج حاکی از آن است که مدل پیشنهادی نسبت به روش‌های قبلی به عملکرد توصیه دقیق‌تری با دقت[5] ۵۸٫۹۵٪ و میانگین رتبه متقابل[6] ۲۸٫۲۷٪ دست یافته است.

سازماندهی مقاله به این صورت است که در بخش ۲ به تشریح و بررسی تحقیقات پیشین در زمینه سیستم‌های توصیه‌گر مبتنی بر جلسه خواهیم پرداخت. در بخش ۳ روش پیشنهادی توضیح داده می‌شود. نتایج پیاده‌سازی و همچنین مقایسه عملکرد روش پیشنهادی با سایر روش‌های مشابه در بخش ۴ ارائه می‌گردد. نهایتاً در بخش ۵ نتیجه‌گیری و مسیر آتی تحقیق بیان می‌شود.

## ۲- مروری بر تحقیقات پیشین

سیستم‌های توصیه‌گر مبتنی بر جلسه توسط پژوهش‌های متعددی مورد مطالعه قرار گرفته‌اند. این پژوهش‌ها را می‌توان به دو دسته روش‌های کلاسیک و روش‌های یادگیری عمیق تقسیم‌بندی کرد. در این بخش به بیان و مرور برخی از پژوهش‌های انجام‌گرفته در این حوزه پرداخته می‌شود.

### ۲-۱ روش‌های کلاسیک

یکی از رایج‌ترین تکنیک‌های کلاسیک در سیستم‌های توصیه‌گر، پالایش مشارکتی از قبیل ماتریس فاکتوربندی (MF) می‌باشد. MF با تجزیه ماتریس تعاملات کاربر- آیتم، کاربران و آیتم‌ها را به یک فضای پنهان با بعد کم نگاشت می‌نماید و با ضرب داخلی بردارها، ترجیحات کاربر روی آیتم‌ها را تخمین می‌زند [۱۷] و [۱۸]. تعدادی از پژوهش‌ها از روش‌های همسایگی برای ایجاد توصیه استفاده کرده‌اند. سرور و همکارانش در [۱۹]، یک سیستم توصیه‌گر مبتنی بر الگوریتم K- نزدیک‌ترین همسایه (KNN) را برای محاسبه شباهت و پیداکردن همسایه‌ها با آخرین آیتم کلیک کاربر ارائه کردند. همچنین جاناچ و لودویگ در [۷]، یک مدل تعمیم‌یافته مبتنی بر الگوریتم KNN را برای محاسبه شباهت با همه آیتم‌های جلسه ارائه کردند. این روش‌ها ترجیحات کلی کاربران را مبتنی بر کل تاریخچه تعاملات کاربران و با در نظر گرفتن اهمیت یکسان مدل می‌کنند و اطلاعات متوالی تعاملات کاربر را نادیده می‌گیرند.

یک رویکرد کلاسیک در سیستم‌های توصیه‌گر که از ترتیب تعاملات برای مدل‌سازی ترجیحات کاربران استفاده می‌کند، مدل‌های مبتنی بر زنجیره مارکوف است. شانی و همکاران [۳] ایجاد توصیه را به عنوان یک مشکل بهینه‌سازی متوالی در نظر گرفتند و از فرایند تصمیم‌گیری مارکوف برای مدل‌کردن آن استفاده کردند. علاوه بر این، مدل ترکیبی FPMC [۴] از ترکیب زنجیره مارکوف و ماتریس فاکتوربندی برای مدل‌کردن دو عمل متوالی کاربر و پیش‌بینی آیتم بعدی کاربر استفاده می‌کند. هی و مک‌آولی [۲۰] مدل‌های شباهت را با زنجیره مارکوف به منظور حل چالش توصیه‌های متوالی اسپارس، ترکیب نموده و همچنین توصیه‌های شخصی‌سازی شده را برای کاربران فراهم می‌کنند. 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 [۲۱].

### ۲-۲ روش‌های یادگیری عمیق

در سال‌های اخیر، یادگیری عمیق با قابلیت استخراج ویژگی‌های عمیق، عملکرد سیستم‌های توصیه‌گر مبتنی بر جلسه را به طور گسترده‌ای بهبود بخشیده است. شبکه‌های عصبی مکرر (RNN) توسط بسیاری از سیستم‌های توصیه‌گر برای مدل‌کردن رفتار متوالی کاربران استفاده شده‌اند. GRU4REC [۵] اولین مدل توصیه‌گر مبتنی بر جلسه است که بر اساس RNN ارائه گردید و در آن از واحدهای برگشتی دروازه‌دار (GRU)[7] به عنوان یک مدل بهبودیافته از شبکه‌های عصبی مکرر به منظور پیش‌بینی آیتم بعدی کاربر در جلسه فعلی استفاده می‌شود. لی و همکارانش [۸] مدل NARM را ارائه کردند که از مکانیزم توجه[8] روی شبکه عصبی مکرر برای به دست آوردن هدف اصلی کاربر در یک جلسه و ایجاد توصیه‌ها استفاده می‌کند. مدل STAMP [۲۲] از یک شبکه عصبی پرسپترون چندلایه (MLP)[9] و یک شبکه توجه استفاده می‌کند که که علایق کلی کاربر را از حافظه بلندمدت یک جلسه و علایق فعلی

1. Graph Neural Network
2. Latent Vector
3. Self-Attention Network
4. Transformer
5. Precision
6. Mean Reciprocal Rank
7. Gated Recurrent Unit
8. Attention
9. Multi-Layer Perceptron

کاربر را از حافظه کوتاه‌مدت آخرین کلیک کاربر به دست می‌آورد. وانگ و همکارانش [۲۳] یک سیستم توصیه مبتنی بر جلسه مشارکتی با استفاده از شبکه‌های عصبی مکرر و مکانیزم توجه پیشنهاد کردند که از اطلاعات همسایگی جلسات به منظور تقویت نمایش جلسه استفاده می‌کند. ژانگ و وانگ [۱۰] یک مدل عصبی مشترک برای توصیه مبتنی بر جلسه به منظور یادگیری هم‌زمان علایق ترتیبی و کلی کاربر ارائه کردند که از یک شبکه GRU با مکانیزم توجه برای استخراج علایق متوالی جلسه و از یک شبکه عصبی چندلایه با اتصالات باقیمانده[1] برای استخراج علایق کلی جلسه استفاده می‌کند.

