A Novel Hotel Recommendation Model

Qingxi Peng [a], Ling Shen [b,*]

Wuhan Donghu University, Jiangxia, Wuhan, Hubei, PRC

[a]pqx@hubu.edu.cn, [b]aleenapple@163.com

* Corresponding Author

Keywords: Recommendation, Trip, Collaborative Filtering, Item-based

Abstract. With the development of Internet, a large amount of users purchase product, book hotels, buy goods from the web. They also share their experiences on the web. Other users can read the reviews or rating scores to choose the products or services. The user is difficult to read the information, since the data on the web is very huge. Recommendation system is an important technique which produces high quality recommendations. However, the amount of work increases with the number of participants in traditional recommendation system. In this paper, we propose a novel recommendation model which employs user-based collaborative filtering techniques to recognize the user-item matrix. Our experiment on different datasets shows that the proposed model acquires satisfactory performance than existing model.

Introduction

Previous users like to choose product and service according to advice from friends or relatives. With the development of the Internet, world has experienced profound changes. Life style of people has been changed. Nowadays, people prefer to get information from the web. While users complete the trip, they usually publish their feeling and rate the hotel. The potential customers tend to making decision according to it. 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 browse large amount of webpages to choose interesting hotel. Fortunately, recommendation system can help the users making decision easily.

Recommendation system is a subclass of information filtering system that seek to predict the rating or preference that a user would give to an item. 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), content-based filtering and hybrid recommendation 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. The basic idea of CF-based algorithms is to provide item recommendations or predictions based on the opinions of the like-minded users. Two main methods are adopted by CF-based algorithms: user-based CF and item-based CF. In hotel recommendation, the whole ratings are usually taken into consideration. The whole score, however, could not represent the fine-grain item features. For example, a user may not satisfy with a hotel. He gives "3" to this hotel. But in feature "convenience" he may give "5" to this hotel. According to existing approaches, potential customers will not be recommended for this hotel. However, if someone only care about the "convenience", it is better to recommend this hotel to him.

In this paper, we propose a novel recommendation model which employs item-based collaborative filtering. Detail features have been investigated to provide recommendation in fine-grain level. Therefore, our recommendation 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 build a hotel dataset for research, and publish it for researchers.

2). We propose a novel item-based collaborative filtering model which captures the fine-grained features for better recommendation.

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 give related work about our research. Section 3 investigates the dataset of hotel websites. Two important datasets are collected. We also analyze the datasets. In section 4, we propose the item-based collaborative filtering which capture the fine-grain features, and give better recommendation. Comparative experiments demonstrate the effectiveness of proposed method in section 5. Section 6 is the conclusion of the paper.

Related Works

Hotel review mining has been studied by many researchers [1-3]. Zhang et al. combine collaborative filtering with content-based method to overcome the sparsity issue from low rating frequency problem [1]. Fumiyo Fukumoto et al. incorporate different aspects of a hotel into account for a hotel recommendation model. The model is effective for finding hotels that have never been strayed but share the same neighborhoods [2]. Sharma et al. present a Multi Criteria recommendation system for hotel recommendations to choose the best suited hotel in a city according to a user's preference and other user's review. The model also addresses the Cold Start Problem for this domain [3]. In [4], a fuzzy and nonlinear programming approach for ubiquitous hotel recommendation has been proposed.

Collaborative Filtering is an extensive studied field [5-8]. CF technology brings together the opinions of large interconnected communities on the web, supporting filtering of substantial quantities of data [5]. In 1994, Resnick et al. propose a entire architecture which predict scores based on the heuristic that people who agree in the past will probably agree again [6]. Breese et al. describe several algorithms, including techniques based on correlation coefficients, vector-based similarity calculations and statistical method [7]. In [8], Greg Linden et al. present recommendation algorithms to personalize the online store for each customer. They also compare common approaches with their algorithms.

Hotel Dataset

We crawl data from several famous hotels website. There are three websites are considered in our experiments: Tripadvisor, Ctrip and booking. In each website, we choose five cities to collect the data: Beijing, Wuhan, Hangzhou, Xi'an and Kunming. The data statistics are as follows. The data are from 15 Jan 2015 to 15 Feb 2015. There are different languages used in hotel reviews. We only choose English reviews for the convenience. The dataset can also serve other researchers.

Table 1 Number of reviews in Dataset

	TripAdvisor	Ctrip	Booking
Beijing	1352	2689	896
Wuhan	857	1034	677
Hangzhou	987	966	853
Xi'an	675	640	459
Kunming	574	638	431

Proposed Method

Hotel Feature Model.

