

# A Hybrid Collaborative Filtering Algorithm for Hotel Recommendation

Ling Shen<sup>1,a</sup>, Qingxi Peng<sup>1,b,\*</sup>

<sup>1</sup>Wuhan Donghu University, Jiangxia, Wuhan, Hubei, China

<sup>a</sup>aleenapple@163.com, <sup>b</sup>pqx@hubu.edu.cn

\* Corresponding Author

**Keywords:** Recommendation, Collaborative Filtering, Item-based, User-based, Hybrid

**Abstract.** Recommendation systems apply knowledge discovery techniques to the problem of making personalized recommendation for information, products or services in the Internet. These works, especially collaborative filtering algorithms acquired relatively satisfactory results. They can millions of users easily search hundreds of millions of items. In tourism industry, potential customers may book hotel just once, which make it very important to recommend them suitable hotel. In this paper, user-based and item-based collaborative filtering methods have been proposed for hotel recommendation. Then we combine them to achieve better effectiveness. Experimental result demonstrates that our hybrid collaborative filtering algorithms outperform existing approaches.

## 1. Introduction

Tourism industry becomes more and more popular in recent years. Previous users choose hotels according to travel guidance. With the development of the Internet, world has experienced profound changes. Life style of people has been also changed. Nowadays, with the help of Internet, users can give rating score to express their opinions. Normally, the scores are from 1 to 5 stars. One star means the user dislike the hotel, while 5 stars mean that the user like the hotel very much. Moreover, an increasing number of websites allow the users express detail score to many aspect of the hotel. In general, the rating score mainly represent the quality of the hotel. However, since the information is very huge, it is difficult for the users to read and compare large amount of webpages to choose favorite hotel. Fortunately, recommendation system can help the users making decision easily.

Recommendation systems have become extremely common in recent years, and are applied in a variety of application. In previous recommendation systems, many models have been employed. Collaborative filtering (CF for short) and content-based filtering system are often adopted. Collaborative filtering is based on collecting and analyzing a large amount of information on users' behaviors, activities or preferences and predicting what users will like based on their similarity to other users. It identifies the potential preference of a consumer/user for a new product/service by using only the information collected from other consumers/users with similar products/services in the database. CF is a simple technique as it is not necessary to apply more complicated content analysis compared to the content-based filtering systems. Item-based collaborative filtering is one of the major recommendation algorithms because of its simplicity. The algorithms assume that the consumers/users are likely to prefer product/service to what they have bought before. In user-based CF, the unknown rating is predicted by averaging the known rating of the test item by similar users.

In this paper, we combine the user-based CF and item-based CF. We proposes a hybrid collaborative filtering algorithm for hotel recommendation. Therefore, our recommendation algorithm can provide accurate prediction to the user, and recommend better hotel for the user.

In this paper, we have made contributions listed below:

- 1) We construct a model which combines the item-based and user-based CF to provide better recommendation.
- 2) We propose a novel similar measure for the hybrid CF recommendation framework.
- 3) Experiments in real life datasets show that our method improves the accuracy of recommendation.

The rest of this paper is listed as follows. Section 2 discuss about related works. Section 3 investigates the item-based and user-based collaborative filtering respectively. In section 4, we propose the hybrid CF and the similarity measures. Comparative experiments demonstrate the effectiveness of proposed method in section 5. Section 6 is the conclusion of the paper.

## 2. Related Works

In recent years, people tend to book hotel on the professional booking websites such as *booking*, *tripadvisor*, *ctrip* and *yelp*. However, it is difficult for the users to compare hotels. Therefore, large amounts of works [1-4] have been devoted in hotel recommendation. In [1], a method of hotel recommendation incorporated different aspects of a hotel for user preference is presented. Collaborative filtering and content-based method has been combined to overcome data sparse problem, and improve the recommendation accuracy in [2]. In [3], aspect-based sentiment analysis has been employed in hotel recommendation. The author also score hotel by Markov Random Walk model to explore transitive association between hotels. In [4], opinion mining has been incorporated in hotel recommendation. A hotel recommendation system has been built with above techniques. In [5], a methodology for the development of the knowledge model is proposed for hotel recommendation.

Collaborative Filtering has been extensively studied in many papers [6-9]. In [6], the core concepts of CF have been discussed. A framework which predicts scores based on the heuristic that people who agree in the past has been proposed in [7]. In [8], an algorithmic framework has been proposed, which perform CF and new algorithmic elements to improve the recommendation accuracy. A model has been presented for explanations based on the user's conceptual model of recommendation process in [9].

## 3. Collaborative Filtering

When the users give ratings to the hotel, a rating consists of the association of two things: user and item (often by means of some value). One way to visualize ratings is as a matrix. Without loss of generality, a rating matrix consists of a table where each row represents a user, each column represents a specific hotel, and the number at the intersection of a row and a column represents the user's rating value. CF-based recommendation systems can predict for a given user and items. In this paper, the item is hotel. Rating in a CF-recommendation system can take on a variety of forms. Automatic collaborative filtering (ACF) systems relieve users of this burden by using a database of historical user opinions to automatically match each individual to others with similar opinions.

CF uses the assumption that people with similar tastes will rate things similarly. Content-based filtering (CBF) uses the assumption that items with similar objective features will be rated similarly. CBF and CF have long been viewed as complimentary. CBF can predict relevance for items without ratings. CF needs ratings for an item in order to predict for it. On the other hand, CBF needs content to analyze. For many domains, it is difficult to obtain the content.

In general, CF algorithms are separated as two classes: memory-based algorithms that require all ratings. Items and users are stored in memory and model-based algorithms that setup rating patterns offline.

