Association for Information Systems
AIS Electronic Library (AISeL)

PACIS 2004 Proceedings

Pacific Asia Conference on Information Systems (PACIS)

December 2004

Collaborative Filtering Algorithm Based on Mutual Information

Li Yu
School of Economics & Management

Lu Liu
School of Economics & Management

Collaborative Filtering Algorithm Based on Mutual Information

Li Yu
School of Economics & Management
BeiHang University, China, 100083
buaayuli@hotmail.com

Lu Liu
School of Economics & Management
BeiHang University, China, 100083
liulu@buaa.edu.cn

Abstract

Recommender systems are used by E-commerce sites to suggest products to their customers and to provide consumers with information to help them determine which products to purchase. Collaborative filtering algorithm is the most extensive personalized recommendation used in recommender systems. Since not being considering the dependence between predicted item and historical item, typical collaborative filtering algorithm is not fit for multiple interests recommendation. The authors analyzed the reason and presented a new algorithm, collaborative filtering based on mutual information. By removing the historical items on which predicted item has not high dependence, the algorithms can deal with personalized recommendation for user's multiple interests. The experiment shows new algorithms is more accurate than other algorithms, especially fitting for the environment where users have many completely different interests.

Keywords: Collaborative filtering, Recommender system, Mutual information, Personalized recommendation,

1. Introduction

The amount of information in the world is increasing far more quickly than our ability to process it. All of us have known the feeling of being overwhelmed by the number of new books, journal articles, and conference proceedings coming out each year. Technology has dramatically reduced the barriers to publishing and distributing information. Now it is time to create the technologies that can help us sift through all the available information to find what is most valuable to us.

One solution to this information overload problem is the use of recommender systems. Recommender systems are used by E-commerce sites to suggest products to their customers and to provide consumers with information to help them determine which products to purchase. The products can be recommended based on the top overall sellers on a site, on the demographics of the consumer, or on an analysis of the past buying behavior of the consumer as a prediction for future buying behavior. The forms of recommendation include suggesting products to the consumer, providing personalized product information, summarizing community opinion, and providing community critiques. Recommender systems enhance E-commerce sales in three ways: helping customers find products they wish to purchase, converting browsers into buyers; improving cross-sell by suggesting additional products for the customer to purchase; improving loyalty by creating a value-added relationship between the site and the customer (Schafer et al 2001).

The underlying techniques used in today's recommendation systems fall into two distinct categories: content-based and collaborative methods. Content-based methods require textual descriptions of the items to be recommended and draw on results from both information retrieval and machine learning research (Pazzani et al 1997). In general, a content-based system analyzes a set of documents rated by an individual user and uses the content of these documents, as well as the provided ratings, to infer a profile that can be used to recommend additional items of interest. In contrast, collaborative methods recommend items based on aggregated user ratings of those items, i.e. these techniques do not depend on the availability of textual descriptions. Both approaches share the common goal of assisting in the user's search for items of interest, and thus attempt to address one of the key research problems of the information age: locating needles in a haystack that is growing exponentially (Breese et al 1998).

In this paper we focus on collaborative filtering techniques. A variety of algorithms have previously been reported in the literature and their promising performance has been evaluated empirically (Shardanand et al 1995, Breese et al 1998; Sarwar et al 2000, HSR95, Resnick et al 1994). These results, and the continuous increase of people connected to the Internet, led to the development and employment of numerous collaborative filtering systems. Virtually all topics that could be of potential interest to users are covered by special-purpose recommendation systems: web pages, news stories, movies, music videos, books, CDs, restaurants, and many more. Some of the best-known representatives of these systems, such as FireFly (www.firefly.com) or WiseWire (www.wisewire.com) have turned into commercial enterprises. Furthermore, collaborative filtering techniques are becoming increasingly popular as part of online shopping sites. These sites incorporate recommendation systems that suggest products to users based on products that like-minded users have ordered before, or indicated as interesting. For example, users can find out which CD they should order from an online CD store if they provide information about their favorite artists, and several online bookstores (e.g. www.amazon.com) can associate their available titles with other titles that were ordered by like-minded people.However, there remain important research questions in overcoming fundamental challenges for collaborative filtering recommender systems.

1.1 Related Work

In this section we briefly present some of the research literature related to collaborative filtering, recommender systems.Tapestry (Goldberg et al 1992) is one of the earliest implementations of collaborative filtering-based recommender systems. This system relied on the explicit opinions of people from a close-knit community, such as an office workgroup. However, recommender system for large communities cannot depend on each person knowing the others. Later, several ratings-based automated recommender systems were developed. The GroupLens research system (Konstan 1997, Resnick et al 1994) provides a pseudonymous collaborative filtering solution for Usenet news and movies. Ringo and Video Recommender (Hill 1995) are email and web-based systems that generate recommendations on music and movies respectively. A special issue of Communications of the ACM (Resnick et al 1997) presents a number of different recommender systems. Other technologies have also been applied to recommender systems, including Bayesian networks, clustering, and Horting (Aggwrwal et al 1999，Breese et al 1998).