شبکه‌های عصبی گراف، با توجه به مدل‌کردن انتقالات پیچیده بین تعاملات یا کلیک‌های کاربران با استفاده از ساختار گراف، در تعدادی از سیستم‌های توصیه‌گر به کار رفته‌اند. وو و همکاران [۱۱] یک سیستم توصیه‌گر مبتنی بر جلسه با شبکه‌های عصبی گراف به نام SR-GNN ارائه کردند. آنها ابتدا توالی‌های جلسه را به صورت ساختارهای گراف مدل کرده و از شبکه‌های عصبی گراف برای یادگیری بردارهای نهفته آیتم‌ها استفاده می‌نمایند. سپس ترجیحات کلی کاربر را توسط تجمیع تمام بردارهای نهفته آیتم‌های جلسه مبتنی بر یک شبکه توجه و همچنین ترجیحات فعلی کاربر را توسط بردار نهفته آخرین آیتم جلسه استخراج می‌کنند. در نهایت از ترکیب ترجیحات فعلی و ترجیحات کلی هر جلسه برای نمایش جلسه و ایجاد توصیه استفاده می‌نمایند. یو و همکاران [۱۲] یک مدل شبکه عصبی گراف آگاه به هدف 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. چن و ونگ [۱۳] با مشکل از دست دادن اطلاعات در مدل‌های مبتنی بر GNN برای توصیه مبتنی بر جلسه از طریق ترکیب دو لایه تجمیع حفظ ترتیب یال و لایه توجه گراف میانبر[2] مقابله می‌کنند. روش‌های مبتنی بر RNN و GNN تنها متکی بر ترتیب توالی هستند و وابستگی کلی بین آیتم‌های جلسه نادیده گرفته شده است.

در سال‌های اخیر شبکه‌های خودتوجه (SAN) به کار رفته در معماری شبکه ترنسفورمرها توسط بسیاری از برنامه‌های پردازش زبان طبیعی و ترجمه ماشین و همچنین سیستم‌های توصیه‌گر استفاده شده‌اند. این شبکه‌ها با مدل‌کردن توالی‌ها بدون استفاده از هیچ بازگشت[3] یا کانولوشنی[4] و همچنین افزایش سرعت آموزش شبکه با موازی‌سازی محاسبات مکانیزم توجه، کارایی و عملکرد بسیار خوبی از خود نشان داده‌اند. شو و همکاران [۲۴] از مکانیزم خودتوجه بعد از شبکه عصبی گراف به منظور یادگیری وابستگی کلی میان آیتم‌های جلسه استفاده کردند. فنگ [۱۵] یک مدل مبتنی بر SAN را برای توصیه مبتنی بر جلسه تحت عنوان SR-SAN پیشنهاد کرد. او به طور مستقیم مکانیزم خودتوجه را بعد از لایه تعبیه[5] قرار داد و از بردار نهفته آخرین آیتم توالی برای نمایش جلسه استفاده کرد. وانگ و همکاران [۲۵] یک شبکه توجه گراف گیتی[6] آگاه به موقعیت را برای توصیه مبتنی بر جلسه ارائه کردند که مبتنی بر ترتیب آیتم‌ها در توالی، یک تعبیه موقعیت را به گره‌ها در گراف جلسه اختصاص می‌دهد و همچنین از یک شبکه خودتوجه برای مدل‌کردن وابستگی‌های کلی بین آیتم‌ها استفاده می‌کند. این پژوهش‌ها فقط از جلسات فعلی کاربر استفاده می‌کنند و فعالیت‌های گذشته کاربر را در نظر نمی‌گیرند و تنها علایق کوتاه‌مدت کاربر را مدل می‌کنند.

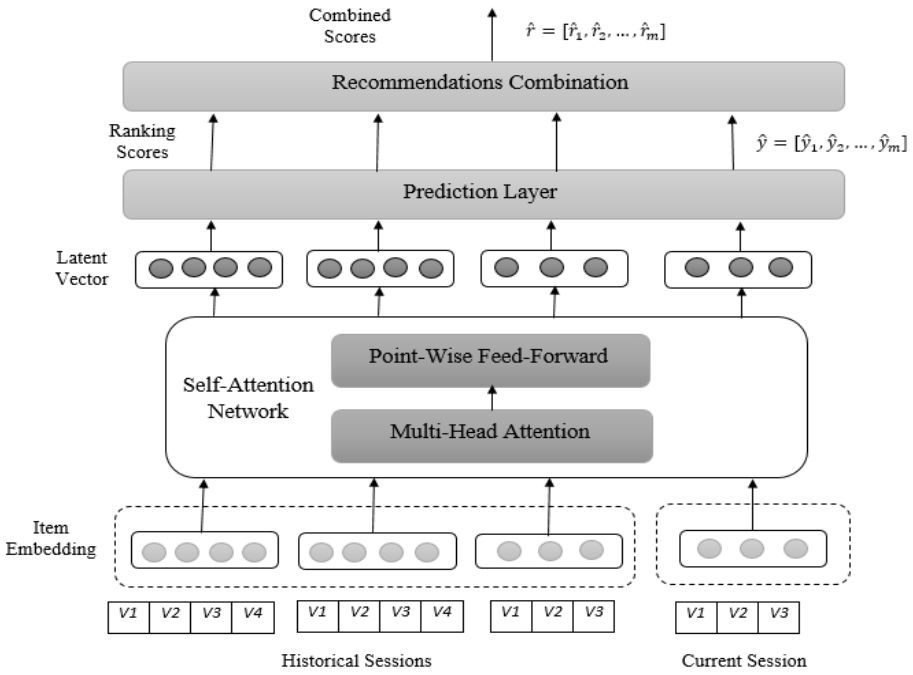
تعدادی از مدل‌ها برای ایجاد توصیه‌ها علاوه بر توالی جلسه فعلی از جلسات قبلی کاربر هم استفاده کرده‌اند. مثلاً [۹] و [۲۶] با استفاده از یک معماری سلسله‌مراتبی مبتنی بر چند شبکه RNN، علایق کاربر را متکی بر جلسه فعلی و جلسات قبلی کاربر مدل می‌کنند. همچنین مدل HierTCN [۲۷] به عنوان یکی از کارهای موفق اخیر ارائه‌شده، از یک معماری سلسله‌مراتبی تشکیل شده که از یک شبکه GRU و یک شبکه کانولوشنال زمانی (TCN)[7] برای یادگیری هر دو نوع علایق بلندمدت و کوتاه‌مدت کاربر استفاده می‌کند. این پژوهش‌ها جلسات گذشته کاربر را به صورت نمایش‌هایی[8]، رمزگذاری[9] و سپس از آن برای پیش‌بینی علاقه کاربر استفاده می‌کنند که این مکانیزم رمزگذاری می‌تواند قابلیت‌های مدل را محدود کند. علاوه بر آن، این مدل‌ها تنها با تکیه بر وابستگی‌های ترتیبی تعاملات جلسه، الگوهای رفتاری کاربران را استخراج می‌کنند. در جدول ۱ به طور خلاصه یک طبقه‌بندی از رویکردهای موجود در این حوزه و نقاط قوت و ضعف آنها ارائه شده است.

همان طور که بیان گردید، اکثر سیستم‌های توصیه‌گر مبتنی بر جلسه موجود تنها متکی بر تعاملات جلسه فعلی کاربر هستند و پژوهش‌های کمی از تعاملات و جلسات گذشته کاربر برای تقویت توصیه‌ها استفاده کرده‌اند. روش پیشنهادی علاوه بر توالی جلسه فعلی از توالی‌های جلسات قبلی کاربر هم استفاده نموده و علاقه و ترجیحات کوتاه‌مدت و بلندمدت کاربر را مدل می‌کند. مدل پیشنهادی تنها از یک شبکه و با هدف یادگیری وابستگی‌های کلی بین آیتم‌های جلسه استفاده می‌کند. یکی از مهم‌ترین مزایای مدل پیشنهادی این است که برخلاف روش‌های قبلی، جلسات گذشته کاربر را رمزگذاری نمی‌کند. به این صورت که مدل ابتدا بر اساس الگوهای رفتاری جلسات ایجادشده توسط کاربران ناشناس ساخته می‌شود و سپس با عبور جلسه فعلی و جلسات قبلی کاربر از مدل شبکه به صورت جداگانه و ترکیب آنها، توصیه‌های شخصی‌سازی شده را برای کاربر فراهم می‌کند.

