International Journal of Web Information Systems

Improving recommender systems' performance on cold-start users and controversial items by a new similarity model

Masoud Mansoury Mehdi Shajari

Received 31 July 2015
Revised 20 November 2015
21 December 2015
Accepted 21 December 2015

Improving recommender systems' performance on cold-start users and controversial items by a new similarity model

Masoud Mansoury and Mehdi Shajari
Department of Computer Engineering and Information Technology, Amirkabir University of Technology, Tehran, Iran

Abstract

Purpose – This paper aims to improve the recommendations performance for cold-start users and controversial items. Collaborative filtering (CF) generates recommendations on the basis of similarity between users. It uses the opinions of similar users to generate the recommendation for an active user. As a similarity model or a neighbor selection function is the key element for effectiveness of CF, many variations of CF are proposed. However, these methods are not very effective, especially for users who provide few ratings (i.e. cold-start users).

Design/methodology/approach – A new user similarity model is proposed that focuses on improving recommendations performance for cold-start users and controversial items. To show the validity of the authors' similarity model, they conducted some experiments and showed the effectiveness of this model in calculating similarity values between users even when only few ratings are available. In addition, the authors applied their user similarity model to a recommender system and analyzed its results.

Findings – Experiments on two real-world data sets are implemented and compared with some other CF techniques. The results show that the authors' approach outperforms previous CF techniques in coverage metric while preserves accuracy for cold-start users and controversial items.

Originality/value – In the proposed approach, the conditions in which CF is unable to generate accurate recommendations are addressed. These conditions affect CF performance adversely, especially in the cold-start users' condition. The authors show that their similarity model overcomes CF weaknesses effectively and improve its performance even in the cold users' condition.

Keywords Communities on the Web, Web search and information extraction, Web-commerce and E-business

Paper type Research paper

1. Introduction

Development of e-commerce has led to behavioral changes in traditional businesses where users increasingly tend to buy products via the internet. However, the proliferation of information by the internet companies has caused information overload that leads to a decline in customer satisfaction. One way to deal with this problem is to create recommender systems that extract information about products which are desired by each customer. A recommender system uses knowledge discovery techniques to solve the problem of recommendation to a user during the

International Journal of Web Information Systems
Vol. 12 No. 2, 2016
pp. 126-149

1744-0084
DOI 10.1108/IJWIS-07-2015-0024

purchase phase. There are many e-commerce websites that utilize the advantages of recommender systems to increase their profits, such as the recommendation of books in Amazon (Brynjolfsson *et al.*, 2003), photo groups in Flickr (Zheng *et al.*, 2010), videos in YouTube (Baluja *et al.*, 2008) and results in the Web search (Zhang and Li, 2008).

A recommender system is a new and powerful system to extract information from the customers' database, which is collected from customers' purchase behavior and has led to the growth of businesses. These systems help the customers to buy their desired items (2005). Hence, a recommender system has become a vital tool in e-commerce. In other words, a recommender system tries to guess the user's way of thinking to identify the closest products to his/her taste. These systems simulate and automatically run the same process that we use in our everyday lives. This is the process by which we try to find people close to our interests and ask them about our choices. There are many types of recommendation methods, such as collaborative filtering technique (Breese *et al.*, 1998), content-based technique (Pazzani and Billsus, 2007), social recommendation (Huang *et al.*, 2010), semantic recommendation (Lu *et al.*, 2010) and so on.

The collaborative filtering works based on a similarity model to generate a recommendation. It tries to find similar users to the active user and uses their opinions as recommendation partners. However, collaborative filtering (CF) has some weaknesses that deteriorate its performance and effectiveness. Shilling attacks are one of the problems that affect the CF to generate biased recommendations. An attacker tries to bold a specific item by injecting faked profiles or ratings in the system (Lam and Riedl, 2004; Burke *et al.*, 2006; Mobasher *et al.*, 2007; Williams *et al.*, 2007; Mehta and Nejdl, 2008). Scalability is another source of problems that makes it difficult, in some conditions, to generate recommendations in real-time, especially when user-item matrix is large. It means that the computing time will grow exponentially with the increase of users and items (Xue *et al.*, 2005; Bell and Koren, 2007). Another problem related to CF is coverage rate which is the portion of items ratings that can be predicted by CF for all users. In some conditions, CF cannot predict ratings for the user-item sets due to inability in finding similar users to the active user, especially in the cold user condition. The last but not the least problem is accuracy of the system in predicting ratings. Accurate predictions are the main goal in a recommender system because low accuracy in rating predictions may lead to wrong recommendations.

According to the mentioned problems for CF, this paper focuses on how to improve accuracy and coverage metrics simultaneously. We believe that in e-commerce systems, users are linked to each other based on their preferences. Moreover, users with similar preferences are linked to each other with higher weight than users who have less similar preferences. Therefore, we use a distance function to determine the similar users (i.e. connected users) and their similarity values (i.e. weights). This function works based on (dis)agreement on commonly rated items between a pair of users. It means that the higher the degree of (dis)agreement between users, more is the (dis)similarity. This function is designed to better utilize the ratings in cold-start users and controversial items conditions as worst-case ones for coverage rate and accuracy metrics, respectively. Therefore, it ensures the overall improvement in recommendation performance and overcomes traditional CF problems.

The contributions of this paper are mainly in two-fold. First, we conduct experiments to verify the validity of our similarity model. We compare the outputs of our model with

Pearson (Resnick *et al.*, 1994) in calculating similarity values on coverage rate and accuracy metrics. Experimental results show the superiority of our model, especially in cold user conditions. Second, we propose a novel CF method which can effectively overcome the weaknesses of traditional CF method. This is achieved by applying our similarity model in the traditional CF algorithm. Experiments are performed on three sets of data: cold-start users, controversial items and general conditions. First two sets are the main source of concern for CF because existing CF systems are not able to produce accurate results in these situations. Therefore, we use these (i.e. cold-start users and controversial items) worst-case data to show the superiority of our model, and also we use general condition to show the overall improvement on CF by our model. Experimental results on two real-world data sets show that our proposed model is superior to the traditional CF and trust-aware recommender system (TARS). The specifications of these data sets are shown in Table I.

2. Background

CF is the most widely used technique for generating personalized recommendations. However, CF suffers from a few problems, for instance, scalability, cold-start problem, data sparsity, coverage problem, and so on. To overcome these problems, different researches are conducted. Liu *et al.* (2014) proposed a new similarity model for CF named proximity, significance and singularity (PSS) model. The basic idea behind this model was from proximity, impact and popularity (PIP) model (Ahn, 2008) and aimed to enhance the PIP to gain better accuracy. The PSS model used Jaccard (Koutrika *et al.*, 2009) idea by considering the proportion of co-rated items along with the absolute values of ratings. It means that PSS differentiated between similarity values for users who have similar rating values on commonly rated items but different in number. The PSS, also, considered both local and global information to calculate similarity values, and finally, the PSS normalized the similarity values to be combined with other similarity models easily. O'Donovan and Smyth (2005) presented a method to calculate the global trust of each user implicitly. The global trust, then, is combined with the similarity metric by a harmonic formula. They showed that the accuracy of the CF will be improved this way. Although these works improved the accuracy metric, improving accuracy may ruin the coverage. There is a trade-off between accuracy and coverage metrics as ignoring one of them can boost the other one significantly. In our model, we consider both accuracy and coverage metrics to improve the CF performance thoroughly.

TARS is another widely used method to improve the recommend performance in CF. TARS uses trust relations between users rather than similarity. Massa and Avesani

Table I.
Summary of the data sets

Data set	Description	Profile	Trust data	Evaluating similarity model	Evaluating CF model	Availability
Epinions	Ratings of items on the scale 1-5	49,290 users, 139,738 items and 664,824 ratings	Yes	Yes	Yes	www.trustlet.org
MovieLens	Ratings of movies in scale 1-5	943 users, 1,682 movies and 100,000 ratings	No	No	Yes	www.grouplens.org

(2004); Massa and Bhattacharjee (2004); Avesani *et al.* (2005); Massa and Avesani (2007) for first time showed the applicability of using trust instead of similarity in traditional CF. they showed that trust can enhance the CF performance in special condition where CF cannot work well. They conducted their experiments on different groups of users and items. Cold-start users and controversial items were the most important groups that they used for experiments. They showed that TARS can improve the results for these groups of data. The main problem related to the cold-start users is lack of enough commonly rated items between users and traditional CF cannot find similar users to the active user, but trust alleviates this issue by using propagation capability of trust. Propagation helps to estimate trust values between two unfamiliar users who do not know each other. In their work, Messa and Avesani only considered positive trust ratings, but (Victor *et al.*, 2006, 2009, 2011) proposed that distrust ratings play an important role as well and can improve the performance of trust-aware recommender systems. Although TARS improve the CF performance by using trust relations, it cannot be applied to all e-commerce systems because it needs additional information (i.e. trust data) that does not exist in most of systems. In our model, in contrast, we do not need additional information as TARS needs, and we show the superiority of our model in comparison to TARS with the least information that exists in all e-commerce system.

