Proceedings of the 6th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/69/20241512

Emerging synergies between large language models and machine learning in e-commerce recommendations

Xiaonan Xu[1,6,*], Yichao Wu[2], Penghao Liang[3], Yuhang He[4], Han Wang[5]

[1]Independent Researcher, Northern Arizona University, Flagstaff, AZ, 86011, USA
[2]Computer Science, Northeastern University, Boston, 02115, MA
[3]Information Systems, Northeastern University, San Jose, 95110, CA
[4]Computer Science and Technology, Tianjin University of Technology, Tianjin, 300384, CN
[5]Financial Mathematics, University of Southern California, Los Angeles, 90089, USA

[6]xiaonanxu5@gmail.com
*corresponding author

Abstract. This paper explores the integration of large language models (LLMs) into collaborative filtering algorithms to enhance recommendation systems in the e-commerce domain. The proposed approach combines user-based and item-based collaborative filtering with LLMs to improve recommendation accuracy and personalization. Specifically, the study introduces a novel framework called PALR, which leverages LLMs to refine user-item interactions and enrich item representations. PALR utilizes historical user behavior data, such as clicks, purchases, and ratings, to guide candidate retrieval and generate recommended items. This study highlights the importance of integrating LLMs into recommendation systems to deliver more accurate and personalized suggestions, ultimately improving user satisfaction and driving sales in e-commerce platforms.

Keywords: Large language models, collaborative filtering, recommendation systems, e-commerce.

1. Introduction

In recent years, the surge in online activity propelled by the Internet's rapid advancement has transformed e-commerce into the primary avenue for consumer transactions. With this growth, the challenge of efficiently guiding users to relevant products amidst an overwhelming array of choices has become increasingly prominent. Recommendation systems play a vital role in addressing this challenge by alleviating information overload and enhancing user experience across various online domains.

Traditionally, recommendation systems relied on collaborative filtering algorithms to analyze user-item interactions and generate personalized recommendations. In response to these challenges, this paper proposes a novel framework, [1]PALR (Personalized Adaptive Language-based Recommendation), which integrates large language models (LLMs) into collaborative filtering algorithms to enhance recommendation accuracy and personalization. PALR leverages LLMs' natural language processing capabilities to refine user-item interactions and enrich item representations, thereby improving the quality of recommendations. By fine-tuning [2]LLMs on a large scale and incorporating user-item

DOI: 10.54254/2755-2721/69/20241512

interaction data, PALR aims to overcome the limitations of traditional recommendation techniques and deliver superior recommendation performance across various sequential recommendation tasks. The core research of this paper is structured as follows: Section 2 provides a theoretical overview, Section 3 details the methodology of PALR framework development, Section 4 presents experimental results, and Section 5 concludes with future research directions.

2. Related Work

2.1. AI-driven Intelligent Recommendation

With the rapid development of the Internet, the e-commerce industry has become the main channel for consumers to shop. How to let consumers quickly and accurately find the goods they need is a problem that the e-commerce industry has been exploring. In recent years, with the gradual application of artificial intelligence technology[3], a new model has emerged in the e-commerce industry - intelligent recommendation. Moreover, for some consumers, may feel that intelligent recommendations are not accurate and limited because the algorithm model may only recommend based on the historical behavior of the user, without taking into account the actual needs of the user. In addition, some consumers may not like smart recommendations because they want to be more independent in their shopping decisions.

2.2. Recommendation System Overview

Recommendation systems [4-5](RecSys) play an important role in reducing information overload and enriching the online experience for users who need to filter through large amounts of information to find what interests them. In several applications such as entertainment, e-commerce, and job matching, the recommendation system provides personalized candidate recommendations based on user preferences. For example, movie recommendation systems, such as[6] IMDB and Netflix, recommend the latest movies based on the content of the movie and the user's past interaction history to help users discover new movies that fit their interests. The basic idea of the recommendation system is to use the interaction between the user and the item and its related side information, especially text information (such as the title or description of the item, the user profile, and the user's comments on the item), to predict the match score between the user and the item (that is, the probability that the user likes the item).

2.3. Collaborative Filtering in RecSys

The recommendation system has become a basic service of the Internet, which recommends personalized products to users by learning their preferences in historical interaction behaviors. At present, collaborative filtering algorithms based on Graph Neural Networks have shown great advantages in the field of recommendation. Generally speaking, in the Collaborative Filtering (CF) [7]scenario, we have a user set U and item set I, and their interactions. So, if we treat each user and item as a node and the interaction record between them as an edge, we can construct a [8]User-Item Interaction Graph. Then, based on the layer of information transmission and aggregation of graph neural networks, we can finally get the Representation learned by each user and commodity node based on the graph structure.