## ۳- مدل پیشنهادی

در این بخش، سیستم توصیه‌گر پیشنهادی مبتنی بر جلسه شخصی‌سازی شده با شبکه‌های خودتوجه (PSR-SAN)[10] توصیف می‌گردد. مدل پیشنهادی بر اساس مدل SR-SAN [۱۵] ارائه شده است. هدف این سیستم، پیش‌بینی آیتم یا کلیک بعدی کاربر بر اساس ترکیبی از توالی جلسه فعلی و توالی‌های جلسات قبلی کاربر که به ترتیب بیانگر علاقه و ترجیحات کوتاه‌مدت و بلندمدت کاربر هستند، می‌باشد. شکل ۱ معماری PSR-SAN را نشان می‌دهد که از سه مؤلفه اصلی تشکیل شده است: یادگیری بردارهای نهفته آیتم‌ها مبتنی بر شبکه خودتوجه، محاسبه احتمال توصیه آیتم‌ها و ترکیب توصیه‌ها. در ادامه ابتدا مسئله توصیه مبتنی بر جلسه بیان می‌گردد و سپس معماری سیستم پیشنهادی با جزئیات توضیح داده می‌شود.

1. Residual
2. Shortcut
3. Recurrence
4. Convolution
5. Embedding
6. Gated
7. Temporal Convolutional Network
8. Representation
9. Encode
10. Personalized Session-Based Recommendation with Self-Attention Networks




شکل ۱: معماری مدل PSR-SAN.

جدول ۱: خلاصه‌ای از نقاط قوت و ضعف رویکردهای سیستم‌های توصیه‌گر مبتنی بر جلسه.

| | رویکرد | ایده اصلی | نقاط قوت | نقاط ضعف |
|---|---|---|---|---|
| کلاسیک | ماتریس فاکتوربندی (MF) | با تجزیه ماتریس تعاملات کاربر–آیتم، کاربران و آیتم‌ها را به یک فضای پنهان نگاشت می‌کند. | – به دست آوردن وابستگی‌های ضمنی در فضای پنهان | – نیاز به حجم زیادی از تعاملات کاربران دارد و از پراکندگی داده‌ها رنج می‌برد.<br>– نادیده‌گرفتن ترتیب آیتم‌ها و وابستگی‌های متوالی |
| | نزدیک‌ترین همسایه (KNN) | با محاسبه شباهت و همسایگی با آیتم‌های جلسه، توصیه‌هایی را فراهم می‌کند. | – شباهت اقلام را می‌توان از قبل محاسبه کرد و پاسخ سریع را در زمان توصیه فراهم می‌کند.<br>– سادگی عملکرد | – نادیده‌گرفتن ترتیب آیتم‌ها و وابستگی‌های متوالی<br>– عدم توانایی مدل‌کردن جلسات پیچیده |
| | زنجیره مارکوف (MC) | احتمال انتقال را بر روی دنباله‌هایی با طول ثابت و محدود، محاسبه می‌کند. | – در نظر گرفتن اطلاعات متوالی | – ترتیب انتقال ثابت و محدود<br>– مستقل در نظر گرفتن تعاملات<br>– بر آخرین تعامل یا چند تعامل اخیر کاربر متکی است و وابستگی‌های کوتاه‌مدت و محلی بین تعاملات را مدل می‌کند. |
| یادگیری عمیق | شبکه عصبی مکرر (RNN) | داده‌های ترتیبی یا دنباله‌ای را توسط وضعیت‌های مخفی فشرده مدل می‌کند. | – مدل‌کردن وابستگی‌های متوالی<br>– یادگیری توالی‌های جلسه با طول متغیر<br>– مدل‌کردن وابستگی‌های طولانی‌مدت | – یادگیری وابستگی‌های محلی بین آیتم‌های مجاور و عدم یادگیری وابستگی‌های کلی بین تعاملات<br>– ساختار پیچیده |
| | شبکه عصبی گراف (GNN) | توالی‌های جلسه را به صورت داده‌های ساختاری گراف پردازش می‌کند. | – مدل‌کردن جلسات با انتقالات و وابستگی‌های پیچیده | – یادگیری وابستگی‌های محلی بین آیتم‌های مجاور و عدم یادگیری وابستگی‌های کلی بین تعاملات<br>– ساختار پیچیده |
| | شبکه خودتوجه (SAN) | به آیتم‌های مهم‌تر توالی ورودی، اهمیت و وزن بیشتری اختصاص می‌دهد. | – یادگیری وابستگی‌های کلی بین تعاملات<br>– مدل‌کردن جلسات با تعاملات ناخواسته و نویزی<br>– شناسایی آیتم‌های بااهمیت در جلسه | – نادیده‌گرفتن ترتیب آیتم‌ها و وابستگی‌های متوالی |

## ۳-۱ توصیه‌گر مبتنی بر جلسه

در توصیه‌گر مبتنی بر جلسه، $V = \{v_1, v_2, \dots, v_m\}$، مجموعه تمام آیتم‌های منحصربه‌فرد در تمام جلسات را نشان می‌دهد. یک توالی جلسه به صورت یک لیست $S = [v_1, v_2, \dots, v_n]$ و $v_i \in V$ نشان داده می‌شود که شامل کلیک‌های کاربر در یک جلسه است که به ترتیب زمانی مرتب شده‌اند. هدف سیستم توصیه‌گر مبتنی بر جلسه، پیش‌بینی آیتم بعدی یعنی $v_{n+1}$ در جلسه $S$ می‌باشد. برای تولید توصیه‌ها در یک مدل توصیه مبتنی بر جلسه، برای هر جلسه $S$، یک امتیاز توصیه $\hat{z}$ و سپس مبتنی بر آن یک احتمال $\hat{y}$ برای همه آیتم‌های مجموعه $V$ تولید می‌شود. سپس $K$ آیتم با مقادیر احتمال بالاتر، به عنوان آیتم‌های کاندیدا به کاربر پیشنهاد می‌شوند.

## ۳-۲ یادگیری بردارهای نهفته آیتم‌ها، مبتنی بر شبکه خودتوجه

ابتدا هر آیتم در یک توالی جلسه ورودی $S = [v_1, v_2, \dots, v_n]$ به یک فضای تعبیه‌شده یکپارچه $d$ بعدی $E = [e_1, e_2, \dots, e_n]$ نگاشت می‌شود. سپس با عبور بردارهای تعبیه آیتم‌ها از شبکه خودتوجه، بردارهای نهفته

آیتم‌ها به دست می‌آیند. شبکه خودتوجه از دو زیرلایه توجه چندسر[1] و شبکه پیش‌خور نقطه‌ای[2] تشکیل شده است.