Various approaches for recommender systems have been developed that utilize either demographic, content, or historical information (Balabanovic et al 1998, Basu et al 1998, Shardanand et al 1995, Terveen et al 1997, Konstan 1997). User-based collaborative filtering is probably the most successful and widely used techniques for building recommender systems (Resnick et al 1994, Konstan 1997). For each user, user based collaborative filtering recommender systems use historical information to identify a neighborhood of people that in the past have exhibited similar behavior (e.g., accessed the same type of information, purchased a similar set of products, liked/disliked a similar set of movies) and then analyze this neighborhood to identify new pieces of information that will be liked by the user. So this method also be called neighbor-based collaborative filtering or nearest neighbor algorithms.

Although these systems have been successful in the past, their widespread use has exposed some of their limitations such as the problems of sparsity in the data set, problems associated with high dimensionality and so on. Sparsity problem in recommender system has been addressed in (Good et al 1999). The problems associated with high dimensionality in recommender systems have been discussed in (Billsus et al 1998), and application of dimensionality reduction techniques to address these issues has been investigated in (Sarwar et al 2000).

But very few work show that classical collaborative filtering is not adaptive to multiple interests recommendation. In fact, the quality of its recommendation is very poor when users in recommender systems have completely different interests. Unfortunately, this situation exists commonly. In this paper, we focus on solving this problem. Once (Hofmann 2001) tried to solve the problem using a probabilistic model from the model-based perspective, present the probabilistic model based collaborative filtering. But the method has the shortcoming of all model-based collaborative filtering. Firstly, the method presented by (Hofmann 2001) cannot explain why classic collaborative filtering cannot adapt to multiple interests recommedation. Secondly, although the computing speed is far than user-based collaborative filtering, recommendation cannot vary with the rating database of recommendation system, which results that users can get on-line recommendation.

Our previous work attacks the problem from user-based perspective (Yu et al 2003a). In order to address this issue we have explored item and user based collaborative filtering techniques by combining item-based and user-based collaborative filtering techniques together (short for CF-IU). Item and user based collaborative filtering techniques analyze the user-item matrix to identify similarity between target items and other items, get similar item to the target item, and determine neighbor users of active user according to the rating of other user for the similar item, then compute recommendations for users. In fact, the algorithm is a kind of collaborative filtering based on clustering of items. But Pearson correlation measuring the similarity of item used in this algorithm, which is very key for determine the neighbors, has many limitations when dealing with multi-interests problem. So in this paper, we present the collaborative filtering based on mutual information (CF-MI) for multi-interest recommendation.

1.2 Contributions

This paper has three primary research contributions: (1) Identification limitation of similarity measurement used in CF-IU algorithms (Yu et al 2003a) for multi-interest; (2). Presentation of the new algorithms, collaborative filtering based on multi Information (CF-MI), to improve the recommendation accuracy for the user with the multi-interest; (3). An experimental comparison of the quality of CF-MI algorithms with the typical collaborative filtering algorithms and CF-IU algorithms.

1.3 Organization

The rest of the paper is organized as follows. The next section provides a brief background in user-based collaborative filtering algorithms. We first formally describe classic collaborative filtering algorithm, then analyze questions of user-based collaborative filtering for multiple interest recommendation. In section 3, we present collaborative filtering based on mutual information (CF-MI), which gives recommendation based on the subset of all items. Section 4 describes our experimental work. It provides the details of data sets, evaluation metrics, procedure and results of different experiments, as well as the discussion of the results. The final section provides some concluding remarks and directions for future research.

2. Typical Collaborative Filtering and Problem Analysis

Recommender systems apply data analysis techniques to the problem of helping users find the items they would like to purchase at E-Commerce sites by producing a predicted likeliness score or a list of top–N recommended items for a given user. Item recommendations can be made using different methods. Recommendations can be based on demographics of the users, overall top selling items, or past buying habit of users as a predictor of future items. Currently, there are two kind of collaborative filtering algorithm, including user based collaborative filtering and item based collaborative filtering.. In this paper, we will mainly discuss the user based collaborative filtering, which is the most typical personalized recommendation techniques researched and used today. In this part, the algorithm is firstly introduced, and we will analyze its limitation for dealing with user's multiple interests.

2.1 Typical Collaborative Filtering (short as CF-TY)

Collaborative filtering (CF) (Resnick et al 1994, Shardanand et al 1995, Sarwar et al 2000) is the most successful recommendation technique to date, and is extensively used in many commercial recommender systems. These schemes rely on the fact that each person belongs to a larger group of similar behaving individuals. Consequently, items (i.e., products) frequently purchased by the various members of the group can be used to form the basis of the recommended items. Collaborative filtering includes the user-based collaborative filtering and item-based collaborative filtering. And user-based collaborative filtering is the most typical schemes, which widely used in personalized recommendation systems. So we note it as CF-TY in this paper.

