Analysis of Developments and Challenges in Dealing with Recommender System Cold-Start Issue

Yue Zhong*
Faculty of humanities, Zhujiang College, South China Agricultural University, Guangzhou, 510900, China
*

Keywords: Cold Start Problem, Meta Learning, Knowledge Graph, Large Language Model.

Abstract: The cold-start problem, caused by a lack of historical interaction data, remains a significant challenge for recommender systems. This study explores three strategies to address this issue: knowledge graphs, meta-learning, and large language models. The meta-learning method, Dual enhanced Meta-learning with Adaptive Task Scheduler (DMATS), improves embedding accuracy and task adaptation through autonomous learning. The knowledge graph method, Knowledge-Enhanced Graph Learning (KEGL), enhances recommendation quality using collaborative embeddings and knowledge-enhanced attention. The large language model method, Automated Dis-entangled Sequential recommendation (AutoDisenSeq-LLM), optimizes recommendations by leveraging text understanding. These methods were tested on datasets like MovieLens, Amazon, and Yelp. The meta-learning method demonstrated strong generalization, the knowledge graph method tackled data sparsity, and the language model method showed potential but needed further improvement. Challenges include high computational costs, lack of standardized evaluation metrics, and dataset issues. Future research should focus on novel techniques, datasets, and metrics.

1 INTRODUCTION

Nowadays, a large amount of information fills people's lives, which makes people's lives more convenient, but at the same time brings a lot of redundant information, resulting in users in the state of information overload. Recommender Systems (RS) have emerged at this time, RS can help users filter information based on the user's own preferences, recommending appropriate information to the user to solve the problem of information overload. Nonetheless, there are still issues with RS, and one of them is the Cold Start Problem (CSP), which is the phenomena whereby the performance of recommendations drastically declines due to a lack of historical interaction data for new users or new objects (Qian et al., 2024). It is primarily divided into three categories: user cold start (Panda et al., 2022), item cold start (Yuan et al., 2023) and system cold start (Wu et al., 2024). These problems are especially prominent in e-commerce, short video, news recommendation and other scenarios. This research can optimize the user experience in RS, meet the practical application requirements of e-commerce platforms (allowing cold-start items to gain exposure and bring transaction growth directly), provide solutions to the core challenges of the current RS and the solution to the cold-start problem can be migrated to other dynamic scenarios such as social media.

The three most popular traditional approaches to recommendation algorithms prior to the complete adoption of RS in the internet world were hybrid recommendation algorithms, content filtering-based recommendation algorithms, and collaborative filtering-based recommendation algorithms (Panda et al., 2022). These three approaches are inexpensive and use straightforward algorithms. However, as RS have evolved, the algorithms for making recommendations have been modified iteratively. The inability of these three conventional approaches to process information while dealing with CSP has become a challenge for RS. A number of approaches have been developed in recent years to address the issue of data sparsity in CSP, including meta-learning, knowledge mapping, etc., to tackle CSP with elevation metrics. All of these methods have more or less solved the problems of low recommendation quality, low effect, and low diversity brought by the CSP. However, these methods also

* Corresponding author

Zhong, Y.
Analysis of Developments and Challenges in Dealing with Recommender System Cold-Start Issue.
DOI: 10.5220/0013990300004916

In *Proceedings of the 2nd International Conference on Public Relations and Media Communication (PRMC 2025)*, pages 289-295
ISBN: 978-989-758-778-8

have their drawbacks includes: Scenario generalization ability, Computational efficiency, Data dependency, etc. To solve the problem of cross-domain migration and generalization ability, using Large Language Models (LLMs) generating user profiles with text can be the way to improve it (Zhang et al., 2025).