Nazemian *et al.* (2012) proposed a model that reconstructs trust networks by removing the trust relations between users when similarity values between those users fall below a certain threshold. They showed that using the trust statements between two users who have low similarity adversely impacts the prediction quality, and removing these trust relations from prediction process will improve the prediction accuracy. Yaun *et al.* (Yuan *et al.*, 2010a) presented a new TARS model which used small-worldness of trust networks to improve the conventional TARS performance. By using small-world topology of the trust network, they determined the optimized maximum trust propagation distance (MTPD) that not only improved the CF performance but also decreased the computational complexity. In another research, Yaun *et al.* (Yuan *et al.*, 2010) introduced the implicit trust-aware recommender system (iTARS) and used implicit trust driven from the similarity values between users. As user similarity suffers from sparseness of ratings matrix, iTARS used the advantage of the transitivity of trust to overcome this shortcoming. This work showed that the iTARS approach outperforms explicit trust-aware recommender system (eTARS) approach of Massa on accuracy metric; however, it was outperformed by eTARS on coverage metric. Although these researches addressed accuracy and coverage metrics as well, their experiments were not conducted on cold-start users and controversial items as worst-case groups of data for CF. In our model, we conduct all of experiments on cold-start users and controversial items to show the real improvement in performance.

3. The drawbacks of existing similarity models

In this section, we compare the output of our model with several state-of-the-art similarity models by an example provided in Liu *et al.* (2014). Liu *et al.* (2014) provides an example to show the main drawbacks of other similarity models and the superiority of their own model. They extracted five drawbacks related to the other models that their model overcome. We will use the same example to calculate the similarity values calculated by our similarity model and to discuss the superiority of our model over the others. For this purpose, according to Table II, first, we extract some of the obvious rules that a similarity model should regard. Then, we calculate the similarity values by each of the similarity models. Finally, according to the extracted rules and calculated similarity values, we create a list of rules for each similarity model which are disregarded. As a result, the best similarity model would be the one that has fewer members in the disregarded rules list.

Table II shows the example user-item matrix provided in Liu *et al.* (2014). There are four items and five users. Dash symbols in the table represent the missing values. Rules extracted from Table II are listed in Table III. Moreover, the calculated similarity values are shown in Figure 1. In Table III, sim(A,B) means the similarity value between user A and user B. To interpret Table III, for instance, Rule 1 says that similarity value between User 1 and User 3 should be high because based on ratings provided by these users in Table II, they have similar ratings on their commonly rated items. And all other rules can be interpreted in a similar fashion. In Figure 1, the values which are calculated by our similarity model and other state-of-the-art similarity models including Pearson correlation coefficient (PCC) (Resnick *et al.*, 1994), constrained Pearson correlation coefficient (CPCC) (Shardanand and Maes, 1995), sigmoid function based Pearson correlation coefficient (SPCC) (Jamali and Ester, 2009), cosine (COS) (Adomavicius and Tuzhilin, 2005), adjusted cosine metric (ACOS) (Ahn, 2008), mean squared difference (MSD) (Cacheda *et al.*, 2011), Jaccard (Koutrika *et al.*, 2009), Jaccard and MSD (JMSD) (Zheng *et al.*, 2010), proximity-impact-popularity (PIP) (Ahn, 2008) and the new heuristic similarity model (NHSM) (Liu *et al.*, 2014) are presented.

As shown in Table IV, although the previously mentioned similarity models calculate similarity values accurately, they have at least one member in their disregarded rules list. It means they have problems in some conditions which affect their accuracy adversely. However, our similarity model regards all rules without any fault. In fact, this example is a sample of customer's behavior in a real world, and our similarity model shows high accuracy with this sample. In the following, we will prove this claim with several experiments.

Table II. An example of the user-item matrix. The missing ratings are represented by the dash symbol

Users	Item 1	Item 2	Item 3	Item 4
User 1	4	3	5	4
User 2	5	3	–	–
User 3	4	3	3	4
User 4	2	1	–	–
User 5	4	2	–	–

Table III. Some of the obvious rules extracted from Table II which a similarity model should regard

Rule no.	
Rule 1	Sim(U1,U3) should be high
Rule 2	Sim(U2,U4) should be low
Rule 3	Sim(U1,U3) should be higher than or equal to Sim(U1,U2)
Rule 4	Sim(U4,U5) should be higher than or equal to Sim(U2,U4)
Rule 5	Sim(U4,U5) should be low
Rule 6	Sim(U1,U2) should be high
Rule 7	Sim(U3,U5) should be higher than or equal to Sim(U4,U5)

$$\begin{array}{cc} & \begin{array}{cccc} u2 & u3 & u4 & u5 \end{array} \\ \begin{array}{c} u1 \\ u2 \\ u3 \\ u4 \end{array} & \begin{bmatrix} 0.707 & 0.0 & 0.707 & 0.707 \\ & 1.0 & 1.0 & 1.0 \\ & & 1.0 & 1.0 \\ & & & 1.0 \end{bmatrix} \end{array}$$

(a) PCC

$$\begin{array}{cc} & \begin{array}{cccc} u2 & u3 & u4 & u5 \end{array} \\ \begin{array}{c} u1 \\ u2 \\ u3 \\ u4 \end{array} & \begin{bmatrix} 1.0 & 0.577 & -0.447 & 0.707 \\ & 1.0 & -0.447 & 0.707 \\ & & -0.447 & 0.707 \\ & & & 0.316 \end{bmatrix} \end{array}$$

(b) CPCC

$$\begin{array}{cc} & \begin{array}{cccc} u2 & u3 & u4 & u5 \end{array} \\ \begin{array}{c} u1 \\ u2 \\ u3 \\ u4 \end{array} & \begin{bmatrix} 0.517 & 0.0 & 0.517 & 0.517 \\ & 0.731 & 0.731 & 0.731 \\ & & 0.731 & 0.731 \\ & & & 0.731 \end{bmatrix} \end{array}$$

(c) SPCC

$$\begin{array}{cc} & \begin{array}{cccc} u2 & u3 & u4 & u5 \end{array} \\ \begin{array}{c} u1 \\ u2 \\ u3 \\ u4 \end{array} & \begin{bmatrix} 0.612 & 0.975 & 0.606 & 0.605 \\ & 0.703 & 0.997 & 0.997 \\ & & 0.696 & 0.695 \\ & & & 1.0 \end{bmatrix} \end{array}$$

(d) COS

$$\begin{array}{cc} & \begin{array}{cccc} u2 & u3 & u4 & u5 \end{array} \\ \begin{array}{c} u1 \\ u2 \\ u3 \\ u4 \end{array} & \begin{bmatrix} -0.363 & 0.0 & -0.316 & -0.316 \\ & 0.171 & 0.077 & 0.077 \\ & & 0.224 & 0.224 \\ & & & 0.1 \end{bmatrix} \end{array}$$

(e) ACOS

$$\begin{array}{cc} & \begin{array}{cccc} u2 & u3 & u4 & u5 \end{array} \\ \begin{array}{c} u1 \\ u2 \\ u3 \\ u4 \end{array} & \begin{bmatrix} 0.98 & 0.96 & 0.84 & 0.98 \\ & 0.98 & 0.74 & 0.96 \\ & & 0.84 & 0.98 \\ & & & 0.9 \end{bmatrix} \end{array}$$

(f) MSD

$$\begin{array}{cc} & \begin{array}{cccc} u2 & u3 & u4 & u5 \end{array} \\ \begin{array}{c} u1 \\ u2 \\ u3 \\ u4 \end{array} & \begin{bmatrix} 0.5 & 1.0 & 0.5 & 0.5 \\ & 0.5 & 1.0 & 1.0 \\ & & 0.5 & 0.5 \\ & & & 1.0 \end{bmatrix} \end{array}$$