Let R be an $n \times m$ user-item matrix containing historical purchasing information of n customers on m items. In this matrix, $R_{i,j}$ is one if the i th customer has purchased the j th item, and zero otherwise. Let U be the set of items that have already been purchased by the customer for which we want to compute the top-N recommendations. We will refer to this customer as the active customer and in order to simplify the presentation we will assume that the active customer does not belong to the n customers stored in matrix R. User-based CF recommender systems compute the top-N recommended items for that customer as follows.
First they identify the k most similar customers in the database. This is often done by modeling the customers and items with the vector-space model, which is widely used for information retrieval (Breese et al 1998, Sal89, Sarwar et al 2000). In this model, each of the n customers as well as the active customer is treated as a vector in the m-dimensional item space, and the similarity between the active and the existing customers is measured by computing the cosine between these vectors or correlation.

Once this set of the k most similar customers have been discovered, their corresponding rows in R are aggregated to identify the set C of items purchased by the group as well as their frequency. Using this set, user-based CF techniques then recommend the N most frequent items in C that are not already in U (i.e., the active user has not already purchased). Note that the frequency of the items in the set C can be computed either by just counting the actual occurrence frequency or by first normalizing each row of R to be of the same length. This latter normalization gives less emphasis to items purchased by customers that are frequent buyers and leads to somewhat better results.

Despite the popularity of user-based CF recommender systems, they have a number of limitations related to scalability and real-time performance. The computational complexity of these methods grows linearly with the number of customers that in typical commercial applications can grow to be several millions. Furthermore, user-based CF recommender system is hard to provide the explanation for the recommendation.

Except for above limitation, CF-TY has another limitation, which provides much poor recommendation if users have many different interests. It means that CF-TY cannot deal with multiple interest recommendation. Unfortunately, this situation exists commonly. The limitation will results into non-confidence of user for recommender system. The fact has shown that users usually do not seek for help by recommender system when they want to find and purchase a costly item. In this paper, we will focus on the limitation of CF-TY and try to resolve it by presenting a novelty recommendation method.

2.2. Problem Analysis

According to CF-TY, prediction for target item is determined by preference for rated item. But it is common that one user have many different interests. For example, a user may be interested in both 'Football' and 'English'. If we predict the 'English' item using interest preference for 'Football', the prediction result is doubtful.
Let's see an example as following.

User/Item	**I1**(English)	**I2**(Football)	**I3**(English)	I4**(Football)**	**I5**(Football)	**I6**(English)
U1	3	1	2	3	5	5
U2	3	1	2	3	5	5
U3	3	1	2	3	5	5
U4	1	5	3	3	1	1
U5	2	5	2	3	2	1
U6	3	5	1	3	2	1
U7	3	5	2	4	2	**?**

Table 1: An example of user/item data matrix

As table 1 shown above, in the user/item data matrix, there are seven users and six items. For analyzing, among the six items, I1(English),I3(English) and I6(English) means their content are on 'English', that is to say the three item is similar, but are different items; I2(Football),I4(Football) and I5(Football) means their content are on 'Football', they are similar, but are different items. Let's see an example, now we are ready to predict rating of user U7 for item I6, R76=?

Here we suppose each user has three neighbor users for analyzing.If prediction is done with classical user-based CF algorithm, then U4, U5, and U6 will be neighbors of U7, It is easy to get the prediction value, R76=1. But we find that the reason why U4, U5, and U6 will be neighbors of U7 is that all of them are interested in Football, Football is their common interest. That is to say that we use the interest preference on 'football' to predict the interest preference on 'English', but 'Football' and 'English' are not related, so prediction is not accuracy and is doubtful.

In order to further expatiate our view, suppose an extremity, in the table 2. If there were no rating of U4, U5 and U6 for I1 and I3, we would get the table. According to user-based CF, U4, U5, U6 will be neighbors to U7 and prediction of rating of U7 for I6 is completely determined by their preference information on Football. Because they are completely different and not related on content, error of prediction will be very high.

User/Item	**I1**(English)	**I2**(Football)	**I3**(English)	I4**(Football)**	**I5**(Football)	**I6**(English)
U1	3	1	2	3	5	5
U2	3	1	2	3	5	5
U3	3	1	2	3	5	5
U4	--	5	--	3	1	1
U5	--	5	--	3	2	1
U6	--	5	--	3	2	1
U7	3	5	2	4	2	**?**

Table 2: Another example of user/item data matrix

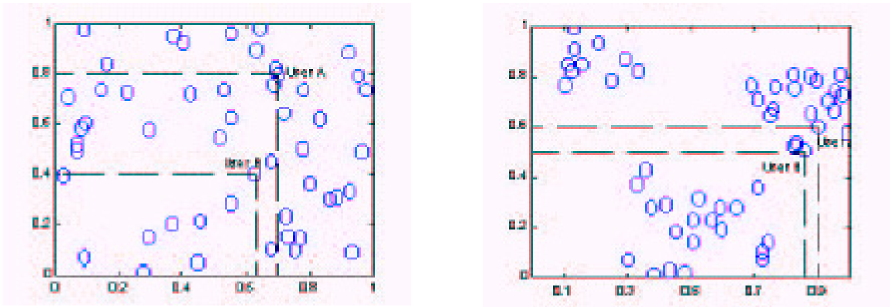
For the above example, if neighbors of U7 is determined according to interest preference for 'English' of U1-U6, that is to say, similarity between U7 and other user is compute according to the rating of user for I1 and I3, not from I1 to I5, then we will find that the users U1, U2. U3 will be neighbors to user U7. So prediction value of item I6 for user U7, p76=5 will more be trustful because the interest preference of U7 and its neighbors U1, U2. U3 for 'English' is similar.

