Transactions on Computational and Scientific Methods | Vo. 4, No. 11, 2024
ISSN: 2998-8780
https://pspress.org/index.php/tcsm
Pinnacle Science Press

Personalized Multimodal Recommendations Framework Using Contrastive Learning

Ankai Liang
Independent Researcher , Newark, USA
liangankai123@gmail.com

Abstract: This study proposes a multimodal cross-domain recommendation framework based on contrastive learning to address the shortcomings of traditional recommendation systems in dealing with multimodal data and cross-domain user preferences. By encoding multimodal features such as text and images of users and items, and using contrastive learning strategies to construct positive and negative sample pairs, the framework can effectively capture the associations between different modalities in high-dimensional feature space. In addition, through the cross-domain alignment method, the migration and integration of user preferences are achieved, so that the model can adapt to the complex needs of users in multiple fields. Experimental results show that the framework outperforms traditional models in evaluation indicators such as NDCG and AUC, and shows high accuracy and robustness in dealing with cross-domain recommendation tasks. This study verifies the application potential of contrastive learning in recommendation systems and provides a new technical approach for future recommendation systems. At the same time, this paper compares and analyzes a variety of classic recommendation models, further highlighting the role of cross-domain information and multimodal fusion in improving recommendation effects. Future research directions include improving contrastive learning methods, exploring the use of more unsupervised data, and improving the real-time response capability of the model, so as to provide users with better personalized recommendation services.

Keywords: Contrastive learning, multimodal recommendation, cross-domain recommendation, personalized recommendation

1. Introduction

In recent years, the multimodal cross-domain recommendation framework based on contrastive learning has gradually emerged in the recommendation system. With the development of the Internet and social media, users have generated a large amount of rich multimodal data on different platforms. These data cover different forms such as text, images, and audio, providing a valuable multimodal information source for the recommendation system. Through the effective fusion of multimodal information, not only can the user's preferences be captured more comprehensively, but also more complex cross-domain relationships can be mined, thereby improving the accuracy of the recommendation system and user experience.

In the traditional recommendation framework, information between different modalities is often processed independently, while the introduction of contrastive learning provides a new idea [1]. By comparing and measuring the similarity of multimodal data, information on different modalities can be optimized collaboratively. Under this method, the system can better understand the implicit associations between the modalities and mine deeper recommendation features [2]. This technology not only expands the data utilization methods of the recommendation system but also improves the generalization ability of the model and provides users with more personalized recommendations [3].

Cross-domain recommendation is a further improvement on the basis of multimodal recommendation [4]. With the diversification of user needs, traditional single-domain recommendation systems make it difficult to meet the personalized needs of users in multiple platforms and scenarios. Through cross-domain recommendation, the system can break the information barriers between different fields and realize cross-platform user preference migration, so that the user behavior of the system in a certain field can provide an effective reference for recommendations in other fields. This cross-domain recommendation method not only broadens the breadth of recommendations but also brings more diverse recommendation content to meet the personalized needs of users in different scenarios [5].

The application of contrastive learning in multimodal and cross-domain recommendations not only improves the accuracy of recommendations but also enhances the robustness of the model. In the face of data imbalance, noise interference and changes in user preferences, through contrastive learning, the model can more stably capture the user's real interests and preferences. This superior robustness enables the recommendation system to maintain high recommendation quality and adaptability under dynamically changing user needs, bringing users a better user experience.

In addition, contrastive learning effectively reduces the dependence on labeled data and saves the cost of data labeling through unsupervised or semi-supervised learning modes [6]. This is particularly important for recommendation systems that need to process a large amount of unlabeled data. Under the contrastive learning framework, the recommendation system can obtain better training results with fewer annotations, while continuously expanding the recommendation scope of the model, thereby improving the overall performance of the system.

2. Method

The construction of a multimodal cross-domain recommendation framework based on contrastive learning requires multiple key steps, including the extraction and fusion of multimodal features [7], the construction of the objective function of contrastive learning [8], and the alignment of cross-domain information [9]. In this method, different modal data are optimized through contrastive learning, and cross-domain associations are used to improve the accuracy of the recommendation system. The network architecture of contrastive learning is shown in Figure 1.