### ۳-۲-۱ مکانیزم توجه چندسر

زیرلایه خودتوجه، هر آیتم خاص را با توجه به ارتباط و وابستگی آن با آیتم‌های دیگر در توالی ورودی رمزگذاری می‌کند. برای این هدف، ابتدا بردارهای تعبیه آیتم‌های یک جلسه به عنوان ورودی به لایه خودتوجه وارد می‌شوند. برای هر آیتم، سه بردار پرس‌وجو ($Q$)، کلید ($K$) و مقدار ($V$) توسط ضرب بردار تعبیه آیتم در سه ماتریس نمایش[3] قابل یادگیری $W^Q$، $W^K$ و $W^V \in R^{d\times d}$ ایجاد می‌گردد و این بردارها به صورت $Q = EW^Q$، $K = EW^K$ و $V = EW^V$ نمایش داده می‌شوند. تابع توجه به صورت نگاشت یک بردار پرس‌وجو و یک مجموعه جفت بردارهای کلید- مقدار به یک بردار خروجی توصیف می‌شود. خروجی تابع به صورت یک مجموع وزن‌دار از مقادیر محاسبه می‌گردد که وزن هر مقدار توسط مکانیزم توجه ضرب نقطه‌ای مقیاس‌بندی شده[4] بردار پرس‌وجو با بردار کلید مربوط به دست می‌آید. تابع به صورت رابطه زیر تعریف می‌شود

$$Attention(Q,K,V) = softmax(\frac{QK^T}{\sqrt{d}})V \quad (۱)$$

شبکه خودتوجه از مکانیزم توجه چندسر استفاده می‌کند که به مدل، اجازه می‌دهد با استفاده از چندین سر توجه[5] به وابستگی‌ها در موقعیت‌های مختلف توجه کند. سپس با ترکیب همه سرهای توجه و ضرب در یک ماتریس وزن $W^O \in R^{d\times d}$ نتایج به دست می‌آیند. مکانیزم به صورت زیر فرموله می‌شود

$$MultiHead(Q,K,V) = Concat(head_1,\dots,head_h)W^O \quad (۲)$$

$$head_i = Attention(EW_i^Q, EW_i^K, EW_i^V) \quad (۳)$$

در این رابطه $h$ تعداد سر توجه است. سپس خروجی این زیرلایه به شبکه پیش‌خور تغذیه می‌شود.

### ۳-۲-۲ شبکه پیش‌خور نقطه‌ای

خروجی زیرلایه خودتوجه از یک شبکه پیش‌خور کاملاً متصل دولایه با تابع فعال‌سازی واحد خطی اصلاح‌شده (Relu)[6] در بین آنها، عبور داده می‌شود به طوری که به هر موقعیت در ورودی، به صورت جداگانه اما یکسان اعمال می‌گردد. در نهایت بردارهای نهفته آیتم‌های جلسه یعنی $H = [h_1, h_2, \dots, h_n]$ به دست می‌آیند

$$FFN(x) = \max(0, xW_1 + b_1)W_2 + b_2 \quad (۴)$$

که در آن $W_1, W_2 \in R^{d\times d}$ ماتریس‌های وزن و $b_1, b_2 \in R^d$ بردارهای بایاس هستند.

برای حفظ اطلاعات ویژگی‌های اصلی یا به عبارت دیگر اطلاعات لایه پایین، بعد از خروجی هر دو زیرلایه توجه چندسر و شبکه پیش‌خور، یک اتصال باقیمانده و نرمال‌سازی اعمال می‌شود. همچنین برای جلوگیری از مشکل بیش‌برازش[7] شبکه از روش منظم‌سازی[8] L2 استفاده می‌گردد.

## ۳-۳ محاسبه احتمال توصیه آیتم‌ها و آموزش مدل

همان طور که در بخش قبل توضیح داده شد، شبکه خودتوجه (SAN)، بردار نهفته هر آیتم در توالی جلسه را با توجه به ارتباط و وابستگی کلی آن آیتم با همه آیتم‌های جلسه و بدون در نظر گرفتن فاصله آنها تولید می‌کند. با توجه به آن که بردارهای نهفته آیتم‌ها تجمیعی از تمام آیتم‌های جلسه هستند، می‌توانند علایق کلی یک جلسه را نمایش دهند. بعد از به دست آوردن بردارهای نهفته آیتم‌ها، برای نشان‌دادن هر جلسه، بردار نهفته آخرین آیتم کلیک‌شده که به طور هم‌زمان هم ترجیح و علاقه فعلی و هم علایق کلی یک جلسه را منعکس می‌کند، در نظر گرفته می‌شود. سپس در لایه پیش‌بینی، امتیاز توصیه $\hat{z}_i$ برای هر آیتم کاندیدای $v_i \in V$ توسط ضرب بردار نهفته آخرین آیتم جلسه در بردار تعبیه هر آیتم به صورت زیر محاسبه می‌شود

$$\hat{z}_i = h_n^T e_i \quad (۵)$$

سپس برای به دست آوردن احتمال هر آیتم کاندیدا $\hat{y}$ از تابع softmax استفاده می‌شود

$$\hat{y} = soft\max(\hat{z}) \quad (۶)$$

که در این رابطه، $\hat{z} \in R^m$ امتیازات توصیه روی همه آیتم‌های کاندیدا و $\hat{y} \in R^m$ احتمال این که آیتم‌ها کلیک بعدی در جلسه $S$ باشند را نشان می‌دهد.

برای آموزش مدل پیشنهادی، از تابع ضرر آنتروپی متقاطع[9] بین مقدار پیش‌بینی شده و مقدار واقعی استفاده می‌گردد که به صورت رابطه زیر تعریف می‌شود

$$L(\hat{y}) = -\sum_{i=1}^{m} y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i) \quad (۷)$$

در این رابطه $y$ بردار رمزگذاری one-hot مقدار واقعی آیتم را نشان می‌دهد. از الگوریتم پس‌انتشار در زمان (BPTT)[10] به منظور بهینه‌سازی این تابع استفاده می‌شود.

## ۳-۴ ترکیب توصیه‌ها

برای هر کاربر، مجموعه‌ای از جلسات شامل جلسه فعلی و تعداد $M$ جلسه قبلی به صورت $S^u = \{S_c^u, S_{h1}^u, S_{h2}^u, \dots, S_{hM}^u\}$ در نظر گرفته می‌شود. $S_c^u$ جلسه فعلی کاربر $u$ و $S_c^u$، $i$امین جلسه قبلی کاربر $u$ می‌باشد. هدف سیستم این است که با توجه به مجموعه جلسات کاربر، آیتم یا کلیک بعدی کاربر را پیش‌بینی کند. بعد از محاسبه احتمال همه آیتم‌های کاندیدا، به ازای هر جلسه کاربر، امتیازات جلسه فعلی و جلسات قبلی کاربر تجمیع می‌شوند. جلسه فعلی و جلسات قبلی کاربر دارای اهمیت و اولویت یکسانی نیستند و به صورت وزن‌دار و با وزن‌دهی متفاوت با هم ترکیب می‌شوند. امتیاز توصیه رتبه‌بندی تجمیع‌شده برای هر آیتم $\hat{r}_i$ از طریق فرمول زیر محاسبه می‌شود

$$\hat{r}_i = \sum_{s=1}^{N_{ses}} \alpha_s \hat{y}_{i,s} \quad (۸)$$

1. Multi-Head Attention
2. Point-Wise Feed-Forward Network
3. Projection Matrix
4. Scaled Dot-Product Attention
5. Attention Head
6. Rectified Linear Unit
7. Overfitting
8. Regularization
9. Cross-Entropy Loss
10. Back-Propagation Through Time