(g) Jaccard

$$\begin{array}{cc} & \begin{array}{cccc} u2 & u3 & u4 & u5 \end{array} \\ \begin{array}{c} u1 \\ u2 \\ u3 \\ u4 \end{array} & \begin{bmatrix} 0.49 & 0.96 & 0.42 & 0.49 \\ & 0.49 & 0.74 & 0.96 \\ & & 0.42 & 0.49 \\ & & & 0.90 \end{bmatrix} \end{array}$$

(h) JMSD

$$\begin{array}{cc} & \begin{array}{cccc} u2 & u3 & u4 & u5 \end{array} \\ \begin{array}{c} u1 \\ u2 \\ u3 \\ u4 \end{array} & \begin{bmatrix} 0.743 & 1.0 & 0.167 & 0.506 \\ & 0.743 & 0.162 & 0.763 \\ & & 0.167 & 0.506 \\ \tau & & & 0.767 \end{bmatrix} \end{array}$$

(i) PIP

$$\begin{array}{cc} & \begin{array}{cccc} u2 & u3 & u4 & u5 \end{array} \\ \begin{array}{c} u1 \\ u2 \\ u3 \\ u4 \end{array} & \begin{bmatrix} 0.020 & 0.055 & 0.005 & 0.024 \\ & 0.018 & 0.005 & 0.035 \\ & & 0.006 & 0.025 \\ & & & 0.015 \end{bmatrix} \end{array}$$

(j) NHSM

$$\begin{array}{cc} & \begin{array}{cccc} u2 & u3 & u4 & u5 \end{array} \\ \begin{array}{c} u1 \\ u2 \\ u3 \\ u4 \end{array} & \begin{bmatrix} 0.75 & 0.75 & 0.0 & 0.75 \\ & 0.75 & -0.25 & 0.5 \\ & & 0.0 & 0.75 \\ & & & 0.25 \end{bmatrix} \end{array}$$

(k) Our similarity model

Source: Liu *et al.* (2014)

Figure 1.
The user similarity matrix in Table II, according to all kinds of similarity models

Similarity model	Disregarded rules
PCC	Rules 1, 2, 3 and 5
CPCC	Rule 3
SPCC	Rules 1, 2 ,3, 5 and 7
COS	Rules 2, 5 and 7
ACOS	Rules 1, 5, 6 and 7
MSD	Rules 2, 3 and 5
Jaccard	Rules 2, 5 and 7
JMSD	Rules 2, 5 and 7
PIP	Rules 5 and 7
NHSM	Rules 1 and 6
Our similarity model	–

Table IV.
Disregarded rules list

4. The proposed similarity model

In this section, we explain an approach to calculate similarity values between users from users' ratings data. We use (dis)agreement on commonly rated items to calculate similarity values. (Dis)Agreements are calculated by a distance function which determines the distance of two users based on their preferences. Our definition of agreement is nearness of ratings given to items by users. We believe that when two users give very similar ratings to items, they think similar to each other.

4.1 Formalization of the new similarity model

In this section, we give the mathematic formalization of the proposed similarity model. To predict a similarity value between two users, first we get commonly rated items between them and then based on their given ratings to these items, we calculate their disagreements. Using disagreement metric, we can specify the degree of similarity between a pair of users (i.e. low disagreement shows similarity, while high disagreement shows dissimilarity). The user similarity can be calculated as follows:

$$Sim(i,j) = \frac{\alpha - D(i,j)}{\alpha} \tag{1}$$

where α is a constant value to adjust the value of $Sim(i, j)$ to a standard interval to be applicable in CF. The value of α depends on the value of $D(i, j)$. We call this factor as normalization factor. $D(i,j)$ is the disagreement degree of users i and j. To calculate disagreement degree of two users, we introduce the average difference between commonly rated items (ADCRI). If a user has a low ADCRI with a target user, he will be similar to the target user, and if a user has a high ADCRI with a target user, he will be dissimilar to the target user. The disagreement is calculated as follows:

$$D(i,j) = \frac{\sum_{k \in I_{ij}} |r_{ik} - r_{jk}|}{N_{ij}} \tag{2}$$

where r_{ik} is the rating that user i have assigned to item k, I_{ij} is defined as the set of common items that have been rated by users i and j, N_{ij} stands for number of members in I_{ij} set and $D(i, j)$ is the disagreement degree of users i and j. As $r_{ik} \in [R_{low} .. R_{\max}]$, $D(i, j)$ would be in $[0 .. R_{\max} - 1]$, where 0 shows the minimum disagreement or high similarity degree and $R_{\max} - 1$ shows the maximum disagreement or low similarity degree. For example, in Epinions and MovieLens websites, $R_{low} = 1$ and $R_{\max} = 5$, therefore, $D(i,j)$ would be in $[0 .. 4]$ and to achieve similarity values in the interval of $[-1, 1]$, normalization factor would be equal to two. Subsequently, as agreement and disagreement are complementary, the similarity equation in equation (1) can be also rewritten using the agreement concept.

4.2 Experimental verification of our similarity model

In this section, we experimentally verify the validity of our similarity model by data extracted from a real application.

4.2.1 Experimental setup. As we discussed, to calculate similarity value between two users, first we calculate agreement/disagreement between those users and then calculate the similarity value using agreement/disagreement value. However, calculating similarity value is impossible for those users who do not have any commonly rated items.

To verify the validity of our similarity model, we use existing trust relation between users as existence of similarity between those users. In many e-commerce websites, trust statement is strongly based on users' reviews and ratings given to items and how much his/her reviews and ratings have been helpful in the past. Naturally, if a user is satisfied with an item, he/she will give a high rate and will write a good feedback about that item. Therefore, rating of items can be used to specify the degree a user is satisfied with a specific item. Furthermore, a user decides to state trust based on other users' reviews and ratings. If reviews and ratings are helpful in successful selection, then the writer of the review or rater of the item will be trusted. Therefore, when trust relation exists between two users, it shows that they have had similar opinion on their commonly rated items. In addition, Abdul-Rahman and Hailes (2000) have experimentally shown the correlation between similarity and trust in an online community named Allconsuming.net.

In the Epinions data set, two separate files exist: trust ratings and item ratings files. Trust ratings file has three fields including trustor, trustee and trust rate. Each record in this file represents a trust relation that a trustor has stated to a trustee. On the other hand, items ratings file has three fields including user, item and item rate. Each record in this file represents the rating given to an item by a user. We used items ratings to calculate the similarity values between each trustor and trustee by our similarity model. To determine the accuracy of our model, existing trust statements in the trust ratings file are used to obtain the number of correctly calculated values. In fact, a calculated similarity value between two users is considered as a correct one when it satisfies one of these conditions:

- If there is trust relation between two users, the calculated similarity value between those users should be more than zero.
- If there is distrust relation between two users, the calculated similarity value between those users should be less than zero.
- If there is not any relation between two users, the calculated similarity value between those users should be equal to zero.

Additionally, the performance of the Pearson model is investigated in the same way. Finally, the performance of our model is compared with the Pearson model.

To show the performance of our similarity model, we use the following equation:

$$precision = \frac{C}{T} \tag{3}$$

where C is the number of correct predictions and T is the total number of calculated similarity values. The calculated similarity value will be more accurate if two users have more commonly rated items.

4.2.2 Experimental results for the proposed similarity model. We evaluate our similarity and Pearson models on Epinions data set based on proposed evaluation criteria in subsection 4.2.1.

As mentioned in the previous section, we used each record in trust ratings file as test case and applied our similarity model on it. In fact, based on this file, we know that each two users in a record are trusted or similar, and an accurate similarity model should consider them as similar one. According to the trust ratings file, we extract some groups of data as follows:

- Users who have only one commonly rated items (i.e. extracting records of file in which its users have only one commonly rated items), and users who have 2-20 commonly rated items.
- Users who have less than or equal to 2-20 commonly rated items. To interpret, consider the group in which users have less than or equal to five commonly rated items. It means that users with 0, 1, 2, 3, 4 or 5 commonly rated items are considered.

Obviously, the groups of data with few commonly rated items are the worst-case for any similarity model because there is the least information for calculating the similarity value. In contrast, the groups of data with high commonly rated items are the best-case for any similarity model. Thus, we can be sure that experiments are comprehensive, and the results will represent the real performance of similarity models. The number of users in each group is briefly shown in Table V. First, we verify that our similarity model has high coverage rate in calculating similarity values. Coverage rate is the per cent of records that a similarity model is able to predict a similarity value for it, and it is calculated as follows:

$$Coverage\ rate = \frac{Number\ of\ predicted\ records}{Total\ number\ of\ records} \times 100 \tag{4}$$

High coverage rate in calculating similarity values would result in higher coverage rate of the CF.

