Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/43/20230811

# Review of collaborative filtering recommendation systems

**Xueyan Wu**

Department of Information Science and Engineering, East China University of Science and Technology, Shanghai, 201424, China

2857199945@qq.com

**Abstract.** In the era of information overload, recommender systems develop rapidly. And because the needs of information consumers are full of diversity and the information data provided by information producers is too large, to enhance the efficiency and quality of recommendations, the research community has introduced numerous approaches to optimize recommendation systems. As collaborative filtering stands as a time-tested technique in recommendation systems, This paper facilitates a swift comprehension of recent advances in collaborative filtering. It does so by examining the techniques presented across the entire collaborative filtering recommendation systems research field in recent years, especially its development in the domain of deep learning, and have a solid understanding of the field of study.

**Keywords:** Collaborative Filtering, Recommendation Systems, deep learning.

## 1. Introduction

In the present era of the rapidly evolving Internet and the exponential expansion of information data, how to achieve high-quality and effective information in an environment of information over- load has become a major problem, and recommendation systems stand as crucial tools in addressing this challenge. The core essence of a recommendation system is its role as an information filtering method, leveraging user historical behavior and other pertinent data, processes them according to different technologies and algorithm models, and makes recommendations to achieve accurate predictions, to better suit the interests and preferences of users [1]. At the same time, the recommendation system permeates all aspects of our lives, from government affairs to learning and tourism [2], continuously promotes business growth, enriches information acquisition, promotes social interaction, promotes innovation and personalized content creation, significantly contributes to social and economic advancement while enhancing people's quality of life.

Collaborative filtering represents the quintessential approach in recommendation system technology. It forecasts a user's preference for a particular item by analyzing the user's historical behavior data alongside other user behavior patterns. The initial collaborative filtering algorithm relies on user behavioral data, such as the user's Score or click behavior [3], subsequently, researchers began to propose different algorithm improvements and recommendation models, neighborhood-based techniques encompass methods such as user-based collaborative filtering, item-based collaborative filtering, and matrix factorization-based approaches like matrix decomposition algorithms, etc [4], owing to the progress in deep learning, several collaborative filtering algorithms built upon neural networks have emerged, such as neural network-based

DOI: 10.54254/2755-2721/43/20230811

collaborative filtering and collaborative filtering with attention mechanisms, etc. [5]. These algorithms leverage the potent expressive capabilities of neural network models to more effectively capture the relationships between users and items.

As an important class of algorithms, collaborative filtering algorithms can predict and make decisions based on group behavior and better achieve diversified and personalized recommendations. If there is larger group size and more user behaviors, it can achieve a better recommendation effect. Therefore, after long-term improvement and optimization, the collaborative filtering algorithm has been well implemented and applied in many large Internet companies at home and abroad, such as Google, Netflix, YouTube, etc. [6-8]. At the same time, because the collaborative filtering algorithm only relies on user behavior, such as purchase records, click records, rating records, etc., it can make recommendations without considering the relevant information of specific users or the relevant information of recommended items, which greatly saves data collection and the time and cost of processing makes it an excellent performer in all walks of life. However, because it mainly makes decisions based on user behavior data, it lacks sufficient relevant data when facing new users and new items, affecting recommendations' accuracy. Therefore, solving its cold start problem is also a hot direction of continuous research [9,10]. Moreover, there are many contemporary Internet products and a huge user base. Most users only operate, browse, and comment on a small part of the subject matter. The resulting behavior matrix is too sparse, which will also lead to inaccurate calculation results. Affecting the accuracy of recommendation results, there are many studies on this problem [11-12]. In summary, the ongoing refinement of collaborative filtering algorithms within recommendation systems significantly influences the continued evolution of recommendation systems and the enhancement of recommendation accuracy.

The goal of this paper is to review the research field of collaborative filtering recommendation systems in recent years, focusing on widely introducing the key ideas and technologies used to solve the recommendation problem based on collaborative filtering, avoiding showing most of its implementation details so that readers can quickly understand the recent years. Developments and prospects in the field.

