Scientific
Research

A Contextual Item-Based Collaborative Filtering Technology*

Xueqing Tan, Pan Pan
School of Information Management, Wuhan University, Wuhan, China
Email: 611pp@163.com

ABSTRACT

This paper proposes a contextual item-based collaborative filtering technology, which is based on the traditional item-based collaborative filtering technology. In the process of the recommendation, user's important mobile contextual information are taken into account, and the technology combines with those ratings on the items in the users' historical contextual information who are familiar with user's current context information in order to predict that which items will be preferred by user in his or her current context. At the end, an experiment is used to prove that the technology proposed in this paper can predict user's preference in his or her mobile environment more accurately.

1. Introduction

Traditional personalized recommendation systems are used to suggest products or information or service to the customers in some E-commerce sites. However, the technology is bounded to the desktop. Now, with the continuous improvement of the wireless communication technology and the emergence of so many advanced mobile terminals, such as smart phone and tablet PC, the recommendation's environment is changing and the recommendation technology should make a change. For instance, in an online tourism site the user's previous visiting experience to a scenery spot in a sunshine day may be quite different from the visit in a rainy day. Consequently, user's contextual information should be taken into account, and it has demonstrated that the contextual information did play an important role in the recommendation technology by Adomavicius *et al*. [1].

Collaborative Filtering (CF) is proved effective in the E-commerce systems, but those traditional collaborative filtering technology did not consider the user's contextual information [2-4]. Some applications in ubiquitous computing environment has used collaborative filtering in their recommendation, but they also did not utilize the user's contextual information [5,6]. However, some context-aware recommendation systems using collaborative filtering have taken some user's contextual information into consideration. Annie Chen [7] proposed a context-aware collaborative filtering system, which mainly is base

on the user-based CF, not the item-based CF. Min Gao *et al*. [8] proposed a personalized context-aware collaborative filtering based on neural network and Slope One algorithm. Slope One is a famous item-based CF algorithm, while it calculates the dissimilarity between items rather than the similarity which is calculated in traditional item-based CF.

In this paper, we propose a contextual item-based collaborative filtering, which is based on the traditional item-based CF calculating the similarity. Our approach calculates the similarity between the active user's current context and the other contexts, finds out the ratings already given by the active user about the items under the contexts which are similar with user's current context, and then predicts that which items the active user will prefer in the current context.

The remainder of this paper is organized as follows. In Section II, we will give an overview of the process of the traditional item-based collaborative filtering. Then, we will introduce the concept of the context and put an emphasis on the process of the contextual item-based collaborative filtering in Section 3. Subsequently, we experimentally evaluate our approach in Section 4. And at last we draw a conclusion in Section 5.

2. Traditional Item-Based Collaborative Filtering

Collaborative Filtering is already applied in E-commerce systems and is proved to be very effective. Its main idea is that people who like the same things are likely to feel

similarly towards other things [7]. CF can be divided into two parts: user-based CF and item-based CF. In this paper, we mainly discuss the process of the item-based CF. Item-based CF computes the similarity between items and then selects the most similar items for prediction [8].

2.1. Building a User Profile

First, we should build a user profile. The basic data source of the technology is a user-items matrix, $A(m,n)$. It stores the ratings which are given by m users for n items. m denotes the users information $U = (u_1, u_2, \cdots, u_m)$, and n denotes the items information $I = (i_1, i_2, \cdots, i_n)$. If a user u rates an item i, it will generate a rating, $R_{u,i}$, which is between 0 and 5. The bigger the rating, the more the user likes the item.

