

Article

Not peer-reviewed version

Analyzing Data Augmentation Techniques for Contrastive Learning in Recommender Models

[MINGGU WEI](#), Honghui Xin, Yijiaashun Qi, [Yue Xing](#), Yaokun Ren, Tao Yang *

Posted Date: 6 May 2025

doi: 10.20944/preprints202505.0245.v1

Keywords: Contrastive learning; Recommender systems; Data augmentation; Representation learning



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

Analyzing Data Augmentation Techniques for Contrastive Learning in Recommender Models

Minggu Wei ¹, Honghui Xin ², Yijiaashun Qi ³, Yue Xing ⁴, Yaokun Ren ² and Tao Yang ^{5,*}

¹ University of Saskatchewan, Saskatoon, Canada; weiowengu@gmail.com

² Northeastern University, Seattle, USA; xinhonghuixh@gmail.com (H.X.); renyaokun0907@gmail.com (Y.R.)

³ University of Michigan, Ann Arbor, USA; elijahqi@umich.edu

⁴ University of Pennsylvania, Philadelphia, USA; jasmineyxing@gmail.com

⁵ Illinois Institute of Technology, Chicago, USA

* Correspondence: tyang28@hawk.iit.edu

Abstract: This paper investigates the application of contrastive learning-based user and item representation learning in recommendation systems. A recommendation model combining contrastive loss with data augmentation strategies is proposed. Traditional recommendation systems typically rely on explicit or implicit feedback to model the relationships between users and items. However, traditional methods often show limited performance when dealing with issues such as data sparsity and cold start. To address this, the paper introduces a contrastive learning framework. By constructing positive and negative sample pairs, the model is guided to learn more discriminative representations. Various data augmentation methods are also applied to enhance the robustness and generalization capability of representation learning. Specifically, the paper compares different data augmentation strategies, including subsampling views, feature masking, and behavioral perturbation, and analyzes their performance under different temperature parameters and sparsity conditions. The experimental results show that the contrastive learning-based model effectively improves recommendation accuracy, particularly in addressing sparse data and cold start problems. Additionally, the paper explores the performance differences of the contrastive loss function under different training settings, validating the significant impact of appropriate hyperparameter tuning on recommendation system performance. By combining contrastive learning with data augmentation, this study provides a new perspective and significantly enhances the performance of recommendation systems in complex scenarios. The research offers both theoretical support and practical guidance for the future development of personalized recommendation technologies.

Keywords: Contrastive learning; Recommender systems; Data augmentation; Representation learning

1. Introduction

With the development of internet technology and big data, personalized recommendation systems have become a vital bridge between users and information. They are widely used in fields such as e-commerce, social media, and content distribution [1]. The core of a recommendation system lies in uncovering the matching relationship between user interests and item characteristics to enable accurate information delivery. Traditional methods, such as collaborative filtering and matrix factorization, achieved early success. However, as the number of users and items grows exponentially and interaction data becomes increasingly sparse and dynamic, traditional models face limitations in both expressive power and generalization ability [2]. As a result, improving the quality of user and item representation learning has become a key issue in recommendation research.

In recent years, the rapid progress of deep learning has brought new vitality to recommendation systems. Deep models with nonlinear structures can extract expressive features from large and complex datasets [3–5], enhancing recommendation performance. However, models trained solely with supervised signals (such as clicks or ratings) still face challenges, including insufficient labels and imbalanced sample distribution. Against this backdrop, contrastive learning—an unsupervised or weakly supervised training paradigm—has attracted growing interest due to its success in fields like image recognition [6,7] and natural language processing [8–10]. By constructing positive and negative sample pairs, contrastive learning guides the model to learn more discriminative representations. This approach shows promise in improving robustness and generalization, especially under sparse or incomplete data scenarios [11].