## 2. Overview of Collaborative Filtering

Collaborative filtering (CF) is a recommendation system algorithm for predicting items or content that users may like. CF makes recommendations by analyzing the similarity between users based on the user's historical behavior and preferences. The process of CF consists of two main steps: user similarity calculation and recommended item prediction. In the calculation of user similarity, CF will calculates user similarity by analyzing a user's historical behavioral data, such as purchase records or ratings. Commonly employed similarity calculation methods encompass Angular similarity and the correlation coefficient. Through the calculation of user similarity, it becomes possible to identify other users who share similar interests with the target user. In the recommended item prediction, CF will predict the items or content that the target user may like based on the behavior and preferences of similar users of the user. According to different prediction methods, CF can be divided into memory-based and model-based categories [13,14]. In the initial category of approaches, there is a maintenance of a database containing previous user preferences, which is computed anew each time a fresh recommendation is required. In the latter approach, a database is also maintained but first used to create a descriptive model, which is then used to make recommendations for active users [15]. The different subclasses in each major class in CF are shown in Figure 1. Due to the vigorous development of research in the field of deep learning in recent years, which has attracted widespread attention and enthusiasm, this article also focuses on the latest developments based on CF in this field, and the rest of the sub-categories only introduce the core without too many details.

DOI: 10.54254/2755-2721/43/20230811

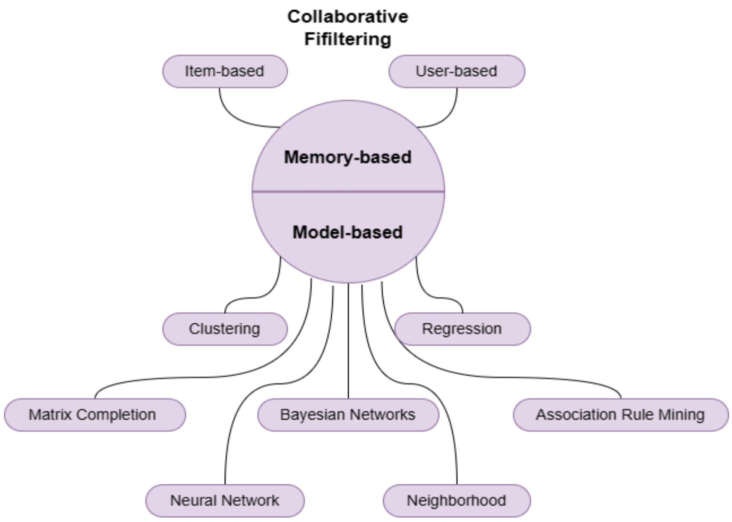


**Figure 1.** Collaborative filtering technology classification

## 3. Methods

### *3.1. Memory-Based Techniques*

Memory-based collaborative filtering technology is a frequently employed approach in recommendation systems, which utilizes existing user-item rating data in the system to predict and recommend users' interest in new items. This approach finds other users with similar interests or similar items based on the user's historical rating data and uses them as the basis for predicting new items. Specifically, the memory-based collaborative filtering method first constructs a user-item rating matrix, which records users' ratings on items. Then, by computing the similarity between users or the similarity between items, the missing ratings are predicted. Ultimately, the item with the highest predicted score is suggested to the user. Memory-based collaborative filtering techniques can be additionally classified into two categories: user-based and item-based methods.

#### *3.1.1. User-Based Method*

This method aims to predict users' interests based on the similarity of behavior or preferences between users. The algorithm calculates the similarity between users by analyzing user behavior data (such as ratings, clicks, purchases, etc.) in the rating matrix, and uses these similarities to make item recommendations. In the user-based collaborative filtering algorithm, similarity functions are utilized to first identify neighboring users of the target user. The similarity function measures the degree of similarity between users, and different measurement methods can be used, such as based on the Pearson correlation coefficient, cosine similarity, etc. To forecast the rating that the target user would assign to an item, it typically involves calculating a weighted average of the ratings provided by nearby users for that particular item. These weights are usually calculated according to the similarity function, indicating the degree of influence of neighboring users on target users [16].

#### *3.1.2. Item-Based Method*

It mainly predicts users' preferences for other items based on their ratings or behavior history on items. The core idea of the algorithm is to make recommendations by mining the similarity between items. In an item-based collaborative filtering algorithm, the initial step involves computing the similarity between items, often using common similarity measures such as the Pearson correlation coefficient and cosine similarity. These similarity measures can be obtained by computing the similarity of ratings or behavior histories between items. Once item similarities are calculated, it becomes possible to predict the user's preferences for candidate items based on the user's ratings of neighboring items. This process is similar to the user-based collaborative filtering algorithm but on the item dimension [17].

