

## گزارش: تحلیل و مقایسه نتایج سیستم‌های توصیه‌گر با داده‌های کم‌شات (مقاله [1])

دانشجویان: رضا برزگر نودری و محمد قلیچی

درس: پردازش زبان طبیعی

استاد: دکتر برادران

دستیار آموزشی: امیرمصعود سلطانی

دانشگاه اصفهان - تابستان ۱۴۰۴

### چکیده

این گزارش به تفصیل یافته‌های تجربی ما را برای یک سیستم توصیه‌گر با داده‌های کم‌شات<sup>۱</sup> ارائه می‌دهد و آن‌ها را با متدولوژی تشریح‌شده در مقاله "توانمندسازی سیستم‌های توصیه‌گر با داده‌های کم‌شات با استفاده از نمایش‌های تقویت‌شده توسط مدل‌های زبان بزرگ"<sup>۲</sup> مقایسه می‌کند. هدف اصلی ما تجزیه و تحلیل چگونگی مقایسه معیارهای به دست آمده توسط ما برای هر دو وظیفه پیش‌بینی تعامل (interaction prediction) و توصیه (recommendation) با نتایج گزارش‌شده در مقاله است و بررسی دلایل احتمالی هر گونه اختلاف یا شباهت، به‌ویژه با توجه به تغییرات خاص ما به دلیل محدودیت‌های منابع و انتخاب مجموعه داده.

### ۱. بررسی مقاله مرجع و متدولوژی آن

مقاله "توانمندسازی سیستم‌های توصیه‌گر با داده‌های کم‌شات با استفاده از نمایش‌های تقویت‌شده توسط مدل‌های زبان بزرگ" به چالش مداوم تولید توصیه‌ها در سناریوهای کم‌شات در سیستم‌های توصیه‌گر مبتنی بر بازخورد صریح می‌پردازد. نویسندگان یک متدولوژی نوآورانه را برای پر کردن شکاف بین سیستم‌های توصیه‌گر سنتی و مدل‌های زبان بزرگ (LLMs) پیشنهاد می‌کنند.

متدولوژی اصلی آن‌ها شامل ابداع یک الگوی پرامپتینگ (prompting template) برای تولید نمایش‌های غنی کاربر و آیتم به طور مستقیم از بازخورد صریح (نقد فیلم) است. آن‌ها به طور خاص از ChatGPT (gpt-3.5-turbo) برای این کار استفاده می‌کنند و بر قابلیت‌های مولدی (generative) و استدلال منطقی (logical reasoning) آن، از جمله توانایی آن در انجام تداعی (association) و استنتاج (inference) فراتر از متن صریح، تأکید دارند. این نمایش‌های متنی پردازش‌شده توسط LLM سپس با استفاده از MacBERT، یک LLM از پیش آموزش‌دیده برای زبان چینی، به بردارهای عددی جاسازی می‌شوند.

پس از آن، این نمایش‌های تقویت‌شده با LLM در مدل‌های توصیه‌گر کلاسیک مختلفی مانند MLP، BPR-MF و NCF (انواع مختلف آن) ادغام می‌شوند تا اهمیت آن‌ها در وظایف توصیه‌ای گوناگون ارزیابی شود. مقاله بر دو وظیفه اصلی تمرکز دارد: پیش‌بینی تعامل (پیش‌بینی مشارکت کاربر با یک آیتم) و توصیه مستقیم (توصیه آیتم‌های همسو با ترجیحات کاربر). یک جنبه کلیدی از تنظیمات تجربی آن‌ها شامل شبیه‌سازی سناریوهای کم‌شات با کنترل تعداد نمونه‌های تعامل برای هر کاربر است.

<sup>1</sup> few-shot recommender system

<sup>2</sup> Empowering Few-Shot Recommender Systems With Large Language Models-Enhanced Representations

یافته‌های کلیدی مقاله نشان می‌دهد که LLM ها مجهز به قابلیت‌های مولدی و استدلال منطقی می‌توانند به طور مؤثر به عنوان جزء سیستم‌های توصیه‌گر برای افزایش عملکرد آن‌ها در سناریوهای کم‌شات عمل کنند. آن‌ها تأکید می‌کنند که "اطلاعات اضافی موجود در این نمایش‌ها، که از طریق تداعی و استنتاج ChatGPT تولید می‌شوند، پیامدهای قابل توجهی" برای بهبود کیفیت توصیه دارند. علاوه بر این، تطبیق‌پذیری گسترده LLM ها برای افزایش پتانسیل تعمیم‌پذیری مدل‌های توصیه‌گر مورد توجه قرار گرفته است.