As shown in Table V, for records in which two users have only one commonly rated items, Pearson cannot calculate any value, however, our similarity model can calculate

No. of commonly rated items	Total no. of records	Our similarity model: No. of predicted records	Our similarity model: Coverage rate (%)	Pearson model: No. of predicted records	Pearson model: Coverage rate (%)
0	310,843	0	0	0	0
1	85,511	85,511	100	0	0
5	6,614	6,614	100	4,981	99.98
10	1,212	1,212	100	1,005	100
15	400	400	100	347	100
20	172	172	100	158	100
Less than or equal to 2 items	432,654	121,811	28.15	16,330	3.77
Less than or equal to 5 items	468,495	157,652	33.65	44,601	9.52
Less than or equal to 10 items	480,851	170,008	35.35	56,427	11.73
Less than or equal to 15 items	483,932	173,089	35.77	87,121	18
Less than or equal to 20 items	485,077	174,234	35.92	88,266	18.2

Table V. Coverage rate of similarity models for different number of commonly rated items similarity value for all pair of users in the record, even when there is only one commonly rated item between users. This results in higher coverage rate. It means that our similarity model can calculate similarity value for any record with at least one commonly rated item. However, Pearson can achieve complete coverage rate only with heavy raters. For users who have 1 and 5 commonly rated items, for example, it achieves 0 and 99.98 per cent coverage rate, respectively. Furthermore, we perform experiments on groups with varied number of commonly rated items. For users who have less than or equal to 2, 5, 10, 15 and 20 commonly rated items, our similarity model works well by achieving 28.15, 33.65, 33.35, 35.77 and 35.92 per cent (against 3.77, 9.52, 11.73, 18 and 18.2 per cent for Pearson) coverage rate, respectively. Figures 2(a) and (b) show the trend of coverage rate with different number of commonly rated items. Figure 2(a) shows the coverage rate for users who have 1-20 commonly rated items, and Figure 2(b) shows the coverage rate for users who have less than or equal to 2-20 commonly rated items.

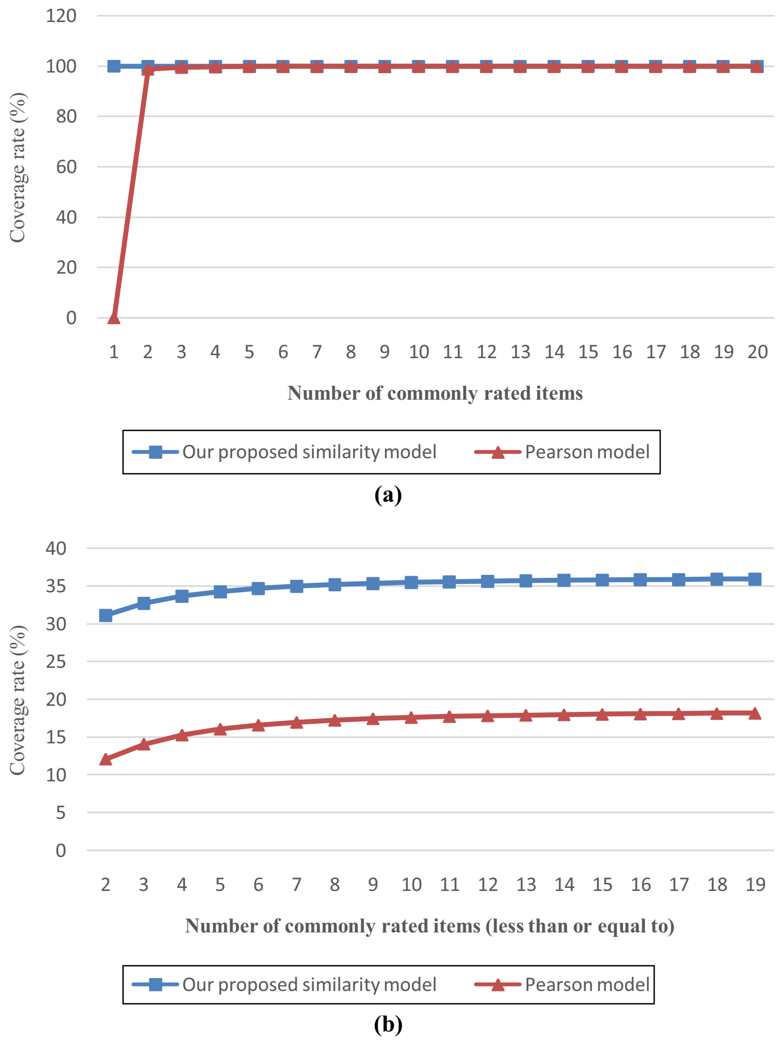
Second, we verify that our similarity model has high accuracy in calculating similarity values. Coverage rate and accuracy have indirect relation, which means that improving one of them may affect the other one adversely. Thus, an effective similarity model should consider both of them simultaneously. Table VI shows the accuracy of Pearson and our similarity models in calculating similarity values.

According to Table VI, for users who only have one commonly rated item, our similarity model works well by achieving 76.55 per cent accuracy (against 0 per cent accuracy for Pearson). Moreover, for users who have less than two commonly rated items, our similarity model achieves 78.58 per cent accuracy (against 71.42 per cent accuracy for Pearson). These two groups of data are the source of concern for traditional CF. When a similarity model works well with them, it will overcome CF's problem. For user who have 5, 10, 15 and 20 commonly rated items, our similarity model achieves 94.74, 98.68, 99.5 and 100 per cent accuracy (against 75.32, 82.92, 86.75 and 91.86 per cent accuracy for Pearson), respectively. Additionally, for users who have less than 5, 10, 15 and 20 commonly rated items, our similarity model achieves 81.65, 82.81, 83.1 and 83.21 per cent accuracy (against 69.82, 71.29, 71.81 and 72.05 per cent accuracy for Pearson), respectively.

Figure 3(a) shows the accuracy of our model and Pearson in calculating similarity values when the number of commonly rated items varies from 1 to 20. It shows that our similarity model calculates similarity values more accurate than Pearson not only for cold users but also for heavy raters. Additionally, Figure 3(b) shows the accuracy of our model and Pearson in calculating similarity values when the number of commonly rated items is less than or equal to 2-20. For instance, when the number of commonly rated items is less than or equal to five, it means that users with 1, 2, 3, 4 or 5 commonly rated items are considered. Similarly, this figure shows the superiority of our model over Pearson in calculating similarity values.

Generally, there are some limitations in Pearson that decline its performance. However, our similarity model overcomes these limitations. Ahn (2008) presented some important limitations related to Pearson as follows:

- Low number of commonly rated items under data sparsity.
- When there is one commonly rated items between two users, Pearson cannot calculate any similarity value.



Notes: (a) Exact number; (b) less than or equal to varied numbers

Figure 2.
Coverage rate for different number of commonly rated items

- When all provided ratings by a user are flat [i.e. all ratings have the same value such as (1, 1 and 1), (2, 2 and 2) or (3, 3 and 3)], Pearson cannot calculate correct similarity value because it always results zero in this condition.
- Pearson sometimes considers very different users as similar one, and vice versa. For example, assume User 1 with ratings of (5, 4 and 5) and User 2 with ratings of (3,5 and 4) on Item 1, Item 2 and Item 3, respectively. These two users seem to be almost similar, but Pearson considers them as dissimilar.

No. of commonly rated items	Our similarity model T	C	Accuracy (%)	Pearson model T	C	Accuracy (%)
0	0	0	0	0	0	0
1	85,511	65,461	76.55	0	0	0
5	6,614	6,267	94.75	6,613	4,981	75.32
10	1,212	1,196	98.68	1,212	1,005	82.92
15	400	398	99.5	400	347	86.75
20	172	172	100	172	158	91.86
Less than or equal to 2 items	121,811	95,723	78.58	16,330	11,664	71.42
Less than or equal to 5 items	157,652	128,730	81.65	44,601	31,142	69.82
Less than or equal to 10 items	170,008	140,784	82.81	56,427	40,227	71.29
Less than or equal to 15 items	173,089	143,841	83.1	87,121	62,564	71.81
Less than or equal to 20 items	174,234	144,985	83.21	88,266	63,593	72.05

Notes: T = number of predicted records; C = number of correctly predicted records

Table VI. Accuracy of similarity models for different number of commonly rated items

5. Improving collaborative filtering using new similarity model

Using our verified similarity model, we propose an enhanced variation of CF model, which improves the traditional model by calculating similarity values with high coverage rate and accuracy. Our proposed model is simple, straightforward and requires little knowledge and computational efforts in comparison to TARS and traditional CF.