DOI: 10.54254/2755-2721/43/20230811

### *3.2. Model-Based Techniques*

#### *3.2.1. Clustering*

A clustering algorithm is an unsupervised learning algorithm used to divide objects in a dataset into groups with similar characteristics. It divides similar objects into the same cluster by calculating the similarity between data and makes the difference between different clusters as large as possible. Clustering algorithms can help us discover hidden patterns and structures in datasets. In collaborative filtering, clustering algorithms can be used to construct groups of users or items. By grouping users or items into the same cluster according to their similarity, we can find groups of users or items with similar preferences or characteristics. Information from these clusters can then be exploited to predict user preferences for items or recommend similar items to users [18].

This article [19] proposes that the clustering collaborative filtering algorithm has some progress in solving the problems in the collaborative filtering algorithm. Specifically, the clustering collaborative filtering algorithm mainly solves the problems of sparse scoring matrix, weak scalability, and deviation of user interest. The algorithm improves the scoring prediction accuracy by improving the Mini Batch K-Means clustering algorithm and introducing time weighting. The article [20] introduces a novel clustering collaborative filtering algorithm known as the Soft K Index Alternative Projection (SKAP) algorithm. This method efficiently addresses high-dimensional soft clustering challenges by generating a sparse partition matrix, ultimately delivering a Top-N recommendation list.

#### *3.2.2. Regression*

Regression analysis is a statistical approach utilized to assess the connection between variables, one being independent and the other dependent. It estimates the value of the dependent variable by fitting a curve or a line to the data points. In collaborative filtering, regression analysis can be used to predict the association between user ratings and communities. By analyzing users' ratings and common patterns in the community, users' ratings of other items can be inferred. In collaborative filtering, regression analysis can be used to model the relationship between users and items. By applying regression analysis to user rating data and item features, one can make predictions for user ratings on new items [21]. This approach can improve the degree of personalization and accuracy of recommendation systems [22]. This study proposes a collaborative filtering method based on high-dimensional regression, which can achieve recommendation performance close to the dense solution while improving computational efficiency.

#### *3.2.3. Matrix Completion*

Matrix Completion is a technique to restore a complete matrix by filling in the missing values in the matrix. In collaborative filtering, since the user's rating data for items is usually incomplete, the use of matrix completion technology can predict missing rating values, thus enhancing the precision of the recommendation system. By using the matrix completion technique, the missing ratings can be inferred from the existing rating data to construct a complete rating matrix. In this way, the recommendation system can make recommendations based on the complete scoring matrix and provide more accurate recommendation results [23].

#### *3.2.4. Bayesian Networks*

Bayesian Networks, a probabilistic graphical model, are employed to depict relationships among variables. They consist of nodes and directed edges, with nodes representing random variables and directed edges indicating dependencies among these variables. Bayesian networks can be used to infer probability distributions among variables for prediction and inference. In collaborative filtering, Bayesian networks can be used to model the relationship between users and items. By analyzing the user's historical behavior and the attributes of the item, a Bayesian network can be constructed to represent the user's preference and evaluation of the item and improve the accuracy and personalization of the recommendation system [24].

#### *3.2.5. Association Rule Mining*

This method is a data mining technique for discovering frequent item sets and association rules in data sets. It analyzes the relationship between items in the data set, finds out the combination of items that often appear at the same time, and generates association rules to describe the relationship between these items. This mining algorithm can predict items based on the association between past and present purchases of items [25].

#### *3.2.6. Neighborhood*

In collaborative filtering, a neighborhood refers to a group of users or items that share similar interests or characteristics with the target user or item. Neighborhood collaborative filtering methods have gained widespread popularity in recommendation systems owing to their notable advantages, including interpretability, robustness, and competitive performance [26].