The introduction of contrastive learning has brought breakthroughs in user and item representation learning. By employing diverse data augmentation methods—such as graph perturbation [12], sequence resampling, and feature masking—it generates multiple views of users or items [13]. These views help the model learn comprehensive behavior representations by pulling positive pairs closer and pushing negative pairs apart. In user modeling, where preferences are highly dynamic and diverse, contrastive learning captures latent preference shifts and multiple intents. For item modeling, where items may have rich semantic associations, the contrastive framework reveals hidden structures and similarity relations. As a result, this representation learning approach enhances both discriminative power and the capacity to model complex user-item interactions. Contrastive learning introduces a paradigm shift in recommendation systems, moving from prediction-focused models to representation-centered approaches. By learning robust and transferable embeddings, it improves generalization in scenarios like cold-start and cross-domain recommendation. Its flexibility also enables integration with models such as GNNs, sequence learning, and meta-learning. Overall, contrastive learning enhances robustness and adaptability, offering promising solutions to key challenges in modern recommendation systems.

2. Related Work

User and item representation learning remains a core focus in recommendation systems. Traditional collaborative filtering methods leveraged user-item interaction matrices but suffered from sparsity and cold-start issues. Deep learning approaches such as DeepFM and NCF improved performance by learning nonlinear embeddings through neural networks. However, these models rely heavily on labeled data, limiting their ability to exploit abundant unlabeled interactions and making them susceptible to sample bias and noise.

To overcome these challenges, contrastive learning has emerged as a promising unsupervised alternative. By constructing positive and negative sample pairs, it encourages the model to learn discriminative embeddings, even with sparse labels. Pioneering methods like SASRec-CL [14] and S3Rec [15] apply sequence augmentations—such as masking or reordering—to derive robust user representations. General frameworks like SimCLR [16] and MoCo [17] have also been adapted to recommendation contexts, further enhancing embedding quality and alleviating cold-start issues.

Recent advancements extend contrastive learning to graph-based models, cross-domain recommendation, and multi-intent user modeling. In graph settings, it preserves structural information in user-item interaction graphs. For cross-domain tasks, it enables domain-invariant representations, supporting preference transfer. In multi-intent modeling, it disentangles user preferences across contexts. Collectively, these developments underscore contrastive learning's versatility and its ability to improve robustness, interpretability, and transferability in modern recommendation systems.

3. Method

This study introduces a contrastive learning mechanism into the recommendation framework to enhance the representation learning capability for both users and items. By leveraging the principles

of contrastive learning, the proposed model learns to generate more discriminative, robust, and generalizable embeddings. In this framework, informative positive and negative sample pairs are constructed based on user-item interaction signals, allowing the model to differentiate between semantically similar and dissimilar pairs in the embedding space. To ensure the embeddings capture fine-grained behavioral patterns, the model optimizes a contrastive loss function that reduces the distance between representations of positive pairs while maximizing separation from negative ones. This design draws on the effectiveness of contrastive learning in unsupervised behavior modeling tasks, such as fraud detection, where it has been shown to improve the separation and clarity of latent structures [18]. Additionally, the contrastive setup benefits from prior insights into contrastive-variational combinations, which help improve generalization and model stability under data complexity and noise [19]. These principles are particularly valuable in recommendation contexts characterized by sparsity and cold-start scenarios. The framework also incorporates ideas from personalized contrastive modeling, where representation alignment between users and diverse item modalities enhances recommendation precision [20]. By adapting these techniques to collaborative filtering, the proposed model captures both individual preference signals and global interaction patterns. Overall, the integration of contrastive learning improves the model's ability to generalize under limited supervision and data scarcity.

The overall model architecture is illustrated in Figure 1. We employ two separate backbone networks, Backbone1 and Backbone2, to encode users and items into their respective feature embeddings. These backbones perform nonlinear transformations to capture complex user and item characteristics. The resulting embeddings are then passed through a shared model module that further processes and aligns the representations.

