CS410: Text Information System Technology Review Collaborative Filtering Algorithms

Xuehao Wang(xuehaow2)
11/15/2020

Overview

This report has reviewed the latest development of Collaborative Filtering (CF) algorithms. It gives the challenges in the development of Collaborative Filtering algorithms and improvement achieved in this process. Also, the current problems and possible solutions are discussed in the conclusion.

Introduction

Recommendation systems have been used a lot in our daily life especially in this information age. They help people to discover information which might be interested in. Recommendation systems can be classified as two main categories. One is content-based filtering algorithms, and another is Collaborative Filtering. Collaborative Filtering Algorithms are the widely used technologies in recommendation systems. They work based on the personalized histories of user's choices such as the ratings of one product or the searching history on the website. There are two types of Collaborative Filtering approaches which are memory-based algorithms and model-based algorithms. Memory based algorithms include user-based collaborative filtering and item-based collaborative filtering. This report mainly focused on the discussion of user-based collaborative filtering also called neighborhood-based algorithms.

The following figure shows the process of the Collaborative Filtering process. The given rating table is taken as the input in which u represents the users in this system and i represents the list of items. The core part of Collaborative Filtering Algorithm includes prediction and recommendation. Prediction gives the numerical number which is computed as the likeness of item belong to each user with the same scale with the rating input of this user. Recommendation is the list of Top-N items chosen and raked based on the prediction results.
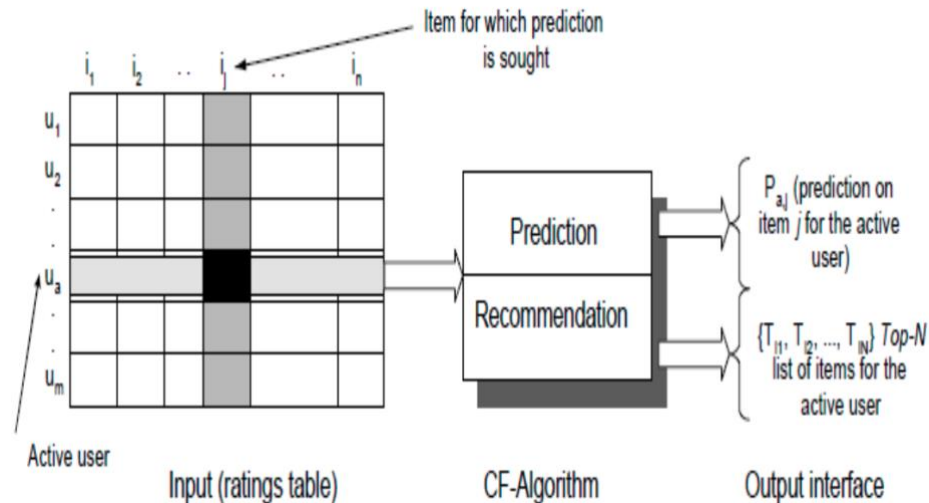
Figure1. Collaborative Filtering process [Badrul Sarwar, et. 2001]

Discussion

The two main issues faced by user-based Collaborative Filtering Algorithms are scalability and sparsity. Recent years, with the development and application of data mining approaches, Collaborative Filtering Algorithms has become much more mature to be used in lots of fields. For example, customers can always see the recommendation gives by Amazon showing that "Customer Who Bought This Item Also Bought", which is a combination of user-based and item-based Collaborative Filtering Algorithms. However, more research is still needed to solve these two main challenges and further problems such as the time complexity issue.

This part discusses the two main challenges sparsity and scalability and the solutions people used to solve these problems and improve the performance of recommendation system.

1. Sparsity
As discussed above, one of the cores of Collaborative Filtering Algorithms is computing the similarity among users and items. However, there is a great possible that most of the ratings in user-item rating table/matrix are missing in practice, especially in lots of commercial recommendation systems, which has large item and user sets. In the other word, the sparsity is mainly caused by high item to user ratio. This results the poor prediction results of Collaborative Filtering algorithms. Kant, S. and Mahara, T. (2018) proposed a hybrid approach that combines the user-based CF algorithm and item-based CF algorithm to overcome the sparsity of input matrix. They also used the biclustering technique to lower down the dimensions. As a result, this approach achieved much better performance compared to traditional CF algorithm and some of state-of-art approaches. Wang, Y. and He, L. (2018) optimized the computation of

similarity using information entropy and data sparseness recommendation. They also compared four different approaches (Pearson correlation, Spearman correlation, Cosion similarity, and WIE algorithms) based on the computation of recommendation precision. By verifying in the experience, they proposed that the elaborated algorithm can elevate the accuracy of prediction caused and alleviate the sparsity problem.

2. Scalability
The growth of items and users is the most common problem facing in the real-world and real-time issues. A traditional recommendation system would suffer from the serious troubles due to the scalability. Nilashi M., etc. (2018) used dimension reduction and ontology techniques to develop a new hybrid CF algorithm. Meanwhile, they take advantage of the Singular Value Decomposition approach improve the scalability of recommendation system significantly. Li, C. and He, K. (2017) proposed an optimized MapReduce approach and experimented on the real-world data. They came to the conclusion that their approach is better in processing large-scale datasets with scalability.

Beside of these two main challenges, there are still many other problems people need to deal with such as the user/item cold start problem, gray sheep problem, ramp-up problem, shared account problem, evaluation problem, etc.

Conclusion

This report has discussed some challenges related to Collaborative Filtering Algorithms. The performance of these algorithms can be improved by incorporating trust, context and time-variant features, etc. Due to the increase of the need for information, the number of users and items in the database are increasing exponentially. The solution of this problem can head to the reducing time complexity of Collaborative Filtering Algorithms. All of these possible solutions can enhance the Collaborative Filtering and increase the quality of the user experience of recommendation systems.

Reference

https://dl.acm.org/doi/10.1145/371920.372071
https://link.springer.com/chapter/10.1007/978-3-540-72079-9_9
https://onlinelibrary.wiley.com/doi/full/10.1002/cpe.4092
https://www.sciencedirect.com/science/article/pii/S0957417417306577
https://ieeexplore.ieee.org/abstract/document/7977363
https://link.springer.com/chapter/10.1007/978-3-319-90092-6_10
https://www.sciencedirect.com/science/article/pii/S0957417412002825
https://www.sciencedirect.com/science/article/pii/S0020025517309702

https://www.researchgate.net/profile/Kumar_Abhishek16/publication/266182161_A_Review_on_Personalized_Information_Recommendation_System_Using_Collaborative_Filtering/links/5620d7b808ae70315b559d06/A-Review-on-Personalized-Information-Recommendation-System-Using-Collaborative-Filtering.pdf