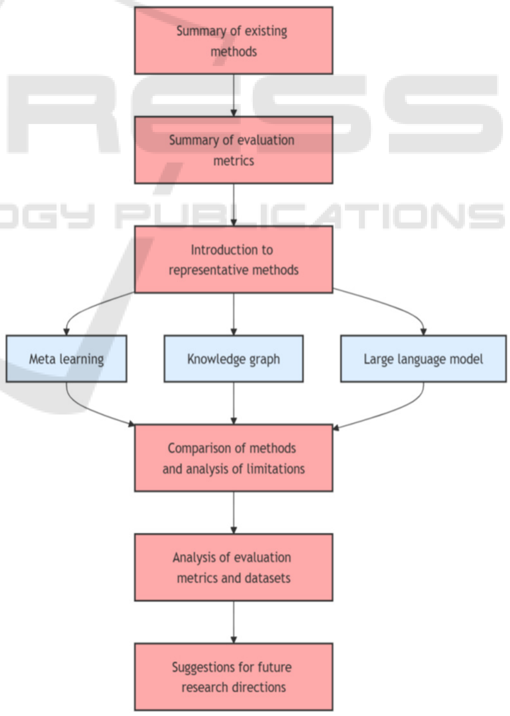
This study's primary goals are to review and summarize the state of the CSP research. The following objectives are the focus of this paper: First, a summary of the common approaches to solving the CSP is given. A thorough technical and methodological examination of CSP solving will follow this introduction. Second, this study looks at the evaluation metrics and datasets that are frequently utilized for the CSP. This analysis helps assess the current progress in addressing cold-start issues and evaluates the effectiveness of different approaches in practical applications. Third, this paper critically examines the limitations of existing methods, datasets, and evaluation metrics. By identifying the weaknesses in current research, areas for improvement are highlighted. Finally, this paper discusses potential solutions to the current problems in the CSP and suggesting future research directions. This paper aims to advance the development and enhancement of RS by suggesting new research directions and possible techniques.

2 METHODOLOGY

This paper summarized the existing methods on solving CSP mainly including: meta-learning, knowledge mapping, machine learning, deep learning, graphical neural networks, LLMs, and so on. In this chapter, meta-learning, knowledge graphs, and LLMs are selected as representative methods to be introduced in terms of principles as well as motivation after an overview of existing methods, and in the next chapter, the performance and limitations of the three methods are compared with, and the limitations of the current evaluation metrics and datasets are also analyzed, so as to provide suggestions and directions for the construction of new datasets and new evaluation methods in the future. The processing is shown in the Figure 1.

2.1 Overview

Previous research has developed new models and techniques to address CSP and increase the accuracy of RS in this context. These emerging methods have their own unique perspectives and performances in dealing with CSP: Meta-learning: meta-learning methods, as one of the key research method directions at present. Its core lies in reducing the amount of data samples needed for learning, which leads to better performance and faster adaptation to new tasks. This is advantageous for enhancing recommendation models' capacity for generalization (WU et al., 2024). Knowledge Graph: Knowledge Graph method includes connection-based methods, embedding-based methods, and propagation-based methods. It centers on combining user information and various elements to build a huge knowledge network graph, thus effectively solving the problem of scarce data and complex computation (Zhang et al., 2024). LLMs: The rise of LLMs as the most emerging technology is a great breakthrough in natural language processing methods. Using LLMs to understand the context, pre-training LLMs to model and represent cold-start users can significantly enhance the recommendation performance of RS in cold-start situations (Zhang et al., 2025). The next part of this paper will introduce these three methods in more detail and describe how they work and what motivates them.



Alt Text for the figure: A flowchart illustrating the process of summarizing existing methods, analyzing meta-learning, knowledge graphs and LLMs, comparing their limitations, and proposing future datasets and evaluation metrics improvements.

Figure 1. Overview of methodology (Picture credit : Original).

2.1.1 Meta-Learning

Meta-learning, as an important research direction in the current field of deep learning, is also known as learn to learn. Traditional machine learning methods for task-specific training require large amounts of data support, so in a new domain with little or no data (i.e., cold start), recommendation performance is often poor. Compared with traditional machine learning, meta-learning provides a new learning method, both through the joint training of multiple tasks to learn prior knowledge, in the face of new tasks, you can use the previously learned prior knowledge to quickly learn, this method significantly enhances the model's adaptability across diverse tasks, particularly boosting its generalization ability in RS (WU et al., 2024).

Broadly categorized, meta-learning techniques can be divided into three main approaches: optimization-based methods, model-based methods, and metric-based methods. Metric-based methods and model-based methods are mainly applied to classification tasks, while optimization-based methods perform better in the face of a wider range of task assignments. Optimization-based methods are also the most widely used methods in RS, as the Model Agnostic Meta-Learning (MAML) framework is the basis of many existing meta-learning methods, which is also an optimization-based method (Finn et al., 2017).