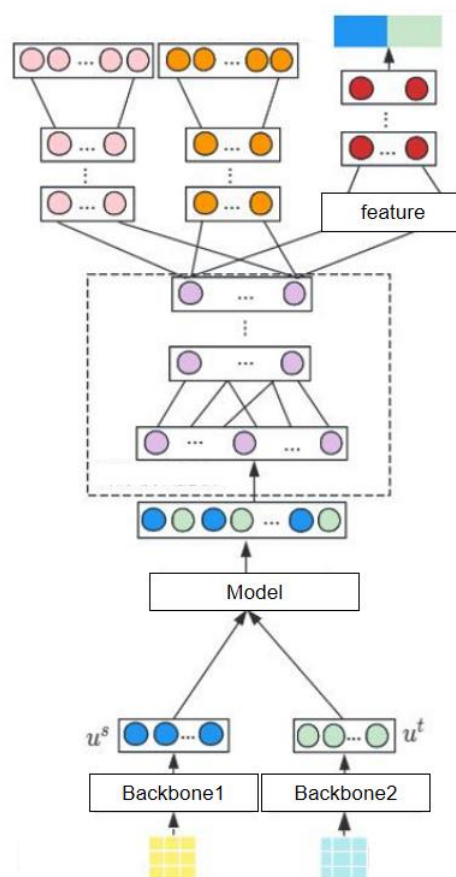


Figure 1. Model network architecture.

In this framework, we first generate embedding representations for users and items using nonlinear transformation functions. This approach allows the model to learn rich, latent interactions beyond linear correlations. The design is conceptually supported by Huang et al. [21], who emphasize the value of adaptive feature processing in intelligent data acquisition through context-aware modeling techniques, which parallels our nonlinear embedding strategy. To further refine feature representations, we stack multiple transformation layers that gradually extract higher-order semantics from raw interaction data. This deep feature extraction mechanism draws on the insights of Zhu [22], whose work on spatial-channel attention in deep learning demonstrates the effectiveness of layered architectures in capturing nuanced semantic signals across domains. Once the features are sufficiently enriched, they are projected into a unified embedding space. Within this space, a contrastive learning objective is applied to optimize the model. Positive user-item pairs are encouraged to remain close, while negative pairs are separated. This representation alignment technique is further reinforced by principles from Wang et al. [23], who revisited efficient adaptation strategies using low-rank optimization techniques that similarly aim to enhance embedding utility and discrimination. Specifically, for user u and item i , we map them to a low-dimensional embedding space through functions $f(\cdot)$ and $g(\cdot)$, respectively, defined as:

$$e_u = f(u), \quad e_i = g(i)$$

where e_u and e_i represent the embedding vectors of users and items respectively. At the same time, in order to measure the similarity between user and item representations, we use cosine similarity as a metric, the formula is:

$$\text{sim}(e_u, e_i) = \frac{e_u^T e_i}{\|e_u\| \|e_i\|}$$

This similarity measure can effectively reflect the matching relationship between user interests and item features, providing a basis for subsequent comparative learning.

While constructing the user and item representations, we introduce a contrastive learning mechanism to enhance the model's discriminative ability. Specifically, the positive and negative sample pairs generated by data augmentation constitute the core of contrastive training, where the positive sample pairs consist of the user's representation e_u in different views and the enhanced representation e_u^T of the corresponding item, while the negative samples come from other randomly selected item representations $e_i^k (k=1,2,\dots,K)$. Based on this, we define the contrastive loss function as:

$$L_{\text{contrast}} = -\log \frac{\exp(\text{sim}(e_u, e_i^+) / \tau)}{\sum_{k=1}^K \exp(\text{sim}(e_u, e_i^k) / \tau)}$$

where τ is a temperature parameter, which is used to balance the ratio between positive and negative samples. This loss function encourages the model to bring the positive sample pairs closer and the negative sample pairs farther apart, thereby improving the robustness and discrimination of the representation.