5.1 Our proposed collaborative filtering model

The previous works (Massa and Avesani, 2004, 2007) have shown that applying explicit trust ratings instead of similarity model can improve the performance of recommender systems. The main challenge of RSs is generating recommendation for cold-start users who provide few ratings. Due to the lack of commonly rated items, calculating similarity values is impossible for these users. On the other hand, trust-based systems improve the results by the propagation capability of trust. Trust propagation allows estimating the hidden relationship between users by inferring new trust relationships among users. The explicit trust ratings improve the performance of the recommender systems, but these ratings are not always available. Many of the e-commerce websites do not provide the opportunity for their users to give trust ratings to each other. To overcome these problems, we suggest an improved CF model which does not have the problem of TARS and traditional CF.

All of these approaches work based on CF algorithm, but each of them uses a specific similarity function to weigh the recommenders' opinions. TARS uses trust ratings between users, traditional CF uses Pearson and the proposed CF method in this paper uses new similarity model introduced in Section 3. Therefore, similarity function is a key element to determine the performance of CF. In Section 3, we showed that our new similarity model calculates similarity values with high performance. In the rest of this paper, also, we perform further experiments with CF to show the superiority of our similarity model when applied to CF. CF predicts the rating that the active user will give to a desired item using formula (5):

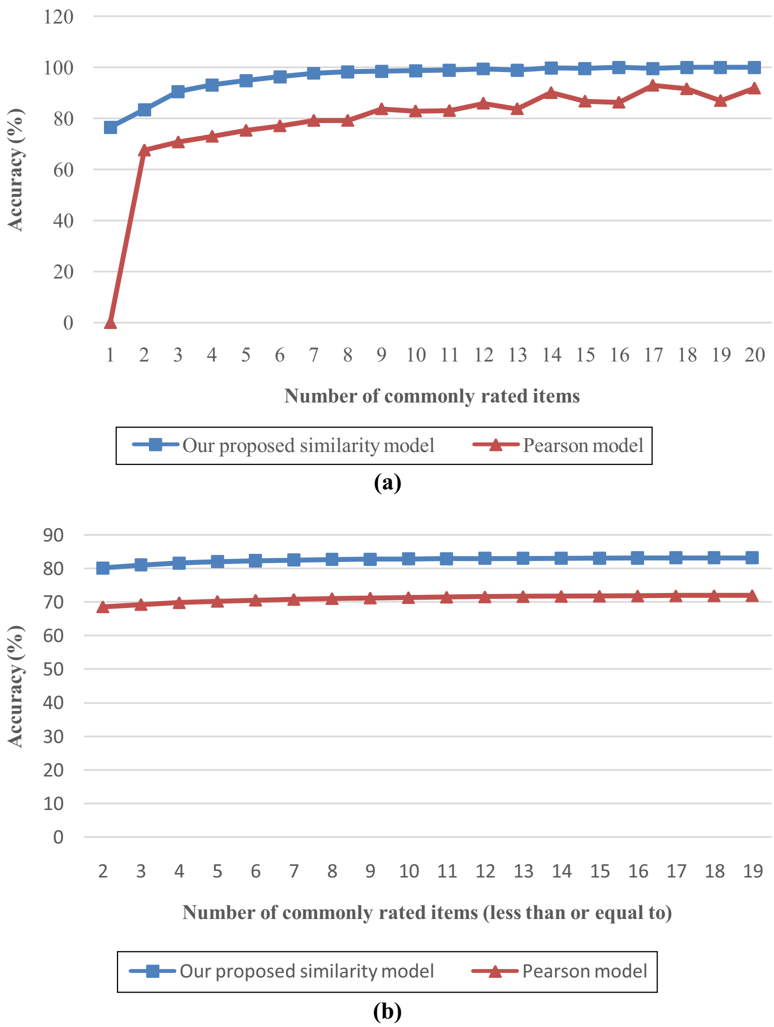


Figure 3.
Accuracy for different number of commonly rated items

Notes: (a) Exact number; (b) less than or equal to varied numbers

$$p_{ai} = \overline{r_a} + \frac{\sum_{u=1}^{k} (r_{ui} - \overline{r_u}) S_{au}}{\sum_{u=1}^{k} S_{au}} \tag{5}$$

where S_{au} is the similarity (or trust) value between user a and user u which is calculated by a similarity function, r_{ui} is the recommender u's recommendation on the item i and k is the number of recommenders.

To show the effectiveness of our CF method, we perform our experiments on cold-start users, controversial items and randomly selected user-item sets (i.e. general condition). Cold-start users are users who have few commonly rated items (i.e. between 1 and 4 ratings) with other users. The few commonly rated items will result in a low coverage rate. Coverage rate is the fraction of items that CF is able to predict rating for them. A good similarity model will increase the coverage rate by finding more users similar to the active user. Improvement of the coverage rate for cold-start users is an indicator of overall coverage rate improvement of the system. Controversial items are items that have ratings with high standard deviation (i.e. greater than 1.5). The high standard deviation will result in a low accuracy. A good similarity model will improve the accuracy by calculating a more accurate similarity value of an active user to other users. Improvement of the accuracy for controversial items is also an indicator of overall accuracy improvement of the system. The last group of data is randomly selected user-item sets, which confirm the performance of CF in general conditions as more common one in real-world and ensures the overall improvement.

To provide a clear analysis, experiments are performed on our CF approach, eTARS1 (direct trust ratings), eTARS2 (combination of direct trust and propagated trust ratings by direct propagation model) and traditional CF using two real data sets, including Epinions and MovieLens. Experimental results show that our CF approach outperforms the other ones.

5.2 Experimental verification of our collaborative filtering model

In this section, we experimentally show the performance of our similarity model by data extracted from the real applications. Also, we compare our model with several well-known researches and show the superiority of our model.

5.2.1 Experimental setup. For experiments, two data sets are used: Epinions and MovieLens. Specifications of data sets are summarized in Table I. We choose these two data sets because they are the most used data sets by researchers in CF domains. To be applicable and comparable, we perform different technique on each data set. On Epinions data set, the technique used for evaluating the CF approaches is based on leave-one-out. Leave-one-out is an offline technique that involves hiding the rating of an active user and trying to predict it. As Epinions data set is so sparse, we extract cold-start users and controversial items from it and perform our experiments on these two groups of data. The first group of data is used to show the coverage rate of CF approaches, and the second one is used to show the accuracy of CF approaches. On the other hand, MovieLens data set is used to show the performance in general conditions as it is common in real-world. MovieLens data set also no longer contains any cold-start user and completely has different conditions with Epinions data set. On MovieLens data set, we perform random selection strategy. This means that we randomly split the data set into two sets of data as train and test sets. Train and test set include 80 and 20 per cent of data, respectively. After splitting, train data is used to predict the ratings of user-items sets in test data. This process is performed 10 times to ensure the validity of outcomes.

To evaluate our CF technique and to compare it with previous works, we use the following four metrics which are used in (Massa and Avesani, 2007):

(1) Mean absolute error (MAE).

(2) Mean absolute user error (MAUE).

(3) Coverage rate.
(4) User coverage rate.

