Heterogeneous Knowledge Fusion: A Novel Approach for Personalized Recommendation via LLM

BIN YIN*, JUNJIE XIE*, YU QIN, ZIXIANG DING, and ZHICHAO FENG, Meituan, China

XIANG LI† and WEI LIN, Unaffiliated, China

The analysis and mining of user heterogeneous behavior are of paramount importance in recommendation systems. However, the conventional approach of incorporating various types of heterogeneous behavior into recommendation models leads to feature sparsity and knowledge fragmentation issues. To address this challenge, we propose a novel approach for personalized recommendation via Large Language Model (LLM), by extracting and fusing heterogeneous knowledge from user heterogeneous behavior information. In addition, by combining heterogeneous knowledge and recommendation tasks, instruction tuning is performed on LLM for personalized recommendations. The experimental results demonstrate that our method can effectively integrate user heterogeneous behavior and significantly improve recommendation performance.

CCS Concepts: • **Information systems** → **Recommender systems**.

Additional Key Words and Phrases: Recommendation, Large Language Models

1 INTRODUCTION

The analysis and mining of user behavior is a crucial aspect in recommendation systems. In the context of Meituan Waimai, user behavior exhibits heterogeneous characteristics, including various behavior subjects, content, scenarios. The current industry approach mostly involves continuously adding various heterogeneous behavior to the traditional recommendation models, which brings two obvious problems. Firstly, the multitude of behavior subjects leads to sparse features that pose challenges to efficient modeling. Secondly, separating the modeling of user, merchant, and commodity behavior ignores the fusion of heterogeneous knowledge among behavior. However, we have noticed that heterogeneous user behavior contain rich semantic knowledge, and using semantics to represent and reason about user behavior can more effectively promote heterogeneous knowledge fusion and capture user interests.

LLMs have shown remarkable capabilities in various fields, thanks to rich semantic knowledge and powerful inferential reasoning [1, 10]. We have designed a new user behavior modeling framework via LLM, which extracts and integrates heterogeneous knowledge from heterogeneous behavior information of users, and transforms structured user behavior into unstructured heterogeneous knowledge. In the field of recommendation, there have been some attempts to use LLM for personalized recommendation. Liu et al [5] have used LLM to express recommendation tasks in natural language based on context cues, but only adopt a single user behavior modeling, ignoring the fusion of user

Fig. 1. The overall framework of HKFR, including heterogeneous knowledge fusion (left), fine-tuning and recommendation (right).

heterogeneous knowledge. In addition, there are differences between the training task of LLMs and the recommendation task, which requires fine-tuning for specific recommendation tasks. Therefore, we propose **H**eterogeneous **K**nowledge **F**usion, a new approach of personalized **R**ecommendation via LLM (**HKFR**). Firstly, we design a new user behavior modeling framework using LLM, which extracts and integrates heterogeneous knowledge from heterogeneous behavior information of users. Then we combine heterogeneous knowledge and design instruction datasets for recommendation tasks to fine-tune LLM. Finally, using the constructed recommendation task instruction and heterogeneous knowledge as input and the fine-tuned LLM for calculation, we output the recommended features and results for users.

In summary, our contributions are as follows:

- We propose a new paradigm for user modeling, which extracts and integrates heterogeneous knowledge from heterogeneous user behavior information, transforming structured user behavior into unstructured heterogeneous knowledge, effectively capturing user interests.
- We construct a personalized recommendation model based on LLM, combining recommendation tasks and heterogeneous knowledge, fine-tuning LLM to make it more effective for tasks in recommendation scenarios.
- We have conducted sufficient offline and online experiments on Waimai dataset, verifying that our approach can fully integrate heterogeneous user behavior and effectively improve recommendation performance.

2 METHODOLOGY

In this section, we will introduce the framework of HKFR, which includes three stages as shown in Figure 1.

2.1 Heterogeneous Knowledge Fusion

In the stage of heterogeneous knowledge fusion, we utilize the rich semantic knowledge and powerful reasoning abilities of LLM to facilitate the integration of heterogeneous knowledge.