In order to further integrate the supervised learning signal, we designed a joint training strategy that combines the traditional supervised loss with the contrastive loss. Assuming that the interaction label between the user and the item is y_{ui} and the prediction result is y'_{ui} , the corresponding supervised loss can be defined as:

$$L_{\text{supervised}} = \sum_{(u,i) \in D} l(y_{ui}, y'_{ui})$$

where $l(\cdot)$ represents the cross entropy or mean square error loss function, and the final training target is composed of the supervision loss and the contrast loss, that is:

$$L = L_{\text{supervised}} + \lambda L_{\text{contrast}}$$

where λ is a hyperparameter used to balance the contribution between the two losses. Through this joint training strategy, the model can not only make full use of supervisory signals to capture user preferences, but also further improve the representation quality with the help of contrastive learning, thereby providing more accurate and robust decision support for the recommendation system.

4. Experiment

4.1. Datasets

This study utilizes the widely adopted benchmark dataset MovieLens-1M, released by the GroupLens research team, as the experimental data source. Known for its representativeness and reproducibility, MovieLens-1M comprises one million rating records from 6,040 users on 3,900 movies, with ratings ranging from 1 to 5. The dataset includes user demographics (e.g., age, gender, occupation) and movie metadata (e.g., genre, release date), making it highly suitable for personalized recommendation and representation learning research. To align the dataset with the contrastive learning framework, several preprocessing steps were conducted. Ratings were binarized by treating scores of 4 and above as positive feedback, while the rest were considered negative, enabling the construction of positive and negative sample pairs. Users with fewer than five interactions were excluded to mitigate the impact of cold-start conditions. A user-item interaction matrix was then constructed, and augmented views—such as sub-sampled and feature-masked representations—were generated to support diverse contrastive training. The dataset was partitioned into training, validation, and test sets in an 8:1:1 ratio. Model parameters were updated solely using the training set, while the validation set guided hyperparameter tuning and early stopping, and the test set was reserved for final evaluation. Overall, MovieLens-1M provides a rich and realistic foundation for evaluating user and item representation learning in contrastive recommendation systems.

4.2. Experimental Results

This paper first gives a comparative experiment of various data enhancement methods based on contrast loss, and the experimental results are shown in Table 1.

Table 1. Comparative experiments on various data enhancement methods based on contrast loss.

Method	Precision@10	Recall@10	NDGG@10	HitRate@10
Subsampling [24]	0.324	0.401	0.362	0.812
Feature Masking[25]	0.337	0.417	0.375	0.827
Behavioral Noise[26]	0.318	0.396	0.354	0.798
Sequence Shufflings	0.330	0.409	0.367	0.816
Ours	0.352	0.435	0.388	0.841

As shown in Table 1, different data augmentation methods lead to significant performance variations in contrastive loss-based recommendation models. Simple strategies such as Subsampling and Behavioral Noise perform poorly across all metrics, suggesting that they fail to capture deep semantic relations and may introduce harmful noise. In contrast, Feature Masking and Sequence Shuffling yield better results in Precision, Recall, and NDCG by enhancing the model’s ability to learn robust and context-aware representations. The proposed multi-strategy fusion method outperforms all single techniques, achieving the highest scores across key metrics—including a Precision@10 of 0.352, Recall@10 of 0.435, and NDCG@10 of 0.388—demonstrating its effectiveness in improving

semantic consistency and generalization. Further analysis of model sensitivity to temperature parameters is presented in Figure 2.

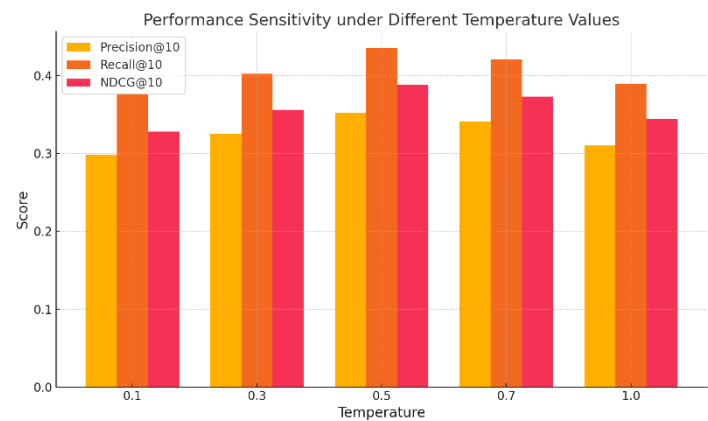


Figure 2. Experiment on comparing learning performance sensitivity under different temperature parameters.