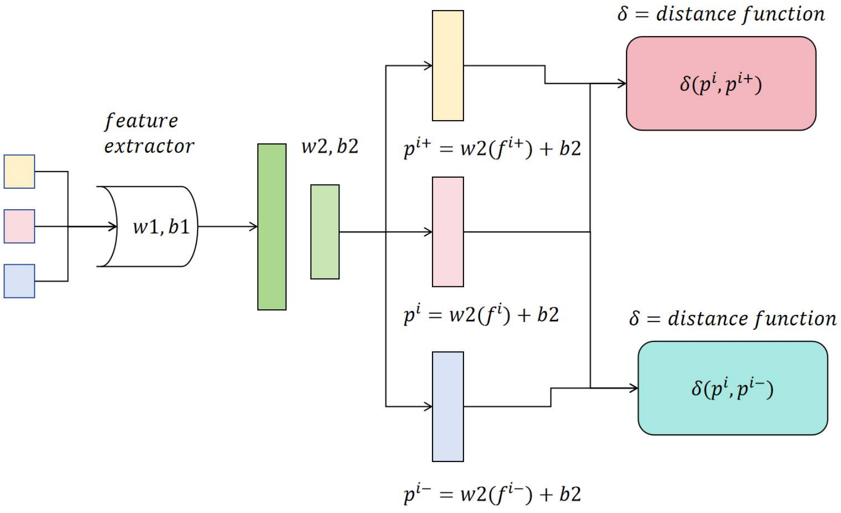
$\delta = distance\ function$
$\delta(p^i, p^{i+})$
$feature$ $extractor$
$w1, b1$
$w2, b2$
$p^{i+} = w2(f^{i+}) + b2$
$p^i = w2(f^i) + b2$
$\delta = distance\ function$
$\delta(p^i, p^{i-})$
$p^{i-} = w2(f^{i-}) + b2$

Figure 1. Overall architecture of the model

First, for feature extraction of multimodal data, assume that the modal data of users and items are X_u and X_i respectively, where X_u represents the features of users, including modalities such as text and images, and X_i represents the multimodal features of items. The data of each modality can be input into its dedicated encoder. For example, the text modality can be encoded using a pre-trained BERT model [10], and the image modality can be encoded using a deep convolutional neural network such as ResNet [11]. We get the multimodal embedding vectors of users and items as E_u and E_i, respectively, that is:

$$E_u = f_{text}(X_u^{text}) + f_{image}(X_u^{image}) + \ldots$$

$$E_i = f_{text}(X_i^{text}) + f_{image}(X_i^{image}) + \ldots$$

Among them, f_{text} and f_{image} represent the feature extraction functions of text and image modalities respectively. In this way, we map the data of different modalities into a unified feature space for subsequent contrastive learning tasks.

Secondly, we construct the objective function of contrastive learning to optimize the quality of multimodal embedding. We use the contrastive loss of positive and negative sample pairs to improve the distinguishability of embedding. Define the positive sample of each user u as i^+ and the negative sample set as $\{i^-\}$. The loss function of contrastive learning can be expressed as:

$$L = -\log \frac{\exp(sim(Eu, Ei+)/\tau)}{\exp(sim(Eu, Ei+)/\tau) + \sum_{i^- \in N} \exp(sim(E_u, E_{i-})/\tau)}$$

Among them, $sim(\cdot)$ is the similarity function, usually cosine similarity, and τ is the temperature coefficient, which is used to control the distribution smoothness in contrastive learning. The role of this loss function is to maximize the similarity between positive sample pairs and minimize the similarity between negative sample pairs.

Next, in order to achieve effective alignment of cross-domain information [12], we map the user's behavior data in different domains into the same space. Let $E_u^{d_1}$ and $E_u^{d_2}$ represent the feature embeddings of the user in domains d_1 and d_2 respectively. Through contrastive learning, we can make the embeddings of the same user in different domains as close as possible, thereby capturing the user's interest preferences across domains. The cross-domain contrast loss is defined as:

$$L_{domain} = \| E_u^{d_1} - E_u^{d_2} \|^2$$

The goal of this loss is to minimize the embedding distance of the same user in different fields so that the system can better understand the user's cross-field preferences and provide more accurate recommendations.