#### *3.2.7. Neural Network*

A neural network is a computational model that emulates the structure and functionality of the neuron network in the human brain. It helps to extract features from input data and make pre- dictions or classifications. In collaborative filtering, neural networks can be used to enhance the performance of recommendation systems, especially in dealing with cold-start problems, sparse data, and scalability. Specifically, multi-layer perceptrons and convolutional neural networks are commonly used neural network models in collaborative filtering. They improve the accuracy of personalized recommendations by learning the interaction patterns between users and items. Currently, the application of neural networks in recommender systems is still a hot area of deep learning research. Future research directions include improving accuracy, and scalability, and using more datasets [24,25].

In this work, the author uses a graph neural network (GNN) to learn the embedded representation of users and items. Specifically, they use GNNs to propagate information, aggregate information from neighbor nodes to target nodes, and utilize intent-aware and global decoupling methods to improve recommendation performance. This method has advantages in solving problems such as data sparsity and over-smoothing. In this paper [27], the author uses Convolutional Neural Network (Convolutional Neural Network) and Recurrent Neural Network (Recurrent Neural Network) to improve the collaborative filtering algorithm. Convolutional neural networks can be used to process image and text data, while recurrent neural networks can process sequential data, such as time series or user behavior sequences, bringing higher performance and flexibility to recommender systems. This work [28] proposed the DCCF algorithm, which improves the accuracy of recommendations by introducing a combination of global intent decoupling and local collaborative filtering signals. It uses a graph neural network (GNN) to learn representations of users and items and propagates and integrates these representations through message passing and information aggregation. At the same time, DCCF also introduces the concept of intent decoupling, which decomposes the preference information of users and items into multiple potential intents, so as to better capture complex user-item interaction behaviors.

## 4. Conclusion

At present, deep learning has exhibited remarkable success in numerous domains, encompassing natural language processing, image processing, and speech recognition. Within the realm of recommendation systems, deep learning technology can adeptly capture the intricate relationships between users and items, thereby improving the accuracy and personalization level of collaborative filtering algorithms. The deep learning model can learn richer feature representations from large-scale data, helps compensate for the limitations of traditional collaborative filtering methods when dealing with sparse data, thereby enhancing the accuracy of recommendations. Meanwhile, the cold-start problem has always been a difficult problem in recommender systems, especially for new users or new items. In the future, with the continuous increase of data, the ability of deep learning can be combined to obtain more features from other information sources (such as user social network, user historical behavior, etc.), so as to effectively alleviate the cold start problem. In the future, collaborative filtering can also be combined with more efficient predictive models (such as time series forecasting, statistical models, etc.) to better handle data dynamics and trends, while collaborative filtering methods can capture similarities between users. Combining these two methods, a more comprehensive and accurate recommendation system can be constructed. By fusing the outputs of different models, higher quality recommendation results can be achieved. And in the future, it can also be combined with a natural language generation model like ChatGPT, which can add more humanized interaction and interpretability to the recommendation system. This approach not only provides users with personalized recommendations, but also explains why those recommendations are made, thereby increasing user trust and satisfaction. This is crucial to the user experience of recommender systems.