According to the above analysis, we may get a conclusion that for a certain user, its neighbors is also related to predicted item. It means that neighbor of different items for identical user be not identical, which requires that, in collaborative rcommendation, predicted item and items which is used to predict for predicted item is similar on content of items. It led to the improvement for user-based collaborative filtering, so we present following a new algorithms, collaborative filtering based on mutual Information (CF-MI).

3、Collaborative Filtering based on Multi Information (CF-MI)

3.1、Measuring the dependence of two items by mutual information

Three reasons let us select mutual information to measure the dependence of the two items: Firstly, Pearson correlation only can be used to reflect the liner relation of the two items, but it is not accurate when using the linear relation to reflect the dependence or similarity of the two items, especially if the relation of two items is not linear. But mutual information can reflect the no-linear relation.Let us consider an example. As shown in Figure 1, 50 consumers give votes for movie 1 and movie 2 for case 1 and case 2. Linear relation, such as Pearson correlation coefficient, cosine etc, is hard to distinguish the relation of the movie1 and movie 2 in case1 and case 2 because the rating for the two movies distributed randomly, that is to say, if we compute the correlation of two movies by Pearson correlation coefficient, correlation coefficient for both case1 and case 2 are almost equal, all close to zero. But in fact, case 1 and case 2 are distinguishing. In case 1, we find consumers are nearly uniformly distributed in the movie-movie vote space. If A and B are two arbitrary consumers who have similar ratings for movie 2, it does not necessarily indicate that they also have similar ratings for movie 1. In case 2, however, we find that those consumers who dislike movie 2 always like movie 1. While those consumers who like movie 2 always rate the other one just above the average. This indicates that movie j in case 2 should play an important role in inferring consumer preference for movie 1, while in case 1 it is not so useful.



CASE1 CASE2

The points distribute random in case 1.The points form some clusterings in case 2.

Figure 1:Two cases of rating

Secondly, since computing Pearson correlation coefficient, cosine etc need more data, items similarity measured by Pearson correlation coefficient, cosine etc is not accurate when data used is little. But mutual information be suit to the situation, that is to say, item similarity or dependency measured by mutual information is more accurate than measured by Pearson correlation coefficient, cosine etc when data which could be used is little. The fact is very important in recommendation application because the user-rating matrix is generally very sparse. The latter experiment had approved it. Finally, mutual information is a fast computation method used to reflect the dependence of items.

The dependence of product i on product j can be formally defined by the following conditional probability:

$$p(\left|v_{j,u_A} - v_{j,u_B}\right| < e \mid \left|v_{i,u_A} - v_{i,u_B}\right| < e) \qquad (1)$$

where A and B represent two arbitrary consumers and e is a threshold. If the difference between two votes is less than e, then the two votes are considered close. The above conditional probability indicates the probability of two arbitrary consumers having close preference for product i given the condition that the two consumers have close preference for product j.

But the above computing for mutual information is complicate. If there are n users and m items in recommender system, them complication of computing is O(n2m2). So we develop an information-theoretic measure that is equivalent to the above probabilistic dependence definition in the case of discrete voting. The conclusion be supported the following theorem.

Theorem . Let P(V_i), P(V_j), and P(V_i, V_j) be the margin and joint distributions of votes for two products i, j, and e = 1 the interval of discrete vote value, 0, 1, ..., N, assume that P(V_i) and P(V_j) are fixed, if A and B are two arbitrary consumers who have voted for both products, then MI(V_i; V_j) increases as dependence increases, which means the differential of dependence defined by (1) with respect to the mutual information MI(V_i; V_j) is always positive.

$$\frac{d\left\lfloor p(|v_{j,u_A}-v_{j,u_B}|<e \mid |v_{i,u_A}-v_{i,u_B}|<e)\right\rfloor}{d\left\lfloor MI(V_j;V_i)\right\rfloor}>0 \qquad (2)$$

Proof:
Since P(Vi) and P (Vj) are given, we have:

$$d\left\lfloor MI(V_j;V_i)\right\rfloor = d\left\lfloor H(V_j)-H(V_j \mid V_i)\right\rfloor = -d\left\lfloor H(V_j \mid V_i)\right\rfloor$$

Inequation (2) can be written as:

$$\frac{d\left\lfloor p(|v_{j,u_A}-v_{j,u_B}|<e \mid |v_{i,u_A}-v_{i,u_B}|<e)\right\rfloor}{d\left\lfloor H(V_j \mid V_i)\right\rfloor}<0 \qquad (3)$$

Next, we have

$$H(V_j \mid V_i)=\sum_{v\in\aleph} p(V_i\equiv v)H(V_j \mid V_i\equiv v) \qquad (4)$$

and

$$d\left\lfloor p(|v_{j,u_A}-v_{j,u_B}|<e \mid |v_{i,u_A}-v_{i,u_B}|<e)\right\rfloor = \frac{\sum_{v\in\aleph} p(V_i\equiv v)^2\, p(|v_{j,u_A}-v_{j,u_B}|<e \mid v_{i,u_A}=v_{i,u_B}=v)}{\sum_{v\in\aleph} p(V_i\equiv v)^2} \qquad (5)$$

where is the set of all discrete votes. From eq. (4) and eq. (5) we can easily derive in eq. (3). Therefore, in eq.(2) holds.

