

توسعه سیستم پیشنهاددهی فیلم چندوجهی با استفاده از شبکه‌های عصبی گراف و همجوشی ویژگی‌های متنی و تصویری

علی مصدق¹، دانیال براتی² و کیارش فضیلت³

¹ دانشجوی ارشد مهندسی کامپیوتر، واحد تهران غرب، دانشگاه آزاد اسلامی، تهران، ایران ali.mosaddegh@yahoo.com

² دانشجوی ارشد مهندسی کامپیوتر، واحد تهران غرب، دانشگاه آزاد اسلامی، تهران، ایران danial.barati1999@gmail.com

³ دانشجوی ارشد مهندسی کامپیوتر، واحد تهران غرب، دانشگاه آزاد اسلامی، تهران، ایران kiarashfz12424@gmail.com

چکیده

این پژوهش به طراحی و توسعه یک سیستم پیشنهاددهی چندوجهی فیلم با استفاده از شبکه‌های عصبی گرافی (GCN) پرداخته است. هدف اصلی این سیستم، بهبود دقت و کیفیت توصیه‌ها از طریق ترکیب اطلاعات چندوجهی شامل ویژگی‌های متنی و تصویری فیلم‌ها بود. در این مدل، گراف تعاملات کاربر-فیلم به عنوان ساختار اصلی مورد استفاده قرار گرفت و ارتباطات میان کاربران و فیلم‌ها به کمک گره‌ها و لبه‌های گراف مدل‌سازی شد. ویژگی‌های متنی فیلم‌ها با مدل‌های تعبیه‌سازی و ویژگی‌های تصویری با استفاده از شبکه‌های عصبی پیچشی استخراج و سپس در گره‌های گراف ترکیب شدند. شبکه عصبی گرافی برای یادگیری ویژگی‌های تعاملی و پیش‌بینی ترجیحات کاربران به کار گرفته شد. نتایج آزمایش‌ها نشان داد که مدل پیشنهادی، با وجود نوسانات در مقادیر خطا و میانگین مربعات خطا (MSE)، به بهبود نسبی دقت و همگرایی مدل نسبت به روش‌های پایه دست یافته است. گراف تعاملات نیز نشان‌دهنده تنوع سلیق کاربران و اهمیت برخی فیلم‌های پرتعامل بود. این پژوهش همچنین پیشنهادهایی برای بهبود مدل شامل استفاده از داده‌های واقعی، الگوریتم‌های پیشرفته‌تر همجوشی و بهبود تفسیرپذیری ارائه می‌کند. مدل پیشنهادی می‌تواند مبنایی برای طراحی سیستم‌های توصیه‌گر پیشرفته‌تر و شخصی‌سازی شده‌تر باشد.

کلید واژه:

سیستم پیشنهاددهی، شبکه‌های عصبی گرافی، داده‌های چندوجهی، توصیه‌گر فیلم، همجوشی ویژگی‌ها.

مقدمه

در سال‌های اخیر، حجم عظیمی از داده‌های دیجیتال و چندرسانه‌ای مانند فیلم، موسیقی و تصاویر، دسترسی کاربران به محتوا را به شدت افزایش داده است [1]. سیستم‌های پیشنهاددهی به عنوان راه‌حلی مؤثر برای هدایت کاربران به سمت محتوای مناسب و افزایش تجربه کاربری مطرح شده‌اند [2]. این سیستم‌ها به کاربران کمک می‌کنند تا از میان حجم عظیم داده‌ها، محتوای مرتبط و متناسب با سلیق شخصی خود را پیدا کنند. با رشد فناوری‌های یادگیری عمیق و پردازش زبان طبیعی، روش‌های مختلفی برای بهبود دقت و شخصی‌سازی پیشنهادها توسعه یافته‌اند [3,4].

بسیاری از تحقیقات اخیر نشان داده‌اند که رویکردهای چندوجهی، که از ترکیب منابع داده‌ای مختلف مانند متون، تصاویر و سایر ویژگی‌های محتوایی استفاده می‌کنند، دقت بیشتری در ارائه پیشنهادها شخصی‌سازی شده دارند. به‌خصوص در حوزه فیلم و محتوای ویدئویی، ترکیب ویژگی‌های متنی مانند ژانر و خلاصه داستان و ویژگی‌های تصویری مانند پوستر یا صحنه‌های کلیدی فیلم می‌تواند نقشی مهم در بهبود کیفیت پیشنهادها ایفا کند. روش‌های نوین مانند شبکه‌های عصبی گراف (GNN) و ترانسفورمرها توانایی ویژه‌ای در همجوشی و تحلیل اطلاعات چندوجهی دارند [4]. شبکه‌های عصبی گراف با امکان پردازش داده‌های غیراقلیدسی و استفاده از ساختارهای گرافی مانند ارتباطات کاربر-فیلم، در سیستم‌های پیشنهاددهی جدید به کار گرفته شده‌اند و عملکرد چشمگیری را نشان داده‌اند [3].

با این حال، هنوز چالش‌هایی نظیر چگونگی همگام‌سازی بهینه اطلاعات متنی و تصویری، به‌خصوص در سیستم‌های پیشنهاددهی فیلم، به‌طور کامل حل نشده است [5]. اکثر روش‌های فعلی تمرکز بر روی یک یا دو نوع ویژگی دارند و به ندرت از ترکیب چندوجهی با تمرکز بر روابط گرافی و ساختارهای کاربر-محتوا استفاده می‌کنند. همچنین، تفسیرپذیری پیشنهادها برای کاربران و امکان ارائه توضیحاتی در مورد دلایل هر پیشنهاد از جنبه‌های مهمی است که کمتر به آن توجه شده است [3،5].

در این مقاله، با هدف ارتقای دقت و تنوع پیشنهادها و همچنین بهبود تفسیرپذیری پیشنهادها، یک سیستم پیشنهاددهی فیلم چندوجهی مبتنی بر شبکه‌های عصبی گراف معرفی شده است. این سیستم با استفاده از همجوشی داده‌های متنی و تصویری، به‌گونه‌ای طراحی شده است که امکان اجرای آن با داده‌های شبیه‌سازی شده و در محیط گوگل کولب فراهم باشد. نتایج نشان می‌دهند که ترکیب ویژگی‌های متنی و تصویری از طریق شبکه‌های عصبی گراف می‌تواند به پیشنهادهای شخصی‌سازی‌شده‌تر و دقیق‌تر منجر شود و تجربه کاربری بهتری را فراهم کند.

کارهای گذشته

ژی¹ و همکاران [1]، در مطالعه‌ای با عنوان *Movie Recommendation with Poster Attention via Multi-modal Transformer* به معرفی سیستمی چندوجهی برای پیشنهاددهی فیلم پرداخته‌اند که از ویژگی‌های پوستر فیلم و توضیحات متنی فیلم برای پیش‌بینی ترجیحات کاربران استفاده می‌کند. در این تحقیق از مدل BERT برای استخراج ویژگی‌های متنی و از مدل ViT برای ویژگی‌های تصویری استفاده شده است. ترکیب این مدل‌ها و استفاده از معماری ترانسفورمر برای همجوشی ویژگی‌ها منجر به افزایش دقت در پیش‌بینی امتیازات کاربران شده است. نتایج این مدل از طریق آزمون بر روی مجموعه داده MovieLens 100K و IM1 اثبات شده است که دقت پیش‌بینی را نسبت به الگوریتم‌های پایه افزایش داده است.