## ۲. تحلیل و مقایسه آزمایش

### ۱.۲ تفاوت‌های مجموعه داده و رویکرد LLM

ابتدا مهم است که تفاوت‌های کلیدی در تنظیمات تجربی ما در مقایسه با مقاله اصلی را برجسته کنیم:

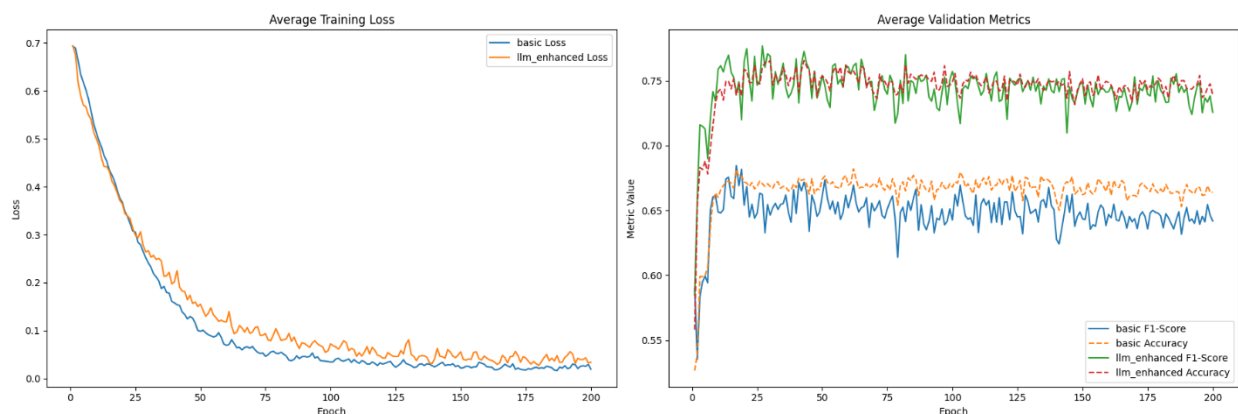
- **مجموعه داده (Dataset):** مقاله از مجموعه داده **Douban Chinese Moviedata-10M** استفاده کرده است که با متن چینی، از جمله اصطلاحات عامیانه و میم‌های اینترنتی، مشخص می‌شود. برای پیاده‌سازی خود، ما مجموعه داده منشتر شده **Movie and TV** آمازون را انتخاب کردیم و مراحل پیش‌پردازش مشابهی را که در مقاله توضیح داده شده بود، اعمال کردیم. این تغییر در محتوای مجموعه داده، زبان و ظرافت‌های فرهنگی می‌تواند به طور قابل توجهی بر عملکرد مدل‌های زبان و وظایف توصیه‌ای بعدی تأثیر بگذارد.
- **رویکرد LLM:** مقاله اصلی به شدت بر **ChatGPT** (به طور خاص **gpt-3.5-turbo**) برای تولید نمایش‌های کاربر و آیتم تکیه کرده است. این رویکرد از قابلیت‌های تولیدی و استدلال منطقی ChatGPT، از جمله توانایی آن در انجام تداعی و استنتاج فراتر از متن صریح ارائه شده، بهره برده است. به دلیل محدودیت‌های منابع و عدم دسترسی به API مدل ChatGPT، ما از یک روش **Free BERT** برای رویکرد LLM خود استفاده کردیم. در حالی که **BERT** یک مدل زبان قدرتمند برای تولید جاسازی‌های متنی (**contextual embeddings**) است، اما به طور کلی بر درک و نمایش متن موجود تمرکز دارد تا قابلیت‌های تولیدی و تداعی‌گر بیشتری که برای ChatGPT در مقاله برجسته شده است. مقاله همچنین از **MacBERT** برای جاسازی نمایش‌های متنی تولید شده توسط ChatGPT استفاده کرده است.