MAE metric calculates the average error between the predicted ratings and the real ratings. MAUE metric calculates the average MAE error for each user. In fact, this metric shows how much each user will be satisfied with the predicted ratings. Accordingly, MAUE will be calculated according to formula (6):

$$MAUE = \frac{\sum_{i \in U} MAE_i}{N_U} \tag{6}$$

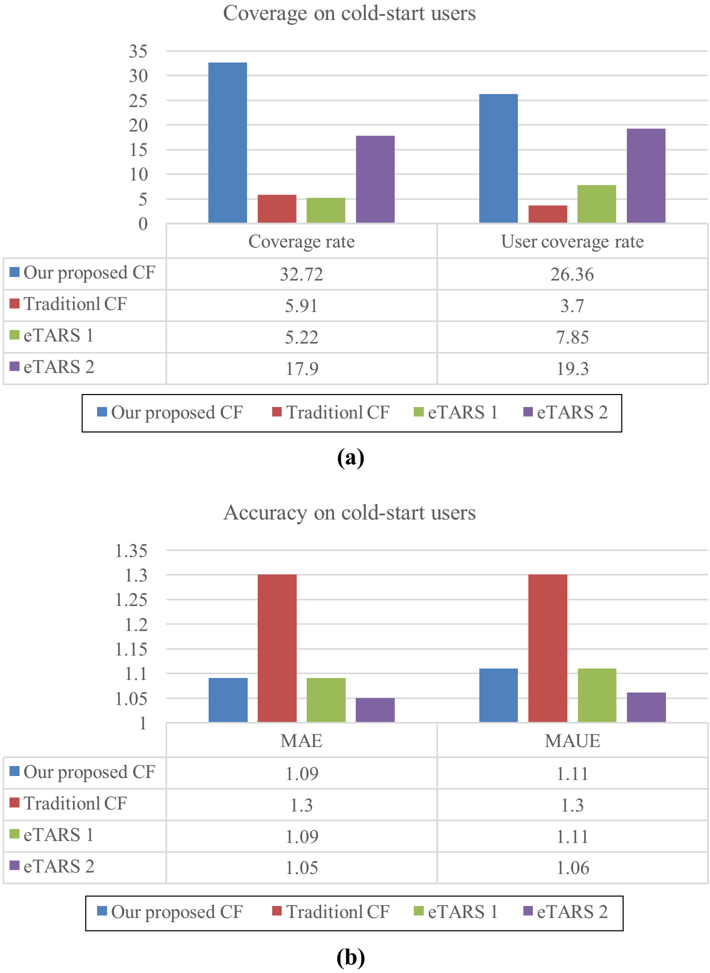
where U is all users that prediction is calculated for them, MAE_i is the MAE error for user i and N_U is the number of U. Coverage metric refers to fraction of items that the system is able to generate a prediction. The user coverage refers to the subset of users for which the CF is able to predict at least one rating for them.

5.2.2 Experimental results for the proposed collaborative filtering model. We examine the performance of our proposed CF model on four aspects including MAE, MAUE, coverage rate and user coverage rate on two publicly available data sets, including Epinions and MovieLens. In the following, we present the experimental results of each data set and discuss the results.

On Epinions data set, experiments are performed on cold-start users and controversial items. Experimental results on cold-start users and controversial items are shown in Figures 4 and 5, respectively. In these figures, our CF approach is compared with eTARS approaches, eTARS1 and eTARS2 (Massa and Avesani, 2007) and traditional CF (Breese *et al.*, 1998). eTARS1 shows direct trust between users without any propagation, and eTARS2 shows the combination of direct trust and direct propagation as proposed in MoleTrust (Avesani *et al.*, 2005).

Figure 4 shows the experimental results on cold-start users of Epinions data set. Figure 4(a) shows the results of coverage metrics for cold-start users. For cold-start users who obviously are the most difficult group to find a proper recommendation (i.e. lower coverage rate), our CF approach achieved 32.72 and 26.36 per cent (against 5.91 and 3.7 per cent of traditional CF, 5.22 and 7.85 per cent of eTARS1 and 17.9 and 19.3 per cent of eTARS2) for coverage and user coverage rate, respectively. Moreover, according to Figure 4(b), which shows the results of accuracy metrics for cold-start users, our CF approach achieved 1.09 and 1.11 (against 1.3 and 1.3 of traditional CF, 1.09 and 1.11 of eTARS1 and 1.05 and 1.06 of eTARS2) for MAE and MAUE, respectively.

Figure 5 shows the experimental results on controversial items of Epinions data set. For controversial items which have high standard deviation, the accuracy of predictions is important for them and is also the most difficult group for accurate predictions. Based on Figure 5(a), which shows the results of coverage metrics for controversial items, our CF approach works well by achieving 74.15 and 71.38 per cent (against 57.75 and 52.25 per cent of traditional CF, 22.71 and 26.85 per cent of eTARS1 and 57.46 and 52.86 per cent of eTARS2) for coverage and user coverage rates, respectively. In addition, according to Figure 5(b), which shows the results of accuracy metrics for controversial items, our CF approach achieved 1.09 and 1.11 (against 1.3 and 1.3 of traditional CF, 1.09 and 1.11 of eTARS1 and 1.05 and 1.06 of eTARS2) for MAE and MAUE, respectively.



Notes: (a) Coverage metric; (b) accuracy metric

Figure 4. Experimental results on cold-start users

According to results in Figures 4 and 5, our CF model outperforms all of previous approaches from the point of coverage rate. Although eTARS2 is the combination of direct trust and direct propagation of trust, our CF model outperforms it without considering any propagation. Moreover, these improvements in coverage rate do not affect the accuracy. According to Figures 4(b) and 5(b), from the point of accuracy, our CF model is as accurate as the other approaches. Regarding to coverage metrics for eTARS, not only explicit trust ratings need extra effort from the users in recording such ratings but also they require the users to have many interactions with each other to build a sense of trust. For this reason, users rarely rate each other explicitly. On the other hand, even in the absence of explicit trust ratings, it is possible for users to have implicit connectivity. For example, assume that user A buys an item X based on the review and rating of user B. Although a degree of trust/

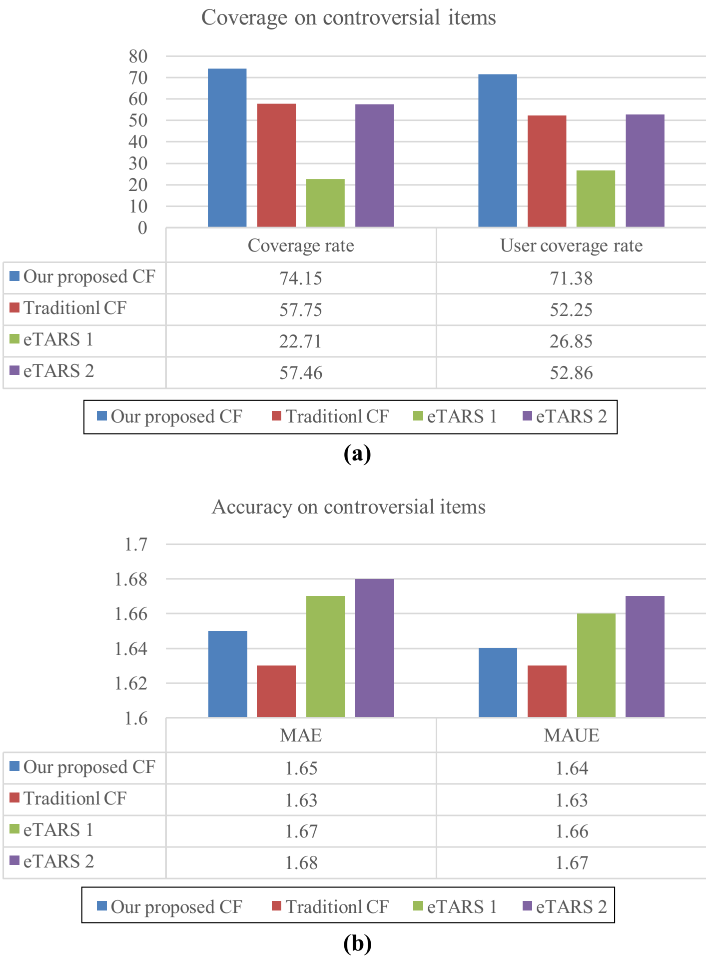


Figure 5.
Experimental results on controversial items

Notes: (a) Coverage metric; (b) accuracy metric

distrust is built between A and B, A needs more interaction with *B* to achieve a full trust/distrust degree. Discovering such connectivity can improve the coverage rates. We used such implicit connectivity between users to calculate their mutual similarity value. This will improve the coverage rates for users who provide few ratings.

Regarding coverage metrics for traditional CF, it suffers from users who have only one rated item because Pearson cannot calculate similarity between these users and the other users. According to Pearson formula in Resnick *et al.* (1994), when a user has only one rated item, the calculation of similarity between this user and the other users is impossible. Because the difference between $r_{u,i}$ and $\overline{r_u}$ will be zero and consequently the output will be undefined. Hong *et al.* (Lee *et al.*, 2007) shed light on the parameters for traditional CF including the number of provided ratings by each user and the total number of ratings for a particular item. They investigated the relationship between these parameters and the traditional CF performance. These parameters have impact on CF performance and should be limited to a minimum threshold. For traditional CF, therefore, we normally ignore users who only provide one rating. Traditional CF cannot predict any ratings for these users, but our CF model predicts ratings for these users as well. In fact, our CF model does not have any weaknesses on users who provide only one rating and is able to predict ratings for all users.