2.2. Selecting the Nearest Neighbors

The key of the CF is selecting the nearest neighbors and utilizing the neighbors' preference to predict the active user's preference. In the item-based CF, we should calculate the similarity between the items at first. Three similarity measures were introduced in [4]: Pearson correlation, Spearman rank correlation, and cosine vector similarity. We use the Pearson correlation to calculate the similarity between the items. The equation calculating the similarity between the item i and the item j is:

$$sim(i,j) = \frac{\sum_{u=1}^{m}\left(R_{u,i} - \overline{R_i}\right)\left(R_{u,j} - \overline{R_j}\right)}{\sqrt{\sum_{u=1}^{m}\left(R_{u,i} - \overline{R_i}\right)^2}\sqrt{\sum_{u=1}^{m}\left(R_{u,j} - \overline{R_j}\right)^2}} \qquad (1)$$

In the equation, $\overline{R_i}$ denotes the mean of the ratings which are given by all users for item i.

When we get all the similarities, we sort by size and select the top K. Then they consist of the Nearest Neighbors Collection of the item i, $KNNC_i$.

2.3. Generating the Prediction

Now, we can utilize the neighbors' ratings to predict the active user's preference by computing a weighted average of the ratings which is formulated as

$$P_{a,i} = \overline{R_i} + \frac{\sum_{j \in KNNC_i} sim(i,j)\left(R_{a,j} - \overline{R_j}\right)}{\sum_{j \in KNNC_i} \left|sim(i,j)\right|} \qquad (2)$$

The three steps cover the basic process of the item-based CF. Then, we will introduce how to generate a prediction for a user when he is in a mobile environment.

3. A Contextual Item-Based Collaborative Filtering Technology

The technology we will introduce is an item-based CF

incorporating user's contextual information. So we should introduce what context is at first.

3.1. What Is Context

Context is a multifaceted concept. Many researchers have given different definitions. In 2001, Dey [9] proposed his definition of context: "Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves". Zimmermann *et al*. [10] utilized Dey's definition and then divided the context into five categories: individuality, activity, location, time, and relations. The categories are very suitable to build context model for the recommendation. So we propose our presentation for context:

$$C = (C_1, C_2, \cdots, C_n) \qquad (3)$$

C_i presents one type of contexts, such as Time. And C_i consists of many different variables. For instance, in the type of Time, there are several values (such as morning, noon, afternoon, and evening) or lots of specific values of time (such as 5:00 pm, 11:00 am, and so on). And a user may have different preference for the same item in different variables of one type of context. So we can give the definition of C_i:

$$C_i = \{C_{i1}, C_{i2}, \cdots, C_{ik}\} \qquad (4)$$

Then, we will introduce the specific process of the contextual item-based CF. Because the technology we proposed is based on traditional item-based CF, the process we will introduce in the following is similar with the content mentioned in the Section 2.

3.2. Building a User Profile

This step is also building a user profile, but we must incorporate some user's contextual information. In the traditional item-based CF, what we build is a two-dimension model which includes users' and items' information. Adomavicius *et al*. [11] proposed a multi-dimension model recommendation space. It is very suitable for us to build a User × Item × Context model. **Figure 1** shows our specific three-dimension rating model.

In the **Figure 1**, there is $R_{u_1, i_1, c_{11}}$, which means the user u_1 gives a rating whose value is 3 for the item i_1 in the variable c_{11} of context C_1. So there are many models which store the ratings given by all users for every item in each context variable.

3.3. Calculating the Similarity between Contexts

Calculating the similarity between contexts is aimed at finding out which ratings given in the other contexts are
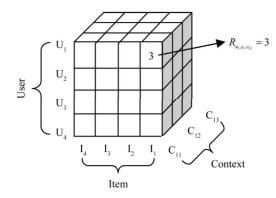
Figure 1. User × Item × Context three-dimension rating model.

more relevant for the current context. However, there are many variables in each context type. Therefore we should find out which variables of the other context are the same with the variables of current context at first. Then, we could calculate the similarity between the two variables using the equation as follows:

$$sim_t(x,y,i) = \frac{\sum_{u=1}^{m}\left(R_{u,i,x} - \overline{R_i}\right)\left(R_{u,i,y} - \overline{R_i}\right)}{\sqrt{\sum_{u=1}^{m}\left(R_{u,i,x} - \overline{R_i}\right)^2}\sqrt{\sum_{u=1}^{m}\left(R_{u,i,y} - \overline{R_i}\right)^2}} \quad (5)$$

This equation is similar with Equation (1), but in Equation (5), t denotes one type of context, $R_{u,i,x}$ means that user u gave a rating for item i on the variable x in the context t, and $\overline{R_i}$ represents the mean of the ratings for the item i.