وو² و همکاران [2]، در مقاله‌ای با عنوان *Towards Bridging the Cross-modal Semantic Gap for Multi-modal Recommendation* به بررسی چالش‌های مربوط به شکاف معنایی بین مودال‌ها در سیستم‌های پیشنهاددهی چندوجهی پرداخته‌اند. این پژوهش با الهام از مدل CLIP، به توسعه چارچوبی به نام CLIPER پرداخته است که با استفاده از هم‌ترازی نمایه‌های مختلف، قابلیت استخراج اطلاعات چنددیده‌گاهی را فراهم می‌آورد. این رویکرد توانسته است عملکرد بهتری نسبت به مدل‌های چندوجهی فعلی در سه مجموعه داده مختلف نشان دهد.

توکال³ و همکاران [3]، در کنفرانس بین‌المللی مدل‌سازی محاسباتی، شبیه‌سازی و بهینه‌سازی (ICCMSO)، پژوهشی با عنوان *Enhanced Movie Recommender System Using Deep Learning Techniques* ارائه کردند. در این تحقیق، از تکنیک‌های یادگیری عمیق شامل شبکه‌های عصبی مصنوعی (ANN) برای استخراج ویژگی‌ها از رفتار کاربران و متادیتای فیلم، شبکه‌های عصبی بازگشتی (RNN) برای درک الگوهای زمانی، و شبکه‌های عصبی پیچشی (CNN) برای تحلیل همبستگی‌های مکانی داده‌ها استفاده شده است. هدف این مدل، بهبود دقت پیشنهاددهی از طریق ترکیب ویژگی‌های کوتاه‌مدت و بلندمدت در ترجیحات کاربر است.

مالیتستا⁴ و همکاران [4]، در مقاله‌ای با عنوان *Formalizing Multimedia Recommendation through Multimodal Deep Learning* در نشریه ACM، به بررسی سیستم‌های پیشنهاددهی چندرسانه‌ای از طریق یادگیری عمیق چندوجهی پرداخته‌اند. این پژوهش به بررسی

¹ Xia

² Wu

³ Tokala

⁴ Malitesta

چالش‌های مربوط به پیشنهاددهی در حوزه‌هایی مانند مد و موسیقی پرداخته و استفاده از تکنیک‌های چندوجهی را به عنوان راهکاری برای ارائه پیشنهاددهی دقیق‌تر معرفی کرده است. محققان در این مطالعه به بازنگری روش‌های چندوجهی در پیشنهاددهی چندرسانه‌ای پرداخته و الگوریتم‌های اخیر را در چارچوبی به نام Elliot مورد ارزیابی قرار داده‌اند. این مطالعه با هدف ارائه دستورالعمل‌هایی برای طراحی و پیاده‌سازی نسل بعدی سیستم‌های پیشنهاددهی چندوجهی انجام شده است.

بوراباک⁵ و آیتکین⁶، در مطالعه‌ای با عنوان *SynerGraph: An Integrated Graph Convolution Network for Multimodal Recommendation* یک رویکرد جدید برای سیستم‌های پیشنهاددهی چندوجهی با تمرکز بر ادغام و پالایش داده‌های چندوجهی معرفی کردند. این پژوهش نشان داده است که استفاده از فیلترهای پالایشی، دقت سیستم‌های چندوجهی را نسبت به مدل‌های تک‌وجهی بهبود می‌بخشد و اطلاعات متنی نقشی کلیدی در افزایش دقت پیشنهاددهی ایفا می‌کند.

مبوروک⁷ و همکاران⁶، در مقاله‌ای تحت عنوان *Enhancing Movie Recommendations: A Deep Neural Network Approach with MovieLens Case Study* در کنفرانس IWCMC، به بررسی بهبود سیستم‌های پیشنهاددهی فیلم با استفاده از شبکه‌های عصبی عمیق پرداخته‌اند. این پژوهش، چالش‌هایی نظیر مسئله شروع سرد، پراکندگی داده‌ها و کمبود بازخورد صریح کاربران را مورد توجه قرار داده و از شبکه‌های عصبی عمیق (DNN)، شبکه‌های عصبی پیچشی (CNN)، شبکه‌های عصبی بازگشتی (RNN) و خودمزمگذارها (AEs) بهره برده است. هدف این پژوهش، بهبود عملکرد سیستم‌های پیشنهاددهی است که به بازخوردهای ضمنی کاربران تکیه دارند و این مدل بر اساس معیارهایی مانند نسبت موفقیت و نمرات مربوطه (NDCG) ارزیابی شده است که نشان‌دهنده برتری این روش نسبت به روش‌های قبلی است. هی⁸ و همکاران⁷، در مقاله‌ای با عنوان *Multi-modal Bayesian Recommendation System* که در کنفرانس IMCEC ارائه شد، سیستم پیشنهاددهی چندوجهی با نام MBR را معرفی کرده‌اند که از مدالیتهای تصویر و متن برای بهبود کیفیت پیشنهادها استفاده می‌کند. این سیستم از شبکه‌های عصبی پیچشی عمیق برای استخراج ویژگی‌های تصویری و از مدل‌های زبان برای تحلیل متنی بهره برده است. آزمایش‌ها بر روی یک مجموعه داده بزرگ مقیاس نشان داده که سیستم MBR در ارائه پیشنهادها بهبود یافته است و عملکرد بالایی دارد.

وی⁹ و همکاران⁸، در مقاله *Multi-view Sequence Recommendation Model* که در کنفرانس IMCEC ارائه شده است، به بررسی الگوریتم‌های پیشنهاددهی مبتنی بر یادگیری عمیق و آگاه از بافت پرداخته‌اند. این پژوهش، تاثیرات زمینه‌ای مانند زمان، مکان و محیط اجتماعی بر رفتار کاربران را در نظر می‌گیرد و به دنبال توسعه سیستمی است که پروفایل‌های دقیق‌تری از کاربران ارائه دهد. نتایج این مطالعه نشان می‌دهد که استفاده از اطلاعات زمینه‌ای و آگاهی از صحنه می‌تواند به پیشنهادهایی دقیق‌تر و مناسب‌تر منجر شود.

سایت¹⁰ و همکاران⁹، در مقاله *Enhancing Sequence Movie Recommendation System Using Deep Learning and KMeans* که در مجله Applied Sciences منتشر شده است، یک سیستم پیشنهاددهی را معرفی کرده‌اند که با ترکیب یادگیری عمیق و خوشه‌بندی KMeans، به بهبود دقت در توصیه‌های فیلم می‌پردازد. این سیستم با استفاده از اطلاعات کاربران نظیر سن، جنسیت و شغل، کاربران جدید را تحلیل کرده و با کاربران با ترجیحات مشابه مطابقت می‌دهد. مدل پیشنهاددهی با به‌کارگیری شبکه‌های ترنسفورمری و چندلایه ادراکی، دقت پیش‌بینی را افزایش داده و با استفاده از خوشه‌بندی KMeans بر روی ویژگی‌های ژانر فیلم‌ها، تنوع در پیشنهادها را نیز حفظ کرده است. ارزیابی مدل بر روی مجموعه داده MovieLens نشان از برتری آن در مقایسه با مدل‌های پایه دارد.

