Proceedings of the 7th International Conference on Economic Management and Green Development
DOI: 10.54254/2754-1169/30/20231485

Collaborative Filtering Uncovered: Exploring Modern Techniques for Personalized Recommendations

Shunuo Shi[1,a,*]

[1]*School of Data Science, Chinese University of Hong Kong, Shenzhen, Shenzhen 518172, China*
a. 120090216@link.cuhk.edu.cn
**corresponding author*

Abstract: The prosperity of the Internet industry also facilitates the development of all kinds of online entertainment. Recommendation systems show their capability to make good recommendations and save people's time. As the most widely used recommendation technique, Collaborative Filtering (CF) shows its superiority in making accurate predictions based on existing data. In this passage, the overall introduction of CF is given, including CF itself, challenges, methodologies, concrete examples, and the limitation of current CF. As CF has developed for a long time, there are specific algorithms to deal with different challenges. However, the separated algorithm may be incapable of dealing with multiple challenges. Moreover, hybrid models have unique effects on data since it combines various models and their advantages. In the future, researchers should pay attention to the combination of different algorithms. This article aims to give newcomers a clear perspective of CF and its development. Hopefully, it can also indicate a possible direction for further studies.

Keywords: collaborative filtering, recommendation system, hybrid, deep-learning

1. Introduction

1.1. Overall Introduction

In people's modern life, surfing the internet for online shopping, watching videos, listening to music, and browsing social media have become a daily routine. In this case, recommendation systems have become integral to people's everyday lives since people always want to get what they need quickly. These sophisticated algorithms, designed to recommend personalized content, have revolutionized how individuals interact with information, products, and services across various domains. From shopping and entertainment to news consumption and social media, recommendation systems shape our choices, preferences, and decision-making processes. It has been more than 30 years since one of the first recommendation systems appeared, which was used by Tapestry to handle mail problems [1]. Also, it was the exact time that the concept of 'collaborative filtering (CF)' was proposed. In 1994, a user-based CF was introduced by researchers at the University of Minnesota. It focuses on recommending Usenet news articles to users based on the ratings of similar users [2]. In 1998, the launch of Amazon's recommendation system utilized item-to-item CF. This approach analyzes user behavior data, such as browsing history and purchase patterns, to recommend products similar to those a user has shown interest in or purchased [3]. In 2006, the competition held by Netflix also facilitated the development of CF, including the practical use of the hybrid model. In 2009, YouTube

DOI: 10.54254/2754-1169/30/20231485

started to integrate deep-learning and CF, and the improvement of it was significant. More details of Netflix and YouTube cases are shown in sections 3 and 4. The mentioned example and the development of the recommendation system and CF indicate that CF contributes to people's daily life, and the field is worth further investigation and research.

1.2. Challenges of CF

The challenges of CF can be concluded in 4 main categories. First is the cold start problem, which means finding similarities when a new user or item is introduced to the system since the historical data does not exist. Second is the data sparsity caused by the limited interactions between users and items because users can only interact with a small fraction of items. Scalability is also a concern. As the number of users and items increases, the computational cost also grows, especially for algorithms like memory-based model that relies on the calculation of similarities. The last problem relates to diversity and serendipity. Collaborative filtering techniques often focus on recommending items similar to those a user has previously interacted with or preferred, which can result in a lack of diverse recommendations, potentially limiting users' exposure to novel and serendipitous content. There are also several non-computation problems, such as shilling attacks and privacy concerns.

1.3. Motivation

To sum up, further study of CF is significant to the future development of recommendation systems and the internet industry. Hence, this article aims to provide a thorough examination of CF techniques, highlighting their key concepts, methodologies, and underlying assumptions, which enables a deeper understanding of this field, especially for those who are new to this subject. To achieve this goal, the framework of this article is as follows. Section 2 will introduce mainstream algorithms in CF and their methodologies, including memory-based, model-based, hybrid, and deep-learning. Sections 3 & 4 provide examples of the Netflix Prize competition and YouTube's deep-learning-based CF. Section 5 will offer insights into potential future directions for collaborative filtering research, exploring new methodologies, and addressing unresolved challenges.

2. Introduction of CF

2.1. General Review of CF

Nowadays, CF has developed various forms since the concept was raised. Conclusively, CF can be divided into two classes, which are known as User-based and Item-based. The former one measures the similarity between different users by their rating on particular products and gathers users with the same interest for further recommendation. The other investigates the similar patterns between different items and recommends suitable products based on existing users' data. To achieve the purpose of recommendation system, there are four main techniques or algorithms that are widely used today, including memory-based, model-based, hybrid CF algorithm and CF regarding deep learning.