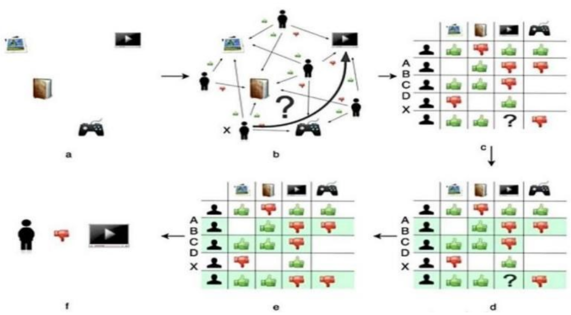

Figure 1. Framework of the basic principles of collaborative filtering

As shown in the frame diagram in Figure 1 above, The evaluation of collaborative filtering algorithm recommendation systems is to assess whether a recommendation system is good. A good recommendation system can not only accurately predict the user's behavior but also expand the user's vision and help the user find things that they may be interested in but are not so easy to find, thus increasing the revenue benefit through the recommendation system.

2.4. Principles of Personality Recommendation Matrix Decomposition

The disadvantage of collaborative filtering is that the co-occurrence matrix is generally sparse, and the process of finding similar users is not accurate in the case of a few user behaviors. The matrix decomposition method improves the ability to deal with sparse matrices. [9]Matrix decomposition is to use the co-occurrence matrix to generate the hidden vector of the item and the user. (Figure 2)The hidden vector can be understood as the same vector space in the same vector similarity to the sort recommended.

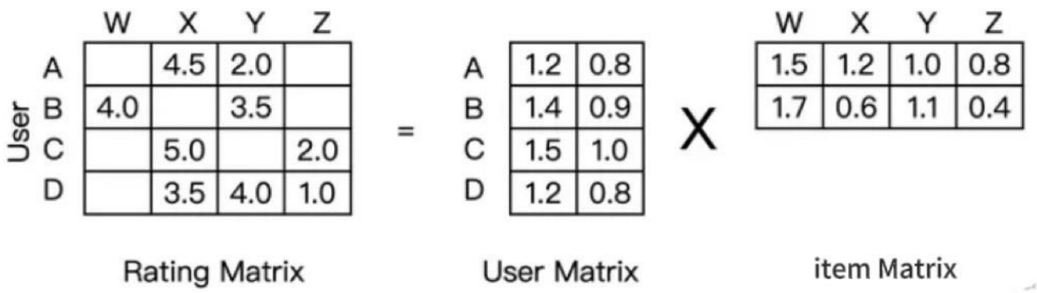

Figure 2. Matrix decomposition structure diagram

In the realm of collaborative filtering (CF) for recommendation systems, integrating large language models (LLMs) presents a novel approach to enhance recommendation accuracy and personalize user experiences. [10]This integration capitalizes on the natural language processing prowess of LLMs to refine the interactions between users and items, and to enrich item representations. By aiming to closely align the product of user and item matrices with the original co-occurrence matrix, the process imitates matrix multiplication, where user-implicit vectors and item-implicit vectors are derived to encapsulate the nuanced dynamics of user-item relationships more accurately.

3. Methodology

The collaborative filtering recommendation algorithm clusters items with similar attributes to suggest them to users, enhancing personalization. It works best when there are more items than users, like in movie, music, or book recommendations. By grouping users, it identifies varied interests and recommends unexplored items, ideal for platforms with more users than items, such as social networks and e-commerce sites. This method boosts user satisfaction and sales by tailoring recommendations.

The recommendation algorithm consists of three steps: collecting user preferences, finding similar users or items, and calculating recommendations.

Collect user preferences → Find similar items or users → Calculate recommendations[11]

3.1. User Similarity Calculation

(1) Build a list of users or items

Assume that four users A, B, C, and D have scored five items a, b, c, d, and e, and a user-item rating table can be created according to the score, as shown in Table 1:

Table 1. List of users and item types

User	Type1	Type2	Type3	type4	Type5
A	3.0	4.0	0	3.5	0
B	4.0	0	4.5	0	3.5
C	0	3.5	0	0	3.0
D	0	4.0	0	3.5	3.0

(2) Calculation of similarity

Based on the collected data, algorithms (such as Pearson correlation coefficient, cosine similarity, etc.) are used to calculate the similarity between the target user and other users. By analyzing the language expression and content preferences of users, the large language model can provide a richer dimension of similarity calculation, thus helping to identify the user groups with truly similar purchasing behaviors and preferences.

First, have to predict user 3's rating for item 4. In a user-based recommendation system, we find three users who are most similar to user 3 and use the ratings of these three users to predict user 3's rating of item 4.