In order to further improve the recommendation performance, we integrate multimodal contrastive learning and cross-field alignment loss [13] into the overall optimization objective function:

$$L_{total} = \alpha L + \beta L_{domain}$$

Among them, α and β are weight parameters used to balance the multimodal contrast loss and cross-domain alignment loss. The optimization of this total loss function can not only enhance the distinguishability of multimodal features but also effectively integrate cross-domain user preference information.

During the training process of the model, positive and negative sample pairs are constructed by random sampling, and the total loss function is minimized using the gradient descent method. After the optimization is completed, the recommendation system can generate multimodal embedding vectors for users and items, and use these vectors to calculate the user's interest in unseen items. Finally, a recommendation list is generated through a sorting algorithm to achieve personalized and cross-domain recommendations.

3. Experiment

3.1 Datasets

The dataset used in this experiment is the Amazon-Beauty cross-domain recommendation dataset, which contains user interaction data in different fields, especially covering the two main fields of "beauty" and "fashion". This dataset comes from real users' purchase records and product evaluation data, including hundreds of thousands of user-product interaction information. Each record contains a user ID, product ID, interaction timestamp, evaluation score, etc., providing rich multimodal data support. Specifically, products in the beauty field include cosmetics and skin care products, while the fashion field includes clothing, shoes, etc. This diverse data source provides strong support for cross-domain recommendations, enabling the model to make recommendations based on the relevance of products in different fields.

In addition, in order to analyze multimodal features, the dataset also provides multimodal information such as product text descriptions and images. Text can be used to obtain product description information, such as ingredients, efficacy and other details, while images show the appearance and design style of the product. These modal features provide a rich foundation for comparative learning. The multimodal information in the dataset helps the model capture user preferences more comprehensively, which can not only improve the accuracy of recommendations but also enhance the promotion and generalization capabilities of the model in different fields.

The diversity and scale of the dataset provide sufficient challenges for the experiment. The purchasing behaviors of users in the two domains are significantly different and cross-domain related, which provides a good basis for testing and optimizing the multimodal contrastive learning framework. This experiment verifies the effectiveness of the model by integrating features of different modalities and aligning user preferences across domains, and achieves a high recommendation accuracy on this dataset, demonstrating the application potential of contrastive learning in multimodal and cross-domain recommendations.

3.2 Experimental setup

The setup of this experiment revolves around the effectiveness of multimodal contrastive learning and the evaluation of cross-domain recommendation performance. First, the dataset is divided into a training set and a test set with an 80-20 ratio to ensure that the model can be evaluated on unseen data. To improve the performance of the model in multimodality and cross-domain [14], we designed a multimodal feature extraction process, where the text description is input into the BERT encoder [15]to extract text features, and the image is passed through the ResNet network to extract visual features [16]. In this way, the multimodal features of each user and item are encoded into a high-dimensional vector to support the optimization process of contrastive learning.

During the experiment, the core of contrastive learning is to construct positive and negative sample pairs, so we randomly sample positive and negative sample pairs in the training set. For each user and their favorite items, other items are randomly selected as negative samples to ensure that the model learns user preferences more accurately in the optimization of positive and negative sample pairs. In terms of optimization, we set the temperature parameter τ of contrastive learning to adjust the smoothness of the similarity metric and select the best temperature value through hyperparameter tuning. The weight parameters of cross-domain alignment are adjusted according to the experimental results to balance the impact of multimodal contrast loss and cross-domain preference alignment loss.

Finally, in order to evaluate the recommendation effect of the model, the experiment uses indicators such as precision, recall, and F1 score to evaluate the performance of the recommendation system on the test set. These indicators can comprehensively measure the accuracy and coverage of the model in the recommendation task [17]. In the experiment, we also conducted a cross-domain effect analysis to observe the recommendation results of the model in the two fields of beauty and fashion to verify the effectiveness of the contrastive learning framework for multimodal and cross-domain information. The experimental results show that the framework has high recommendation accuracy in the two major fields, especially when dealing with cross-domain preferences.