2.1.1. Memory-based

The most basic technique is the memory-based algorithm, also known as the neighborhood-based algorithm, which utilizes the whole user-item data to generate the prediction. The advantage is obvious since it is reasonable, which was attached of great importance in recommendation system. Moreover, it is easy to compute and use. However, the computation of it is usually costly, and scientists have been working on it to optimize it in past decades. The team of Yu succeeded in reducing the computational costs by using a probabilistic framework [4]. The performance of

DOI: 10.54254/2754-1169/30/20231485

memory-based was also improving. For example, similarity updating and prediction modulation were included for improvement [5].

2.1.2. Model-based

The second technique is the model-based algorithm. It is a general technique with various branches since the 'model' here means any suitable models can be used to develop a CF, including models from data mining and machine learning. Compared with the memory-based algorithm, one significant difference is model-based algorithm will usually reduce the dimension of objective problems. Methods like singular value decomposition or principal component analysis are used to compress the user-item matrix. Even though it abandons the abundance of known data, the accuracy, computational time and robustness of the recommendation system are increased [6].

2.1.3. Hybrid

As for the hybrid algorithm, it is a mixture of the memory-based and model-based algorithms. It can help CF overcome some problems like sparsity and deficiency in information, which may often cause bad results in naïve algorithms like memory-based ones. The way of combination is various. In the paper of Tan, Ye, and Gong, they used memory-based CF to generate a user-item matrix and used model-based CF to find out the nearest neighbors for every item, which allowed clients to get the prediction of target items on target users [7]. In the latter paragraph of this paper, a more concrete example of the integration of memory-based and model-based will be introduced. Su and Khoshgoftaar have already investigated the three main techniques of CF [8]. In the paper, the methodologies of three different algorithms were clearly stated, and concrete examples are given.

2.1.4. Deep-learning

In recent years, scientists started to integrate deep-learning and CF. It has significantly advanced the field of Collaborative Filtering (CF) by capturing complex patterns in user-item interactions. By utilizing deep neural networks, CF algorithms can better model user preferences and item characteristics, resulting in more accurate and personalized recommendations. Notable examples include neural collaborative filtering, which combines matrix factorization with multi-layer perceptrons, and autoencoders, which employ unsupervised learning for latent feature extraction. These approaches have demonstrated substantial improvements in recommendation quality across various domains. An early experiment on two real-world datasets showed that neural CF could significantly improve the accuracy of prediction [9].

2.2. Evaluating Metrics

The evaluating metrics of CF are various. The recommendation system is widely used in rating needed fields such as e-commerce and video platforms. Since people need to compare the prediction and actual values, the most commonly used are Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Receiver Operating Characteristic (ROC) sensitivity, and so on.

3. CF in Improving the Netflix Recommendation System

3.1. Introduction of Netflix Competition

In 2006, Netflix launched a competition on performance improvement in movies recommendation. The race attracted thousands of researchers, students, and mathematicians to participate. The competition aimed to find out a team that could make the improvement in Root Mean Square Error

DOI: 10.54254/2754-1169/30/20231485

(RMSE) score by 10% [10]. The competition ended in July 2009, when the team of BellKor's reached the winning threshold with a 10.05% improved RMSE over the recommendation system of Netflix. The data given by Netflix was largescale, consisting of 480,000 users and 17,770 movies.

3.2. Adopted Methods and Their Extension

BellKor's team got their inspiration from a published article by Koren, who indicated a merged style of CF regarding neighborhood-based and latent factor approach. It was also the first time that two algorithms were combined in a single model [11]. The first concern was the high sparsity (around 99%) of the data. Hence, the model was regularized, and the regularization was controlled by two constants obtained from cross-validation. In the baseline estimation process, in order to avoid overfitting problems, the author adopts a penalizing method by using a regularizing term. The first algorithm adopted was neighborhood models. In the previous models, they mostly paid attention to the relationship between two items but neglected the relationship between full sets of neighbors. It was slightly changed to secure the models can focus more on the correlation between items. Then, latent factor models regarding singular value decomposition (SVD) are used for further computation. The SVD applied here was retrieved from Paterek with some innovative extensions [12]. After extending the above SVD by Paterek, Koren obtained a new SVD model called the Asymmetric-SVD model. It outperformed other models by having fewer parameters, handling new users better, and possessing a better explainability, and even showed a greater accuracy compared with the original SVD (usually, explainability is more important than accuracy in CF). With further improvement on Asymmetric-SVD, the new model SVD++ did not show a better explanatory property but gained a higher score in RMSE evaluation (shown in Table 1). The data retrieved from Netflix was both implicit and explicit, while the data in implicit form only provided rentals history, and the explicit form of data included rates from users.