2.1.2 Knowledge Graph

In essence, a knowledge network that records the links between elements is called a knowledge graph. By structuring the words of the relationships between entities, and assigning the meaning of information to each point and each edge in the network, the entire knowledge graph contains a large amount of information, so that it can go on to capture potentially possible connections between individuals, and get the result of efficient storage and utilization of knowledge and thus enhance the effectiveness of recommendations.

Knowledge graphs can also be differentiated in three types: hybrid approaches, path-based methods, and embedding-based methods. In item cold start, the main problem faced by the knowledge graph method is that the embedding generated in data scarcity is usually unreliable, leading to the reduction of recommendation performance, so in item cold start, the path-based method is often used as an auxiliary method to raise the quality of recommendations (Zhang et al., 2025).The knowledge graphs construction is often based on extracting and integrating knowledge in different formats from multiple databases, so that knowledge graphs can be used in RS to analyze the relationship between the user and the item through the rich semantic information they contain, and more accurately predict the user's interests and needs, thus solving the data sparsity problem and also giving recommendation results interpretability, which allows the system to explain the reason for the recommendation (Zhang et al., 2024).

2.1.3 Large Language Models

LLMs are deep learning-trained generative artificial intelligences, with excellent text comprehension and text generation capabilities, and can complete complex conversations. In recent years, LLMs have garnered significant attention in the field of RS due to their effectiveness in natural language processing. Numerous related models have been developed, achieving notable results. Leveraging their powerful text processing capabilities, LLMs are frequently employed in RS to address cold-start challenges, including small-sample and zero-sample scenarios.

There are two categories of research on LLMs in RS: LLMs as RS and LLMs as knowledge enhancers. LLMs as RS: In this type of research, LLMs are used directly to generate recommendation results. The textual descriptions of users and items allow the LLMs to understand user preferences and item characteristics, thus generating high-quality recommendations despite the lack of data.

LLMs as knowledge enhancers: In this type of research, LLMs are used to enhance the knowledge representation of RS. The RS acquires richer knowledge information by pre-training LLMs, which helps to optimize and enhance the recommender model and enhance the RS's performance while dealing with sparse data. (Zhang et al., 2025)

2.2 Datasets and Evaluation Metrics Description

The evaluation metrics and datasets are important components of RS research and development, and they have a significant influence on how recommender models are constructed and trained. The datasets used in the 72 papers screened in the 2019-2023 RS and the evaluation metrics are examined in SLRRS. Datasets frequently utilized in recommender systems (RS) research span a wide range of domains and include well-known examples such as the MovieLens series (100K, 1M, and 10M), Netflix, and Yahoo Music for media recommendations. Other commonly used datasets include FilmTrust, Epinions, and BookCrossing, which focus on trust-based or book-related recommendations, as well as MovieTweetings and Yelp, which cater to social media and business reviews, respectively. Additionally, platforms like Ciao and Amazon provide datasets that are widely adopted for e-commerce recommendation studies. These datasets collectively offer diverse scenarios for evaluating RS models. The evaluation of recommender systems (RS) relies on a variety of metrics designed to assess different aspects of performance. Commonly used metrics include Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) for measuring prediction accuracy, as well as Normalized Discounted Cumulative Gain (NDCG), Precision, Recall, and F1 value for ranking and relevance evaluation. The majority of evaluation metrics referenced in the literature for this thesis are based on these established metrics (Saifudin et al., 2024).