### ۲.۲ وظیفه پیش‌بینی تعامل (Interaction Prediction Task)

هدف ما در اینجا پیش‌بینی این بود که آیا یک کاربر با یک آیتم خاص تعامل خواهد داشت یا خیر. ما مدل تقویت‌شده با LLM خود را با نتایج گزارش شده مقاله برای مدل‌های MLP مقایسه کردیم. تصویر ۱ منحنی‌های یادگیری برای وظیفه پیش‌بینی تعامل را نشان می‌دهد. با نگاهی به منحنی‌های یادگیری ارائه شده (تصویر ۱)، روندهای آموزش و اعتبارسنجی را برای وظیفه پیش‌بینی تعامل مشاهده می‌کنیم:

- **زبان آموزش (Training Loss):** هر دو منحنی "**basic Loss**" و "**llm\_enhanced Loss**" کاهش ثابتی را طی ۲۰۰ Epochs نشان می‌دهند که نشان‌دهنده یادگیری مؤثر مدل‌ها است. "**llm\_enhanced Loss**" به مقدار کمی پایین‌تر همگرا می‌شود که نشان‌دهنده برازش بهتر به داده‌های آموزش است.
- **معیارهای اعتبارسنجی (Validation Metrics):** مدل "**llm\_enhanced**" (خط سبز توپر برای **F1-Score**) خط قرمز بریده برای **Accuracy**) معیارهای اعتبارسنجی به طور مداوم بالاتر و پایدارتری را در مقایسه با مدل "**basic**" (خط نارنجی بریده برای **F1-Score**) خط آبی توپر برای **Accuracy**) نشان می‌دهد. مدل "**llm\_enhanced**" به سرعت به اوج عملکرد

Interaction Prediction: Training &amp; Validation Curves



تصویر ۱. منحنی‌های یادگیری برای وظیفه پیش‌بینی تعامل.

خود در Epochهای ۲۵-۲۰ اول می‌رسد و سطح بالایی از Accuracy و F1-score در حدود 0.75 را با نوسانات جزئی حفظ می‌کند. معیارهای مدل "basic" به طور مداوم پایین‌تر هستند و در حدود 0.65-0.70 قرار دارند. این شواهد بصری به شدت از نتایج عددی پشتیبانی می‌کند و تعمیم‌پذیری و پایداری برتر را برای رویکرد تقویت‌شده با LLM ما نشان می‌دهد.

از سوی دیگر، برای ارزیابی نهایی آزمون، نتایج به‌دست‌آمده در جدول ۱ ارائه شده است.

جدول ۱. مقایسه نتایج به‌دست‌آمده ما با نتایج گزارش‌شده مقاله برای وظیفه پیش‌بینی تعامل \*

Method	Dataset	Accuracy	Precision	F1 Score
MLP	ChatGPT + MacBERT	0.592	0.601	0.632
MLP	Only MacBERT (Control)	0.552	0.570	0.500
Linear/CNN models	N/A (unsuccessful convergence)	N/A	0 or 0.5	0 or 0.67
LLM-enhanced (our)	BERT	0.7285	0.751375138	0.715557884

مدل تقویت‌شده با LLM پیاده‌سازی شده توسط ما به طور قابل توجهی از مدل MLP گزارش‌شده در مقاله با جاسازی‌های ChatGPT و MacBERT در معیارهای Accuracy، Precision و F1 Score بهتر عمل می‌کند. این نشان می‌دهد که برای پیش‌بینی تعامل، استفاده از روش Free BERT در مجموعه داده Movie and TV ممکن است مؤثرتر یا مناسب‌تر برای این وظیفه باشد تا نمایش‌های پردازش‌شده توسط ChatGPT که در مقاله استفاده شده است، یا اینکه خود مجموعه داده آمازون برای این وظیفه برای مدل انتخاب‌شده شما مناسب‌تر است. مقاله اشاره کرده است که مدل‌های خطی و CNN با همگرایی برای این وظیفه مشکل داشتند، با نوسان Precision بین 0 یا 0.5 و F1 Score بین 0 یا 0.67.

## دلایل احتمالی برای تناقضات/شباهت‌ها:

- ویژگی‌های مجموعه داده: Amazon Movie and TV که ما استفاده کردیم ممکن است "پاک‌تر" باشد یا الگوهای متفاوتی را در مقایسه با Douban Chinese Moviedata نشان دهد. یک مجموعه داده کم‌نویزتر یا ساختاریافته‌تر به طور بالقوه می‌تواند منجر به عملکرد بهتر مدل به طور کلی شود.