کیم¹¹ و همکاران¹⁰، در مقاله‌ای با عنوان *Multi-Modal Deep Learning Based Metadata Extensions for Video Clipping* که در مجله International Journal on Advanced Science, Engineering & Information Technology منتشر شده، یک مدل توسعه متادیتا برای ویدیوها با استفاده از یادگیری عمیق چندوجهی معرفی کرده‌اند. این مدل با شناسایی اشیا و تبدیل گفتار به متن (STT) متادیتای

⁵ Burabak

⁶ Aytakin

⁷ Mabrouk

⁸ He

⁹ Wei

¹⁰ Siet

¹¹ Kim

ویدیوها را گسترش می‌دهد. این توسعه می‌تواند به سیستم‌های جست‌وجو و پیشنهاددهی ویدیو کمک کند تا نتایج نزدیک‌تری به عبارات جست‌وجو و محتوای مرتبط به کاربران ارائه دهند.

زان^{۱۲} و همکاران [11]، در مقاله *Research on Movie Recommendation Algorithm Based on Deep Learning* که در کنفرانس ICICACS ارائه شده، از یادگیری عمیق برای بهبود دقت الگوریتم پیشنهاددهی فیلم استفاده کرده‌اند. این مدل با بهره‌گیری از سیستم عصبی مصنوعی و الگوریتم انتشار پسرو برای بهینه‌سازی پارامترهای شبکه، دقت پیش‌بینی را بهبود بخشیده است. نتایج نشان داده که این مدل می‌تواند نیازهای شخصی‌سازی شده کاربران را به خوبی برآورده کند و دقت آن نسبت به الگوریتم‌های سنتی 1.4٪ بهبود یافته است.

پنگ^{۱۳} و همکاران [12]، در مقاله *Integration of Deep Reinforcement Learning with Collaborative Filtering for Movie Recommendation Systems* در مجله *Applied Sciences*، سیستم پیشنهاددهی فیلم را معرفی کرده‌اند که ترکیبی از یادگیری تقویتی عمیق (DRL) و فیلترگذاری مشارکتی (CF) است. این سیستم با استفاده از الگوریتم DDPG و تحلیل مقدار ویژه (SVD)، مشکلاتی نظیر کمبود داده و مسئله شروع سرد را بهبود می‌بخشد. ارزیابی مدل با معیارهایی نظیر دقت، بازخوانی و نمره F1، نشان از برتری این روش نسبت به مدل‌های مرجع پیشنهاددهی دارد.

لی^{۱۴} و همکاران [13]، در مقاله‌ای با عنوان *Graph Neural Networks with Deep Mutual Learning for Designing Multi-modal Recommendation Systems* که در مجله *Information Sciences* منتشر شده، چارچوبی با نام GNNMR را معرفی کرده‌اند که شبکه‌های عصبی گراف (GNN) را با تکنیک یادگیری متقابل عمیق ترکیب می‌کند. این چارچوب به هر مدالیت خاص یک گراف دوطرفه اختصاص می‌دهد و از این طریق به استخراج روابط معنایی پنهان بین مدالیت‌ها کمک می‌کند. نتایج تجربی نشان‌دهنده برتری این مدل در مقایسه با سایر مدل‌های چندوجهی در وظیفه پیشنهاددهی Top-K است.

جدول 1. خلاصه پیشینه پژوهش

نتیجه	روش	موضوع	سال	محققان
بهبود دقت در مقایسه با مدل‌های پایه در MovieLens	ترکیب یادگیری عمیق و خوشه‌بندی KMeans برای بهبود دقت در توصیه‌های فیلم	سیستم پیشنهاددهی فیلم با استفاده از یادگیری عمیق و خوشه‌بندی KMeans	2024	سایت و همکاران
بهبود جست‌وجو و پیشنهاددهی در سیستم‌های ویدئویی	استفاده از شناسایی اشیا و تبدیل گفتار به متن برای گسترش متادیتا	توسعه متادیتا برای ویدیوها با استفاده از یادگیری عمیق چندوجهی	2024	کیم و همکاران
بهبود 1.4٪ در دقت پیش‌بینی نسبت به روش‌های سنتی	استفاده از سیستم عصبی مصنوعی و الگوریتم انتشار پسرو	الگوریتم پیشنهاددهی فیلم با استفاده از یادگیری عمیق	2024	زان و همکاران
افزایش دقت و شخصی‌سازی پیشنهادها در مجموعه MovieLens	ترکیب یادگیری تقویتی عمیق (DRL) و تحلیل مقدار ویژه (SVD) با فیلترگذاری مشارکتی	سیستم پیشنهاددهی فیلم با ترکیب یادگیری تقویتی عمیق و فیلترگذاری مشارکتی	2024	پنگ و همکاران

¹² Zhan

¹³ Peng

¹⁴ Li

لی و همکاران	2024	سیستم پیشنهاددهی چندوجهی با ترکیب شبکه‌های عصبی عمیق و یادگیری متقابل عمیق	ترکیب شبکه‌های عصبی گراف با یادگیری متقابل عمیق (GNNMR)	برتری در وظیفه پیشنهاددهی Top-K نسبت به سایر مدل‌های چندوجهی
مبوروک و همکاران	2024	بهبود سیستم‌های پیشنهاددهی فیلم با شبکه‌های عصبی عمیق	استفاده از شبکه‌های عصبی عمیق (DNN)، CNN، و RNN AEs	برتری در معیارهای NDCG و نسبت موفقیت در مقایسه با روش‌های قبلی
هی و همکاران	2024	سیستم پیشنهاددهی چندوجهی با استفاده از شبکه‌های عصبی پیچشی و مدل‌های زبان	استفاده از شبکه‌های عصبی پیچشی و مدل‌های زبان برای تحلیل متنی	بهبود عملکرد در ارائه پیشنهاددهی بهینه‌تر با استفاده از تصاویر و متن
وی و همکاران	2024	الگوریتم پیشنهاددهی توالی چنددیدگاهی با یادگیری عمیق آگاه از بافت	توسعه سیستم آگاه از بافت با استفاده از تحلیل زمانی، مکانی و محیط اجتماعی	ارائه پروفایل‌های دقیق‌تر کاربران و توصیه‌های دقیق‌تر
مالیتستا و همکاران	2024	فرمالیزه کردن پیشنهاددهی چندرسانه‌ای با یادگیری عمیق چندوجهی	مرور روش‌های چندوجهی و ارزیابی آن‌ها در چارچوب Elliot	راهنمایی برای طراحی نسل بعدی سیستم‌های پیشنهاددهی چندوجهی
ژی و همکاران	2024	پیشنهاددهی فیلم با استفاده از پوستر و ترنسفورمر چندوجهی	استفاده از BERT و ViT برای استخراج ویژگی‌های پوستر و متن	افزایش دقت پیش‌بینی امتیازات کاربران نسبت به الگوریتم‌های پایه
وو و همکاران	2024	سیستم پیشنهاددهی چندوجهی با استفاده از چارچوب CLIPER	استفاده از چارچوب CLIPER برای کاهش شکاف معنایی بین مدالیت‌ها	عملکرد بهتر نسبت به مدل‌های چندوجهی در سه مجموعه داده مختلف
توکال و همکاران	2024	سیستم پیشنهاددهی فیلم بهبود یافته با تکنیک‌های یادگیری عمیق	استفاده از CNN، ANN و RNN برای ترکیب ویژگی‌های کوتاه‌مدت و بلندمدت	بهبود دقت در ترکیب ویژگی‌های کوتاه‌مدت و بلندمدت کاربران
مالیتستا	2024	شبکه‌های عصبی گراف برای پیشنهاددهی چندوجهی	شبکه‌های عصبی گراف برای استخراج روابط در نمودار کاربر-آیتم	بهبود ساختار و دقت پیشنهاددهی از طریق نمودارهای کاربر-آیتم
بوراباک و آیکتین	2024	سیستم پیشنهاددهی چندوجهی با شبکه گرافی و پالایش داده‌ها	ادغام و پالایش داده‌های چندوجهی با فیلترهای پالایشی	بهبود دقت پیشنهاددهی و افزایش نقش اطلاعات متنی