According to above analysis, we use multi information to measure probabilistic dependence definition in the case of discrete voting. Mutual Information is important concept in information theory, which was used to show the overlapping of the two variables, as shown in figure 2. In information theory, mutual information represents a measure of statistic dependence between two random variables *X* and *Y* with associated probability distributions $p(x)$ and $p(y)$ respectively(Yu et al 2003). Following Shannon, the mutual information between (Yu et al 2003) *X* and *Y* is defined as:

$$MI(X;Y)=\sum_x\sum_y p(x,y)\log\left(\frac{p(x,y)}{p(x)p(y)}\right) \qquad (6)$$

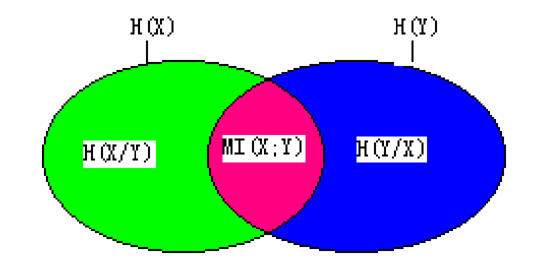


Figure 2：Mutual Information

Furthermore, mutual information can be equivalently transformed into the following formulas:

$$MI(X;Y)=H(X)-H(X \mid Y) \qquad (7)$$

$$MI(X;Y)=H(Y)-H(Y \mid X) \qquad (8)$$

$$MI(X;Y)=H(X)+H(Y)-H(X,Y) \qquad (9)$$

where H(X) is the entropy of X, H(X|Y) is the conditional entropy of X given Y and H(X,Y) is the joint entropy of two random variables, where

$$H(X)=-\sum_{j=1}^{k} p(x_j)\log p(x_j) \qquad (10)$$

$$H(X \mid Y) = -\sum_i \sum_j p(x_i, x_j) \log p(x_i \mid x_j) \tag{11}$$

$$H(X, Y) = -\sum_i \sum_j p(x_i, x_j) \log p(x_i, x_j) \tag{12}$$

In summary, according to above the definition and theorem, we will use multi information to measure the dependence of two items and multi information can be computed by fellow equation:.

$$MI(V_j; V_i) = H(V_j) + H(V_i) - H(V_j, V_i) \tag{13}$$

where V_i and V_j denoting the rating vector respectively for item i and j .

3.2、CF-MI Algorithm

Definition1: Dependent items ----- Given item *i* and threshold value of multi information, $\overline{MI}$ (or constant *N*), for any item *j*, if $MI(V_j; V_i) > \overline{MI}$ （or $MI(V_j; V_i)$ is the one of the *N* biggest $MI(V; V_i)$ ），then item *j* is called dependent item of item *i*. , noted as $DI(i)$.

Definition2: Dependent item set ----- Dependent item set of item *i* consist of all dependent item of item *i*., noted as $DIS(i)$.

Computing Procedure：

⑴(1) For target item $j \in I$，Compute $MI(V_j; V_k)$ (k =1, 2, …);

⑵(2) According to $MI(V_j; V_k)$ (k =1, 2, …), determine $DIS(j)$;

⑶(3) Based on the $DIS(j)$, compute the similarity between user *a* and other user *u* ($u \neq a$)—$w_j(a,i)$；

⑷(4) According the $w_j(a,i)$, determine neighbor users of user *a* for target item *j*--$Neighbor_{a,j}$；

⑸(5) By performing a weighted average of deviations from the neighbor's mean , compute prediction value of rating of user *a* for item *j* ---$P_{a,j}$；

$$P_{a,j} = \overline{R_a} + k \sum_{i \in Neighnor_{a,j}} w_j(a,i)(R_{i,j} - \overline{R_i}) ,$$

$$\frac{1}{k} = \sum_{i \in Neighbor_{a,j}} w_j(a,i) \tag{14}$$

⑹(6) Determine the most interesting items of user *a* as recommendation according to $p_{a,j}$ ($j \in I_{P,a}$)。

4. Experimental Evaluation

4.1 Datasets

We ran experiments using data from the *EachMovie* collaborative filtering service. The *EachMovie* service was part of a research project at the Systems Research Center of Digital Equipment Corporation. The service was available for a period of 18 months and was shut down in September 1997. During that time the database grew to a fairly large size, containing ratings from 72916 users on 1628 movies. User ratings were recorded on a numeric six-point scale (0.0, 0.2, 0.4, 0.6, 0.8, 1.0). The data set is publicly available and can be obtained from Digital Equipment Corporation (McJ97).