3.3 Experiments

This experiment selected five different recommendation models for comparison to comprehensively evaluate the performance of the multimodal cross-domain recommendation framework based on contrastive learning. First, the traditional matrix factorization (MF) [18] model is a collaborative filtering method based on user-item interaction data, which obtains the implicit features of users and items by decomposing the matrix. Secondly, neural collaborative filtering (NCF) introduces neural networks into collaborative filtering, captures nonlinear feature relationships through multi-layer perceptrons [19], and enhances the recommendation effect. Then there is the recommendation model based on graph neural network (GNN-based), which uses graph structure to learn the complex relationship between users and items and is particularly suitable for modeling multimodal and complex network relationships. The fourth model is a cross-domain recommendation based on graph embedding (GCE) [20], which realizes the transmission and fusion of user preferences through cross-domain graph structure embedding [21]. Finally, the cross-modal contrastive learning model (CMC) is used. Our model optimizes the relationship between modalities through contrastive learning of multimodal data to achieve more accurate multimodal recommendations. By comparing with these models, this experiment can more accurately verify the effectiveness of the proposed framework, and its experimental results are shown in Table 1.

Table 1. Experiment result

Model	NDCG	AUC
MF	0.612	0.733
NCF	0.648	0.759
GNN-based	0.673	0.783
GCE	0.694	0.802
CMC(ours)	0.725	0.826

According to the specific indicators in the experimental results, the CMC model achieved the highest scores in both NDCG and AUC, reaching 0.725 and 0.826, respectively. These values are significantly superior to those of other models, demonstrating that CMC can more effectively capture users' cross-domain preferences and deliver higher-quality recommendations within recommendation systems. In contrast, the MF model scored 0.612 in NDCG and 0.733 in AUC, highlighting its limitations in cross-domain recommendation tasks when relying solely on single-modal information, making it challenging to thoroughly capture users' cross-domain interests and features.

The NCF model uses the multi-layer perceptron structure of the neural network to improve the NDCG and AUC indicators, reaching 0.648 and 0.759 respectively. However, although its nonlinear expression enhances the recommendation effect of the model in specific fields, it is still difficult to adapt to the data fusion needs of multiple fields, and the effect of cross-field recommendation is limited. In contrast, the GNN-based model uses graph neural networks to model the complex relationship between users and items, improving recommendation quality. Its NDCG and AUC are 0.673 and 0.783 respectively, but the capture of cross-domain preferences is still insufficient.

In the GCE model, NDCG and AUC increased to 0.694 and 0.802 respectively, indicating that the strategy of user preference transfer through cross-domain graph embedding can help improve the recommendation effect. However, although the GCE model can capture cross-domain interest transfer to a certain extent, it still has shortcomings in the fusion of multi-modal features and fails to achieve the effect of CMC.

Finally, relying on the cross-modal contrastive learning framework, the CMC model achieved significant improvements in NDCG and AUC, allowing users' interests and preferences in different fields to be better integrated and utilized. This shows that the CMC model has stronger generalization ability and accuracy in multi-modal and cross-domain recommendation tasks, and can effectively improve the overall performance of the recommendation system.

In addition, we also give the decline graphs of two loss functions, as shown in Figure 2.