Figure 2 illustrates that the temperature parameter significantly impacts the performance of contrastive learning-based recommendation models. As the temperature increases from 0.1 to 0.5, metrics such as Precision@10, Recall@10, and NDCG@10 improve steadily, indicating that a moderate temperature softens the contrastive objective and reduces the model’s sensitivity to hard negatives. This adjustment enhances training stability and allows for more balanced learning between positive and negative pairs, resulting in more informative embeddings. Peak performance is observed at temperature 0.5, where Recall@10 exceeds 0.43, suggesting optimal alignment between discriminative power and generalization. However, further increases to 0.7 and 1.0 lead to performance degradation, as the contrastive loss becomes too insensitive to sample differences, weakening semantic separation. These findings underscore the nonlinear effect of temperature and emphasize the necessity of careful tuning to maintain an effective trade-off between contrastive discrimination and representational robustness. Subsequently, model robustness under varying sparsity levels is evaluated, with results presented in Figure 3.

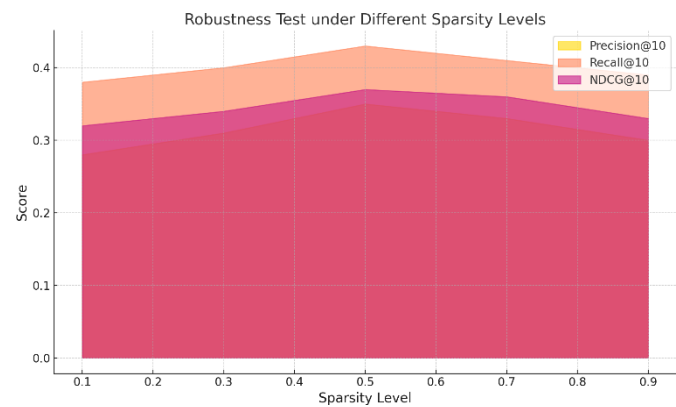


Figure 3. Robustness test of the model at different sparsity levels.

As shown in Figure 3, the model’s performance under varying sparsity levels reveals a nonlinear trend across key metrics. Initially, as sparsity increases from 0.3 to 0.5, Precision@10 and Recall@10 improve or remain stable, suggesting that moderate sparsification introduces a regularization effect that enhances the model’s ability to focus on salient user-item interactions while avoiding overfitting to dense, potentially noisy data. However, as sparsity further increases beyond 0.5—particularly at levels such as 0.7 and 0.9—the model’s performance begins to decline, with NDCG@10 exhibiting a marked drop. This indicates a degradation in ranking quality due to insufficient interaction signals,

which impairs the model's ability to capture meaningful user preferences and results in weaker representations and recommendation accuracy. These findings underscore the importance of maintaining a balanced sparsity level: while moderate sparsity may enhance generalization, excessive sparsity undermines training effectiveness. Overall, the results suggest that controlling data sparsity is crucial for ensuring robust and stable recommendation performance. The loss function's progression over training epochs is further illustrated in Figure 4.