Although the data from 72916 users is available, we restrict the analysis to the first 450 users in the database. These 450 users provided ratings for 250 different movies. We restricted the number of users considered, because we are only interested in the performance of the algorithm under conditions where the number of users and items is low. This is a situation that every collaborative filtering service has to go through in its startup-phase, and in many domains we cannot expect to have many users rating for many items. We also believe that the deficiencies of CF-IU, CF-MI and CF will be more noticeable under these conditions, because it is less likely to find users with considerable overlap of rated items.

Data Sets	Train Count	Test Count	All Count	Train Data Density	Test Data Density	Users	Train Movies	Test Movies
U50M250	512	74	586	5.120%	2.960%	50	200	50
U150M250	1130	173	1303	3.767%	2.307%	150	200	50
U250M250	1598	270	1868	3.196%	2.160%	250	200	50
U350M250	2166	389	2555	3.094%	2.223%	350	200	50
U450M250	2772	585	3357	3.080%	2.600%	450	200	50

Table 3: Experiment Datasets

4.2 Evaluation Metrics

Recommender systems research has used several types of measures for evaluating the quality of a recommender system. They can be mainly categorized into two classes: *Statistical accuracy metrics* evaluate the accuracy of a system by comparing the numerical recommendation scores against the actual user ratings for the user-item pairs in the test dataset. *Mean Absolute Error* (MAE) between ratings and predictions is a widely used metric. MAE is a measure of the deviation of recommendations from their true user-specified values. For each ratings-prediction pair $<p_i, q_i>$ this metric treats the absolute error between them i.e., $|p_i - q_i|$ equally. The MAE is computed by first summing these absolute errors of the N corresponding ratings-prediction pairs, then compute the average. Formally,

$$MAE = \frac{\sum_{i=1}^{N} |p_i - q_i|}{N} \tag{15}$$

The lower the MAE, the more accurately the recommendation engine predicts user ratings. Root Mean Squared Error (RMSE), and Correlation are also used as statistical accuracy metric.*Decision support accuracy* metrics evaluate how effective a prediction engine is at helping a user select high quality items from the set of all items. These metrics assume the prediction process as a binary operation-either items are predicted (good) or not (bad). With this observation, whether a item has a prediction score of 1.5 or 2.5 on a five-point scale is irrelevant if the user only chooses to consider predictions of 4 or higher. The most commonly used decision support accuracy metrics are reversal rate, weighted errors and ROC sensitivity (SKBHMR98). We used MAE as the choice of evaluation metric to report prediction experiments because it is most commonly used and easiest to interpret directly. In the paper (SKBHMR98), the experiments have shown that MAE and ROC provide the same ordering of different experimental schemes in terms of prediction quality.

4.3 Experimental Procedure

The whole experiment is used to test CF-TY, CF-IU and CF-MI algorithm. So we make a comparing for performance of the three algorithms on the whole.Since accuracy of CF-TY is related to user correlation threshold (UCT), it of CF-IU is related to UCT and movie correlation threshold (MCT) while accuracy of CF-MI is related to UCT and mutual information threshold (MIT). In order to compare the three algorithms, we varied UCT and MCT from 0 to 0.9 with step 0.1 and MIT from 1 to 2.8 with step 0.1 (because average of the mutual information is close to 1.5, we varied MIT from 1 to 2.8), then compute their MAE respectively for different UCT, MCT and MIT, and select their least MAE respectively to compare.

4.4 Experimental Results

In this section we present our experimental results of applying the three collaborative filtering techniques for generating predictions.As shown in figure 3, on the while, both of CF-IU and CF-MI have higher accurate than CF-TY. It shows that the algorithms we presented are advantage over the CF-TY algorithm. And we believe that, more different content of items in the dataset is, and more advantage the two new algorithms will have than CF-TY. The experiment also shows that CF-IU and CF-MI have similar recommendation quality. But CF-MI algorithm has higher accuracy than CF-IU when the user number is small which is in our expectation. It validated the conclusion that mutual information has more ability to reflect the relation of two items than pearson correlation used widely in the collaborative filtering, especially when data is sparse. It often occur when the recommendation system start to run. The merit can be used to resolve the "cold-start" problem that is a very key issue in recommendation system based collaborative filtering. Unexpectedly, the advantage of CF-MI over CF-IU is not obvious when data is more.