Regarding accuracy metrics for eTARS, in many e-commerce websites (e.g. Epinions), explicit trust ratings are based on only the full trust relationship. It means that when a user expresses his trust to another user, it will be considered as a full trust. For example, if user A has 0.8 of trust to user B, the system considers it as a complete trust (i.e. the value of 1). Using this unreal trust value to predict items ratings will damage the system accuracy. However, our recommender systems use the continuous similarity values (i.e. calculated similarity values) to predict items ratings. Other weakness of eTARS approach is the lack of distrust ratings. We believe that distrust can play important role in improvement of the accuracy of prediction. In our approach, we used the similarity value between users calculated based on our similarity model as discussed in Section 4. Also, we considered dissimilarity values beside the similarity values to preserve the accuracy.

Regarding accuracy metrics for traditional CF, it, as mentioned, suffers from users who provide few ratings. For these users, either system cannot find enough number of recommendation partners, or calculated similarity between active user and his recommendation partners will be inaccurate. As a result, system will generate items ratings with low accuracy. However, our CF model works well even with a sparse data set owing to the fact that our similarity model calculates similarity between a pair of users with high accuracy as shown in Section 3.

On MovieLens data set, experiments are performed on 10 groups of randomly selected user-item sets. In fact, the data set is split into two groups of train and test sets as 80 per cent for train set and 20 per cent for test set. Train set is used to predict ratings for each record in test set. This process is repeated 10 times on different randomly selected train and test sets to ensure the validity of results. As MovieLens data set does not have trust data, experiments are performed by our CF model and traditional CF model. Experimental results on 10 groups of randomly selected user-item sets from MovieLens data set are shown in Figures 6 and 7.

Based on Figures 6(a) and (b), our CF model outperformed traditional CF model in coverage rate in all groups, while both of our CF and traditional CF models achieved complete user coverage rate as 100 per cent. Achieving complete user coverage rate is due to the fact that in MovieLens data set, each user provides enough ratings. Moreover, according to Figure 7(a), our CF model outperformed traditional CF model by achieving lower MAE value than traditional CF model in all groups. Also, as it is shown in Figure 7(b), in the point of MAUE metric, our CF model had lower errors than traditional CF model in all randomly selected groups.

For future researches by other researchers and to be comparable with our CF model, specifications of each group are presented in Tables VII and VIII. Table VII listed the number of users in each group based on their provided ratings in the group. Consider the third row in Table VII, for instance, 10-15 means only users are counted in each group who provide between 10 and 15 ratings. Also, Table VIII listed the number of items in

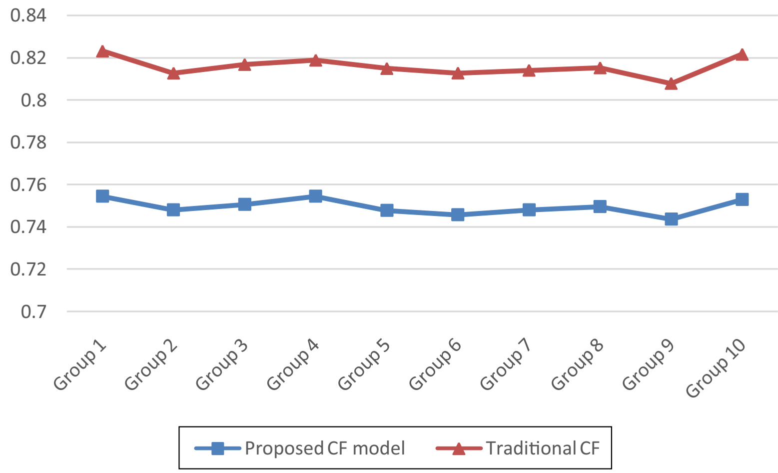
Notes: (a) Coverage rate; (b) user coverage rate

Figure 6.
Experimental results on coverage metric

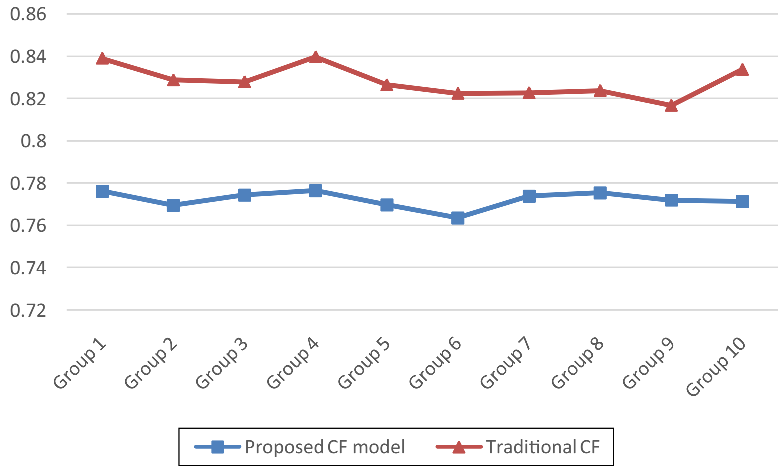
each group based on their standard deviation of ratings in the group. Consider the third row in Table VIII, for instance, [0.5-1) means only items are counted in each group which standard deviation of their ratings are between 0.5 and 1. Bracket ("[") means that 0.5 is considered and parenthesis (")") means 1 is not considered in this interval.

6. Conclusion

In this paper, first we analyze the drawbacks of the existing state-of-the-art similarity models. To overcome these drawbacks, a new similarity model is proposed, which is based on agreement/disagreement on commonly rated items. In fact, this similarity



(a)



(b)

Notes: (a) MAE; (b) MAUE

Figure 7. Experimental results on accuracy metric

model focuses on improving accuracy and coverage as serious problems for traditional CF model. To verify the effectiveness of the proposed similarity model, several experiments are conducted on a popular used data set. From the experimental results, we see that our similarity model is more effective than other models in calculating similarity values. Furthermore, we applied our similarity model in CF algorithm and performed several experiments on two popular used data sets. To show the effectiveness of our model, several experiments were performed on three types of data including cold-start users' data, controversial items data and randomly selected user-item data as follows:

Table VII.
Number of users in each group based on provided ratings

Number of ratings	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7	Group 8	Group 9	Group 10
Less than or equal to 9	0	0	0	0	0	0	0	*1*	0	0
10-15	24	26	23	23	*29*	*29*	25	22	22	21
16-20	117	111	114	120	118	114	*122*	114	113	112
21-25	80	88	90	84	75	83	76	87	92	*93*
26-30	65	65	54	59	*66*	*66*	63	65	57	59
31-35	49	42	*56*	*56*	52	44	49	48	47	54
36-40	47	49	44	44	46	*54*	49	41	45	44
41-45	45	*52*	44	33	44	36	47	46	50	41
46-50	36	29	35	*43*	39	36	29	42	29	41
51-55	31	34	36	34	27	31	35	28	*41*	31
56-60	24	30	23	19	24	27	*31*	25	23	26
61-65	21	15	15	*23*	19	16	14	*23*	21	21
66-70	16	15	21	15	16	*22*	18	12	18	10
71-75	14	12	17	16	10	11	14	13	13	*18*
76-80	16	21	18	21	17	15	12	*23*	21	14
More than 80	358	354	353	353	*361*	359	359	353	351	358

Note: The values of italic data are maximum

Table VIII.
Number of items in each group based on standard deviation of ratings

SD	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7	Group 8	Group 9	Group 10
[0-0.5)	36	30	33	*37*	27	37	35	36	27	34
[0.5-1)	625	626	615	619	624	628	617	*636*	630	621
[1,1.5)	738	771	*780*	758	766	753	776	740	760	764
[1.5,2)	*75*	63	67	*75*	74	71	62	74	69	68
[2,2.5)	*23*	19	18	20	15	16	18	21	18	16
[2.5-3)	*5*	1	2	*5*	3	4	2	3	1	4

Note: The values of italic data are maximum

(1) On cold-start users which are the source of concern from the point of coverage metric, experimental results showed the superiority of our approach on coverage metric while preserving the accuracy compared to previous researches.

(2) On controversial items which are the source of concern from the point of accuracy, our approach approximately achieved the same accuracy as other approaches.

(3) On randomly selected user-item sets, our approach was more effective than traditional CF on both accuracy and coverage metrics, demonstrating the effectiveness of the proposed similarity model on the performance of CF algorithm.