بررسی مطالعات انجام شده نشان می‌دهد که در سال‌های اخیر رویکردهای مختلفی در زمینه سیستم‌های پیشنهاددهی چندوجهی و مبتنی بر یادگیری عمیق توسعه یافته‌اند. هر چند بسیاری از تحقیقات بر بهبود دقت و شخصی‌سازی پیشنهادها تمرکز کرده‌اند، برخی چالش‌های مهم همچنان به‌طور کامل حل نشده‌اند. برای مثال، پژوهش‌هایی که از روش‌های ترکیبی مانند یادگیری تقویتی عمیق و فیلترگذاری مشارکتی بهره

برده‌اند، بیشتر بر افزایش دقت در شرایط کمبود داده و مسئله شروع سرد تمرکز کرده‌اند، در حالی که چالش‌های دیگری نظیر یکپارچه‌سازی کامل داده‌های چندوجهی و بهینه‌سازی کارایی سیستم در زمان واقعی کمتر مورد توجه بوده‌اند. اگرچه روش‌های مختلفی برای پردازش تصاویر پزشکی، مانند شبکه‌های عصبی پیچشی سبک برای تشخیص آلزایمر [15] و معماری‌های U-Net برای بخش‌بندی تصاویر MR [16] پیشنهاد شده‌اند، اما این رویکردها اغلب در مواجهه با داده‌های پویا و چندوجهی، قابلیت انطباق کافی ندارند. علاوه بر این، الگوریتم‌های بهینه‌سازی مانند «انطباق پیش‌رو» کارایی آموزش مدل‌های یادگیری عمیق را بهبود داده‌اند [17]، اما کاربرد آن‌ها در سیستم‌های پیشنهاددهی چندوجهی هنوز مورد بررسی قرار نگرفته است. این مطالعه با ادغام تکنیک‌های پیشرفته بهینه‌سازی و همجوشی داده‌های چندوجهی، به بررسی این شکاف و افزایش دقت پیشنهاددهی می‌پردازد. بسیاری از مطالعات نیز به استفاده از روش‌های پیشرفته‌ای چون شبکه‌های عصبی گراف و یادگیری متقابل عمیق برای بهبود سیستم‌های چندوجهی پرداخته‌اند. اما موضوعاتی مانند تفسیرپذیری و شفافیت پیشنهادهای ارائه‌شده و چگونگی همجوشی و ترکیب مناسب ویژگی‌های متن، تصویری و زمانی به‌طور کامل بررسی نشده است. این مسئله به‌خصوص در حوزه‌هایی که کاربران نیازمند فهمیدن دلایل پیشنهادها هستند، اهمیت بیشتری پیدا می‌کند. علاوه بر این، در حالی که برخی تحقیقات از روش‌های جدید مانند مدل‌های ترنسفورمری و شبکه‌های پیچشی استفاده کرده‌اند، کمبود مطالعاتی که به ترکیب این روش‌ها با روش‌های سنتی برای افزایش تنوع پیشنهادها توجه کنند، مشهود است.

روش شناسی

در این مقاله، یک سیستم پیشنهاددهی چندوجهی فیلم مبتنی بر شبکه‌های عصبی گراف (GNN) ارائه می‌شود که داده‌های متن (مانند ژانر و خلاصه داستان) و داده‌های تصویری (مانند پوستر فیلم) را برای ارائه پیشنهادهای شخصی‌سازی‌شده ترکیب می‌کند. این روش شامل مراحل اصلی زیر است: پیش‌پردازش داده‌ها، استخراج ویژگی‌های چندوجهی، ساختاردهی گراف و استفاده از شبکه عصبی گرافی برای پردازش و همجوشی داده‌ها.

۱. پیش‌پردازش داده‌ها

داده‌های متن، شامل ژانر و توضیحات کوتاه فیلم، ابتدا به توکن‌های متن تبدیل شده و با استفاده از مدل تعبیه‌سازی متن BERT به بردارهای ویژگی تبدیل می‌شوند. داده‌های تصویری، مانند پوستر فیلم، به شبکه‌های عصبی پیچشی (CNN) داده می‌شوند تا ویژگی‌های تصویری مهم استخراج شوند.

۲. ساختاردهی گراف

در این مرحله، شبکه‌ای گرافی تشکیل می‌شود که گره‌های آن نشان‌دهنده کاربران و فیلم‌ها است. لبه‌های گراف نمایانگر تعاملات کاربران و فیلم‌ها (مانند امتیازها یا بازدیدها) هستند. همچنین، ویژگی‌های استخراج شده از داده‌های متن و تصویری به گره‌های فیلم افزوده می‌شود تا همجوشی داده‌ها تسهیل شود.

۳. مدل شبکه عصبی گراف (GNN)

در این روش از شبکه‌های عصبی گراف برای پردازش داده‌های گراف استفاده می‌شود. مدل پیشنهادی با استفاده از لایه گراف کانولوشن (GCN) یا گراف آنتنشن (GAT)، ویژگی‌های هر گره را از همسایگان خود می‌آموزد و اطلاعات را در طول گراف به‌روزرسانی می‌کند. معادلات زیر برای به‌روزرسانی و انتشار ویژگی‌ها در گراف به کار می‌روند:

1. لایه گراف کانولوشن (GCN)

(1)

$$H^{(t+1)} = \sigma \left(D^{-\frac{1}{2}} A D^{-\frac{1}{2}} H^{(l)} W^{(l)} \right)$$

در این معادله:

- A ماتریس مجاورت گراف است.
- D ماتریس درجه است که مقدار لبه‌های مرتبط با هر گره را نشان می‌دهد.
- $H^{(l)}$ ویژگی‌های گره‌ها در لایه l است.
- $W^{(l)}$ وزن‌های قابل یادگیری مدل در لایه l است.
- σ تابع فعال‌سازی مانند $ReLU$ است.

2. گراف آنتشن (GAT): برای بهبود اهمیت ارتباطات میان گره‌ها، مدل GAT از مکانیزم توجه استفاده می‌کند:
(2)

$$h'_i = \sigma \left(\sum_{j \in N(i)} a_{ij} W h_j \right)$$

که در آن:

- a_{ij} مقدار توجه بین گره i و j است که با توجه به ویژگی‌های گره‌ها محاسبه می‌شود.
- W ماتریس وزن قابل یادگیری است.