We can utilize the Equation (5) to calculate the similarity between the active user's current context and the other contexts in which some variables are the same with ones of current context. We take the $sim_t(C,s,i)$ as the similarity where C means the current context, and s denotes the variables of the current context who are the same with the ones in the other contexts.

3.4. Selecting the Nearest Neighbors

Before we select the nearest neighbors of the active item, we should calculate the similarity between the items. However, the method calculating the similarity is based on the ratings for these items. So we must know the ratings which are rated by the active user for the items in the current context at first. By using the Equation (5), we define the ratings as follows:

$$R_{u,i,C} = \frac{\sum_{s \in C} R_{u,i,x}\left(\sum_{t=1}^{n} sim_t(C,s,i)\right)}{\sum_{t=1}^{n}\left|sim_t(C,s,i)\right|} \quad (6)$$

Then, we can utilize the ratings and combine with the Equation (1) to calculate the similarity between the item i

and the item j. After that, we do the same work introduced in the part B of the Section 2, and consequently we find out the Nearest Neighbors Collection of the item i.

3.5. Generating the Prediction

After we get the nearest neighbors, we can utilize the Equation (2) to predict the active user's ratings for the item i in the context C. We define the ratings as $P_{a,i,C}$.

4. Experiment and Results

We evaluate our approach through an experiment in which we compare our approach with the traditional item-based CF. The specific process is as follows:

4.1. Datasets

The datasets offered for the experiment are from Movie-Lens which is a famous recommender system. We use the 100 k rating data set which consists of 100,000 ratings from 943 users on 1682 movies. In the datasets, each user has rated at least 20 movies and the rating is from 0 to 5.

We use the u1.base of the data set as our base data, and use the u1.test as the test data to prove if our approach is more effective in recommendation. Because we must incorporate some contextual information into the process, we can consider users' age, gender and occupation as the contextual information since these three types belongs to the category of individuality which is already introduced in the part A of Section 3.

4.2. Metrics

We use the Mean Absolute Error (MAE) as the metrics to measure the prediction quality of our approach.

The MAE is defined as [12]:

$$MAE = \frac{\sum_{i=1}^{N}\left|p_i - q_i\right|}{N} \quad (7)$$

where p_i denotes the value of the prediction of i by our approach, q_i denotes the value of the real ratings of i, and N denotes the number of the tested ratings.

4.3. Comparative Results in Terms of MAE

In the process of the experiment, we calculate the respective MAE of our proposed approach (Proposed) and the traditional item-base CF (TCF). The number of nearest neighbors is increased from 30 to 100 by 10. Hence, there are eight values of MAE. We compare these values in order to judge which prediction is more close to the real ratings. The results are shown in the **Figure 2** as follows:
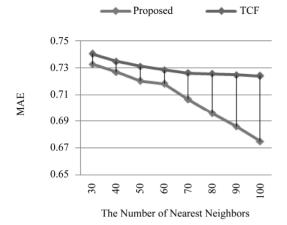
Figure 2. MAE of the two approaches.

The results shown in the **Figure 2** reveal that our approach (Proposed) is better than the traditional item-based CF (TCF). Because with the number of nearest neighbors increasing, the values of MAE of our approach are smaller than the ones of the TCF, and the values of MAE of our approach are decreasing more and more.