## 4. Proposed Method

### 4.1 Hybrid Collaborative Filtering

User-based CF predicts a test user's interest in a test item based on rating information from similar user profiles. In our approach, each user profile is formulated as a row vector, and sorted by its dissimilarity towards the test user's profile. We employ a ranking model to predict the test item rating. The set of similar users can be identified by a threshold or a selecting top- $N$  policy.

Item-based CF recommends the similar items to one user. The unknown rating of a test item by a test user can be predicted by averaging the ratings of other similar items rated by this test user. Each item

is regarded as a column vector, and sorted according to its dis-similarity towards the test item in the user-item matrix.

Our method combines user-based CF and item-based CF as to minimize the expectation and the real user and item matrix. Suppose  $u \in U$  is a user and  $h \in H$  is a hotel. We get the equation 1 as follows.

$$\arg \min_{h \in H, u \in U} (I \times S_u(u_k) + \lambda J \times S_h(h_i)) \quad (1)$$

Where  $S_u(u_k)$  and  $S_h(h_i)$  are the user matrix and hotel matrix. Users and hotels in these matrixes have been pre-processed through a selecting top- $N$  algorithm. The parameters come from experiences.

#### 4.2 A Novel Similarity Measure

The important step in the user-based and item-based CF algorithms is to compute the similarity between items and users then to choose the most similar items. By taking advantage of feature matrix, our hybrid CF looks into the set of items and users and computes how similar they are to the target hotel  $h$ , and user  $u$ . Then we select the most similar hotels  $\{h_1, h_2, \dots, h_l\}$ . Once the most similar hotels are found, the prediction is then computed equation 1. The similarity between two hotels and users can be calculated by many methods. We adopt the common used method, namely Pearson correlation. Pearson correlation is calculated for the similarity between two hotels and users with following equation.

$$\text{sim}(X, Y) = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 \sum (Y - \bar{Y})^2}} \quad (2)$$

In Equation 2, vector  $X$  represent user vector or hotel matrix. We apply gradient descent to acquire an optimal solution.

## 5. Experiments

### 5.1 Experiment Setup

In order to evaluate the proposed method, we collect three datasets from real-world. The dataset has been preprocessed in the experiment. The rating scores are all from the original websites. Some of the records are removed for trustiness. We adopt  $MAP$ ,  $p@k$  to evaluate the effectiveness of the proposed hybrid CF method. The definition of two measures is as follows. Precision at  $k$  ( $P@k$ ) has the advantage of not requiring any estimate of the size of the set of relevant documents.  $MAP$  is mean average precision, which provides a single figure measure of quality across recall levels. Table 1 shows the statistics of the users and rating scores in our dataset.

Table 1 Number of Users and Rating Scores

#Users	#Rating Scores
1852	4634

### 5.2 Experiment Result

In this section, we perform the experiment to verify the proposed methods. We choose two approaches as baselines, namely item-based CF and user-based CF. The hybrid CF predicts the user's interest based on the rating information from similar user profiles.

As shown in Table 2, hybrid CF outperforms user-based CF and Item-based CF both in  $MAP$ ,  $p@5$ ,  $p@10$ . It acquires higher performance in recommendation.

Table 2 Comparison of different method

Model	MAP	p@5	p@10
User-based CF	65.3%	84.6%	54.2%
Item-based CF	67.2%	85.7%	56.4%
Hybrid CF	<b>76.4%</b>	<b>88.9%</b>	<b>58.6%</b>

## 6. Conclusion

In this paper, we propose a Hybrid CF which combines item-based CF and user-based CF for hotel recommendation. We firstly analyze the item-based CF and user-based CF respectively. And then propose our method. Moreover, we present a novel similarity measure for the method. Experimental result demonstrates the effectiveness of our method.

## Acknowledgements

This work is supported by “Science and Technology Research Project of Hubei Province”, No. B2016315.

## References

- [1] F. Fukumoto, et al. "Incorporating guest preferences into collaborative filtering for hotel recommendation." International Conference on Knowledge Discovery and Information Retrieval 2014.
- [2] Zhang, Kai, et al. "Hotel recommendation based on user preference analysis." IEEE International Conference on Data Engineering Workshops IEEE, 2015:134-138.
- [3] Fukumoto, Fumiyo, et al. Exploiting Guest Preferences with Aspect-based Sentiment Analysis for Hotel Recommendation. Knowledge Discovery, Knowledge Engineering and Knowledge Management. Springer International Publishing, 2014.
- [4] Aamir, Mohammad, and Bhusry, Mamta. "Building Recommendation System for Hotel." International Journal of Computer Applications 134(2016).
- [5] Gobin, B. A., and R. K. Subramanian. "Knowledge Modelling for a Hotel Recommendation System." Enformatika (2007):244.
- [6] Schafer, J. Ben, et al. "Collaborative Filtering Recommender Systems." Expert Systems with Applications. An International Journal 9.3(2015):46-45.
- [7] Resnick, Paul, et al. "GroupLens: an open architecture for collaborative filtering of Netnews." Proceedings of the 1994 ACM conference on Computer supported cooperative work ACM, 1994:175--186.
- [8] Herlocker, Jonathan L., et al. "An algorithmic framework for performing collaborative filtering." International ACM SIGIR Conference on Research and Development in Information Retrieval ACM, 1999:230-237.
- [9] Herlocker, Jonathan L., J. A. Konstan, and J. Riedl. "Explaining collaborative filtering recommendations." ACM Conference on Computer Supported Cooperative Work ACM, 2001:5-53.