THE USER INFORMATION-BASED MOBILE APPLICATION RECOMMENDER SYSTEM

So-Young Yun, Sung-Dae Youn
Dept. of Computer Engineering, Pukyong National University
Pusan, Republic of Korea
ysmallzero@pknu.ac.kr
Dept. of Computer Engineering, Pukyong National University
Pusan, Republic of Korea
sdyoun@pknu.ac.kr

ABSTRACT
Smart phone and tablet PC users would like the variety of application contents in addition to web searching. Most of the mobile application stores tend to recommend other users applications merely because they have been chosen by many users. Thus, users who would like to choose applications that they personally prefer from hundreds of thousands of applications in mobile application stores need to spend a lot of time and efforts. This paper suggests a Collaborative Filtering (CF) method in utilization of users’ device ID information and the weight of certain demography information in order to minimize problems stated above and enhance the accuracy of recommendations. This method extracts preferred category information for each age bracket based on application categories and users’ age information, and decides the ranks in consideration of the similarity reflected in the weight of certain applications within the category based on the users’ information. The experiment shows that the proposed method improved the performance, which made accuracy of recommendation better than ranking-based method.

KEY WORDS
Recommender system, demography information, mobile application, collaborative filtering

1. Introduction
For the last a few years, the use of mobile devices such as smart phones and tablet PCs has been rapidly increasing around the world. Most of the users of such mobile devices would use them for applications as well as web searching. Accordingly, the use of contents by means of mobile applications is also increasing.
In 2011, the number of downloads of mobile applications around the world is estimated to 18 billion, which shows 144% increase compared to that of the previous year[1], and the smart phone application market is estimated to expand from 6.8 billion dollars in 2010 to 92.5 billion by 2013, which is about four times increase. Even the new term, ‘App economy,’ which means that mobile applications create a new economic structure, has emerged [2].
In expectation that application stores will become an influential service just as the web itself, contents businesses are making efforts to develop applications while companies are seeking to improve business achievements in various perspectives such as expanding the business scale through application stores, providing differentiated services, and diversified marketing strategies [2].
Hundreds of thousands of applications with various and convenient functions have been already registered in mobile application stores, and the number is rapidly increasing. Currently, these application stores recommend applications to other users based on the popularity of certain applications in each category or in general. Thus, users have to refer to reviews, recommendations from acquaintances, or the ranking of purchased items in order to find applications that they prefer from a number of applications in application stores.
Such a mere ranking-based recommender method causes users a lot of time and effort to find and select applications that they are looking for.
Thus, this paper suggests a CF recommender system based on users’ device ID information and the weight of demography information instead of recommendation based on mere information by word of mouth or popularity. The suggested method gathers application category information preferred by certain age groups of users, and uses similarity that reflects how users’ information affects the ranking of applications in each category. We only use age and gender as demography information. Because App stores don't require a lot of user information when user sign up app store. Upon access to an application store, the user can see application categories and application ranking in each category based on the user’s device ID information so that he or she can find the preferred application more accurately and promptly.
2. Related Work

In this section, we describe recommender system and collaborative filtering.

2.1 Recommender System

The Recommender System is a new personalized system that finds and recommends appropriate products or items depending on users’ preference [3]. The Recommender System has been successfully adopted to e-commerce companies such as Amazon.com and CDnow and facilitated purchases. This system has been proved to improve sales in a number of e-commerce sites [4].

The Recommender System takes advantage of various methods to process data and generate the list of recommendations such as contents-based recommendation, rule-based reasoning, case-based reasoning, and collaborative filtering. Among these, collaborative filtering is adopted most successfully [5].

2.2 Collaborative Filtering

The collaborative filtering method analyzes other users’ preferences similar to a certain user’s preferred patterns and recommends related items or services. This method has been most successful in the sector of recommendation systems, especially among e-commerce companies [6]. Collaborative filtering is classified to the user-based algorithm [7] and item-based algorithm [8] based on what to be the basis for predicting preferences between ‘user’ and ‘item.’ The user-based algorithm determines the similarity among users, and recommends items to be preferred by a certain user based on items recommended by other users with similar preferences. The item-based algorithm determines the similarity among items and predicts which user would prefer which items [9].

3. User Information-based Mobile Application Recommender System

In this section, we describe our proposed method using application categories and weight of demography information, taking three steps as in the following:

3.1 Step 1: Extracting application categories preferred by a certain user group

In this step, the database is generated based on application category and user information(age, gender) to determine the application category preferred by a certain user group. The database classifies seven age groups and creates sub-database for each group. From each group, information on application categories preferred by each gender is extracted.