In the context of Meituan Waimai, user heterogeneous behavior includes: multiple behavior subjects, such as merchants and products; multiple behavior contents, such as exposure, clicks, and orders; multiple behavior scenarios, such as APP homepage and mini-programs. To address diverse user heterogeneous behavior, we extract and structure

user behavior data from the database with users as the core. Then, in the prompt engineering module, we design and construct different structured templates for diverse user behavior, expressing heterogeneous behavior as templated text language. Next, in the knowledge fusion module, we employ ChatGPT [6] to perform heterogeneous knowledge fusion on the behavior text, obtaining heterogeneous knowledge text. The heterogeneous knowledge generated based on user behavior will be used for the fine-tuning and recommendation stages of LLM.

2.2 Instruction Tuning

The fine-tuning process aims to help LLM better understand heterogeneous knowledge and further improve its accuracy and adaptability in recommendation tasks [7]. We construct an instruction dataset based on the recommendation task and heterogeneous knowledge, which includes input, instruction, and output. The input is the heterogeneous knowledge generated in the Section 2.1. The instruction and output are a series of task descriptions and target results generated specifically for recommendation. The instruction includes recommendations for user preferences on categories, price, and merchants, among others, while the output is the true label of the user's next order. Based on the constructed instruction dataset, we perform instruction tuning on LLM. Specifically, we choose the open-source model ChatGLM-6B [2] as our base LLM and adopt the Lora method for fine-tuning [4].

2.3 Recommendation

Given a user, retrieve user behavior heterogeneous knowledge from the database as input for LLM. Then, design instruction based on the recommendation task, perform reasoning and calculations, and output the recommended results for the user. These instructions include predicting user preferences for categories, prices, and other features, as well as their next click on a merchant. The predicted results can be output as direct recommendations in natural language form and can also be used as semantic features to enhance the recommendation effect by concatenating with the existing features in traditional recommendation models.

3 EXPERIMENT

In order to evaluate the performance of the model, we select two different recommendation tasks: recommending categories and POIs (points of interest) to user.

3.1 Implementation

For offline experiment, we select the dataset from March to April 2023 of Meituan Waimai, design 20 recommendation task instructions, and construct a total of 100,000 users and 1 million instruction data. The testing set is selected from the samples on May 9th, 2023, consisting of 10,000 instruction data with two tasks, recommending POIs and categories. Each data includes user statistical features, POI features, user heterogeneous behavior sequence features, label, etc. Due to the limitation of the input length, the user sequence length is limited to 300. Additionally, user and POI data are anonymized before being inputted into LLM. To evaluate the recommendation effectiveness, we select top-k HR and top-k NDCG, with k=5, 10. To demonstrate the effectiveness of our method, we compare it with traditional recommendation methods Caser [9], BERT4Rec [8], and language models P5 [3], ChatGLM-6B [2].

3.2 Results and Anaysis

Table 1 displays the experimental results, which demonstrate that our model outperforms multiple baselines on the Waimai dataset with significant improvements. Compared with traditional sequential recommendation methods such

Table 1. Experimental results on two recommendation tasks, with bold indicating the best results and italic indicating ablation experiments."no-HKF" means removing heterogeneous knowledge fusion, and "no-IT" means removing instruction tuning.

Methods	Category				POIs			
	HR@5	NDCG@5	HR@10	NDCG@10	HR@5	NDCG@5	HR@10	NDCG@10
Caser	0.1152	0.1063	0.2147	0.1320	0.0897	0.0770	0.1842	0.1012
BERT4Rec	0.1217	0.1140	0.2196	0.1440	0.0875	0.0744	0.1811	0.0995
P5	0.1416	0.1384	0.2477	0.1589	0.1218	0.1159	0.2187	0.1260
ChatGLM-6B	0.1074	0.1019	0.2038	0.1254	0.0785	0.0720	0.1702	0.0872
$HKFR_{no-IT}$	0.1241	0.1175	0.2267	0.1415	0.1014	0.0952	0.2050	0.1165
$HKFR_{no-HKF}$	0.1813	0.1308	0.2825	0.1580	0.1421	0.0975	0.2432	0.1270
HKFR	**0.2160**	**0.1586**	**0.3007**	**0.1840**	**0.1726**	**0.1243**	**0.2610**	**0.1525**

as Caser and BERT4Rec, our model demonstrates superior performance. This is mainly attributed to the effective fusion of heterogeneous knowledge by LLM. Our approach also achieves the best performance compared to language models such as P5 and ChatGLM. This is mainly due to the fine-tuning of professional datasets, which effectively improves the mismatch between LLM and recommendation tasks.