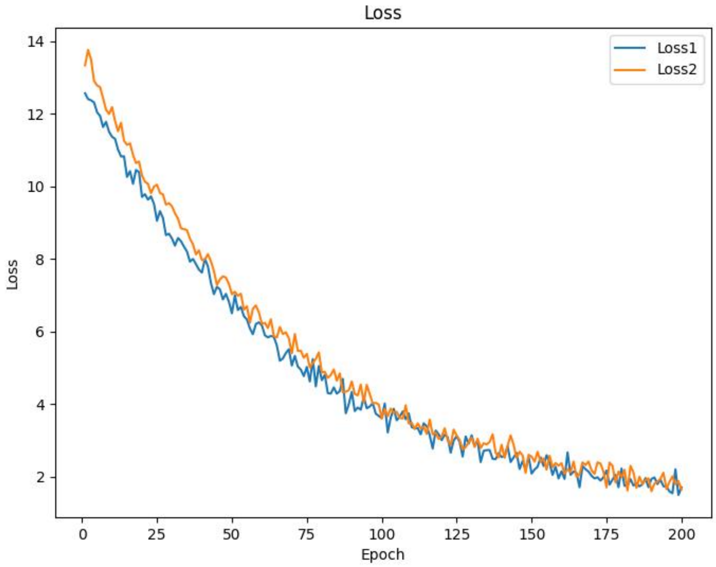

Figure 2. The curves of the two loss functions decreasing during training

As can be seen from Figure 2, the curves of the two loss functions show a similar downward trend, and overall they gradually decrease as training progresses. The initial and final values of the two curves are slightly different, but both show a gradual convergence process. This shows that during the model training process, the loss values of the two experiments gradually became stable, achieving the expected optimization effect. Such loss changes indicate that the model is able to gradually adapt to the data and reduce prediction errors during training.

Judging from the details, there are some small fluctuations in the curve. During the deep learning training process, fluctuations in the loss value are normal and are usually caused by factors such as data diversity and random initialization. Although the two curves will cross or separate at certain epochs, the overall trend is consistent. This consistency indicates that the model is robust and is not prone to significantly affecting the training results due to small differences in the initial state.

Finally, the gradual convergence of these two loss curves, especially the leveling off near the final value, indicates that the model gradually approaches the optimal state during the longer training process. The difference in the final loss value is not significant, indicating that the effects of the two sets of experiments are similar, and the model can achieve a good performance level under different settings. Such analysis is of great significance for understanding the stability and convergence of the model, helping to select appropriate training parameters in practical applications, and ensuring consistent performance of the model under different initial conditions.

Finally, we also conducted relevant ablation experiments on the modules added in the article. The experimental results are shown in Table 2.

Table 2. Ablation Experiment Results

Different experimental settings	**NDCG**	**AUC**
All modules (complete model)	0.725	0.826
No contrastive learning module	0.691	0.805
No cross-domain feature fusion	0.673	0.783
No Image Mode	0.702	0.810
No multimodal feature fusion	0.658	0.764

It can be seen from the ablation experiment results in Table 2 that the changes in model performance under different experimental settings fully demonstrate the contribution of each module to the overall performance of the model. By analyzing these results, the importance of contrastive learning modules, cross-domain feature fusion, multi-modal feature fusion, and specific modality data in recommendation systems can be revealed. These experimental results provide strong support for the rationality of the model design, and are analyzed in detail below from the perspective of different modules.

First of all, the role of the comparative learning module is obvious. When the contrastive learning module is removed, NDCG drops from 0.725 to 0.691, AUC drops from 0.826 to 0.805, and the performance drops significantly. This shows that contrastive learning plays a key role in optimizing the relationship between multimodal data. By introducing pairs of positive and negative samples, contrastive learning can enhance the consistency between features of different modalities and improve the model's ability to understand the commonalities and differences between modalities. After removing this module, it is difficult for the model to effectively capture the deep correlation between multi-modal data, resulting in a significant reduction in recommendation effect. Therefore, this module is crucial to improve the performance of multi-modal recommendation systems.

Secondly, the contribution of cross-domain feature fusion cannot be ignored. After removing the cross-domain feature fusion module, NDCG and AUC dropped to 0.673 and 0.783 respectively, which was second only to the removal of the contrastive learning module. This shows that cross-domain feature fusion is of great significance for capturing user preference migration in different domains. Users' interests usually have cross-domain correlations, and through cross-domain feature fusion, the model can more comprehensively integrate users' preference information in different fields, thereby improving the accuracy of recommendations. Once this module is removed, the model's performance on data in a single domain will be limited and it will be unable to fully utilize the cross-domain behavioral characteristics of users.