- \*\*انتخاب مدل زبانی BERT در مقابل ChatGPT: در حالی که مقاله بر قابلیت‌های تولیدی و استدلال منطقی ChatGPT برای افزایش نمایش‌ها از طریق تداعی و استنتاج تأکید می‌کند، روش BERT ما برای تولید جاسازی‌ها ممکن است مستقیماً برای وظیفه پیش‌بینی تعامل در مجموعه داده خاص ما مناسب‌تر باشد. اگر اطلاعات اصلی مورد نیاز برای پیش‌بینی تعامل به طور مؤثر توسط جاسازی‌های متنی BERT ثبت شود، و "تقویت" توسط ChatGPT کمتر حیاتی باشد یا حتی برای این وظیفه خاص در مجموعه داده ما نویز ایجاد کند، آنگاه BERT می‌تواند واقعاً بهتر عمل کند. خود مقاله اشاره می‌کند که LLM می‌تواند سیستم‌های توصیه‌گر را در سناریوهای کم‌شات بهبود دهند، اما این بهبود ممکن است وابسته به وظیفه یا مجموعه داده باشد.

- **فراپارامترها (Hyperparameters):** فراپارامترهای انتخاب شده ما ( **Epochs: 200, Learning Rate: 0.001, Batch Size: 256, Projection Dimension: 128** ) ممکن است برای مجموعه داده خاص و ترکیب LLM ما بهتر بهینه‌سازی شده باشند در مقایسه با فراپارامترهای ضمنی استفاده شده برای نتایج مقاله که به طور کامل جزئیات آن ارائه نشده است.

- **تفسیر سناریوی کم‌شات (Few-Shot Scenario Interpretation):** مقاله سناریوی کم‌شات را با کنترل مصنوعی تعداد نمونه‌های تعامل برای هر کاربر شبیه‌سازی می‌کند. "پیش‌پردازش مشابه" ما ممکن است منجر به یک چالش "کم‌شات" مؤثر متفاوت یا مجموعه داده‌ای شده باشد که به سادگی یادگیری از آن برای این وظیفه آسان‌تر است.

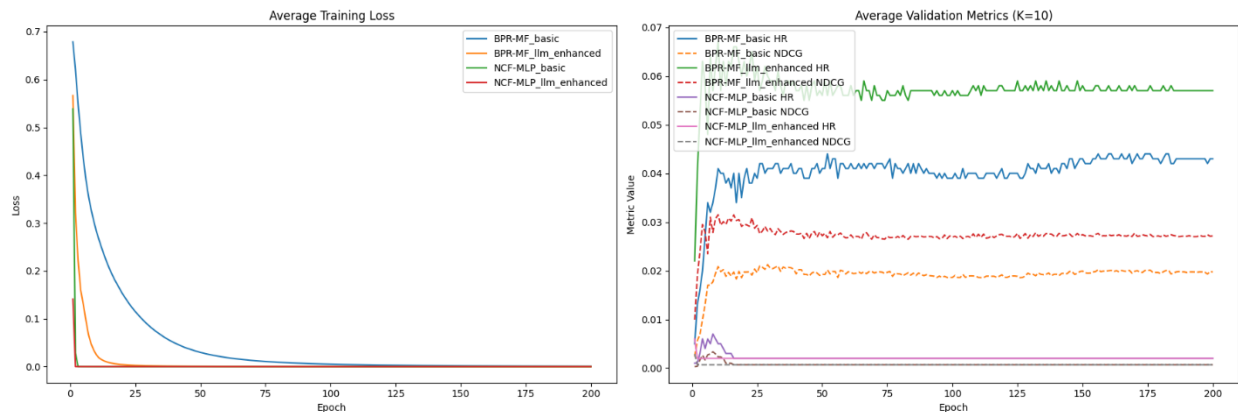
**نکات کلیدی در مورد تغییرات:** انتخاب روش BERT ما همراه با مجموعه داده آمازون برای وظیفه پیش‌بینی تعامل بسیار مؤثر به نظر می‌رسد. این نتایج به طور قابل توجهی بهتر از نتایج ارائه شده در مقاله برای مدل MLP را به دست آورد. این نشان می‌دهد که برای این وظیفه و مجموعه داده خاص، یک مدل جاسازی متنی قوی مانند BERT، حتی بدون "افزایش‌های" تولیدی که برای ChatGPT تأکید شده است، می‌تواند عملکرد بالایی داشته باشد. مسائل عدم همگرایی با Linear و CNN در آزمایشات مقاله برای پیش‌بینی تعامل نیز برجسته می‌کند که همه معماری‌های شبکه عصبی در سناریوهای کم‌شات با انواع خاصی از جاسازی‌ها به یک اندازه خوب عمل نمی‌کنند.