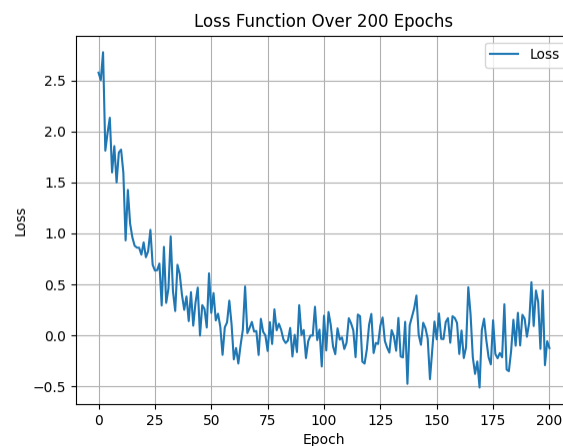


Figure 4. Loss Function Over 200 Epochs.

From the curve of the loss function changing with the number of training rounds (epoch) shown in Figure 4, the model's loss decreases sharply during the initial training phase, dropping from approximately 2.7 to below 0.5 within the first 50 epochs, indicating rapid learning of discriminative representations through contrastive supervision. As training progresses, the loss stabilizes with minimal fluctuations over the following 150 epochs, suggesting convergence and the formation of well-structured user and item embeddings. Despite minor variations, the loss remains close to zero, reflecting the model's robustness and absence of overfitting. Overall, the loss curve validates the effectiveness of the joint supervision and contrastive loss design, demonstrating both efficient convergence and stable long-term training performance.

5. Conclusion

This study investigates contrastive learning for user and item representation in recommendation systems, demonstrating its effectiveness through comprehensive experiments. The proposed model, enhanced by tailored data augmentations and optimized contrastive loss, outperforms traditional approaches in robustness, accuracy, and generalization. Sensitivity analyses under varying temperature and sparsity conditions offer practical insights for hyperparameter tuning. Despite these strengths, challenges remain in scaling to large datasets and adapting to cold-start and rapidly changing user scenarios. Future research should explore integrating graph neural networks and self-supervised techniques to improve adaptability and resilience, particularly in cross-domain and multimodal contexts. Overall, this work highlights the potential of contrastive learning while pointing to key directions for further advancement.

References

1. J. Smith and A. Brown, "A Comparative Study of Contrastive Learning Approaches for User and Item Representation in Recommender Systems," *Journal of Machine Learning Applications*, vol. 15, no. 2, pp. 145-162, 2023.
2. Y. Zhang, "Social Network User Profiling for Anomaly Detection Based on Graph Neural Networks," arXiv preprint arXiv:2503.19380, 2025.