Furthermore, significance analysis for multi-modal data shows that when image modalities are removed, NDCG and AUC decrease to 0.702 and 0.810, respectively. Although the performance decrease is smaller, it still shows that image modality provides supplementary information to the recommendation system to a certain extent. The introduction of multi-modal data enables the model to understand the characteristics of users and items from the perspectives of different modalities. Even if a certain modality (such as text or image) has less information in some scenes, it may still be useful in certain situations. Enhance the richness and diversity of recommendations. Therefore, removing the image modality leads to a slight decrease in performance, indicating that the comprehensiveness of multi-modal information has a positive effect on the improvement of the recommendation system.

Finally, the core role of multi-modal feature fusion has also been verified. When the multi-modal feature fusion module is removed, NDCG and AUC drop to 0.658 and 0.764 respectively, showing the worst performance. Compared with the removal of the contrastive learning module, the reduction caused by the removal of multi-modal feature fusion is more significant. This shows that it is difficult for a single modal feature expression to fully capture the complex relationship between users and items, while the fusion of multi-modal features can effectively make up for the shortcomings of single modal data and provide the model with a more comprehensive and rich ability to express user interests. Therefore, multi-modal feature fusion is a core component of the entire model, and its removal directly leads to a serious decrease in recommendation performance.

Based on comprehensive analysis, the results of this ablation experiment further verify the rationality and necessity of the complete model design. The contrastive learning module and cross-domain feature fusion are key parts to improve the cross-domain and multi-modal capabilities of the recommendation system, while the completeness of modal information and the fusion of multi-modal features provide richer data support for the model. Through the synergy of these modules, the model can show higher accuracy and robustness in multi-modal cross-domain recommendation tasks.

4. Conclusion

Through this experiment, we successfully verified the advantages of the multi-modal cross-domain recommendation framework based on contrastive learning in improving recommendation effects. Experimental results show that this framework is significantly better than traditional single-modality and non-cross-domain models in indicators such as NDCG and AUC, and is especially good at handling user cross-domain preferences and multi-modal feature fusion. This achievement demonstrates the potential of contrastive learning in complex recommendation tasks, providing strong support for the personalization and accuracy of recommendation systems.

It can be seen from the experimental analysis that the fusion of multi-modal information and the alignment of cross-domain preferences are crucial to improving the performance of recommendation systems. By comprehensively considering users' interests and preferences in different fields, we can understand user needs more comprehensively and provide more personalized recommendation content. This research also provides a new idea for the field of recommendation systems, which is to effectively combine modal information and cross-domain data through comparative learning to break through the limitations of traditional recommendation methods and enhance the adaptability and generalization capabilities of the system.

In the future, with the increasing types of multi-modal data and the diversification of user needs, recommendation systems will face more complex challenges. We can further explore more efficient contrastive learning methods and combine them with other deep learning techniques to improve the accuracy and real-time performance of the model in cross-domain recommendations. In addition, how to use more unlabeled data for unsupervised learning and reduce data labeling costs will also become an important research direction. These prospects will help promote the development of recommendation system technology and bring users a richer and more personalized experience.

References

[1] K. Sevegnani, A. Seshadri, T. Wang, A. Beniwal, J. McAuley, A. Lu and G. Medioni, "Contrastive Learning for Interactive Recommendation in Fashion," arXiv preprint arXiv:2207.12033, 2022.

[2] W. Yang, H. Zhang and L. Zhang, "Variational Invariant Representation Learning for Multimodal Recommendation," Proceedings of the 2024 SIAM International Conference on Data Mining (SDM), Society for Industrial and Applied Mathematics, 2024.

[3] X. Liu, et al., "Heterogeneous Graph-Based Framework with Disentangled Representations Learning for Multi-Target Cross Domain Recommendation," arXiv preprint arXiv:2407.00909, 2024.

[4] I. Fernández-Tobías, I. Cantador, M. Kaminskas and F. Ricci, "Cross-Domain Recommender Systems: A Survey of the State of the Art," Proceedings of the Spanish Conference on Information Retrieval, vol. 24, Valencia, Spain, ACM, June 2012.