جدول ۲: ویژگی‌های آماری دیتاست REDDIT.

| ویژگی | مقدار |
|---|---|
| تعداد کاربران | ۱۸۲۷۱ |
| تعداد آیتم‌ها | ۲۷۴۵۲ |
| تعداد جلسات | ۱۱۳۵۴۸۸ |
| طول متوسط جلسه | ۳٫۰۲ |
| تعداد جلسات کاربر | ۶۲٫۱۵ |
| تعداد جلسات مجموعه آموزش | ۹۰۱۱۶۱ |
| تعداد جلسات مجموعه آزمون | ۲۳۴۳۲۷ |

که در این رابطه $\alpha_s$ ضریب اهمیت هر جلسه، $\hat{y}_{i,s}$ احتمال کلیک آیتم $i$ در جلسه $s$ و $N_{ses}$ برابر با تعداد جلسات کاربر شامل جلسه فعلی و $M$ جلسه قبلی می‌باشد. از آنجایی که جلسه فعلی بیشترین اهمیت را دارد و جلسات قبلی بر اساس فاصله‌ای که از جلسه فعلی دارند دارای اهمیت متفاوتی هستند، ضریب اهمیت هر جلسه از طریق فرمول زیر محاسبه می‌شود

$$\alpha_s = \frac{N_{ses} - discurses}{N_{ses}} \quad (۹)$$

که $N_{ses}$ برابر با تعداد جلسات کاربر و $discurses$ فاصله جلسه از جلسه فعلی کاربر را نشان می‌دهد.

## ٤- پیاده‌سازی و ارزیابی

در این بخش ابتدا مجموعه داده استفاده‌شده در پژوهش توصیف می‌گردد و سپس به بیان روش‌های مقایسه‌شده، معیارهای ارزیابی و تنظیم پارامترها پرداخته می‌شود. در نهایت مدل پیشنهادی PSR-SAN با روش‌های دیگر مقایسه و نتایج حاصل، تجزیه و تحلیل می‌شود.

### ٤-۱ دیتاست

برای ارزیابی مدل پیشنهادی از مجموعه داده واقعی Reddit[1] استفاده شده که این مجموعه داده از وب‌سایت اخبار اجتماعی Reddit استخراج گردیده است. داده خام دیتاست شامل نظرات کاربران به شکل رکوردهایی از [username, subreddit, timestamp] می‌باشد. با استفاده از رویکرد استفاده‌شده در [۲۶]، رکوردهای هر کاربر به صورت دستی به جلساتی تقسیم گردیده‌اند. در این رویکرد از یک محدوده زمانی استفاده می‌شود و فعالیت‌های متوالی که در این محدوده زمانی رخ می‌دهند، متعلق به یک جلسه در نظر گرفته می‌شوند. محدوده زمانی برای این دیتاست ۶۰ دقیقه تنظیم گردیده است.

پیش‌پردازش‌های زیر بر روی دیتاست همانند [۲۶] انجام شده است: حداکثر طول هر جلسه ($\max length$) برابر ۲۰ در نظر گرفته شده و جلساتی با طول بزرگ‌تر از $\max length$ شکسته می‌شوند. جلسات طولانی با طول بزرگ‌تر از دو برابر $\max length$ و همچنین جلسات با طول کمتر از ۲، فیلتر می‌شوند. حداقل تعداد جلسات هر کاربر، ۳ جلسه در نظر گرفته شده و کاربرانی با کمتر از ۳ جلسه حذف شده‌اند. برای هر کاربر، ۸۰ درصد از جلساتش برای مجموعه آموزش و ۲۰ درصد مابقی برای مجموعه آزمون انتخاب گردیده و ویژگی‌های آماری دیتاست پس از پیش‌پردازش در جدول ۲ ارائه شده است.

به منظور تقویت[2] و افزایش داده‌ها، همانند [۱۱] توالی‌های ورودی هر دو مجموعه آموزش و آزمون به مجموعه‌ای از توالی‌ها و برچسب متناظر تقسیم می‌شوند. به عنوان مثال در مجموعه آموزش یک جلسه ورودی $S=[v_1, v_2, \ldots, v_n]$ به توالی‌ها و برچسب‌های متناظر $([v_1], v_2)$، $([v_1, v_2], v_3)$، ... و $([v_1, v_2, \ldots, v_{n-1}], v_n)$ تقسیم می‌شود، به طوری که $[v_1, v_2, \ldots, v_{n-1}]$ توالی ایجادشده و $v_n$ برچسب متناظر آن می‌باشد. همچنین در داده‌های آزمون، مجموعه جلسات کاربر $u$، $S^u = \{\{v_{11}, v_{12}, v_{13}\}, \{v_{21}, v_{22}\}, \{v_{31}, v_{32}, v_{33}\}\}$ به مجموعه‌ای از جلسات قبلی و جلسه فعلی و برچسب‌های متناظر به صورت زیر تقسیم می‌شود

$$S^u_{h1} = \{\{v_{11}, v_{12}, v_{13}\}\}, S^u_{c1} = \{\{v_{21}\}\}, label_1 = v_{22}$$
$$S^u_{h2} = \{\{v_{11}, v_{12}, v_{13}\}, \{v_{21}, v_{22}\}\}, S^u_{c2} = \{\{v_{31}\}\}, label_2 = v_{33}$$
$$S^u_{h3} = \{\{v_{11}, v_{12}, v_{13}\}, \{v_{21}, v_{22}\}\}, S^u_{c3} = \{\{v_{31}, v_{32}\}\}$$
$$, label_3 = v_{33}$$

به طوری که $S^u_h$ مجموعه جلسات قبلی کاربر، $S^u_c$ جلسه فعلی کاربر و آخرین آیتم در جلسه فعلی به عنوان آیتم برچسب می‌باشد.

### ٤-۲ روش‌های مقایسه‌شده

عملکرد مدل پیشنهادی با مدل‌های پایه زیر مقایسه شده است:

- Pop: آیتم‌های محبوب با بیشترین تعداد تعاملات در مجموعه آموزش را توصیه می‌کند.
- Item-KNN [۱۹]: آیتم‌هایی را شبیه به آیتم‌هایی که در جلسات کلیک شده‌اند، توصیه و برای محاسبه شباهت از شباهت کسینوسی استفاده می‌کند.
- FPMC [۴]: از ترکیب زنجیره مارکوف و ماتریس فاکتوربندی برای پیش‌بینی سبد بعدی کاربر استفاده می‌کند.
- SKNN [۷]: امتیاز آیتم‌های کاندیدا را مبتنی بر مجموعه اقدامات کاربر در جلسه فعلی و $k$ جلسه از مجموعه آموزش که بیشترین شباهت را با جلسه فعلی دارند، محاسبه می‌کند.
- VSKNN [۲۸]: مدل تعمیم‌یافته مبتنی بر SKNN که توالی و موقعیت آیتم‌ها در جلسه را برای محاسبه شباهت در نظر می‌گیرد.
- GRU4Rec [۵]: از یک مدل بهبودیافته شبکه‌های عصبی مکرر، GRU برای توصیه مبتنی بر جلسه استفاده می‌کند.
- SR-GNN [۱۱]: از شبکه‌های عصبی گراف به منظور مدل‌کردن انتقالات پیچیده بین آیتم‌ها و ایجاد توصیه مبتنی بر جلسه استفاده می‌کند.