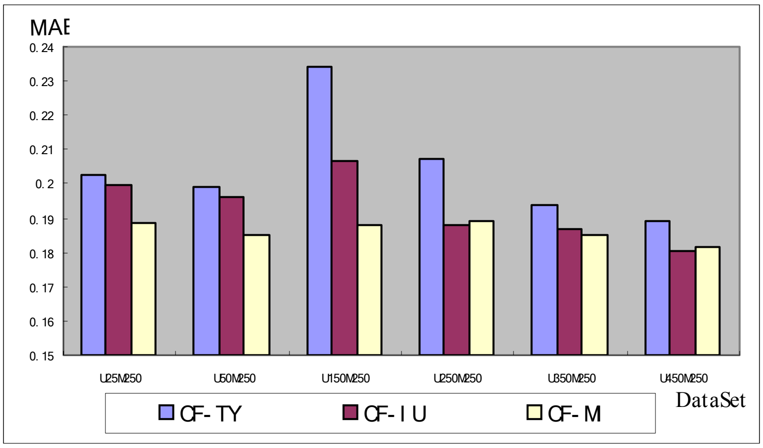


Figure 3: Performance comparing of CF-TY, CF-IU and CF-MI

4.5 Discussion

From the above experimental evaluation we have got some important observations.First, CF-IU and CF-MI provide better quality of predictions than CF-TY The improvement in quality is almost consistent over different threshold user correlation and training/test ratio. The second observation is that CF-MI has a greater advantage over CF-IU when data is small because mutual information has greater ability to reflect the dependence or relation of two items than pearson correlation. The later advantage is very important, especially recommend systems run firstly.One important point is need to state, although the above experiments have shown the advantage of CF-MI over CF-IU and CF-TY, we believe that the advantage is greater if content of item is completely different in user/item matrix and dataset is small. In EachMovie datasets, all of items are movies, which have less interest discrepancy for different users. Supposed the items in datasets is different commodity, such as book, CD etc., in which users have completely different interests, and their content is very discrepant, then CF-MI be more accurate than CF-TY, more accurate than CF-IU when dataset is small. For example, a footballer cannot read English book, a novelist cannot buy book through Internet. For this case, CF-MI has a greater advantage because the algorithm is able to filter the dissimilar item for target item and to engender the neighbor users of active user, which guarantee the target item is consist with the common interest of neighbor users, but CF-TY cannot. In addition, Because CF-IU is based on similar item and neighbor user, the algorithm can not only provide novelty recommendation for user like user-based collaborative filtering, also give an explain for recommendation result as like item-based collaborative filtering.

Also we admit that CF-MI increase computing time for recommendation comparing with CF-TY. For a fixed user, neighbor users must be computed again for different predicted item, so speed of predicting is lower than CF-TY, but accuracy of prediction is improved, especially for multiple interest recommendation. In most case, user would like get more accuracy recommendation by sacrifice of time. If recommendation is poor, users will not believe the recommendation and will not use the recommender systems anymore. So it is worthwhile in order to get more accurate recommendation.

5. Conclusions and Directions for Future Research

Recommender system is a powerful new technology for extracting additional value for a business from its user databases. These systems help users find items they want to buy from a business. Recommender systems benefit users by enabling them to find items they like. Conversely, they help the business by generating more sales. Recommender systems are rapidly becoming a crucial tool in E-commerce on the Web. Recommender systems are being stressed by multi-interest of user and the huge volume of user data in existing corporate databases, and will be stressed even more by the increasing volume of user data available on the Web. New technologies are needed that can dramatically improve the quality of recommender systems.

In this paper we presented and experimentally evaluated a new algorithms for multiple interest recommendation. Our results show that the new algorithm hold the promise of allowing CF-based algorithms to be adaptive to data sets in which users have many different interests and at the same time produce high-quality recommendations.

Future work will address two directions. One is to apply the new algorithms to the application in E-Commerce. In addition, the problem how to give an explanation for recommender systems based on CF-MI is imperative for improving the confidence of recommender systems based on collaborative filtering. It is also a key issue of decrease the computation time and guarantee user getting on-line recommendation.

6 Acknowledgments

Funding for this research was supported by the National Science Foundation of China under grants No.70371004 and PhD program Foundation of Education Ministry of China contract number 2000000601. We would like to thank the System Research Center of Digital

Equipment Corporation for making the Eachmovie database available for research. We also like to thank anonymous reviewers for their valuable comments.

7 References

Aggarwal. C. , Joel L.Wolf, Kun-LungWu, and Philip S. Yu. Horting hatches an egg: A new graph-theoretic approach to collaborative filtering. In Knowledge Discovery and Data Mining, pages 201–212, 1999.

Basu, C., Hirsh, H., and Cohen, W. (1998). Recommendation as Classification: Using Social and Content-based Information in Recommendation. In Recommender System Workshop '98. pp. 11-15.

Breese,J. David Heckerman, and Carl Kadie. Empirical analysis of predictive algorithms for collaborative filtering. In Proceedings of the 14th conference on Uncertaintly in Artificial Intelligence, pages 43¨C52, 1998.

Billsus, D. & Pazzani, M. Learning Collaborative Information Filters, in Proceedings of the International Conference on Machine Learning (Madison WI, July 1998), Morgan Kaufmann Publishers.

Balabanovic. M and Yoav Shoham. FAB: Content-based collaborative recommendation. Communications of the ACM, 40(3), March 1997.

Burke. R, Hybrid Recommender Systems:Survey and Experiments, Department of Information Systems and Decision Sciences, California State University, Fullerton

Goldberg, D. Nichols, D. Oki, B. M. and Terry, D.: Using collaborative filtering to weave an information apestry.Commun. ACM 35, 12, 1992.