[5] R. Zhao, et al., "ASR-Enhanced Multimodal Representation Learning for Cross-Domain Product Retrieval," arXiv preprint arXiv:2408.02978, 2024.

[6] J. T. Springenberg, "Unsupervised and Semi-Supervised Learning with Categorical Generative Adversarial Networks," arXiv preprint arXiv:1511.06390, 2015.

[7] Z. Liu, M. Wu, B. Peng, Y. Liu, Q. Peng and C. Zou, "Calibration Learning for Few-shot Novel Product Description," Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 1864-1868, July 2023.

[8] J. Cao, R. Xu, X. Lin, F. Qin, Y. Peng and Y. Shao, "Adaptive Receptive Field U-Shaped Temporal Convolutional Network for Vulgar Action Segmentation," Neural Computing and Applications, vol. 35, no. 13, pp. 9593-9606, 2023.

[9] J. Song and Z. Liu, "Comparison of Norm-Based Feature Selection Methods on Biological Omics Data," Proceedings of the 5th International Conference on Advances in Image Processing, pp. 109-112, November 2021.

[10] J. Du, Y. Cang, T. Zhou, J. Hu and W. He, "Deep Learning with HM-VGG: AI Strategies for Multi-modal Image Analysis," arXiv preprint arXiv:2410.24046, 2024.

[11] J. Yao, J. Wang, B. Wang, B. Liu and M. Jiang, "A Hybrid CNN-LSTM Model for Enhancing Bond Default Risk Prediction," Journal of Computer Technology and Software, vol. 3, no. 6, 2024.

[12] B. Chen, F. Qin, Y. Shao, J. Cao, Y. Peng and R. Ge, "Fine-Grained Imbalanced Leukocyte Classification with Global-Local Attention Transformer," Journal of King Saud University - Computer and Information Sciences, vol. 35, no. 8, Article ID 101661, 2023.

[13] S. Lu, Z. Liu, T. Liu and W. Zhou, "Scaling-up Medical Vision-and-Language Representation Learning with Federated Learning," Engineering Applications of Artificial Intelligence, vol. 126, Article ID 107037, 2023.

[14] Y. Zi, X. Cheng, T. Mei, Q. Wang, Z. Gao and H. Yang, "Research on Intelligent System of Medical Image Recognition and Disease Diagnosis Based on Big Data," Proceedings of the 2024 IEEE 2nd International Conference on Image Processing and Computer Applications (ICIPCA), pp. 825-830, June 2024.

[15] S. Liu, G. Liu, B. Zhu, Y. Luo, L. Wu and R. Wang, "Balancing Innovation and Privacy: Data Security Strategies in Natural Language Processing Applications," arXiv preprint arXiv:2410.08553, 2024.

[16] D. Sun, M. Sui, Y. Liang, J. Hu and J. Du, "Medical Image Segmentation with Bilateral Spatial Attention and Transfer Learning," Journal of Computer Science and Software Applications, vol. 4, no. 6, pp. 19-27, 2024.

[17] X. Yan, Y. Jiang, W. Liu, D. Yi, and J. Wei, "Transforming Multidimensional Time Series into Interpretable Event Sequences for Advanced Data Mining", arXiv preprint, arXiv:2409.14327, 2024.

[18] D. Bokde, S. Girase and D. Mukhopadhyay, "Matrix Factorization Model in Collaborative Filtering Algorithms: A Survey," Procedia Computer Science, vol. 49, pp. 136-146, 2015.

[19] S. Rendle, W. Krichene, L. Zhang and J. Anderson, "Neural Collaborative Filtering vs. Matrix Factorization Revisited," Proceedings of the 14th ACM Conference on Recommender Systems, pp. 240-248, September 2020.

[20] L. A. Vigo del Rosso, "Context-Aware Recommender Systems with Graph Convolutional Embeddings (CARS-GCE)," 2021.

[21] Z. Wang, W. Wei, G. Cong, X. L. Li, X. L. Mao and M. Qiu, "Global Context Enhanced Graph Neural Networks for Session-Based Recommendation," Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 169-178, July 2020.