Common similarity measures are cosine, Pearson, Euclid, and so on. We will use cosine similarity here, which is defi$IIII$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}}$ (1)

Moreover, Pearson correlation is definedI$r = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i-1}^{n}(y_i-\bar{y})^2}}$ (2)

3.2. Item-Based Collaborative Filtering

Similar to the UserCF algorithm, a user-item inversion table is created when ItemCF is used to calculate item similarity. Then, for each user, each pair of items in his item list can be added by 1 to the co-occurrence matrix.In this method, the cosine similarity measure is used to calculate the similarity between a pair of goods. You can predict target user a's rating of target item i by using a simple weighted average:

$$p_{a,i} = \frac{\sum_{j \in K} r_{a,j} w_{i,j}}{\sum_{j \in K} |w_{i,j}|} \tag{3}$$

To implement Adjusted Cosine similarity in Python, I defined a simple function called computeAdjCosSim, which returns the adjusted cosine similarity matrix, giving the score matrix. The functions findksimilaritems_adjcos and predict_itembased_adjcos use the adjusted cosine similarity to findksimilaritems and calculate the predicted score.

3.3. Assessing Recommendation Algorithm Accuracy

Python 3.9.0 is used to implement the above recommendation algorithm, and Precision and Recall are used to evaluate the accuracy of the algorithm. It is calculated as follows:

$$\text{Precision} = \frac{TP}{TP+FP} \times 100\% \tag{4}$$

$$\text{Recall} = \frac{TP}{TP+FN} \times 100\% \tag{5}$$

Where: TP is the number of positive samples predicted by the model to be positive; FP is the number of negative samples predicted by the model to be positive; FN is the number of positive samples predicted by the model to be negative. These techniques facilitate personalized recommendations based on user behavior and preferences, enhancing the overall effectiveness and personalization of recommendation algorithms in e-commerce platforms.

4. Experiment and Result

The performance of the collaborative filtering algorithm combined with the large language model is compared in the intelligent recommendation experiment of e-commerce products. The first group is the traditional UserCF algorithm and the improved T-UserCF algorithm. In the experiment, their performance is evaluated by comparing their MAE values(Figure 3).

(1) The performance of the UserCF algorithm and the TUserCF algorithm is compared in the experiment. The number of neighbors K starts from 5 and gradually increases with 5 as the basic unit. Then, the average absolute error (MAE) values of these two algorithms are compared under different numbers of neighbors.

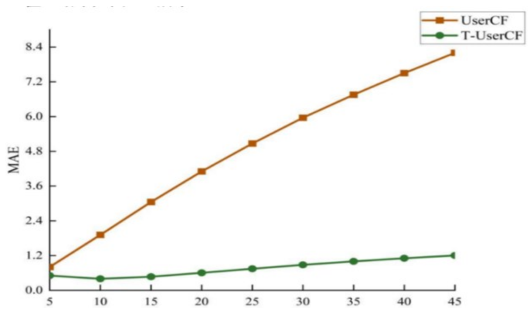

Figure 3. The MAE value of the UserCF algorithm and T-UserCF algorithm

When the K value is 10, the MAE value of the T-UserCF algorithm reaches the lowest point. When K is 5, the MAE value of the UserCF algorithm reaches the lowest point and then continues to rise. According to the experimental results, it can be concluded that the accuracy of the UserCF algorithm and TUserCF algorithm increases with the increase of K value. When the number of neighbors K is the same, the TUserCF algorithm has better-recommended performance than the UserCF algorithm.

1)Evaluation Results of UserCF and T-UserCF Algorithms:

The table presents the MAE values of the UserCF and T-UserCF algorithms for various numbers of neighbors.

It demonstrates that the T-UserCF algorithm consistently outperforms the UserCF algorithm across different neighbor counts.

Table 2. UserCF and T-UserCF recommend algorithm value

Number of neighbors	NAME	
	UserCF	T-UserCF
5	0.807	0.508
10	1.915	0.398
15	3.050	0.471
20	4.107	0.605
25	5.070	0.746
30	5.960	0.880

This can be seen in Table 2, in experiments comparing traditional and improved collaborative filtering algorithms (UserCF vs. T-UserCF and ItemCF vs. T-ItemCF), it was found that the enhanced versions, T-UserCF and T-ItemCF, consistently outperformed their predecessors. By evaluating performance through Mean Absolute Error (MAE) across varying neighbor counts, the improved algorithms demonstrated higher accuracy and better recommendation performance.

This section evaluates the performance of collaborative filtering algorithms combined with large language models in an intelligent recommendation experiment involving e-commerce products. The comparison is made between the traditional UserCF algorithm and the improved T-UserCF algorithm. The evaluation metric used is the Mean Absolute Error (MAE) value.