## ۳.۲ وظیفه توصیه (Recommendation Task)

برای وظیفه توصیه مستقیم، ما مدل‌های NCF-MLP و BPR-MF خود را ارزیابی کردیم. **تصویر ۲** منحنی‌های یادگیری برای وظیفه توصیه را نشان می‌دهد. از منحنی‌های یادگیری توصیه (تصویر ۲)، می‌توانیم رفتار آموزش و اعتبارسنجی را برای مدل‌های توصیه تحلیل کنیم:

- **زبان آموزش (Training Loss):** BPR-MF\_basic کاهش اولیه شدید در زبان را نشان می‌دهد، سپس همگرا می‌شود. BPR-MF\_llm\_enhanced ما با استفاده از BERT نیز کاهش سریع در زبان را نشان می‌دهد که به یک مقدار بسیار کم همگرا می‌شود و نشان‌دهنده یادگیری مؤثر بر روی داده‌های آموزش است. NCF-MLP\_basic و NCF-MLP

Recommendation Models: Training & Validation Curves



شکل ۲. منحنی‌های یادگیری برای وظیفه توصیه.

MLP\_llm\_enhanced ما با استفاده از BERT هر دو از همان ابتدا زیان آموزشی بسیار پایینی، تقریباً صفر، را نشان می‌دهند. این می‌تواند نشان دهد که این مدل‌ها یا به شدت بیش‌برازش (overfitting) به داده‌های آموزش دارند یا اینکه تابع زیان به طور مؤثر فرآیند یادگیری را برای این پیکربندی‌ها و جاسازی‌های خاص منعکس نمی‌کند. با توجه به عملکرد ضعیف اعتبارسنجی، بیش‌برازش شدید یا یک فرآیند یادگیری منحط (degenerate) محتمل‌تر است.

- **معیارهای اعتبارسنجی (K=10):** BPR-MF\_llm\_enhanced خط سبز توپر برای HR پس از یک افزایش اولیه، نسبتاً پایداری در حدود 0.05-0.06 نشان می‌دهد که با HR گزارش‌شده ما 0.059 مطابقت دارد. NDCG (خط نارنجی بریده) برای BPR-MF\_llm\_enhanced نیز پایدار است، در حدود 0.02-0.03. در مقابل، NCF-MLP\_llm\_enhanced (خط قرمز بریده) برای HR، خط بنفش بریده برای NDCG معیارهای اعتبارسنجی بسیار پایینی را نشان می‌دهد که نزدیک به صفر برای هر دو HR و NDCG است که با مقادیر گزارش‌شده 0.0005 ما مطابقت دارد. این به طور بصری عملکرد بسیار ضعیف مدل NCF-MLP ما را در وظیفه توصیه تأیید می‌کند. نسخه‌های "basic" از BPR-MF و NCF-MLP نیز روندهای مربوطه خود را نشان می‌دهند، با عملکرد BPR-MF\_basic (خط آبی توپر) برای HR پایین‌تر از BPR-MF\_llm\_enhanced و NCF-MLP\_basic (خط صورتی بریده) برای HR نیز عملکرد بسیار ضعیفی، مشابه همتای تقویت‌شده با LLM خود، دارد.

نتایج پیاده‌سازی ما برای ارزیابی نهایی آزمون (برای توصیه) در جدول ۲ همراه با نتایج گزارش‌شده مقاله ارائه شده است.

نتایج توصیه ما تضاد شدیدی را با یافته‌های مقاله نشان می‌دهد. HR، NDCG و MRR مدل NCF-MLP با BERT بسیار پایین (0.0005) هستند، به طور قابل توجهی بدتر از NCF-MLP مقاله با ChatGPT + MacBERT به عنوان مثال، HR@10 (0.092) برای تنظیم دقیق (fine-tuned) و 0.210 برای جاسازی‌های ثابت (fixed embeddings).