3.3. Integrated Model

The integrated model needed to deal with the integration of implicit and explicit data, which were represented in R(u) and N(u). During the integrating part, Koren provided a 3-tier model. In the first tier, the model is concerned about the movies and users independently, such as the general reputation of a movie and the rating habit of a user. In the second tier, it considers the relationship between the movies and users, such as the label of a movie and whether a user enjoys such kinds of movies. In the last tier, the algorithm considers the in a "neighborhood" way, such as the users' rates on a similar movie. Eventually, the newly built neighborhood model, SVD-based model, and integrated model were used to compare with former models and showed better results. The final integrated model reached an RMSE of 0.8877 (shown in Table 2), which was higher than any existing model in the competition so far. After the adjustment by BellKor's team, their merged model reached an RMSE of 0.8556 and won the final prize of the competition.

Table 1: RMSE for different SVD [8].

Model	50 factors	100 factors	200 factors
SVD	0.9046	0.9025	0.9009
Asymmetric-SVD	0.9037	0.9013	0.9000
SVD++	0.8952	0.8924	0.8911

DOI: 10.54254/2754-1169/30/20231485

Table 2: RMSE for different iterations [8].

	50 factors	100 factors	200 factors
RMSE	0.8877	0.8870	0.8868
time/iteration	17min	20min	25min

4. YouTube Recommendation System

4.1. Introduction of YouTube

Serving as the largest video platform, YouTube has more than two billion active users in the world. It is a crucial task for YouTube to improve the experience of users. If the recommendation contents are attractive to users, they will probably spend more time discovering what attracts them. YouTube started in an early stage to combine CF and deep-learning to provide a better prediction on recommendations. In the paper by Covington, Adams, and Sargin, they give a detailed inspection of the mechanism of deep neural networks and how they can improve the recommendation system [13]. The CTR (click-through rate) improved significantly by using deep-learning CF, and the overall watch time increased.

4.2. Challenging Perspectives

The first challenge is that the scale of datasets from YouTube is enormous since there are more than two billion users. Most existing algorithms fit problems of small groups well but find it hard to cope with largescale problems. The second challenge is freshness. There are tons of new videos every day, and users' preference is dynamic, making it hard for the platform to make decisions between new videos and well-established videos. Eventually, historical behavior can hardly be used to predict due to its sparsity. Hence, the algorithm needs to be robust enough.

4.3. Mechanism

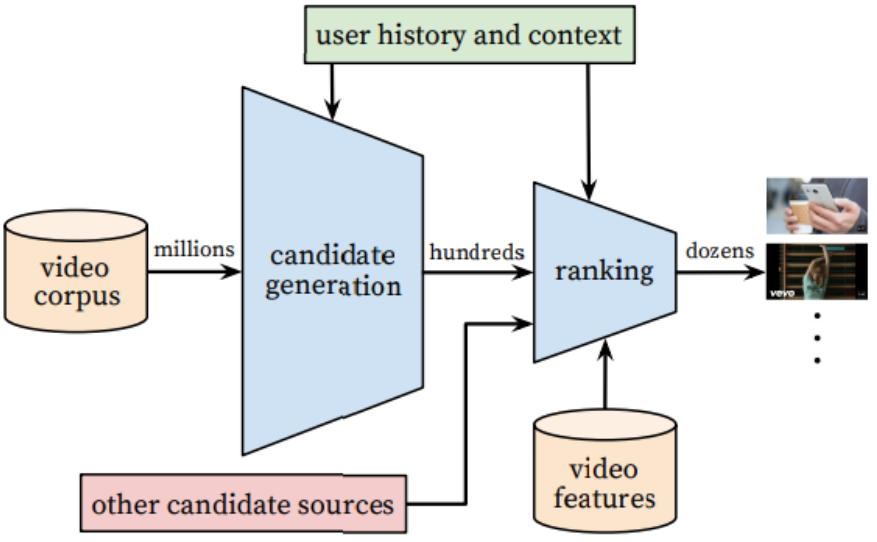

Figure 1: Mechanism of YouTube recommendation system [13].