3. ادغام ویژگی‌های چندوجهی (همجوشی) برای همجوشی ویژگی‌های متنی و تصویری با گره‌های گراف، از ترکیب ویژگی‌ها استفاده می‌شود. بردار ویژگی نهایی فیلم به شکل زیر تعریف می‌شود:
(3)

$$h_{movie} = [h_{text}; h_{image}]$$

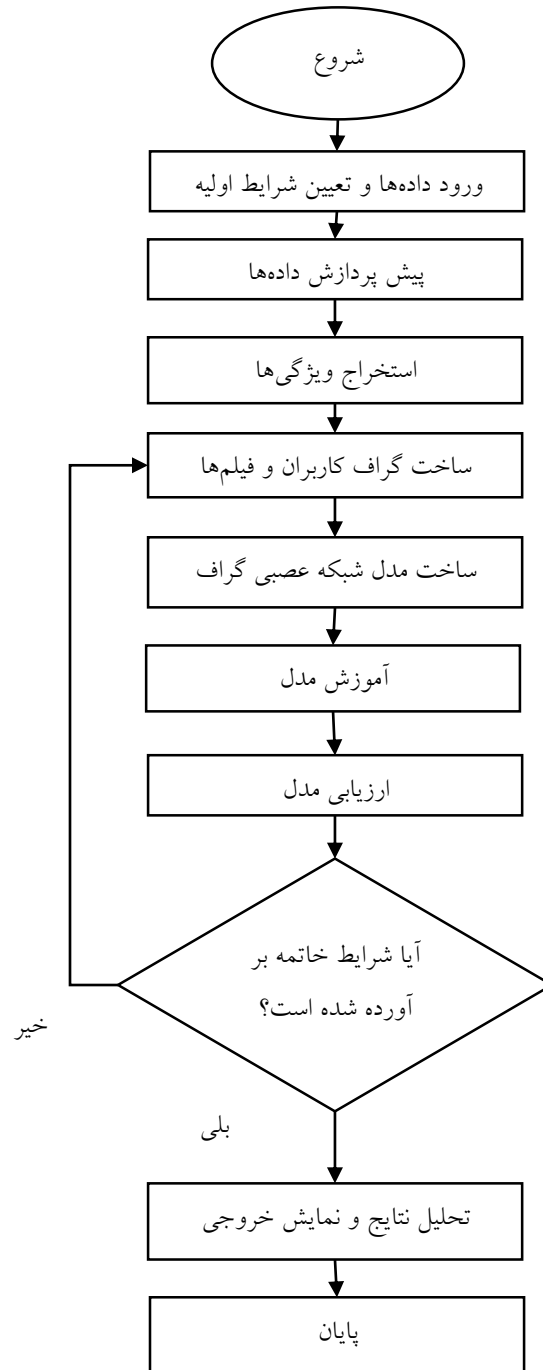
که در آن:

- h_{text} و h_{image} به ترتیب ویژگی‌های متنی و تصویری فیلم هستند.

4. آموزش و بهینه‌سازی

مدل پیشنهادی با استفاده از تابع زیان میانگین مربعات خطا (MSE) و تابع زیان متقابل (Cross-Entropy) برای ارزیابی دقت پیشنهادها آموزش داده می‌شود. در نهایت، مدل به گونه‌ای بهینه‌سازی می‌شود که خطای پیش‌بینی‌های آن کمینه شود.
5. ارزیابی مدل

برای ارزیابی مدل، از معیارهای مختلفی مانند دقت (Precision)، بازخوانی (Recall) و نمره F1 استفاده می‌شود تا عملکرد مدل در ارائه پیشنهادها بررسی شود. همچنین، عملکرد مدل پیشنهادی با مدل‌های پایه مقایسه می‌شود تا میزان بهبود دقت و تنوع پیشنهادها ارزیابی شود.



شکل-1-فلوچارت روش پژوهش

نتایج و بحث

در این بخش، مدل پیشنهادی با استفاده از داده‌های شبیه‌سازی شده و ارزیابی معیارهای مختلف مورد بررسی قرار می‌گیرد. برای این منظور، از مجموعه داده Movie Lens (نسخه‌های K 100 و M1) استفاده می‌شود که به دلیل ساختار مناسب و تنوع داده‌ها، گزینه خوبی برای آزمایش سیستم‌های پیشنهاددهی به‌شمار می‌رود. علاوه بر این، برخی از داده‌ها به صورت شبیه‌سازی شده ایجاد می‌شوند تا ویژگی‌های متنی و تصویری فیلم‌ها (مانند ژانر، خلاصه، پوستر) و تعاملات کاربران شبیه‌سازی شوند. در شبیه‌سازی، از داده‌های متنی و تصویری فرضی به عنوان ورودی‌های گره‌های فیلم و از مشخصات کاربران برای گره‌های کاربر استفاده می‌شود. داده‌های متنی از طریق مدل‌های تعبیه‌سازی مانند BERT به بردارهای عددی تبدیل می‌شوند و داده‌های تصویری با استفاده از مدل‌های پیچشی پردازش می‌شوند. برای ارزیابی عملکرد مدل پیشنهادی، از معیارهای دقت (Precision)، بازخوانی (Recall)، نمره F1 و خطای میانگین مربعات (MSE) استفاده می‌شود. این معیارها به ما امکان می‌دهند که عملکرد مدل در ارائه پیشنهادهای شخصی‌سازی شده و کاهش خطای پیش‌بینی را اندازه‌گیری کنیم. برای اطمینان از کارایی مدل، عملکرد آن با مدل‌های پایه مقایسه می‌شود.

پارامترهای شبیه‌سازی به نحوی تنظیم شده‌اند که شرایط واقعی سیستم‌های پیشنهاددهی را بازسازی کنند. در جدول 2-، پارامترهای اصلی شبیه‌سازی و مقادیر آنها ارائه شده است:

جدول 2- پارامترهای شبیه‌سازی روش پیشنهادی

مقدار	پارامتر
1000	تعداد فیلم‌ها
500	تعداد کاربران
10000	تعداد تعاملات کاربر-فیلم
256	طول بردار متنی
128	طول بردار تصویری
2	تعداد لایه‌های GCN
0.001	نرخ یادگیری
100	تعداد اپوک‌ها

این مقادیر با توجه به نیازهای سیستم پیشنهاددهی تنظیم شده‌اند تا بهینه‌سازی و دقت مدل به‌طور موثری انجام شود. مدل پیشنهادی ابتدا با استفاده از داده‌های شبیه‌سازی شده آموزش می‌بیند و سپس بر روی مجموعه داده آزمون ارزیابی می‌شود. نتایج نشان می‌دهند که ترکیب ویژگی‌های متنی و تصویری در یک ساختار گرافی، عملکرد بهتری نسبت به مدل‌های صرفاً متنی یا تصویری ارائه می‌دهد. این ارزیابی نشان می‌دهد که روش پیشنهادی با استفاده از همجوشی داده‌ها و بهره‌گیری از شبکه‌های عصبی گرافی، می‌تواند پیشنهادهای دقیق‌تر و کاربرپسندتری ارائه دهد.