We conduct several experiments to validate the effectiveness of key modules in our HKFR. Compared to $HKFR_{no-HKF}$, which removes the heterogeneous knowledge fusion stage, HKFR demonstrates superior performance. This is mainly attributed to the accurate capture of user interests by fusing heterogeneous knowledge. Compared to $HKFR_{no-IT}$, which removes the instruction tuning stage, HKFR achieves better results, indicating that instruction tuning can effectively promote LLM to adapt to downstream recommendation tasks.

HKFR was applied to the Meituan Waimai recommendation system for online A/B testing, by utilizing the computed features from the previous day's search query of users and inputting them into the real-time computation on the current day. The experiment ran for a period of May 9, 2023 to May 19, 2023. The results indicate that HKFR achieved improvements of 2.45% in CTR and 3.61% in GMV for cold start users, while there was no significant effect found on other users. This is attributed to the insufficient catering expertise of LLM, which makes it challenging to fully comprehend and integrate heterogeneous behavior. Further training of LLM in the catering domain is necessary to address this limitation.

4 CONCLUSION

In this article, we propose HKFR, a new approach of personalized recommendation via LLM, which is based on the fusion of heterogeneous knowledge. By leveraging the two stages of heterogeneous knowledge fusion and instruction tuning, HKFR can effectively model user heterogeneous behavior. Extensive experiments on the Waimai dataset have validated the ability of HKFR to improve recommendation performance. In the future, we will focus on further training HKFR in the catering domain to better integrate heterogeneous knowledge and enhance recommendation performance.

REFERENCES

[1] Aakanksha Chowdhery, Sharan Narang, Jacob Devlin, Maarten Bosma, Gaurav Mishra, Adam Roberts, Paul Barham, Hyung Won Chung, Charles Sutton, Sebastian Gehrmann, et al. 2022. Palm: Scaling language modeling with pathways. *arXiv preprint arXiv:2204.02311* (2022).

[2] Zhengxiao Du, Yujie Qian, Xiao Liu, Ming Ding, Jiezhong Qiu, Zhilin Yang, and Jie Tang. 2022. GLM: General language model pretraining with autoregressive blank infilling. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 320–335.

[3] Shijie Geng, Shuchang Liu, Zuohui Fu, Yingqiang Ge, and Yongfeng Zhang. 2022. Recommendation as language processing (rlp): A unified pretrain, personalized prompt & predict paradigm (p5). In *Proceedings of the 16th ACM Conference on Recommender Systems*. 299–315.

[4] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2021. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685* (2021).

[5] Junling Liu, Chao Liu, Renjie Lv, Kang Zhou, and Yan Zhang. 2023. Is ChatGPT a Good Recommender? A Preliminary Study. *arXiv preprint arXiv:2304.10149* (2023).

[6] OpenAI. 2023. GPT-4 Technical Report. *CoRR* abs/2303.08774 (2023). https://doi.org/10.48550/arXiv.2303.08774 arXiv:2303.08774

[7] Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. 2022. Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems* 35 (2022), 27730–27744.

[8] Fei Sun, Jun Liu, Jian Wu, Changhua Pei, Xiao Lin, Wenwu Ou, and Peng Jiang. 2019. BERT4Rec: Sequential recommendation with bidirectional encoder representations from transformer. In *Proceedings of the 28th ACM international conference on information and knowledge management*. 1441–1450.

[9] Jiaxi Tang and Ke Wang. 2018. Personalized top-n sequential recommendation via convolutional sequence embedding. In *Proceedings of the eleventh ACM international conference on web search and data mining*. 565–573.

[10] Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, et al. 2023. A survey of large language models. *arXiv preprint arXiv:2303.18223* (2023).