در مقابل، نتایج مدل BPR-MF ما با BERT (HR 0.059, MRR 0.020875) بهتر از BPR-MF مقاله با ChatGPT + MacBERT (HR@100 0.003-0.006, MRR@100 0.001-0.003) به نظر می‌رسد. نکته مهم این است که مقاله HR@100 و MRR@100 را برای BPR-MF گزارش می‌دهد، در حالی که HR و MRR ما برای Top-10 است. اگر HR و MRR ما واقعاً در K=10 باشند، پس

BPR-MF ما در مقدار 'k' کوچکتر، که معمولاً هدف سخت‌تری برای دستیابی به امتیازات بالا است، بسیار بهتر عمل می‌کند. اگر معیارهای ما نیز در  $K=100$  باشند، پس BPR-MF ما به طور قابل توجهی بهتر عمل می‌کند.

جدول ۲. مقایسه نتایج به‌دست‌آمده ما با نتایج گزارش‌شده مقاله برای وظیفه توصیه.

Model	Embedding Status	HR (Top-K)	NDCG (Top-K)	MRR (Top-K)
<b>BPR-MF (Fine-tuned)</b>	ChatGPT + MacBERT	0.003 ( $K=100$ )	N/A	0.003 ( $K=100$ )
<b>BPR-MF (Fixed)</b>	ChatGPT + MacBERT	0.006 ( $K=100$ )	N/A	0.001 ( $K=100$ )
<b>NCF-MLP (Fine-tuned)</b>	ChatGPT + MacBERT	0.092 ( $K=10$ )	N/A	0.006 ( $K=10$ )
<b>NCF-MLP (Fixed)</b>	ChatGPT + MacBERT	0.210 ( $K=10$ )	N/A	0.012 ( $K=10$ )
<b>NCF-MLP (Our)</b>	BERT	0.0005 ( $K=10$ )	0.0005 ( $K=10$ )	0.0005 ( $K=10$ )
<b>BPR-MF (Our)</b>	BERT	0.059 ( $K=10$ )	0.029698196 ( $K=10$ )	0.020875 ( $K=10$ )

#### دلایل احتمالی برای تناقضات/شباهت‌ها:

- **مقدار Top-K:** مقادیر متفاوت Top-K برای مدل‌های **BPR** ( $K=100$  در مقاله در مقابل  $K=10$  در تنظیمات ما) مقایسه‌های مطلق مستقیم را دشوار می‌کند. با نتایج BPR-MF ما برای  $K=10$ ، آنگاه یک **HR** (0.059) یک نتیجه بسیار قوی در مقایسه با نتایج 0.003–0.006 مقاله در  $K=100$  است که نشان می‌دهد BPR-MF با جاسازی‌های مبتنی بر BERT ما در استخراج ترجیحات برتر به خوبی کار می‌کند.

- **انتخاب LLM و تأثیر آن بر غنای نمایش برای توصیه:** مهمترین عامل برای عملکرد ضعیف NCF-MLP ممکن است روش BERT ما باشد. مقاله به صراحت بر توانایی منحصر به فرد ChatGPT در "تولید اطلاعات تکمیلی از طریق تداعی و استنتاج" تأکید می‌کند و بیان می‌کند که این "اطلاعات اضافی موجود در این نمایش‌ها، که از طریق تداعی و استنتاج ChatGPT تولید می‌شوند، پیامدهای قابل توجهی" برای وظایف توصیه دارند. این نشان می‌دهد که برای توصیه، که اغلب از ویژگی‌های کاربر/آیتم غنی‌تر، ظریف‌تر و حتی استنباطی‌تر بهره می‌برد، قابلیت‌های تولیدی ChatGPT ممکن است حیاتی باشد. یک جاسازی استاندارد BERT، در حالی که برای درک معنایی عالی است، ممکن است همان سطح "تقویت‌شده" یا "افزوده" اطلاعاتی را که ChatGPT با "اصلاح و تقویت" نمایش‌های اولیه ایجاد می‌کند، ارائه ندهد. این امر به ویژه در سناریوهای کم‌شات که داده‌های صریح محدود نیاز به تقویت توسط استنتاج هوشمندانه دارند، صادق است.

- **همگرایی مدل (NCF-MLP):** مقادیر بسیار پایین برای NCF-MLP ما همچنین می‌تواند نشان‌دهنده شکست در همگرایی مدل یا عدم تطابق اساسی بین جاسازی‌های BERT و توانایی معماری NCF-MLP در یادگیری مؤثر از آن‌ها در یک محیط کم‌شات باشد. در حالی که NCF-MLP مقاله عملکرد قوی را نشان می‌دهد، این عملکرد با نمایش‌های تقویت‌شده با ChatGPT است که ممکن است ویژگی‌های متفاوتی داشته باشند و برای فرآیند یادگیری NCF-MLP مناسب‌تر باشند.