3. G. Cai, J. Gong, J. Du, H. Liu and A. Kai, "Investigating Hierarchical Term Relationships in Large Language Models," *Journal of Computer Science and Software Applications*, vol. 5, no. 4, 2025.
4. Y. Wang, "Optimizing Distributed Computing Resources with Federated Learning: Task Scheduling and Communication Efficiency," *Journal of Computer Technology and Software*, vol. 4, no. 3, 2025.
5. T. An, W. Huang, D. Xu, Q. He, J. Hu and Y. Lou, "A deep learning framework for boundary-aware semantic segmentation," *arXiv preprint arXiv:2503.22050*, 2025.
6. F. Guo, X. Wu, L. Zhang, H. Liu and A. Kai, "A Self-Supervised Vision Transformer Approach for Dermatological Image Analysis," *Journal of Computer Science and Software Applications*, vol. 5, no. 4, 2025.
7. X. Wang, "Medical Entity-Driven Analysis of Insurance Claims Using a Multimodal Transformer Model," *Journal of Computer Technology and Software*, vol. 4, no. 3, 2025.
8. X. Wang, G. Liu, B. Zhu, J. He, H. Zheng and H. Zhang, "Pre-trained Language Models and Few-shot Learning for Medical Entity Extraction," *arXiv preprint arXiv:2504.04385*, 2025.
9. A. Kai, L. Zhu and J. Gong, "Efficient Compression of Large Language Models with Distillation and Fine-Tuning," *Journal of Computer Science and Software Applications*, vol. 3, no. 4, pp. 30-38, 2023.
10. Z. Yu, S. Wang, N. Jiang, W. Huang, X. Han and J. Du, "Improving Harmful Text Detection with Joint Retrieval and External Knowledge," *arXiv preprint arXiv:2504.02310*, 2025.
11. Y. Deng, "A Reinforcement Learning Approach to Traffic Scheduling in Complex Data Center Topologies," *Journal of Computer Technology and Software*, vol. 4, no. 3, 2025.
12. S. Duan, "Human-Computer Interaction in Smart Devices: Leveraging Sentiment Analysis and Knowledge Graphs for Personalized User Experiences", *Proceedings of the 2024 4th International Conference on Electronic Information Engineering and Computer Communication (EIECC)*, pp. 1294-1298, 2024.
13. J. Zhan, "Single-Device Human Activity Recognition Based on Spatiotemporal Feature Learning Networks," *Transactions on Computational and Scientific Methods*, vol. 5, no. 3, 2025.
14. W. C. Kang and J. McAuley, "Self-attentive sequential recommendation", *Proceedings of the 2018 IEEE International Conference on Data Mining (ICDM)*, pp. 197-206, Nov. 2018.
15. K. Zhou, H. Wang, W. X. Zhao, Y. Zhu, S. Wang, F. Zhang, et al., "S3-rec: Self-supervised learning for sequential recommendation with mutual information maximization", *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pp. 1893-1902, Oct. 2020.
16. S. Chakraborty, A. R. Gosthipaty and S. Paul, "G-SimCLR: Self-supervised contrastive learning with guided projection via pseudo labelling", *Proceedings of the 2020 International Conference on Data Mining Workshops (ICDMW)*, pp. 912-916, Nov. 2020.
17. Z. Wei, N. Wu, F. Li, K. Wang and W. Zhang, "MoCo4SRec: A momentum contrastive learning framework for sequential recommendation," *Expert Systems with Applications*, vol. 223, 119911, 2023.
18. X. Li, Y. Peng, X. Sun, Y. Duan, Z. Fang and T. Tang, "Unsupervised Detection of Fraudulent Transactions in E-commerce Using Contrastive Learning," *arXiv preprint arXiv:2503.18841*, 2025.
19. Y. Liang, L. Dai, S. Shi, M. Dai, J. Du and H. Wang, "Contrastive and Variational Approaches in Self-Supervised Learning for Complex Data Mining," *arXiv preprint arXiv:2504.04032*, 2025.
20. A. Liang, "Personalized Multimodal Recommendations Framework Using Contrastive Learning," *Transactions on Computational and Scientific Methods*, vol. 4, no. 11, 2024.
21. W. Huang, J. Zhan, Y. Sun, X. Han, T. An and N. Jiang, "Context-Aware Adaptive Sampling for Intelligent Data Acquisition Systems Using DQN," *arXiv preprint arXiv:2504.09344*, 2025.
22. L. Zhu, "Deep Learning for Cross-Domain Recommendation with Spatial-Channel Attention," *Journal of Computer Science and Software Applications*, vol. 5, no. 4, 2025.
23. Y. Wang, Z. Fang, Y. Deng, L. Zhu, Y. Duan and Y. Peng, "Revisiting LoRA: A Smarter Low-Rank Approach for Efficient Model Adaptation," 2025.
24. R. Adams and F. Green, "Contrastive Learning for Cross-Domain Recommender Systems: A Comparative Analysis," *Data Science and Engineering Journal*, vol. 10, no. 2, pp. 88-100, 2023.
25. S. Clark and E. Nelson, "Enhancing Collaborative Filtering with Contrastive Loss: An Experimental Study," *Journal of Information Retrieval and Data Mining*, vol. 17, no. 5, pp. 301-316, 2023.

26. E. Wright and J. Roberts, "The Role of Contrastive Learning in Improving Recommendation Systems with Sparse Data", Proceedings of the IEEE Conference on Machine Learning and Data Mining (MLDM), vol. 14, no. 2, pp. 127-135, 2024.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.