1. Introduction

With widely spreading of internet, great number of web service has been involved in people’s life. In order to promote this web service, recommender system is embedded in these web. This system helps the user to easily access information, which is most relevant to their interest. In this way, the user is not necessary to spend great of time in large space. Nowadays, recommender system has be used in a great community of online content, including movies[1], book[2], news [3], friends[4], and jobs[5].

This system greatly enhances the user’s experience in the web service. This convenience provided by recommender system will increase the frequency of web service’s usage, which is the goal of the majority of companies. Nowadays, more and more companies are aware of the importance of the recommender system, which has been incorporated in their web page. For example, Facebook recommends some other people sharing the same friends with the user. Google recommends some other search words relevant to the original one provided by the user. Amazon recommends some other products sharing the same function of the item in the users’ shopping history.

The key problem in the recommender system is how to select the correct item for the users. The algorithm and its implementation, outputting the correct selection for the user, are the key components in this system. Netflix [6], a widely-unknown provider of internet stream media, has supplied 1 Million dollars for a team, who improves the accuracy of predictions about the enjoyable movie. As is described in Campos’ paper [7], the rating system in the webpage is a great tool to solve this problem. The recommended items for each user are some products with high rate from this user. It means the recommender system can be established based on the inference for the rate of the products.

2. Related Work

As is mentioned in last section, the key component in this recommender system is to develop an efficient algorithm to select several items highly-rated by the users. Generally speaking, there are three kinds of filtering algorithm based on different information.

The first algorithm is content-based algorithm. The recommender system with this algorithm has stored content information of both items and users. The items with great similarity of user’s preference, which is indicated by this user’s rating history, are recommended to him. In the idea of probabilistic, there is high probability of the item fitting user’s preference, to obtain high rate from this user. One important technique to implement this algorithm is to create a probabilistic model for a user based on his rating history. The system can use this model to predict the rate of new item for this user, and recommend an item with possible high rate for him [8].

The second algorithm is collaborative-filtering algorithm. This foundation of this algorithm is information of a group of people who share the same preference with a specific user. The item fitting preference of the community will be recommended to this user. In this algorithm, there are two computational approaches in the collaborative-filtering algorithm [9]: neighborhood methods and latent factor models. The neighborhood method involves user-based computation method or items-based computation method. User-based neighborhood method means to directly predict the score of an item based on the score of user’s community. Item-based neighborhood method, also be called item-to-item method [10], is an indirect prediction method. When the rate of one item for a specific user needs to be predicted, the computation is performed based on the corresponding user group’s historical rate of other items similar to the target item. Latent factor models perform the inference in terms of the factors from the rating history. For example, in order to predict the score of a book, the factor about category (e.g. scientific VS fiction) may needs to be taken into consideration.

Now, we can make comparison between the two methods. The content-based algorithm aims at establishing a joint distribution of all items, and then uses it to predict rate for the new items. The benefit of this algorithm is unique characterization in user’s recommendation, for foundation information of this recommendation is preference information of the user himself. The limitation in content-based algorithm is hardly applicable for the new users, whose information is not great enough for the system to provide this recommendation.

The collaborative-filtering algorithm aims at establishing some conditional models, which is used to predict the score based on the score of the same item or the score of the similar item from other users. For this algorithm, information is great enough for our prediction, but there is still some shortcomings in the output of this algorithm. There is great possibility that some popular items are recommended to the user, but this user is not interested in this item.

A common idea is that greater performance exists in the collaborative-filtering recommender system, for it relies on sufficient data in the rating system. Based on the weakness of this system, the content-based algorithm can be used as an alternative method in the system. In this way, the 3rd algorithm is generated, hybrid approach. This approach takes the advantage of both systems. This is the reason we applies the hybrid recommender system in this project. More details of the hybrid system will be discussed in the next section.

1. IMDb. <http://www.imdb.com>

2. Amazon. <http://www.amazon.com>

3. Google news. <http://news.google.com>

4. Facebook. <https://www.facebook.com>

5. Indeed. <http://www.indeed.com>

6. NetflixPrize. <http://www.netflixprize.com>

7. L. Campos, J. Fernandez-Luna, J. Huete, M. Reuda-Mrales. Combining content-based and collaborative recommendations: A hybrid approach based on Bayesian networks, 2010.

8. R.J.Mooney, L.Roy. Content-based book recommending using learning for text categorization, in: DL’00: Proceedings of the Fifth ACM Conference on Digital Libraries, ACM Press, 2000, pp. 195–204.

9. Mike Gartrell, Ulrich Paquet, Ralf Herbrich. A Bayesian Treatment of Social Links in Recommender Systems, Technical Report, CU-CS-1092-12, May 2012

10. Linden, G; Smith, B; York, J. "Amazon.com recommendations: item-to-item collaborative filtering". IEEE Internet Computing (IEEE) 7 (1): 76–80, January 2003