### ٤-۳ 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:

- P@K (دقت): به منظور اندازه‌گیری دقت پیش‌بینی در سیستم‌های توصیه‌گر مبتنی بر جلسه استفاده می‌شود و نسبت آیتم‌های درست پیش‌بینی شده در $K$ تا آیتم بالا در لیست رتبه‌بندی را نشان می‌دهد

$$P@K = \frac{n_{hit}}{N} \quad (۱۰)$$

1. https://www.kaggle.com/colemaclean/subreddit-interactions

2. Augmentation

جدول ۳: پارامترهای استفاده‌شده در مدل پیشنهادی.

| پارامتر | مقدار |
|---|---|
| ابعاد بردار تعبیه آیتم‌ها ($d$) | ۹۶ |
| اندازه دسته[1] | ۱۰۰ |
| بیشینه طول هر جلسه ($\max length$) | ۲۰ |
| تعداد جلسات قبلی | ۵ |
| بهینه‌ساز وزن‌های شبکه | Adam |
| نرخ حذف تصادفی[2] | ۰/۱ |
| نرخ یادگیری[3] | ۰/۰۰۱ |
| کاهش نرخ یادگیری[4] | ۰/۱ |
| تعداد گام‌های کاهش نرخ یادگیری | ۳ |
| منظم‌سازی $L2$ | $10^{-5}$ |
| تعداد سر توجه ($h$) | ۸ |
| ابعاد شبکه پیش‌خور | ۴ برابر ابعاد بردار تعبیه آیتم |
| تعداد لایه‌های SAN | ۱ |

1. Batch Size
2. Dropout
3. Learning Rate
4. Learning Rate Decay

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

- MRR@K (میانگین رتبه متقابل): میانگین رتبه متقابل آیتم‌های هدف را نشان می‌دهد. اگر رتبه بزرگ‌تر از $K$ باشد، مقدار $MRR$ صفر در نظر گرفته می‌شود

$$MRR@K = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{rank_i} \quad (11)$$

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. به طوری که هرچه آیتم هدف در رتبه‌بندی بالاتری از لیست قرار بگیرد، مقدار $MRR$ بیشتر خواهد بود.

## ٤-٤ تنظیم پارامترها

تنظیم پارامترها بر نحوه عملکرد سیستم و دقت آن تأثیر می‌گذارند. جدول ۳ پارامترهای استفاده‌شده توسط مدل پیشنهادی را نشان می‌دهد و پارامترها با استفاده از روش آزمون و خطا تنظیم شده‌اند.

همه پارامترهای مدل پیشنهادی با استفاده از یک توزیع گوسی با میانگین ۰ و انحراف معیار $1/\sqrt{d}$ مقداردهی اولیه می‌شوند. به دلیل محدودیت‌های محاسباتی برای هر کاربر، حداکثر تعداد جلسات قبلی ۵ در نظر گرفته شده است. ابرپارامترهای مربوط به معماری شبکه مدل شامل ابعاد بردار تعبیه آیتم‌ها، تعداد سرهای توجه، ابعاد شبکه پیش‌خور و تعداد لایه‌های SAN روی ۱۰ درصد از مجموعه آموزش که به عنوان مجموعه اعتبارسنجی انتخاب شده‌اند، تنظیم گردیده است.

## ٤-٥ نتایج آزمایش‌ها

در این بخش، عملکرد مدل پیشنهادی با مدل‌های پایه روی مجموعه داده مقایسه شده و برای پیاده‌سازی مدل پیشنهادی از کتابخانه Pytorch[1] استفاده گردیده است. در جدول ۴ مقادیر $P@K$ و $MRR@K$ به دست آمده توسط روش‌ها به ازای $K = 5,10,20$ نمایش داده شده، هر کدام از روش‌ها پنج بار اجرا گردیده‌اند و نتایج نهایی از میانگین مقادیر به دست آمده، گزارش شده است. همان طور که نتایج نشان می‌دهند، مدل پیشنهای توانسته که با بهره‌گیری از معماری SAN و همچنین استفاده از جلسات قبلی کاربر، مقادیر $P@K$ و $MRR@K$ بالاتر و عملکرد توصیه بهتری نسبت به سایر روش‌ها داشته باشد. در ادامه به تحلیل نتایج پرداخته می‌شود.

سیستم‌های توصیه‌گر مبتنی بر جلسه به روش‌های کلاسیک و روش‌های یادگیری عمیق تقسیم‌بندی می‌شوند. روش‌های کلاسیک از ویژگی‌های کم‌عمق و ساده برای آموزش مدل توصیه و پیش‌بینی آیتم هدف استفاده می‌کنند و دقت پایینی در ایجاد توصیه 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. همچنین Item-KNN از شباهت بین آیتم‌ها برای پیش‌بینی استفاده می‌کند و با توجه به این که ترتیب تعاملات جلسه را نادیده می‌گیرد، عملکرد چندان رضایت‌بخشی ندارد. FPMC با در نظر گرفتن اطلاعات متوالی جلسه با استفاده از زنجیره مارکوف نسبت به Pop و Item-KNN نتیجه بهتری به دست آورده که اهمیت اطلاعات توالی جلسه را نشان می‌دهد. اما این روش به دلیل فرض مستقل‌بودن تعاملات متوالی کاربر در یک جلسه، دقت پایینی در ایجاد توصیه‌ها دارد. در بین روش‌های کلاسیک، ۲ روش SKNN و VSKNN نسبت به بقیه روش‌ها بهتر عمل می‌کنند و عملکردی نزدیک به مدل‌های شبکه عصبی عمیق دارند. زیرا این روش‌ها جلسات مشابه با جلسه فعلی کاربر را که شامل تعاملات مشابه با تعاملات جلسه فعلی هستند در نظر می‌گیرند، اما ترتیب آیتم‌ها و وابستگی‌های متوالی را نادیده می‌گیرند.

مدل‌های توصیه‌گر مبتنی بر یادگیری عمیق نسبت به روش‌های کلاسیک عملکرد بهتری به دست آورده‌اند. این روش‌ها قادر به استخراج ویژگی‌های عمیق آیتم‌ها در توالی‌های جلسه هستند. GRU4Rec با استفاده از شبکه‌های عصبی مکرر، 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. همچنین SR-GNN با نمایش توالی‌های جلسه به صورت داده ساختاری گراف، انتقالات پیچیده بین آیتم‌های جلسه را مدل می‌کند و نسبت به GRU4Rec نتیجه بهتری به دست آورده است. به طور کلی روش‌های مبتنی بر GNN، وابستگی بین آیتم‌های مجاور در گراف جلسه را در نظر می‌گیرند و آیتم‌های دورتر را نادیده می‌گیرند.