DOI: 10.54254/2754-1169/30/20231485

To recommend appropriate videos to users, the system has to select videos from the video corpus, which contains millions of videos. The overview of the selection is shown in Figure 1. The system is made up of two neural networks. The first network generates a candidate set based on the users' history, reducing millions of videos to hundreds of videos. The second one ranks the priority of candidates and recommends dozens of videos as the final recommendation. The main task for neural networks is to analyze and learn from the historical context of users. Not only does the explicit data matter (thumbs up/down) but metrics like whether users finish the video are also viewed as an implicit user history.

4.4. Improvement by Neural Network

First, in the traditional CF system, the matrix factorization technique assumes that the relationship between users and items is linear. However, in many cases, it will lead to inaccurate results since the relationship is more complex. Neural Collaborative Filtering (NCF) employs a multi-layer network to predict a non-linear model between user preferences and item characteristics, which enables it to provide a more accurate result. Second, the NCF network is designed as a generalization of traditional matrix factorization techniques. By incorporating different types of layers and activation functions in the neural network, the NCF can model various interaction functions. In the paper, the authors propose two aspects of the NCF network, which are Generalized Matrix Factorization (GMF) and Multi-Layer Perceptron (MLP). GMF is more similar to traditional matrix factorization, which provides a linear computation. MLP can help to capture more non-linear interactions. Eventually, the authors integrate the two methods and propose a hybrid model, offering the benefits of both linear and non-linear modeling. Third, the flexibility of NCF is higher. This allows researchers and practitioners to experiment with different layer configurations, activation functions, and optimization techniques to build models tailored to specific datasets and domains. For example, the number of hidden layers and the variety of activation functions can be adjusted freely. Moreover, the NCF captures latent factors better, which typically relies on the linear combination of latent factors [9]. All these reasons lead to the long-term success of YouTube. To be specific, the Chief product officer of YouTube mentioned that 70% of watch time on YouTube was related to this algorithm [14].

5. Limitation and Future Direction

5.1. Limitation and Endeavor

In section 1, it has been claimed that there are several problems that need to be solved. In the current state, different methods have been developed to deal with certain problems. For the cold start problem, combining CF and content-based problems in a hybrid way can help, and active learning also has a significant impact since no historical data can be used in this case [15]. For scalability problems, methods like matrix factorization and SVD can be combined in model-based problems to lower the computational cost, which is still a mainstream method to lower the dimensionality of the user-item matrix [16]. Deep-learning CF also shows its capability of dealing with both cold start and scalability problems [17]. Besides dealing with scalability problems, some model-based problems can also help with sparsity problems. For diversity and serendipity problems, if new methods can be combined with ranking algorithms, some novel recommendations can be recommended to users.

5.2. Future Direction

From the above limitation and endeavor so far, it can be concluded that some methods have remarkable effects on different aspects, which means that researchers have been working on specific aspects to cope with present problems. From the example of the Netflix Prize competition and the

DOI: 10.54254/2754-1169/30/20231485

current mainstream recommendation system, it can be concluded that hybrid CF can combine the advantages of different algorithms. For example, the winner of the Netflix Prize competition used hybrid methods to cope with sparsity and scalability problems. Therefore, a suggestion for future direction will be how different methods can be combined in a harmonious way. Moreover, as deep-learning CF shows its superiority, it should be a field that needs to be further studied. As reinforcement learning becomes a trendy topic in machine learning, it also has the potential to become fundamental in CF as well. A passage in 2018 claimed that traditional recommendation systems focus on explicit properties like CTR, but few pay attention to how frequently users return after their clicks [18]. Eventually, as machine learning is still a newly-developed subject, besides combining existing methods in a hybrid model, new methods can constantly be adopted to CF and recommendation systems to improve current endeavors.