Previous studies pay attention to the total rating of hotel. If one hotel gets high rating score, it will get high probability to be recommended to other users. The total rating score, however, don't represent the true situation. For example, one hotel has transport facilities. But its rating score may be

common, since its service is not so good. This hotel may not be recommended to the customers with existing recommendation system. This observation inspired us to provide a fine-grain analysis. We propose a hotel feature model (HMF) which represents the fine-grain requirement of users. We studied the main features of hotel, and manually choose five features: price, food, location, service, clearness. We find these five features can represent the requirement of users. We represent the fine-grain features as a matrix, and then use a Matrix Factorization (MF) to acquire great popularity. There are many techniques in MF. In this work, we use Singular Value Decomposition (SVD) to learn item-based fine-grain features. The key step for HFM model is feature extraction, since the quality of prediction mainly depends on those extracted features. We use a bootstrap algorithm to extraction. Firstly, we define the seed features (price, food, location, service, clearness). It is important to note that many synonyms may occur. If we define the fixed keywords, much important information may lose. In our HFM model, the words near our predefined feature word are collected. With the bootstrap algorithm, the features and the users' preferences have been collected.

Feature-based Collaborative Filtering (FBCF).

CF mainly includes two types: user-based CF and item-based CF. User-based CF focus on the social networks, while item-based CF emphasize on personalization. The important step in the item-based collaborative filtering algorithms is to compute the similarity between items and then to choose the most similar items. From the HFM model mentioned above, we get the feature matrix. By taking advantage of feature matrix, our feature-based CF looks into the set of items the target user has rated and computes how similar they are to the target item i, and then we select the most similar items $\{i_1, i_2, ... i_k\}$. At the same time their corresponding similarities $\{s_{i1}, s_{i2}, ..., s_{ik}\}$ are also computed. This vector can be regarded as a product from the feature matrix and the preference vector. Once the most similar features are found, the prediction is then computed by taking a weight average of the target features. The similarity between two items and features can be calculated by many methods. We adopt the common used method, namely Pearson correlation. Pearson correlation is calculated for the similarity between two features based on the HFM model.

$$sim(X,Y) = \frac{\sum (X - \overline{X})(Y - \overline{Y})}{\sqrt{\sum (X - \overline{X})^2 \sum (Y - \overline{Y})^2}} \tag{1}$$

In Equation 1, vector X represent feature vector such as price. We apply gradient descent to acquire an optimal solution.

Experiments

Experiment Setup.

In order to evaluate the proposed method, we collect three datasets from real-world. The dataset has been preprocessed in the experiment. Reviews with less than 10 words have been erased to ensure the trustiness. We adopt precision and recall to measure the effectiveness of the proposed fine-grained item-based collaborative filtering. The definition of two measures is as follows.

$$Precision = \#(relevant\ items\ retrieved)/\#(retrieved\ items) \tag{2}$$

$$Recall = \#(relevant\ items\ retrieved)/\#(relevant\ items) \tag{3}$$

Experiment Result.

In this section, we perform the experiment to verify the proposed methods. We choose two approaches as baselines. One is course-grained item-based collaborative filtering (IBCF). The other is user-based collaborative filtering (UBCF). The UBCF predicts the user's interest based on the rating information from similar user profiles.

As shown above, Feature-based CF outperforms user-based CF and Item-based CF both in precision and in recall. For two baseline methods, item-based CF is better than User-based CF. Since our proposed method captures the fine-grained feature for the recommendation, it acquires higher performance in recommendation.

Table 2 Comparison of different method

Model	Score(%)	TripAdvisor	Ctrip	Booking
UBCF	Precision	67.4	66.5	63.9
	Recall	68.3	64.2	59.6
IBCF	Precision	68.5	54.9	64.1
	Recall	66.5	62.4	**59.0**
FBCF	Precision	**72.1**	**68.3**	**65.9**
	Recall	**72.3**	**66.4**	58.7

Conclusions

In this paper, we propose a fine-grain method which employs feature-based collaborative filtering for hotel recommendation. The proposed method firstly utilize a HFM model which to calculate the feature matrix. Then a feature-based collaborative filtering is present to capture the fine-grain preference of users. experimental result demonstrate the effectiveness of our method.

Acknowledgements

References

[1] Zhang, Kai, et al. "Hotel recommendation based on user preference analysis." IEEE International Conference on Data Engineering Workshops IEEE, 2015:134-138.

[2] F. Fukumoto, et al. "Incorporating guest preferences into collaborative filtering for hotel recommendation." International Conference on Knowledge Discovery and Information Retrieval 2014.

[3] Sharma, Yashvardhan, J. Bhatt, and R. Magon. "A Multi-criteria Review-Based Hotel Recommendation System." IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing IEEE, 2015.

[4] Chen, Toly, and H. C. Yu. "Fuzzy and nonlinear programming approach for optimizing the performance of ubiquitous hotel recommendation." Journal of Ambient Intelligence & Humanized Computing 4.1(2015):1-10.

[5] Schafer, J. Ben, et al. "Collaborative Filtering Recommender Systems." Expert Systems with Applications An International Journal 9.3(2015):46-45.

[6] 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.

[7] Breese, John S., Heckerman, David, and Kadie, Carl. "Empirical analysis of predictive algorithms for collaborative filtering." Fourteenth Conference on Uncertainty in Artificial Intelligence 2013:43--52.

[8] Linden, Greg, B. Smith, and J. York. "Amazon.com Recommendations: Item-to-Item Collaborative Filtering." IEEE Internet Computing 7.1(2010):76-80.