5. Conclusions and Future Work

This study proposed a contextual item-based collaborative filtering technology which incorporated users' important contextual information into the traditional item-based collaborative filtering in order to cope with that users have different preference for the items in different context. The results of the experiment showed that our approach surpasses the traditional item-based CF in providing accurate items for the users in the proper time. That is to say in the mobile environment, when we take user's contextual information into consideration, the results of the recommendation will be more accurate.

The limitation of our approach is that we only selected three attributes namely age, gender and occupation as user's contextual information in the experiment. But in fact, sometimes user's time, location, relations and the other contextual information may be more important.

Therefore, in the future, we are planning to design a better experiment which will own more contextual types and more ratings to evaluate our approach. Meanwhile, we will continue doing research on incorporating context into personalized recommendation systems in E-commerce.

REFERENCES

[1] G. Adomavicius, R. Sankaranarayanan, S. Sen and A. Tuzhilin, "Incorporating Contextual Information in Recommender Systems Using a Multidimensional Approach," *ACM Transactions on Information System*, Vol. 23, No. 1, 2005, pp. 103-145. doi:10.1145/1055709.1055714

[2] S. Perugini, M. A. Goncalves and E. A. Fox, "Recommender Systems Research: A Connection-Centric Survey," *Intelligent Information Systems*, Vol. 23, No. 2, 2004, pp. 107-143. doi:10.1023/B:JIIS.0000039532.05533.99

[3] R. Burke, "Hybrid Recommender Systems: Survey and Experiments," *User Modeling and User-Adapted Interaction*, Vol. 12, No. 4, 2002, pp. 331-370. doi:10.1023/A:1021240730564

[4] R. C. Blatterg, B. Kim and S. A. Neslin, "Database Marketing," Springer Science and Business Media, New York 2008.

[5] M. Setten, S. Pokraev and J. Koolwaajj, "Context-Aware Recommendations in the Mobile Tourist Application COMPASS," *Adaptive Hypermedia*, Vol. 3137, 2004, pp. 235-244. doi:10.1007/978-3-540-27780-4_27

[6] K. Kabassi, "Personalizing Recommendations for Tourists," *Telematics and Informatics*, Vol. 27, No. 1, 2010, pp. 51-66. doi:10.1016/j.tele.2009.05.003

[7] A. Chen, "Context-Aware Collaborative Filtering System: Predicting the User's Preference in the Ubiquitous Computing Environment," In: T. Strang and C. Linnhoff-Popien, Eds., *Location- and Context-Awareness*, Springer-Velag, Berlin, 2005, pp. 75-81.

[8] M. Gao and Z. Wu, "Personalized Context-Aware Collaborative Filtering Based on Neural Network and Slope One," In: H. Yu, Ed., *Cooperative Design*, *Visulization*, *and Engineering*, Springer-Velag, Berlin, 2009, pp. 109-116.

[9] A. K. Dey, "Understanding and Using Context," *Personal and Ubiquitous Computing*, Vol. 5, No. 1, 2001, pp. 4-7. doi:10.1007/s007790170019

[10] A. Zimmermann, A. Lorenz and R. Oppermann, "An Operational Definition of Context," In: B. Kokinov, D. C. Richardson and T. R. Roth-Berghofer, Eds., *Modeling and Using Context*, Springer-Velag, Berlin, 2007, pp. 558-571.

[11] G. Adomavicius and A. Tuzhilin, "Context-Aware Recommender Systems," In: F. Ricci, L. Rokach, B. Shapira and P. B. Kantor, Eds., *Recommender Systems Handbook*, *Part* 1, Springer Science and Business Media, New York, 2011, pp. 217-256.

[12] Z. Huang, X. Lu and H. Duan, "Context-Aware Recom Mendation Using Rough Set Model and Collaborative Filtering," *Artificial Intelligence Review*, Vol. 35, No. 1, 2011, pp. 85-99. doi:10.1007/s10462-010-9185-7

IIM