6. Conclusion

This paper focuses on introducing the development of CF so far, including its methodology and two examples. In the start, the challenges of CF are stated. The methodology covers memory-based, model-based, hybrid, and deep-learning CF. Each algorithm has its advantages and disadvantages. To have a clearer understanding, the Netflix Prize competition and YouTube cases are given. The Netflix Prize competition shows the advantages of the hybrid model, which combines memory-based and model-based algorithms. The example of YouTube gives an example of how neural networks can improve CF. The challenges of CF still exist. In the current state, specific algorithms can cope with certain problems. For future directions, researchers should focus on combining different algorithms in a hybrid model to deal maximize their superiority of them. Moreover, after new algorithms of machine learning come out, more methods can be used to improve current problems. This paper aims to help beginners of CF to have a general understanding of CF. Furthermore, it concentrates on recent CF, and hopefully, it can point out some future endeavor that can be made by researchers.

References

[1] *Goldberg, D., Nichols, D., Oki, B. M., & Terry, D. (1992). Using collaborative filtering to weave an information tapestry. Communications of the ACM, 35(12), 61-70.*

[2] *Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., & Riedl, J. (1994, October). Grouplens: An open architecture for collaborative filtering of netnews. In Proceedings of the 1994 ACM conference on Computer supported cooperative work (pp. 175-186).*

[3] *Linden, G., Smith, B., & York, J. (2003). Amazon. com recommendations: Item-to-item collaborative filtering. IEEE Internet computing, 7(1), 76-80.*

[4] *Yu, K., Schwaighofer, A., Tresp, V., Xu, X., & Kriegel, H. P. (2004). Probabilistic memory-based collaborative filtering. IEEE Transactions on Knowledge and Data Engineering, 16(1), 56-69.*

[5] *Jeong, B., Lee, J., & Cho, H. (2010). Improving memory-based collaborative filtering via similarity updating and prediction modulation. Information Sciences, 180(5), 602-612.*

[6] *Aditya, P. H., Budi, I., & Munajat, Q. (2016, October). A comparative analysis of memory-based and model-based collaborative filtering on the implementation of recommender system for E-commerce in Indonesia: A case study PT X. In 2016 International Conference on Advanced Computer Science and Information Systems (ICACSIS) (pp. 303-308). IEEE.*

[7] *Gong, S., Ye, H., & Tan, H. (2009, May). Combining memory-based and model-based collaborative filtering in recommender system. In 2009 Pacific-Asia Conference on Circuits, Communications and Systems (pp. 690-693). IEEE.*

[8] *Su, X., & Khoshgoftaar, T. M. (2009). A survey of collaborative filtering techniques. Advances in artificial intelligence, 2009.*

[9] *He, X., Liao, L., Zhang, H., Nie, L., Hu, X., & Chua, T. S. (2017, April). Neural collaborative filtering. In Proceedings of the 26th international conference on world wide web (pp. 173-182).*

[10] *Chaturvedi, M. (2021, July 5). How useful was the Netflix Prize really. Analytics in diamag. https://analyticsindiamag.com/how-useful-was-the-netflix-prize-really/*

DOI: 10.54254/2754-1169/30/20231485

[11] *Koren, Y. (2008, August). Factorization meets the neighborhood: a multifaceted collaborative filtering model. In Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 426-434).*

[12] *Paterek, A. (2007, August). Improving regularized singular value decomposition for collaborative filtering. In Proceedings of KDD cup and workshop (Vol. 2007, pp. 5-8).*

[13] *Covington, P., Adams, J., & Sargin, E. (2016, September). Deep neural networks for youtube recommendations. In Proceedings of the 10th ACM conference on recommender systems (pp. 191-198).*

[14] *Solsman, J. E., (2018, Jan 10). YouTube's AI is the puppet master over most of what you watch. CNET. https://www.cnet.com/tech/services-and-software/youtube-ces-2018-neal-mohan/*

[15] *Volkovs, M., Yu, G. W., & Poutanen, T. (2017). Content-based neighbor models for cold start in recommender systems. In Proceedings of the Recommender Systems Challenge 2017 (pp. 1-6).*

[16] *Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. Computer, 42(8), 30-37.*

[17] *Yuan, J., Shalaby, W., Korayem, M., Lin, D., AlJadda, K., & Luo, J. (2016, December). Solving cold-start problem in large-scale recommendation engines: A deep learning approach. In 2016 IEEE International Conference on Big Data (Big Data) (pp. 1901-1910). IEEE.*

[18] *Zheng, G., Zhang, F., Zheng, Z., Xiang, Y., Yuan, N. J., Xie, X., & Li, Z. (2018, April). DRN: A deep reinforcement learning framework for news recommendation. In Proceedings of the 2018 world wide web conference (pp. 167-176).*