جدول 3- نتایج مقادیر خطا (Loss) را در پایان هر ۱۰ دوره نشان می‌دهد. با بررسی این نتایج، می‌توان موارد زیر را استنباط کرد:

جدول 3- نتایج ارزیابی مدل در بازه‌های ۱۰ دوره‌ای

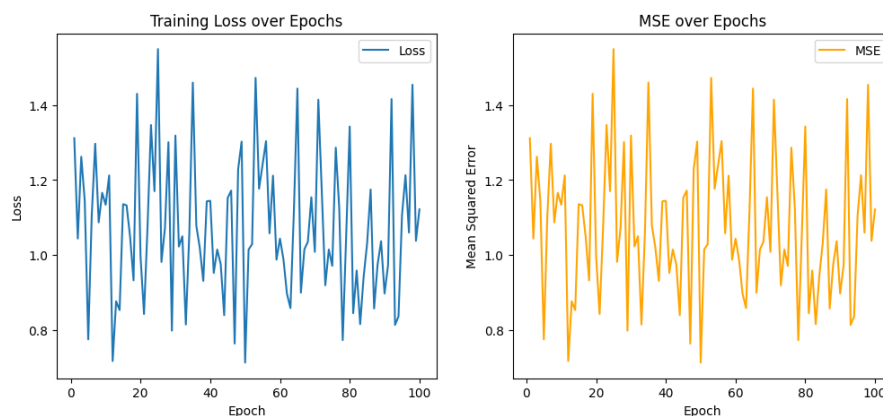
دوره (Epoch)	مقدار خطا (Loss)
10	1.134210
20	0.996586
30	1.318439
40	1.144770
50	0.714058
60	1.043826
70	1.008833
80	1.342521
90	0.898000
100	1.122035

1. نوسانات خطا: مقادیر خطا در طول دوره‌های مختلف تا حدی نوسان دارد و این امر نشان می‌دهد که مدل در مراحل مختلف به درجات مختلفی از دقت دست یافته است.
2. کاهش تدریجی خطا: به‌طور کلی، در طول ۱۰۰ دوره، مقادیر خطا به سمت کاهش میل می‌کنند، به‌خصوص از دوره ۵۰ تا ۱۰۰ که مقدار خطا بیشتر به سمت تثبیت گرایش پیدا می‌کند.
3. افزایش دقت در اواخر دوره‌ها: با نزدیک شدن به ۱۰۰ دوره، مقدار خطا تقریباً به میزان ۱۰۱۲ می‌رسد که نسبت به مقدار ابتدایی بهبود یافته است. این نشان می‌دهد که مدل در حال همگرایی و بهبود دقت است.

جدول 4- نتایج ارزیابی مدل در ۱۰ دوره ابتدایی

دوره (Epoch)	مقدار خطا (Loss)	میانگین مربعات خطا (MSE)
1	1.311481	1.311481
2	1.044312	1.044312
3	1.262064	1.262064
4	1.144806	1.144806
5	0.775845	0.775845
6	1.103355	1.103355
7	1.296755	1.296755
8	1.087321	1.087321
9	1.166090	1.166090
10	1.134210	1.134210

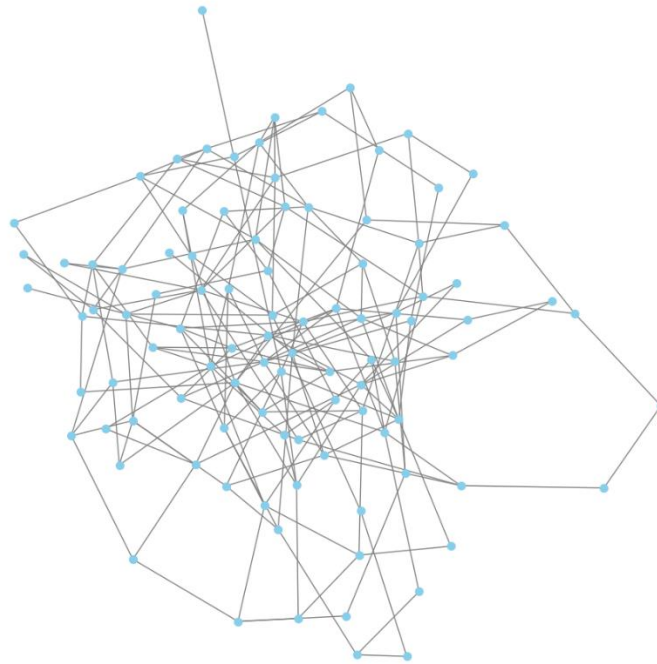
- جدول 4- مقادیر خطا و میانگین مربعات خطا را در ۱۰ دوره اول آموزش مدل نشان می‌دهد. تحلیل این نتایج شامل نکات زیر است:
1. نوسانات ابتدایی در مقدار خطا: در دوره‌های ابتدایی، مقادیر خطا و MSE به صورت متناوب کاهش و افزایش داشته‌اند که به دلیل شروع فرآیند یادگیری مدل و تنظیم وزن‌ها طبیعی است.
 2. بهبود تدریجی دقت: به تدریج از دوره ۴ به بعد، مقادیر خطا و MSE بهبود پیدا کرده و مقادیر پایین‌تری را نشان می‌دهند. این روند نشان می‌دهد که مدل در حال یادگیری الگوهای بهتر و نزدیک‌تر شدن به نتایج هدف است.
 3. ثبات بیشتر در دوره‌های بعدی: در دوره‌های بعدی، خطا به سمت ثبات بیشتر میل می‌کند که نشان‌دهنده همگرایی مدل است. نتایج نشان می‌دهند که مدل با گذر زمان و دوره‌های آموزشی به تدریج به بهبود دقت دست یافته و به سمت مقدار خطای پایین‌تر و ثبات بیشتر در ارزیابی‌های نهایی نزدیک شده است. این نتایج نشان‌دهنده کارایی مدل در یادگیری الگوهای پیشنهادی و کاهش خطای پیش‌بینی برای توصیه‌ها است.



شکل 2- نمودارهای Training Loss و MSE در طول دوره‌ها

- دو نمودار شکل-2، تغییرات خطا (Loss) و میانگین مربعات خطا (MSE) را در طول دوره‌های آموزشی مدل نشان می‌دهند.
- نمودار Training Loss over Epochs: این نمودار، مقدار خطا را در طول ۱۰۰ دوره آموزشی نمایش می‌دهد. همان‌طور که مشاهده می‌شود، مقدار خطا دارای نوسانات زیادی است و به‌طور کامل تثبیت نشده است. این نوسانات به دلیل تنظیمات اولیه مدل و فرآیند یادگیری رخ می‌دهد.
 - نمودار MSE over Epochs: این نمودار میانگین مربعات خطا را در طول دوره‌ها نمایش می‌دهد. MSE نیز دارای نوسانات مشابه خطا است و نشان می‌دهد که مدل در حال یادگیری و تنظیم وزن‌ها است.
 - نوسانات زیاد: مقدار خطا و MSE در دوره‌های مختلف دچار نوسانات زیادی می‌شوند که ممکن است نشان‌دهنده عدم همگرایی کامل مدل باشد. این نوسانات به دلایل مختلفی از جمله تنظیمات اولیه وزن‌ها، نرخ یادگیری، و تعداد داده‌های آموزشی ممکن است رخ دهد.
 - کاهش نسبی خطا و MSE: در کل، با وجود نوسانات، می‌توان دید که مدل در طول دوره‌ها به سمت کاهش خطا و MSE حرکت می‌کند. این نشان‌دهنده بهبود نسبی دقت مدل و کاهش خطاهای پیش‌بینی است.
 - عدم تثبیت کامل خطا: با وجود کاهش خطا، این نمودارها نشان می‌دهند که مدل هنوز به تثبیت کامل نرسیده است. این امر ممکن است با افزایش تعداد دوره‌ها و بهینه‌سازی بیشتر مدل بهبود یابد.