**References**

[1] Bobadilla, J., Ortega, F., Hernando, A. & Guti´errez, A. Recommender systems survey. Knowledge-based systems 46, 109–132 (2013).
[2] Lu, J., Wu, D., Mao, M., Wang, W. & Zhang, G. Recommender system application developments: a survey. Decision support systems 74, 12–32 (2015).
[3] Goldberg, D., Nichols, D., Oki, B. M. & Terry, D. Using collaborative filtering to weave an information tapestry. Communications of the ACM 35, 61–70 (1992).
[4] Hu, Y., Koren, Y. & Volinsky, C. Collaborative filtering for implicit feedback datasets. In 2008 Eighth IEEE international conference on data mining, 263–272 (IEEE, 2008).
[5] Sarwar, B., Karypis, G., Konstan, J. & Riedl, J. Item-based collaborative filtering recommendation algorithms. In Proceedings of the 10th international conference on World Wide Web, 285–295 (2001).
[6] Brin, S. & Page, L. The anatomy of a large-scale hypertextual web search engine. Computer networks and ISDN systems 30, 107–117 (1998).
[7] Bennett, J., Lanning, S. et al. The Netflix prize. In Proceedings of KDD cup and workshop, vol. 2007, 35 (New York, 2007).
[8] Covington, P., Adams, J. & Sargin, E. Deep neural networks for YouTube recommendations. In Proceedings of the 10th ACM conference on recommender systems, 191–198 (2016).
[9] Bobadilla, J., Ortega, F., Hernando, A. & Bernal, J. A collaborative filtering approach to mitigate the new user cold start problem. Knowledge-based systems 26, 225–238 (2012).
[10] Natarajan, S., Vairavasundaram, S., Natarajan, S. & Gandomi, A. H. Resolving data sparsity and cold start problem in collaborative filtering recommender system using linked open data. Expert Systems with Applications 149, 113248 (2020).
[11] Papagelis, M., Plexousakis, D. & Kutsuras, T. Alleviating the sparsity problem of collaborative filtering using trust inferences. In International conference on trust management, 224–239 (Springer, 2005).
[12] Huang, Z., Chen, H. & Zeng, D. Applying associative retrieval techniques to alleviate the sparsity problem in collaborative filtering. ACM Transactions on Information Systems (TOIS) 22, 116–142 (2004).
[13] Bai, T., Wen, J.-R., Zhang, J. & Zhao, W. X. A neural collaborative filtering model with interaction-based neighborhood. In Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, 1979–1982 (2017).
[14] Isinkaye, F. O., Folajimi, Y. O. & Ojokoh, B. A. Recommendation systems: Principles, methods and evaluation. Egyptian Informatics Journal 16, 261–273 (2015).
[15] Vucetic, S. & Obradovic, Z. Collaborative filtering using a regression-based approach. Knowledge and Information Systems 7, 1–22 (2005).
[16] Al-Bashiri, H., Abdulgabber, M. A., Romli, A. & Kahtan, H. An improved memory-based collaborative filtering method based on the topsis technique. PloS one 13, e0204434 (2018).
[17] Pareek, J., Jhaveri, M., Kapasi, A. & Trivedi, M. Snetrs: Social networking in recommendation system. In Meghanathan, N., Nagamalai, D. & Chaki, N. (eds.) Advances in Computing and Information Technology, 195–206 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2013).

DOI: 10.54254/2755-2721/43/20230811

[18] Khatwani, S. & Chandak, M. Building personalized and non personalized recommendation systems. In 2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT), 623–628 (IEEE, 2016).
[19] Han, X., Wang, Z. & Xu, H. J. Time-weighted collaborative filtering algorithm based on improved mini batch k-means clustering. Advances in science and technology 105, 309–317 (2021).
[20] Li, M., Wen, L. & Chen, F. A novel collaborative filtering recommendation approach based on soft co-clustering. Physica A: Statistical Mechanics and its Applications 561, 125140 (2021).
[21] Ge, X., Liu, J., Qi, Q. & Chen, Z. A new prediction approach based on linear regression for collaborative filtering. In 2011 Eighth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), vol. 4, 2586–2590 (IEEE, 2011).
[22] Steck, H. Collaborative filtering via high-dimensional regression. arXiv preprint arXiv:1904.13033 (2019).
[23] Bojnordi, E. & Moradi, P. A novel collaborative filtering model based on combination of correlation method with matrix completion technique. In The 16th CSI International Symposium on Artificial Intelligence and Signal Processing (AISP 2012), 191–194 (IEEE, 2012).
[24] Papadakis, H., Papagrigoriou, A., Panagiotakis, C., Kosmas, E. & Fragopoulou, P. Collaborative filtering recommender systems taxonomy. Knowledge and Information Systems 64, 35–74 (2022).
[25] Amin, S. A., Philips, J. & Tabrizi, N. Current trends in collaborative filtering recommendation systems. In Services–SERVICES 2019: 15th World Congress, Held as Part of the Services Conference Federation, SCF 2019, San Diego, CA, USA, June 25–30, 2019, Proceedings 15, 46–60 (Springer, 2019).
[26] Khojamli, H. & Razmara, J. Survey of similarity functions on neighborhood-based collaborative filtering. Expert Systems with Applications 185, 115482 (2021).
[27] RahmatAbadi, A. F. & Mohammadzadeh, J. Leveraging deep learning techniques on collaborative filtering recommender systems. arXiv preprint arXiv:2304.09282 (2023).
[28] Ren, X., Xia, L., Zhao, J., Yin, D. & Huang, C. Disentangled contrastive collaborative filtering. arXiv preprint arXiv:2305.02759 (2023).