Figure. 1 shows the classification of age, and Figure. 2 shows an example of database of age group

3.2 Step 2: Computing user similarity for each category

In this step, the similarity in user information is computed to decide the ranking of application recommendations in application categories. In the similarity computation, information on users’ age group and gender is used as weight factors.

Weight \( W \) is computed as in the following.

\[
W = 1 + w_a + w_g \tag{1}
\]

\( w_a \) is the weight of age group of user, \( w_g \) is the weight of gender of user respectively. \( w_a \in [0,1], w_g \in [0,1] \).

The following Pearson correlation is used to compute similarity between two users:

\[
sim(a,b) = \frac{\sum_{i=1}^{n} (R_{ai} - \bar{R}_a)(R_{bi} - \bar{R}_b)}{\sqrt{\sum_{i=1}^{n} (R_{ai} - \bar{R}_a)^2 \sum_{i=1}^{n} (R_{bi} - \bar{R}_b)^2}} \tag{2}
\]

where \( R_{ai}, R_{bi} \) are the ratings of the item \( i \) by user \( a \) and user \( b \), \( \bar{R}_a, \bar{R}_b \) are average ratings of user \( a \) and user \( b \) for all the co-rated items. \( I_{ab} \) is the items set both rating by user \( a \) and user \( b \).

The following shows how weight \( W \) is used for Pearson coefficient of (2) to compute similarity:

\[
sim(a,b) = \frac{\sum_{i=1}^{n} W^2(R_{ai} - \bar{R}_a)(R_{bi} - \bar{R}_b)}{\sqrt{\sum_{i=1}^{n} (R_{ai} - \bar{R}_a)^2 \sum_{i=1}^{n} W^2(R_{bi} - \bar{R}_b)^2}} \tag{3}
\]
3.3 Step 3: Extracting application

After the similarity is computed based on the weight, the upper 10% users of high similarity to target user are selected as neighbors. Lastly, among the applications chosen by neighbors, Top-N applications are designated as recommended applications at the top of the category. Based on the user’s device id information, the application category and ranking in the category are arranged and displayed on the screen.

4. Experimental Evaluation

4.1 Data set

To evaluate the performance of the proposed method, we used the MovieLens dataset. MovieLens dataset consist of 100,000 ratings from 943 users on 1682 movies with every user having at least 20 ratings and demographic information for the users is included [10]. We use 80% of rating dataset as training data, and use 20% of dataset as the test data. The experiment is done using the Android-based mobile device.

4.2 Performance measurement

The metrics for evaluating the accuracy of a prediction algorithm can be divided into two main categories [11]: statistical accuracy metrics and decision-support metrics. In this paper, we use the statistical accuracy metrics [12]. As statistical accuracy measure, mean absolute error (MAE) is employed.

If n is the number of actual ratings in an item set, MAE is defined as the average absolute difference between the n pairs. Assume that p₁, p₂, p₃,…,pₙ is the prediction of users’ ratings, and the corresponding real ratings data set of users is q₁, q₂, q₃,…,qₙ. MAE definition as following:

\[ MAE = \frac{\sum_{i=1}^{n} | p_i - q_i |}{n} \] (4)

The lower the MAE, the more accurate the recommendation engine predicts user ratings [8].

4.3 Comparing the proposed method with ranking-based method

The method combining the proposed in this paper user’s device ID information and the user-based CF using weight of demographic information is compared with ranking-based method. Table 1 shows the results of our experiment.

The result indicates that the accuracy of the proposed method performs better than ranking-based method.

<table>
<thead>
<tr>
<th>Method</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ranking-based method</td>
<td>0.932</td>
</tr>
<tr>
<td>Proposed method</td>
<td>0.858</td>
</tr>
</tbody>
</table>

The similarity threshold has a significant effect on the prediction quality. We performed an experiment to determine the sensitivity of the similarity threshold. Figure 3 shows the result of the comparison of prediction quality based on changing threshold of similarity.

5. Conclusion

As the use of applications among smart phone and tablet PC users has been drastically increasing, application stores recommend applications so that users can readily find applications that they want. However, recommending applications only based on categories and number of downloads has limitations in reducing the time of searching for certain items. In this paper, therefore, suggests a method that combines item-based and user-based CF to reduce the searching time and enhance the accuracy of recommendations. This method is to gather information on application categories preferred by each age group, and decide the ranking of recommended applications by means of the weight that is based on demography information. After the ranking of recommendation is decided based on the users’ age and gender, the categories and applications appropriate for the user are shown at the top of the list upon access to the application store through a mobile device, which reduces the searching time and enhances the accuracy of recommendations.

The result of the experiment indicates that the accuracy of the proposed method performs better than a simple ranking-based method.

References