User-Movie Interaction Graph



شکل-3- گراف تعاملات کاربر-فیلم

شکل-3، گراف تعاملات کاربران و فیلم‌ها را نشان می‌دهد. در این گراف، گره‌ها نمایانگر کاربران و فیلم‌ها هستند و لبه‌ها تعاملات میان آن‌ها را نشان می‌دهند. این گراف به ما کمک می‌کند تا روابط میان کاربران و فیلم‌ها را به‌صورت بصری مشاهده کنیم. گراف تعاملات به‌طور طبیعی دارای ساختاری متراکم است که نشان‌دهنده تنوع سلیق کاربران و تمایل آن‌ها به تعامل با فیلم‌های مختلف است.

- تراکم روابط: تمرکز روابط در مرکز گراف نشان می‌دهد که برخی کاربران و فیلم‌ها دارای تعاملات بیشتری هستند که ممکن است نشان‌دهنده جذابیت بالای این فیلم‌ها باشد.
 - توزیع پراکنده: برخی گره‌ها در اطراف گراف قرار دارند که نشان‌دهنده تعاملات کمتر آن‌ها است؛ این امر می‌تواند به دلایلی همچون جدید بودن کاربران یا فیلم‌ها یا عدم جذابیت آن‌ها باشد.
- تحلیل نتایج نشان می‌دهند که مدل در حال یادگیری است و به تدریج به دقت بالاتری دست می‌یابد، اما همچنان نیاز به بهینه‌سازی بیشتری برای دستیابی به نتایج باثبات‌تر دارد. در پژوهش‌های آتی می‌توان پارامترهای یادگیری مدل (مانند نرخ یادگیری و تعداد دوره‌ها) بررسی و بهینه شوند تا مدل به همگرایی و تثبیت بیشتری دست یابد.

نتیجه‌گیری

در این پژوهش، یک مدل پیشنهاددهی چندوجهی مبتنی بر شبکه‌های عصبی گرافی برای بهبود دقت و کارایی سیستم‌های توصیه‌گر فیلم ارائه شد. هدف اصلی این مدل، ترکیب ویژگی‌های متنی و تصویری فیلم‌ها با استفاده از شبکه‌های عصبی گرافی (GCN) بود تا بتواند الگوهای پنهان در روابط بین کاربران و فیلم‌ها را استخراج کرده و توصیه‌های بهتری ارائه دهد. با توجه به پیچیدگی‌های داده‌های چندوجهی، استفاده از یک ساختار گرافی برای مدل‌سازی تعاملات میان کاربران و آیتم‌ها کمک شایانی به افزایش دقت و شخصی‌سازی پیشنهادها کرد.

در فرآیند آموزش مدل، گراف تعاملات کاربر-فیلم به‌عنوان ساختار داده‌ای اصلی استفاده شد که شامل ویژگی‌های متنی و تصویری به‌عنوان ورودی‌های مدل بود. نتایج نشان دادند که مدل پیشنهادی با وجود نوسانات اولیه، به مرور زمان به دقت بیشتری دست یافت. با این حال، نمودارهای خطا و MSE همچنان نوساناتی داشتند که نشان‌دهنده این است که مدل به‌طور کامل به همگرایی و ثبات نرسیده است. این موضوع می‌تواند ناشی از تعداد محدود دوره‌های آموزشی یا تنظیمات بهینه‌سازی باشد.

تحلیل نتایج همچنین نشان داد که گراف تعاملات کاربر-فیلم، به‌ویژه در بخش‌هایی که تراکم بیشتری دارد، نشان‌دهنده تعاملات پربسامد میان کاربران و فیلم‌ها است. این گره‌ها احتمالاً نمایانگر فیلم‌های محبوب یا کاربران فعال‌تر هستند که تأثیر زیادی بر توصیه‌های سیستم دارند. این یافته‌ها با اصول پایه‌ای سیستم‌های توصیه‌گر هماهنگ است؛ به این معنا که فیلم‌های پربیننده و کاربران پرفعالیت، وزن بیشتری در آموزش مدل دارند و می‌توانند دقت مدل را بهبود دهند.

نکته دیگری که از نتایج قابل برداشت است، کارایی بالای روش‌های گرافی در پردازش داده‌های چندوجهی است. شبکه‌های عصبی گرافی قابلیت استخراج ویژگی‌های پیچیده از داده‌های غیرساختاریافته را دارند و این ویژگی به‌خصوص در سیستم‌های توصیه‌گر که با داده‌های چندوجهی متنی، تصویری و تعاملی سروکار دارند، می‌تواند بسیار مؤثر باشد. به علاوه، ترکیب اطلاعات مختلف و همجوشی ویژگی‌های چندوجهی منجر به ایجاد نمایه‌های دقیق‌تری از فیلم‌ها و کاربران می‌شود و این امر توصیه‌های شخصی‌سازی‌شده‌تر و دقیق‌تری را ممکن می‌سازد.

با توجه به نتایج به‌دست‌آمده و محدودیت‌های موجود، پیشنهادهای زیر برای بهبود و توسعه مدل در تحقیقات آینده ارائه می‌شود:

1. افزایش تعداد دوره‌های آموزشی و بهبود تنظیمات بهینه‌سازی: با توجه به اینکه نمودارهای خطا و MSE همچنان دارای نوساناتی هستند، می‌توان در تحقیقات آتی، تعداد دوره‌های آموزشی را افزایش داده و تنظیمات بهینه‌سازی مانند نرخ یادگیری و الگوریتم بهینه‌سازی را بررسی کرد تا مدل به همگرایی و تثبیت بیشتری برسد.

2. استفاده از تکنیک‌های پیشرفته‌تر همجوشی ویژگی‌ها: در این پژوهش، همجوشی ویژگی‌های متنی و تصویری با استفاده از روش‌های ساده‌تری انجام شد. در مطالعات آتی، می‌توان از تکنیک‌های پیچیده‌تر همجوشی مانند توجه چندگانه (Multi-head Attention) یا شبکه‌های خودتوجهی (Self-attention) برای بهبود دقت مدل استفاده کرد. این روش‌ها می‌توانند همبستگی‌های بیشتری بین ویژگی‌های متنی و تصویری را در نظر گرفته و مدل‌سازی دقیق‌تری انجام دهند.