3 RESULTS AND DISCUSSION

This paper analyzes the principles and the effects of meta-learning, knowledge graph and LLMs in solving CSP. The principles of the three methods in solving the CSP have been analyzed in the previous section, and the next section will summarize the effectiveness of the methods based on the experimental results of the related literature:

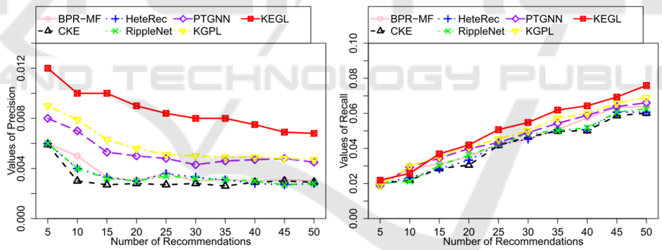
3.1 Results Analysis

The representative method selected for meta-learning in this paper is Dual enhanced Meta-learning with Adaptive Task Scheduler (DMATS) (He et al., 2025) created by D He et al. The datasets selected for this method are DBook, MovieLens and Yelp The evaluation metrics are MAE, RMSE and NDCG.Table1 shows the performance of DMATS on the datasets DBook, MovieLens and Yelp in the user's cold-start problem, as shown in Table 1.

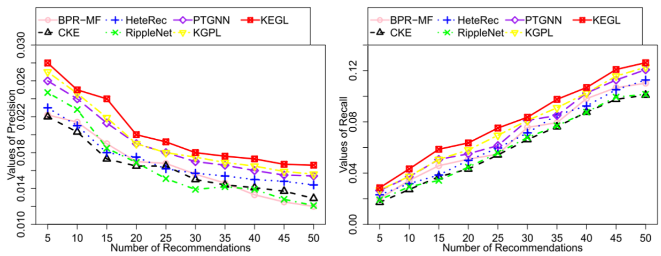
Table 1. Performance of DMATS on three datasets.

Evaluation index	DBook	MovieLens	Yelp
MAE	0.5653	0.8283	0.8515
RMSE	0.6461	0.9490	1.0004
nDCG@5	0.8807	0.7465	0.8618

Though DMATS did not perform as well as the other two (user-item cold start, item cold start) in the user cold start, its performance was still excellent, reaching either optimality or sub-optimality in the comparison with the other baseline methods.



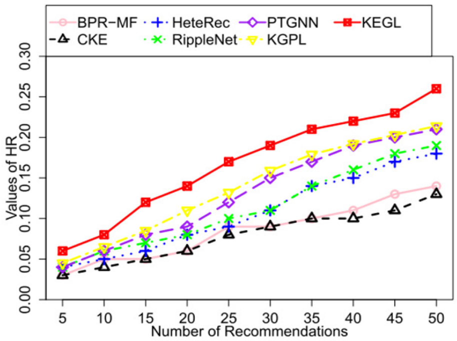
(a) Precision in Amazon-book (b) Recall in Amazon-book



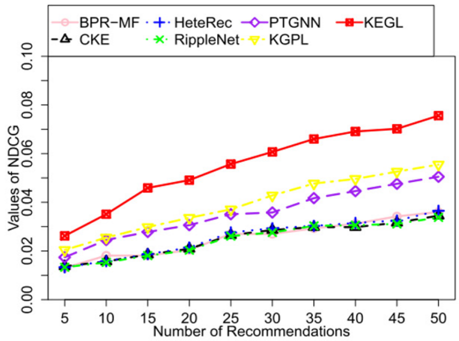
(e) Precision in Yelp2018 (f) Recall in Yelp2018

Alt Text for the figure: KEGL performs better than baseline methods in Precision and Recall on AB and Yelp.

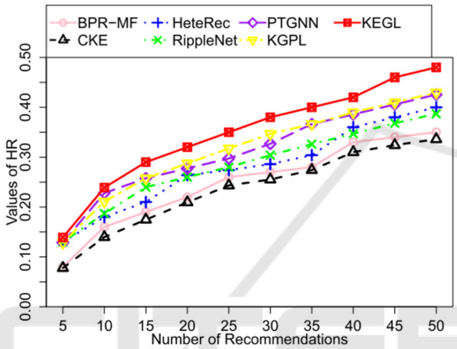
Figure 2. Performance of KEGL on two datasets.