Good, N.; Schafer, J. B.; Konstan, J. A.; Borchers, A.; Sarwar, B.; Herlocker, J.; and Riedl, J. 1999. Combining collaborative filtering with personal agents for better recommendations. In Proceedings of the Sixteenth National Conference on Artificial Intelligence (AAAI-99), 439–446.

Hofmann, T. Learning What People (Don't) Want. In Proceedings of the European Conference on Machine Learning, 2001.

Hill, W., Stead, L., Rosenstein, M. and Furnas, G.: 1995, 'Recommending and evaluating choices in a virtual community of use'. In: CHI '95: Conference Proceedings on Human Factors in Computing Systems, Denver, CO, pp. 194-201.

Kitts.B, David Freed, and Martin Vrieze. Cross-sell: A fast promotion-tunable customer-item recommendation method based on conditional independent probabilities. In Proceedings of ACM SIGKDD International Conference, pages 437–446, 2000.

Konstan.J, B. Miller, D. Maltz, J. Herlocker, L. Gordon, and J. Riedl. GroupLens: Applying collaborative filtering to Usenet news. Communications of the ACM, 40(3):77–87, 1997.

McJones, P. (1997). EachMovie collaborative filtering data set. DEC Systems Research Center. http://www.research.digital.com/SRC/eachmovie/.

Mobasher.B, H. Dai, T. Luo and M. Nakagawa. Improving the effectiveness of collaborative filtering on anonymous Web usage data. In Proceedings of the IJCAI 2001 Workshop on Intelligent Techniques for Web Personalization (ITWP01), August 2001, Seattle.

Mild.A, Martin Natter, A critical view on recommendation systems, Working Paper Series ,Vienna University of Economics and Business Administration, Austria, http://www.wu-wien.ac.at/am

Pazzani, M. J.: 1999, 'A Framework for Collaborative, Content-Based and Demographic Filtering'. Artificial Intelligence Review, 13 (5/6), 393-408.

Pazzani M., and Billsus, D. (1997). Learning and Revising User Profiles: The identification of interesting web sites. Machine Learning 27, 313-331.

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P. and Riedl, J.: 1994, 'GroupLens: An Open Architecture for Collaborative Filtering of Netnews'. In: Proceedings of the Conference on Computer Supported Cooperative Work, Chapel Hill, NC, pp. 175-186.

Resnick and Varian. Recommender systems. Communications of the ACM, 40(3):56¨C58, 1997.

Sarwar, B. M., Karypis, G., Konstan, J. A., and Riedl, J. (2000). Analysis of Recommendation Algorithms for E-Commerce. In Proceedings of the ACM EC'00 Conference. Minneapolis, MN. pp. 158-167

Schafer.J.B, Joseph A. Konstan, John Riedl, E-Commerce Recommendation Applications, University of Minnesota,2001

Schafer, J. B., Konstan, J. and Riedl, J.: 1999, 'Recommender Systems in E-Commerce'. In: EC '99: Proceedings of the First ACM Conference on Electronic Commerce, Denver, CO, pp. 158-166.

Shardanand.U, P. Maes. Social information filtering: algorithms for automating "word of mouth." In Proceedings of the ACM CHI Conference (CHI95), 1995.

Terveen.L, Will Hill, Brian Amento, David McDonald, and Josh Creter. PHOAKS: A system for sharing recommendations. Communications of the ACM, 40(3):59–62, 1997.

Yu L., Liu. L, Personalized Recommendation in E-Commerce and Its Application in China, to appear in *The Seventh International Conference on Industrial Management*, 2004.

Yu L., Liu L, Survey on Personalized Recommendation in E-Commerce, to appear in *Computer Integrated Manufacture System* (In Chinese), 2004.

Yu L., Liu. L, Luo Z.H. Comparison and Analysis on E-Commence Recommendation Method in China, to appear in *Systems Engineering——Theory & Practice* (In Chinese), 2004.

Yu L., Liu. L, Li X. F., Combining Item-based and User-based Collaborative Filtering for Multi-Interests Recommendation in E-Commerce, Submitted to *Knowledge and Information System*, 2004.

Yu L., Liu L, Personalized Recommendation Algorithm for Multiple Interests, to appear in *Computer Integrated Manufacture System* (In Chinese), 2004.

Yu. K, X. Xu, M. Ester, H.-P. Kriegel: Feature Weighting and Instance Selection for Collaborative Filtering: An Information-Theoretic Approach, Knowledge and Information Systems (KAIS), Springer-Verlag, Vol. 5, No. 2, April, 2003.

Yu. K, A. Schwaighofer, V. Tresp, W.-Y. Ma, H.J. Zhang, Collaborative Ensemble Learning: Combining Collaborative and Content-Based Information Filtering via Hierarchical Bayes, to appear in Proceedings of 19th International Conference on Uncertainty in Artificial Intelligence (UAI'03), Acapulco, Mexico, August 7-10, 2003.