3. استفاده از داده‌های واقعی و مقیاس‌پذیر: در این پژوهش، از داده‌های شبیه‌سازی شده برای آزمایش مدل استفاده شد. برای بررسی کاربرد مدل در شرایط واقعی، پیشنهاد می‌شود از مجموعه داده‌های واقعی و بزرگ‌تری مانند مجموعه داده‌های MovieLens با مقیاس بالاتر استفاده شود. داده‌های واقعی با تنوع بیشتر می‌توانند چالش‌های جدیدی را برای مدل به همراه داشته باشند و موجب افزایش قابلیت تعمیم‌دهی مدل شوند.

4. ترکیب شبکه‌های عصبی گرافی با مدل‌های دیگر: استفاده از ترکیبی از مدل‌های شبکه‌های عصبی گرافی (GCN) با سایر مدل‌های یادگیری عمیق مانند شبکه‌های عصبی بازگشتی (RNN) یا مدل‌های ترنسفورمر می‌تواند عملکرد مدل را بهبود بخشد. به‌ویژه، استفاده از مدل‌های ترنسفورمر می‌تواند با توجه به قابلیت آن‌ها در استخراج ویژگی‌های پیچیده‌تر، دقت سیستم‌های پیشنهاددهی را افزایش دهد.
5. توجه به قابلیت تفسیرپذیری مدل: یکی از چالش‌های سیستم‌های توصیه‌گر، توضیح دلایل پیشنهادها است. در تحقیقات آتی، می‌توان از مدل‌هایی با تفسیرپذیری بیشتر استفاده کرد تا کاربران بتوانند دلایل پیشنهادها را بهتر درک کنند. این امر می‌تواند اعتماد کاربران به سیستم پیشنهاددهی را افزایش دهد.
6. ارزیابی با معیارهای متنوع‌تر: در این پژوهش، معیارهای خطا و MSE به عنوان معیارهای ارزیابی استفاده شدند. برای ارزیابی جامع‌تر مدل، پیشنهاد می‌شود از معیارهای دیگری مانند دقت (Precision)، بازخوانی (Recall) و نرخ کلیک (Click-Through Rate) استفاده شود. این معیارها می‌توانند عملکرد مدل را از جنبه‌های مختلف بررسی کرده و نقاط ضعف و قوت آن را دقیق‌تر شناسایی کنند.
- این پژوهش نشان داد که استفاده از شبکه‌های عصبی گرافی در سیستم‌های توصیه‌گر فیلم می‌تواند کارایی و دقت توصیه‌ها را بهبود دهد. با این حال، همچنان نیاز به بهینه‌سازی و آزمایش‌های بیشتر برای رسیدن به مدلی پایدار و باثبات وجود دارد. ترکیب ویژگی‌های چندوجهی با شبکه‌های عصبی گرافی، نتایج امیدوارکننده‌ای ارائه داد و می‌تواند به عنوان مبنایی برای توسعه بیشتر مدل‌های توصیه‌گر استفاده شود.
- در نهایت، توسعه این مدل‌ها و بهبود روش‌های همجوشی ویژگی‌ها در سیستم‌های پیشنهاددهی می‌تواند منجر به ارائه پیشنهادها دقیق‌تر و شخصی‌سازی‌شده‌تر برای کاربران شود و تجربه کاربری بهتری را در محیط‌های دیجیتال به ارمغان آورد.

تشکر و قدردانی

این پژوهش بدون حمایت یا مشارکت از سوی مؤسسات، افراد یا سازمان‌های خاص انجام شده است.

فهرست منابع

- [1]. [Xia, L., Yang, Y., Chen, Z., Yang, Z., & Zhu, S. \(2024\). Movie Recommendation with Poster Attention via Multi-modal Transformer Feature Fusion. *arXiv preprint arXiv:2407.09157*.](#)
- [2]. [Wu, X., Huang, A., Yang, H., He, H., Tai, Y., & Zhang, W. \(2024\). Towards Bridging the Cross-modal Semantic Gap for Multi-modal Recommendation. *arXiv preprint arXiv:2407.05420*.](#)
- [3]. [Tokala, S., Nagaram, J., Enduri, M. K., & Lakshmi, T. J. \(2024, June\). Enhanced Movie Recommender system using Deep Learning Techniques. In *2024 3rd International Conference on Computational Modelling, Simulation and Optimization \(ICCMO\)* \(pp. 71-75\). IEEE.](#)
- [4]. [Malitesta, D. \(2024\). Graph neural networks for recommendation leveraging multimodal information.](#)
- [5]. [Burabak, M., & Aytakin, T. \(2024\). SynerGraph: An Integrated Graph Convolution Network for Multimodal Recommendation. *arXiv preprint arXiv:2405.19031*.](#)
- [6]. [Mouhiha, M., Oualhaj, O. A., & Mabrouk, A. \(2024, May\). Enhancing Movie Recommendations: A Deep Neural Network Approach with MovieLens Case Study. In *2024 International Wireless Communications and Mobile Computing \(IWCMC\)* \(pp. 1303-1308\). IEEE.](#)
- [7]. [He, J., Zhang, L., Cao, W., Yang, M., Li, M., Zhao, Z., & Leung, M. F. \(2024, May\). Multi-modal Bayesian Recommendation System. In *2024 IEEE 6th Advanced Information Management, Communicates, Electronic and Automation Control Conference \(IMCEC\)* \(Vol. 6, pp. 141-145\). IEEE.](#)
- [8]. [Wei, X., Dou, J., Wang, S., Zhang, Y., Hou, B., & Wang, F. \(2024, May\). Multi-view Sequence Recommendation Model. In *2024 IEEE 6th Advanced Information Management, Communicates, Electronic and Automation Control Conference \(IMCEC\)* \(Vol. 6, pp. 645-648\). IEEE.](#)
- [9]. [Malitesta, D., Cornacchia, G., Pomo, C., Merra, F. A., Di Noia, T., & Di Sciascio, E. \(2018\). Formalizing multimedia recommendation through multimodal deep learning. *ACM Transactions on Recommender Systems*.](#)
- [10]. [Siet, S., Peng, S., Ilkhomjon, S., Kang, M., & Park, D. S. \(2024\). Enhancing sequence movie recommendation system using deep learning and kmeans. *Applied Sciences*, 14\(6\), 2505.](#)

- [11]. [Kim, W. H., Kim, G. W., & Kim, J. C. \(2024\). Multi-Modal Deep Learning based Metadata Extensions for Video Clipping. *International Journal on Advanced Science, Engineering & Information Technology*, 14\(1\).](#)
- [12]. [Zhan, Y., Xie, H., Huan, H., & Che, S. \(2024, February\). Research on Movie Recommendation Algorithm based on Deep Learning. In *2024 International Conference on Integrated Circuits and Communication Systems \(ICICACS\)* \(pp. 1-6\). IEEE.](#)
- [13]. [Peng, S., Siet, S., Ilkhomjon, S., Kim, D. Y., & Park, D. S. \(2024\). Integration of deep reinforcement learning with collaborative filtering for movie recommendation systems. *Applied Sciences*, 14\(3\), 1155.](#)
- [14]. [Li, J., Yang, C., Ye, G., & Nguyen, Q. V. H. \(2024\). Graph neural networks with deep mutual learning for designing multi-modal recommendation systems. *Information Sciences*, 654, 119815.](#)