DOI: 10.54254/2755-2721/69/20241512

2)Performance Comparison of UserCF and T-UserCF Algorithms:

The experiment compares the performance of the UserCF and T-UserCF algorithms across different numbers of neighbors (K).

The MAE values of both algorithms are compared, showing that the accuracy increases with an increasing number of neighbors.

Notably, the T-UserCF algorithm achieves better recommendation performance than the UserCF algorithm when the number of neighbors is the same.

3)Comparison of Traditional and Improved Collaborative Filtering Algorithms:

In experiments comparing traditional (UserCF and ItemCF) and improved (T-UserCF and T-ItemCF) collaborative filtering algorithms, the enhanced versions consistently outperformed their predecessors.

By evaluating performance through MAE across varying neighbor counts, the improved algorithms demonstrated higher accuracy and better recommendation performance.

These findings suggest that the modifications in the T-UserCF and T-ItemCF algorithms, potentially including the integration of large language models, significantly enhance the precision and personalization of e-commerce product recommendations.

5. Conclusion

In the realm of e-commerce, the integration of large language models (LLMs) with collaborative filtering algorithms—both user-based and item-based—has emerged as a transformative approach, significantly enhancing the personalization and accuracy of recommendation systems. This study underscores the pivotal role of LLMs in identifying nuanced similarities among users and items through analysis of purchase history and user behavior, thereby elevating the quality of recommendations.Enhanced algorithms, exemplified by T-UserCF and T-ItemCF, surpass traditional methods by harnessing LLMs for deeper feature analysis and more refined similarity calculation, resulting in more precise recommendations. Experimental findings validate the superiority of LLM-augmented collaborative filtering algorithms in E-commerce recommendation systems, offering superior performance over their conventional counterparts.

This advancement highlights the intrinsic value of integrating LLMs with collaborative filtering, where the advanced language understanding capabilities of LLMs enrich the recommendation process. By delivering more accurate and personalized suggestions, this integration not only enhances user satisfaction but also drives sales, underscoring the potential of LLMs to revolutionize e-commerce platforms through improved recommendation accuracy and personalization.

References

[1] Emanuele Barabino,and Giuseppe Cittadini. "From Search Engines to Large Language Models: A Big Leap for Patient Education!." Cardiovascular and interventional radiology (2024):

[2] Liu, Bo, et al. "Integration and Performance Analysis of Artificial Intelligence and Computer Vision Based on Deep Learning Algorithms." arXiv preprint arXiv:2312.12872 (2023).

[3] Zhiyu Li, et al. "BookGPT: A General Framework for Book Recommendation Empowered by Large Language Model." Electronics 12. 22 (2023):

[4] Yu, L., Liu, B., Lin, Q., Zhao, X., & Che, C. (2024). Semantic Similarity Matching for Patent Documents Using Ensemble BERT-related Model and Novel Text Processing Method. arXiv preprint arXiv:2401.06782.

[5] Liu, Bo, et al. "Integration and Performance Analysis of Artificial Intelligence and Computer Vision Based on Deep Learning Algorithms." arXiv preprint arXiv:2312.12872 (2023).

[6] Agathokleous Evgenios, et al. "One hundred important questions facing plant science derived using a large language model.." Trends in plant science (2023):

[7] Møller Lynge Asbjørn. "Recommended for You: How Newspapers Normalise Algorithmic News Recommendation to Fit Their Gatekeeping Role." Journalism Studies 23. 7 (2022):

[8] Yu, Liqiang, et al. “Research on Machine Learning With Algorithms and Development”. Journal of Theory and Practice of Engineering Science, vol. 3, no. 12, Dec. 2023, pp. 7-14, doi:10.53469/jtpes.2023.03(12).02.
[9] Huang, J., Zhao, X., Che, C., Lin, Q., & Liu, B. (2024). Enhancing Essay Scoring with Adversarial Weights Perturbation and Metric-specific AttentionPooling. arXiv preprint arXiv:2401.05433.
[10] Tan, Kai, et al. “Integrating Advanced Computer Vision and AI Algorithms for Autonomous Driving Systems”. Journal of Theory and Practice of Engineering Science, vol. 4, no. 01, Jan. 2024, pp. 41-48, doi:10.53469/jtpes.2024.04(01).06.
[11] Liu, B., Zhao, X., Hu, H., Lin, Q., & Huang, J. (2023). Detection of Esophageal Cancer Lesions Based on CBAM Faster R-CNN. Journal of Theory and Practice of Engineering Science, 3(12), 36–42. https://doi.org/10.53469/jtpes.2023.03(12).06