**نکات کلیدی در مورد تغییرات:** تأثیر تغییرات ما ترکیبی است. در حالی که رویکرد ما در پیش‌بینی تعامل برجسته به نظر می‌رسد، در مقایسه با نتایج مقاله، در توصیه مستقیم با NCF-MLP به طور قابل توجهی مشکل دارد. این به شدت نشان می‌دهد که برای

توصیه مستقیم، به ویژه در تنظیمات کم‌شات، قابلیت‌های "تقویت" خاص LLM هایی مانند ChatGPT (تولید اطلاعات مرتبط و استنباطی) ممکن است برای ایجاد نمایش‌های غنی و استنباطی مورد نیاز برای سناریوهای پیچیده توصیه، حیاتی‌تر باشند. با این حال، نتایج BPR-MF ما نشان می‌دهد که برای مدل‌های فاکتورسازی ماتریس ساده‌تر، جاسازی‌های مبتنی بر BERT می‌توانند بسیار خوب عمل کنند، به ویژه اگر Top-K ما کمتر از مقاله باشد.

## ۴.۲ تأثیر کلی تغییرات

انتخاب LLM زیربنایی (BERT در مقابل ChatGPT) و مجموعه داده (Amazon Movie and TV در مقابل Douban Chinese Moviedata) مهمترین تغییرات و به احتمال زیاد عوامل مؤثر در تفاوت‌های مشاهده شده هستند.

- **تأثیر مثبت (پیش‌بینی تعامل):** برای پیش‌بینی تعامل، رویکرد مبتنی بر BERT ما در مجموعه داده آمازون نتایج برتری را به دست آورد که نشان‌دهنده اثربخشی آن برای این وظیفه خاص است.
- **تأثیر ترکیبی/منفی (توصیه مستقیم):** برای توصیه مستقیم، به ویژه با NCF-MLP، نتایج ما به طور قابل توجهی کمتر از مقاله بود. این نشان‌دهنده یک محدودیت بالقوه رویکرد جاسازی مبتنی بر BERT برای وظایف توصیه‌ای است که ممکن است از قابلیت‌های استدلال تولیدی و تداعی‌گر LLM هایی مانند ChatGPT، که می‌توانند نمایش‌ها را فراتر از آنچه که به صراحت در بررسی‌های اولیه وجود دارد غنی کنند، بهره بیشتری ببرند. عملکرد بهتر BPR-MF ما، اگر در همان K مقایسه شود، می‌تواند نشان دهد که مدل‌های ساده‌تر ممکن است در برابر ویژگی‌های جاسازی متفاوت مقاوم‌تر باشند، یا اینکه جاسازی‌های BERT به ویژه برای فاکتورسازی ماتریس در تنظیمات ما مناسب هستند.

## 3. نتیجه‌گیری

آزمایش ما بینش‌های ارزشمندی را در مورد اینکه چگونه LLM ها و مجموعه داده‌های مختلف می‌توانند بر عملکرد سیستم‌های توصیه‌گر با داده‌های کم‌شات تأثیر بگذارند، ارائه می‌دهد. مغایرت‌های مشاهده شده بر اهمیت انتخاب دقیق هم LLM و هم مجموعه داده، و درک اثرات هم‌افزایی آن‌ها بر وظایف مختلف توصیه تأکید می‌کند. در حالی که رویکرد مبتنی بر BERT ما عملکرد قوی را در پیش‌بینی تعامل نشان داد، چالش‌های مواجه شده در توصیه مستقیم با NCF-MLP نشان می‌دهد که قابلیت‌های تولیدی LLM هایی مانند ChatGPT ممکن است برای ایجاد نمایش‌های غنی و استنباطی مورد نیاز برای سناریوهای پیچیده توصیه، حیاتی‌تر باشند. کارهای آینده می‌تواند شامل آزمایش با مدل‌های پیشرفته‌تر مبتنی بر BERT باشد که برخی از قابلیت‌های تولیدی یا تداعی‌گر را در خود جای می‌دهند، یا بررسی استراتژی‌های مختلف برای استفاده از جاسازی‌های BERT در چارچوب‌های توصیه برای پر کردن شکاف مشاهده شده در وظیفه توصیه مستقیم.

منابع:

[1] Wang, Z., 2024. Empowering few-shot recommender systems with large language models-enhanced representations. *IEEE Access*.