استفاده از شبکه‌های خودتوجه در مدل پیشنهادی منجر به افزایش دقت پیش‌بینی نسبت به سایر روش‌ها شده است. در واقع مدل پیشنهادی با استفاده از شبکه‌های خودتوجه می‌تواند وابستگی کلی بین همه آیتم‌های یک جلسه یا به عبارت دیگر علایق کلی کاربران در جلسه را مدل کند.

از طرف دیگر GRU4Rec و SR-GNN تنها از اطلاعات جلسه فعلی کاربر استفاده می‌کنند، اما مدل پیشنهادی PSR-SAN علاوه بر جلسه فعلی کاربر که بیانگر علاقه کوتاه‌مدت کاربر است، از جلسات قبلی کاربر به عنوان علاقه بلندمدت کاربر برای پیش‌بینی و ایجاد توصیه استفاده می‌کند که منجر به تقویت توصیه‌ها و بهبود قابل توجه نتایج به خصوص از نظر معیار $P@K$ شده است. این واقعیت، اهمیت اطلاعات جلسات قبلی کاربر را در ایجاد توصیه‌ها نشان می‌دهد که می‌توانند توسط ترکیب با جلسه فعلی کاربر منجر به بهبود چشم‌گیر دقت توصیه شوند.

1. https://pytorch.org/

جدول ۴: مقایسه دقت روش پیشنهادی با سایر روش‌ها.

| روش/ معیار | $P@۵$ | $P@۱۰$ | $P@۲۰$ | $MRR@۵$ | $MRR@۱۰$ | $MRR@۲۰$ |
|---|---|---|---|---|---|---|
| Pop | ۱۳٫۲۲ | ۱۹٫۴۶ | ۲۶٫۴۷ | ۸٫۵۰ | ۹٫۳۲ | ۹٫۸۲ |
| Item-KNN | ۲۱٫۷۱ | ۳۰٫۳۲ | ۳۸٫۸۵ | ۱۱٫۷۴ | ۱۲٫۸۸ | ۱۳٫۴۹ |
| FPMC | ۲۹٫۹۱ | ۳۴٫۳۱ | ۴۴٫۳۲ | ۸٫۷۸ | ۶٫۵۶ | ۴٫۵۴ |
| SKNN | ۳۴٫۲۹ | ۴۲٫۱۷ | ۴۹٫۶۸ | ۱۹٫۱۱ | ۲۰٫۱۶ | ۲۰٫۶۸ |
| VSKNN | ۳۴٫۲۵ | ۴۲٫۱۷ | ۴۹٫۶۷ | ۱۹٫۰۹ | ۲۰٫۱۴ | ۲۰٫۶۷ |
| GRU4Rec | ۳۳٫۷۲ | ۴۱٫۷۳ | ۵۰٫۰۴ | ۲۴٫۳۶ | ۲۵٫۴۲ | ۲۶٫۰۰ |
| SR-GNN | ۳۴٫۹۶ | ۴۲٫۳۸ | ۵۰٫۳۳ | ۲۵٫۹۰ | ۲۶٫۸۸ | ۲۷٫۴۴ |
| **PSR-SAN** | **۴۰٫۱۲** | **۴۹٫۴۶** | **۵۸٫۹۵** | **۲۶٫۳۱** | **۲۷٫۵۴** | **۲۸٫۲۷** |

## ۵- نتیجه‌گیری و پیشنهادها

در این پژوهش، یک مدل توصیه‌گر مبتنی بر جلسه شخصی‌سازی شده با شبکه‌های خودتوجه ارائه گردیده است. مدل پیشنهادی با استفاده از شبکه‌های خودتوجه، وابستگی کلی بین همه آیتم‌های یک جلسه را مدل می‌کند و از بردار نهفته آخرین آیتم در جلسه که بیانگر علاقه فعلی و همچنین علاقه کلی کاربر در آن جلسه است، به عنوان نماینده جلسه استفاده می‌کند. علاوه بر این، در مدل پیشنهادی برای تقویت توصیه‌ها و استخراج علایق کوتاه‌مدت و بلندمدت کاربر از ترکیب وزنی جلسه فعلی و جلسات قبلی اخیر کاربر استفاده می‌شود. مدل پیشنهادی روی یک مجموعه داده واقعی ارزیابی شده است. نتایج آزمایش‌ها و مقایسه مدل پیشنهادی با دیگر روش‌ها نشان می‌دهد که مدل پیشنهادی توانسته که با استفاده هم‌زمان از اطلاعات جلسه فعلی و جلسات قبلی کاربر، دقت بسیار بالایی در پیش‌بینی اقدام بعدی کاربر و ارائه توصیه‌ها به دست آورد.

در ادامه پژوهش برای به دست آوردن دقت و اثربخشی بالاتر می‌توان علاوه بر بردار تعبیه خام آیتم‌ها، ویژگی‌های محتوایی آیتم‌ها از قبیل اطلاعات توصیف و دسته‌بندی آیتم را هم در نظر گرفت و با بردار تعبیه خام آیتم‌ها ترکیب کرد. همچنین می‌توان به بررسی تأثیر طول جلسه و تعداد جلسات قبلی متفاوت کاربر روی بهبود کارایی و دقت مدل توصیه‌گر مبتنی بر جلسه پرداخت.

## مراجع


[1] R. Salakhutdinov, A. Mnih, and G. Hinton, "Restricted Boltzmann machines for collaborative filtering," in *Proc. of the 24th Int. Conf. on Machine Learning*, *ICML'07*, pp. 791-798, Corvallis, OR, USA, 20-24 Jun. 2007.

[2] 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, May 2021.

[3] G. Shani, D. Heckerman, R. I. Brafman, and C. Boutilier, "An MDP-based recommender system," *J. of Machine Learning Research*, vol. 6, no. 9, pp. 1265-1295, 2005.

[4] S. Rendle, C. Freudenthaler, and L. Schmidt-Thieme, "Factorizing personalized markov chains for next-basket recommendation," in *Proc. of the 19th Int.Conf. on World Wide Web*, *WWW '10,* pp. 811-820, Raleigh, NC, USA, 26-30 Apr. 2010.

[5] B. Hidasi, A. Karatzoglou, L. Baltrunas, and D. Tikk, "Session-based recommendations with recurrent neural networks," in *Proc. of the 4th Int. Conf. on Learning Representations, ICLR'16*, 10 pp., San Juan, Puerto Rico, 2-4 May 2016.

[6] Y. K. Tan, X. Xu, and Y. Liu, "Improved recurrent neural networks for session-based recommendations," in *Proc. of the 1st Workshop on Deep Learning for Recommender Systems*, DLRS'16, pp. 17-22, Boston, MA, USA, 15-15 Sept. 2016.

[7] D. Jannach and M. Ludewig, "When recurrent neural networks meet the neighborhood for session-based recommendation," in *Proc. of the 11th ACM Conf. on Recommender Systems*, *RecSys'17*, pp. 306-310, Como, Italy, 27-31 Aug. 2017.