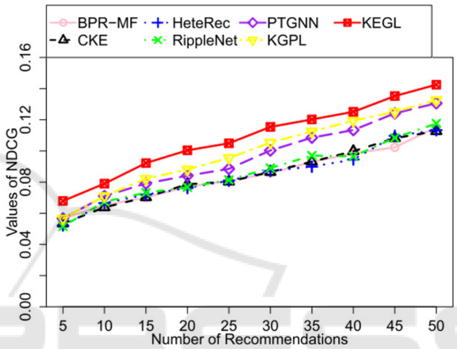
(c) HR in Amazon-book



(d) NDCG in Amazon-book



(g) HR in Yelp2018



(h) NDCG in Yelp2018

Alt Text for the figure: KEGL performs better than baseline methods in HR and NDCG on AB and Yelp.

Figure 3. Performance of KEGL on two datasets.

The representative method selected for knowledge graphs in this paper is Knowledge-Enhanced Graph Learning (KEGL) (Zhang et al., 2025). The datasets selected for this method are Amazon-book and Yelp2018. the selected evaluation criteria are Precision, Recall, HR, NDCG. Because the original article does not show the specific data, this paper only quotes its comparison picture with the selected baseline method (Figures 2 and 3). Figure 2 and Figure 3 clearly show the superiority of KEGL compared with its selected benchmark method in HR, NDCG, Precision and Recall in Amazon-book and Yelp2018 in the case of new item complete cold-start.

The representative method selected for LLMs in this paper is Automated Disentangled Sequential recommendation with large language models (AutoDisenSeq-LLM) (Wang et al., 2025). The datasets selected for this method are MovieLens-1m (ML-1m), Amazon Beauty (AB), Amazon Game (AG), and Steam. The selected evaluation criteria are HR and NDCG. Table 2 shows the recommendation performance of this method on the four datasets.

Table 2. performance of AutoDisenSeq-LLM on four datasets.

Evaluation index	ML-1m	AB	AG	Steam
HR@10	0.7125	0.3995	0.6634	0.6605
NDCG@10	0.4770	0.2601	0.4277	0.4176

3.2 Discussion

As one of the mainstream cold-start problem solving methods in recent years, meta-learning methods perform well in dealing with cold-start problems. Taking DMATS as an example, the autonomous embedding learning mechanism and adaptive task scheduler of DMATS helps the model to learn better node representations and improve the embedding accuracy, Consequently, the data sparsity issue is resolved and the long-tail issue is effectively handled, so the recommendation performance is improved. However, the meta-learning approach represented by DMATS has its drawbacks. First, its meta-learning framework is complex, which leads to the complexity of model training and optimization, that is, it requires a lot of time and computational resources. Second, although DMATS considers the effectiveness of tasks in task scheduling, it still assumes that there is some correlation between different tasks, which may not be fully valid in practical applications, especially in scenarios where user preferences are very diverse, which means the lack of scenario generalization ability.

Knowledge graph methods are also widely used in cold-start recommendation, taking KEGL as an example. KEGL proposes a collaborative-enhanced guaranteed embedding generator to guarantee the quality of embedding and constructs a knowledge-enhanced gated attention aggregator to adaptively control the weights of guaranteed embedding and neighbor embedding, which effectively improves the recommendation quality and solves the items CSP. However, even so, this method still fails to avoid the common problems faced by knowledge graph methods. First, its effectiveness is highly dependent on the quality and completeness of the knowledge graph, and when the quality of the knowledge graph itself is not high and complete enough, the recommendation effect will be reduced. Second, the introduction of two additional operators also increases the computational complexity, which is already not low, and raises the time cost of model training. Finally, although KEGL considers cold-start neighbors, the lack of interaction information in a completely cold-start scenario also reduces the recommendation efficiency.

The LLMs is discussed at the end. On the one hand, compared with other methods, the LLMs is nascent, take AutoDisenSeq-LLM as an example, the performance of facing the CSP is not good enough compared with other methods. But on the other hand, because the youth of LLMs brings more possibilities for this method, LLMs is likely to become the trend of CSP research in the future. The advantages of LLMs are obvious, with its excellent text comprehension in AutoDisenSeq-LLM, the recommendation list is sorted and optimized, which improves the accuracy of the recommendation well. Equally obvious are the drawbacks, which is that the model is too complex and requires a lot of computational resources and time, which is prevalent in current methods.