In cold users' situation, which is usual in the real world, traditional CF cannot work well. Experimental results showed the superiority of our model on other CF models. These results demonstrate the effectiveness of the proposed similarity model, and it can improve the performance of the CF algorithm. For future research, considering the subjectivity measure of user ratings in the similarity model can further improve the performance.

References

Abdul-Rahman, A. and Hailes, S. (2000), "Supporting trust in virtual communities", *Proceedings of the 33rd Hawaii International Conference on System Sciences,IEEE Computer Society, Washington, DC*.

Adomavicius, G. and Tuzhilin, A. (2005), "Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 17 No. 6, pp. 734-749.

Ahn, H.J. (2008), "A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem", *Information Sciences*, Vol. 178 No. 1, pp. 37-51.

Avesani, P., Massa, P. and Tiella, R. (2005), "A trust-enhanced recommender system application: moleskiing", *Proceedings of the ACM symposium on Applied Computing,ACM, New York, NY*.

Baluja, S., Seth, R., Sivakumar, D., Jing, Y., Yagnik, J., Kumar, S., Ravichandran, D. and Aly, M. (2008), "Video suggestion and discovery for youtube: taking random walks through the view graph", *Proceedings of the 17th international conference on World Wide Web*, *ACM*, *New York, NY*, pp. 895-904.

Bell, R.M. and Koren, Y. (2007), "Scalable collaborative filtering with jointly derived neighborhood interpolation weights", *Proceedings of the 2007 Seventh IEEE International Conference on Data Mining*, *IEEE Computer Society*, *Washington, DC*, pp. 43-52.

Breese, J.S., Heckerman, D. and Kadie, C. (1998), "Empirical analysis of predictive algorithms for collaborative filtering", *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, *Morgan Kaufmann Publishers*, *San Francisco, CA*, pp. 43-52.

Brynjolfsson, E., Hu, Y.J. and Smith, M.D. (2003), "Consumer surplus in the digital economy: estimating the value of increased product variety at online booksellers", *Management Science*, Vol. 49 No. 11, pp. 1580-1596.

Burke, R., Mobasher, B., Williams, C. and Bhaumik, R. (2006), "Classification features for attack detection in collaborative recommender systems", *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *ACM*, *New York, NY*, pp. 542-547.

Cacheda, F., Carneiro, V., Fernández, D. and Formoso, V. (2011), "Comparison of collaborative filtering algorithms: limitations of current techniques and proposals for scalable, high-performance recommender systems", *ACM Transactions on the Web (TWEB)*, Vol. 5 No. 1.

Huang, J., Cheng, X.Q., Guo, J., Shen, H.W. and Yang, K. (2010), "Social recommendation with interpersonal influence", *19th European Conference on Artificial Intelligence*, IOS Press, Amsterdam, pp. 601-606.

Jamali, M. and Ester, M. (2009), "TrustWalker: a random walk model for combining trust-based and item-based recommendation", *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *ACM*, *New York, NY*, pp. 397-406.

Koutrika, G., Bercovitz, B. and Garcia-Molina, H. (2009), "FlexRecs: expressing and combining flexible recommendations", *Proceedings of the 2009 ACM SIGMOD International Conference on Management of Data*, *ACM*, *New York, NY*, pp. 745-758.

Lam, S.K. and Riedl, J. (2004), "Shilling recommender systems for fun and profit", *Proceedings of the 13th International Conference on World Wide Web*, *ACM*, New York, NY, pp. 393-402.

Lee, H.J., Kim, J.W. and Park, S.J. (2007), "Understanding collaborative filtering parameters for personalized recommendations in e-commerce", *Electronic Commerce Research*, Vol. 7 Nos 3/4, pp. 293-314.

Liu, H., Hu, Z., Mian, A., Tian, H. and Zhu, X. (2014), "A new user similarity model to improve the accuracy of collaborative filtering", *Knowledge-Based Systems*, Vol. 56 No. 1, pp. 156-166.

Lu, J., Shambour, Q., Xu, Y., Lin, Q. and Zhang, G. (2010), "BizSeeker: a hybrid semantic recommendation system for personalized government-to-business e-services", *Internet Research*, Vol. 20 No. 3, pp. 342-365.

Massa, P. and Avesani, P. (2004), "Trust-aware collaborative filtering for recommender systems", *On the Move to Meaningful Internet Systems: CoopIS, DOA, and ODBASE*, Vol. 3290 No. 2, pp. 492-508.

Massa, P. and Avesani, P. (2007), "Trust-aware recommender systems", *Proceedings of the ACM Conference on Recommender Systems*, *ACM*, *New York, NY*.

Massa, P. and Bhattacharjee, B. (2004), "Using trust in recommender systems: an experimental analysis", *Trust Management*, Vol. 2995 No. 1, pp. 221-235.

Mehta, B. and Nejdl, W. (2008), "Unsupervised strategies for shilling detection and robust collaborative filtering", *User Modeling and User-Adapted Interaction*, Vol. 19 Nos 1/2, pp. 65-97.

Mobasher, B., Burke, R., Bhaumik, R. and Williams, C. (2007), "Toward trustworthy recommender systems: an analysis of attack models and algorithm robustness", *ACM Transactions on Internet Technology*, Vol. 7 No. 4, pp. 1-41.

Nazemian, A., Gholami, H. and Taghiyareh, F. (2012), "An improved model of trust-aware recommender systems using distrust metric", *Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining*, *IEEE Computer Society*, *Washington, DC*, pp. 1079-1084.

O'Donovan, J. and Smyth, B. (2005), "Trust in recommender systems", *Proceedings of the 10th International Conference on Intelligent User Interfaces*, *ACM*, *New York, NY*.

Pazzani, M.J. and Billsus, D. (2007), "Content-based recommendation systems", *The Adaptive Web*, Springer, Berlin Heidelberg, Vol. 4321, pp. 325-341.

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P. and Riedl, J. (1994), "GroupLens: an open architecture for collaborative filtering of netnews", *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, *ACM*, *New York, NY*.

Shardanand, U. and Maes, P. (1995), "Social information filtering: algorithms for automating 'word of mouth'", *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, *ACM Press/Addison-Wesley Publishing*, *New York, NY*, pp. 210-217.

Victor, P., Cornelis, C. and Cock, M.D. (2006), "Enhanced recommendations through propagation of trust and distrust", *Proceedings of the 2006 IEEE/WIC/ACM international conference on Web Intelligence and Intelligent Agent Technology*, *IEEE Computer Society*, *Washington, DC*.

Victor, P., Cornelis, C., Cock, M.D. and Silva, P.P. (2009), "Gradual trust and distrust in recommender systems", *Fuzzy Sets and Systems*, Vol. 160 No. 10, pp. 1367-1382.

Victor, P., Cornelis, C., Cock, M.D. and Teredesai, A.M. (2011), "Trust- and distrust-based recommendations for controversial reviews", *IEEE Intelligent Systems*, Vol. 26 No. 1, pp. 48-55.

Williams, C.A., Mobasher, B. and Burke, R. (2007), "Defending recommender systems: detection of profile injection attacks", *Service Oriented Computing and Applications*, Vol. 1 No. 3, pp. 157-170.

Xue, G.R., Lin, C., Yang, Q., Virginia, W.X., Zeng, H.J., Yu, Y. and Chen, Z. (2005), "Scalable collaborative filtering using cluster-based smoothing", *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, *ACM*, *New York, NY*, pp. 114-121.

Yuan, W., Guan, D., Lee, Y.K., Lee, S. and Hur, S.J. (2010a), "Improved trust-aware recommender system using small-worldness of trust networks", *Knowledge-Based Systems*, Vol. 23 No. 3, pp. 232-238.

Yuan, W., Shu, L., Chao, H.C., Guan, D., Lee, Y.K. and Lee, S. (2010b), "ITARS: trust-aware recommender system using implicit trust networks", *Communications, IET*, Vol. 4 No. 10, pp. 1709-1721.

Zhang, X. and Li, Y. (2008), "Use of collaborative recommendations for web search: an exploratory user study", *Journal of Information Science*, Vol. 34 No. 2, pp. 145-161.

Zheng, N., Li, Q., Liao, S. and Zhang, L. (2010), "Which photo groups should I choose? A comparative study of recommendation algorithms in Flickr", *Journal of Information Science*, Vol. 36 No. 6, pp. 733-750.

Corresponding author
Mehdi Shajari can be contacted at: mshajari@aut.ac.ir