[8] J. Li, P. Ren, Z. Chen, Z. Ren, T. Lian, and J. Ma, "Neural attentive session-based recommendation," in *Proc. of the ACM Conf. on Information and Knowledge Management, CIKM'17*, pp. 1419-1428, Singapore, 6-10 Nov. 2017.

[9] M. Quadrana, A. Karatzoglou, B. Hidasi, and P. Cremonesi, "Personalizing session-based recommendations with hierarchical recurrent neural networks," in *Proc. of the 11th ACM Conf. on Recommender Systems*, *RecSys'17*, pp. 130-137, Como, Italy, 27-31 Aug. 2017.

[10] Z. Zhang and B. Wang, "Learning sequential and general interests via a joint neural model for session-based recommendation," *Neurocomputing*, vol. 415, pp. 165-173, 20 Nov. 2020.

[11] S. Wu, et *al.*, "Session-based recommendation with graph neural networks," in *Proc. of the AAAI Conf. on Artificial Intelligence*, pp. 346-353, Honolulu, HI, USA, 27 Jan.-1 Feb. 2019.

[12] F. Yu, Y. Zhu, Q. Liu, S. Wu, L. Wang, and T. Tan, "TAGNN: target attentive graph neural networks for session-based recommendation," in *Proc. of the 43rd Int ACM SIGIR Conf. on Research and Development in Information Retrieval*, pp. 1921-1924, Xian, China, 25- 30 Jul. 2020.

[13] T. Chen and R. C. W. Wong, "Handling information loss of graph neural networks for session-based recommendation," in *Proc. of the 26th ACM SIGKDD Inte. Conf. on Knowledge Discovery & Data Mining*, pp. 1172-1180, San Diego, California, 6-10 Jul. 2020.

[14] A. Vaswani, *et al.*, "Attention is all you need," in *Proc. of the 31st Conf. on Neural Information Processing Systems, NIPS'17*, pp. 5998-6008, Long Beach, CA, USA, 4-9 Dec. 2017.

[15] J. Fang, *Session-Based Recommendation with Self-Attention Networks*, arXiv preprint arXiv:2102.01922, Feb. 2021.

[16] P. H. Anh, N. X. Bach, and T. M. Phuong, "Session-based recommendation with self-attention," in *Proc. of the 10th Int. Symp. on Information and Communication Technology*, *SoICT'19*, 8 pp., Ha Long Bay, Vietnam, 4-6 Dec. 2019.

[17] R. Mehta and K. Rana, "A review on matrix factorization techniques in recommender systems," in *Proc. of the 2nd Int. Conf. on Communication Systems, Computing and IT Applications, CSCITA'17*, pp. 269-274, Mumbai, India, 7-8 Apr. 2017.

[18] R. Salakhutdinov and A. Mnih, "Probabilistic matrix factorization," in *Proc. of the 20th Int. Conf. on Neural Information Processing Systems, NIPS'07,* pp. 1257-1264,Vancouver,Canada, 3-6 Dec. 2007.

[19] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms," in *Proc. of the 10th Int. Conf. on World Wide Web*, *WWW'01*, pp. 285-295, Hong Kong, China, 1-5 May 2001.

[20] R. He and J. McAuley, "Fusing similarity models with markov chains for sparse sequential recommendation," in *Proc. of the 16th Int. Conf. on Data Mining, ICDM'16*, pp. 191-200, Barcelona, Spain, 12-15 Dec. 2016.

[21] S. Wang, L. Hu, Y. Wang, L. Cao, Q. Z. Sheng, and M. Orgun, "Sequential recommender systems: challenges, progress and prospects," in *Proc. of the Int. Joint Conf. on Artificial Intelligence, IJCAI'19*, pp. 6332-6338, Macao, China,10-16 Aug. 2019.

[22] Q. Liu, Y. Zeng, R. Mokhosi, and H. Zhang, "STAMP: short-term attention/memory priority model for session-based recommendation," in *Proc. of the 24th ACM SIGKDD Int. Conf. on Knowledge Discovery & Data Mining*, pp. 1831-1839, London, United Kingdom, 19-23 Aug. 2018.

[23] M. Wang, *et al*., "A collaborative session-based recommendation approach with parallel memory modules," in *Proc. of the 42nd Int. ACM SIGIR Conf. on Research and Development in Information Retrieval*, pp. 345-354, Paris, France, 21-25 Jul. 2019.

[24] C. Xu, *et al.*, "Graph contextualized self-attention network for session-based recommendation," in *Proc. of the 28th Int. Joint Conf.*

*on Artificial Intelligence, IJCAI'19*, pp. 3940-3946, Macao, China, 10-16 Aug. 2019.

[25] J. Wang, Q. Xu, J. Lei, C. Lin, and B. Xiao, "PA-GGAN: session-based recommendation with position-aware gated graph attention network," in *Proc. of the IEEE Int. Conf. on Multimedia and Expo, ICME'20*, 6 pp., London, UK, 6-10 Jul. 2020.

[26] M. Ruocco, O. S. L. Skrede, and H. Langseth, "Inter-session modeling for session-based recommendation," in *Proc. of the 2nd Workshop on Deep Learning for Recommender Systems*, DLRS17, pp. 24-31, Como, Italy, 27-27 Aug. 2017.

[27] J. You, Y. Wang, A. Pal, P. Eksombatchai, C. Rosenburg, and J. Leskovec, "Hierarchical temporal convolutional networks for dynamic recommender systems," in *Proc. of the World Wide Web Conf., WWW'19*, pp. 2236-2246, San Francisco, CA, USA, 13-17 May 2019.

[28] M. Ludewig and D. Jannach, "Evaluation of session-based recommendation algorithms," *User Modeling and User-Adapted Interaction*, vol. 28, no. 4-5, pp. 331-390, Dec. 2018.

**اعظم رمضانی** در سال ۱۳۹۱ مدرک کارشناسی مهندسی کامپیوتر خود را از دانشگاه صنعتی بیرجند و در سال ۱۳۹۳ مدرک کارشناسی ارشد مهندسی فناوری اطلاعات خود را از دانشگاه بیرجند دریافت نمود. وی در سال ۱۳۹۸ به دوره دکترای مهندسی کامپیوتر گرایش نرم‌افزار در دانشگاه یزد وارد گردید و هم اکنون مشغول به تحصیل می‌باشد. زمینه‌های علمی مورد علاقه ایشان شامل موضوعاتی مانند وب‌کاوی و موتورهای جستجو، سیستم‌های توصیه‌گر، شخصی‌سازی، یادگیری ماشین و یادگیری عمیق می‌باشد.

**علی محمد زارع بیدکی** تحصیلات خود را در مقطع کارشناسی در سال ۱۳۷۸ از دانشگاه صنعتی اصفهان و مقاطع کارشناسی ارشد و دکتری کامپیوتر به‌ترتیب در سال‌های ۱۳۸۱ و ۱۳۸۸ از دانشکده فنی دانشگاه تهران به پایان رسانده است و هم اکنون عضو هیأت علمی دانشکده مهندسی کامپیوتر دانشگاه یزد می‌باشد. زمینه‌های تحقیقاتی مورد علاقه ایشان شامل بازیابی اطلاعات، موتورهای جستجو، رتبه‌بندی و پردازش زبانهای طبیعی می‌باشد.