Current evaluation metrics for cold-start problems, such as MAE, RMSE, Precision, Recall, and NDCG, provide a multidimensional assessment of method performance. However, the lack of standardized metrics makes it difficult to make a comprehensive comparison of methods. A more standardized and comprehensive assessment framework is needed to address this problem. In terms of datasets, while public datasets such as MovieLens, Amazon, and Yelp support cold-start research, they tend to suffer from low timeliness, slow updates, and single-domain limitations. These issues hinder the assessment of method generalizability and cross-domain capabilities. In addition, some datasets contain sensitive user information, raising privacy concerns in an era when data security is increasingly important. Future datasets should prioritize timeliness, multi-domain coverage, and privacy protection to better support the development and evaluation of cold-start recommendation methods.

4 CONCLUSION

This paper focuses on the current state of the CSP in RS and analyzes the principles, performance, and limitations of three advanced approaches: meta-learning, knowledge graphs, and LLMs. Specifically, DMATS improves embedding accuracy and automation, KEGL enhances recommendation quality through high-quality embeddings, and AutoDisenSeq-LLM leverages the text comprehension ability of LLMs to optimize recommendation lists, thereby improving accuracy. Experiments with datasets such as Amazon, MovieLens, and Yelp demonstrate good performance; however, challenges such as weak scene generalization, low computational efficiency, and issues related to datasets timeliness and privacy still persist. Future research will focus on integrating LLMs with other advanced methods to leverage their strengths, improving algorithm efficiency while reducing computational costs. Additionally, efforts will be made to develop privacy-preserving datasets to enhance cross-domain generalization and real-world applicability. These studies aim to improve the performance further and generalizability of RS in dynamic contexts.

REFERENCES

Finn, C., Abbeel, P., & Levine, S. 2017. Model-agnostic meta-learning for fast adaptation of deep networks. In International conference on machine learning, 1126-1135.

He, D., Cui, J., Wang, X., Song, G., Huang, Y., & Wu, L. 2025. Dual Enhanced Meta-learning with Adaptive

Task Scheduler for Cold-Start Recommendation. IEEE Transactions on Knowledge and Data Engineering.
Panda, D. K., & Ray, S. 2022. Approaches and algorithms to mitigate cold start problems in recommender systems: a systematic literature review. Journal of Intelligent Information Systems, 59(2), 341-366.
Qian, M., Wei, X., Yi, Q. 2024. Survey on Solving Cold Start Problem in Recommendation Systems. Journal of Frontiers of Computer Science & Technology, 18(5).
Saifudin, I., & Widiyaningtyas, T. 2024. Systematic literature review on recommender system: Approach, problem, evaluation techniques, datasets. IEEE Access, 12, 19827-19847.
Wang, X., Chen, H., Pan, Z., Zhou, Y., Guan, C., Sun, L., & Zhu, W. 2025. Automated disentangled sequential recommendation with large language models. ACM Transactions on Information Systems, 43(2), 1-29.
WU, G. D., LIU, X. X., BI, H. J., FAN, W. C., & TU, L. J. 2024. Review of personalized recommendation research based on meta-learning. Computer Engineering & Science, 46(02), 338.
Wu, X., Zhou, H., Shi, Y., Yao, W., Huang, X., & Liu, N. 2024. Could small language models serve as recommenders? towards data-centric cold-start recommendation. In Proceedings of the ACM Web Conference, 3566-3575.
Yuan, H., & Hernandez, A. A. 2023. User cold start problem in recommendation systems: A systematic review. IEEE access, 11, 136958-136977.
Zhang, J. C., Zain, A. M., Zhou, K. Q., Chen, X., & Zhang, R. M. 2024. A review of recommender systems based on knowledge graph embedding. Expert Systems with Applications, 123876.
Zhang, W., Bei, Y., Yang, L., Zou, H. P., Zhou, P., Liu, A., ... & Yu, P. S. 2025. Cold-Start Recommendation towards the Era of Large Language Models (LLMs): A Comprehensive Survey and Roadmap. arXiv preprint arXiv:2501.01945.
Zhang, Z., Zhu, Y., Dong, M., Ota, K., Zhang, Y., & Ren, Y. 2025. Embedding Guarantor: Knowledge-Enhanced Graph Learning for New Item Cold-Start Recommendation. IEEE Transactions on Emerging Topics in Computational Intelligence.