Merging Similarity and Trust Based Social Networks to Enhance the Accuracy of Trust-Aware Recommender Systems

Leily Sheugh, Sasan H. Alizadeh[*]

Faculty of Computer and Information Technology Engineering, Qazvin Branch, Islamic Azad University, Qazvin, Iran

Abstract

In recent years, collaborative filtering (CF) methods are important and widely accepted techniques are available for recommender systems. One of these techniques is user based that produces useful recommendations based on the similarity by the ratings of likeminded users. However, these systems suffer from several inherent shortcomings such as data sparsity and cold start problems. With the development of social network, trust measure introduced as a new approach to overcome the CF problems. On the other hand, trust-aware recommender systems are techniques to make use of trust statements and user personal data in social networks to improve the accuracy of rating prediction for cold start users. In addition, clustering-based recommender systems are other kind of systems that to be efficient and scalable to large-scale data sets but these systems suffer from relatively low accuracy and especially coverage too. Therefore to address these problems, in this paper we proposed a multi-view clustering based on Euclidean distance by combination both similarity view and trust relationships that is including explicit and implicit trusts. In order to analyze the effectiveness of the proposed method we used the real-world FilmTrust dataset. The experimental results on this data sets show that our approach can effectively improve both the accuracy and especially coverage of recommendations as well as in the cold start problem.

Keywords: cold start, coverage, accuracy, trust-aware recommender system, multi-view clustering.

1. Introduction

Recommender systems have been created to provide a list of items suitable for user that rely on the opinions of individuals, to find information of interest to them [1, 2]. Many kinds of methods have been proposed for recommender systems, but collaborative filtering (CF) is one of the most well-known techniques [3, 4]. The essence of CF-based recommender systems is to discover similar users based on their rating profiles therefore Similarity plays an important role in CF techniques. Although CF-based recommender systems gained popularity due to its simplicity, however it suffers from data sparsity and cold start [4, 5]. To better model user preferences for the cold-start users who only rated a few items and sparsity problem, additional user information is often adopted. Comparing with membership and friendship, trust information is of less ambiguity and more relevant to user similarity; therefore formally, trust is strongly and positively correlated with user similarity. Hence trust is able to

* Corresponding author. Email: sasan.h.alizadeh@qiau.ac.ir

mitigate the issues of traditional CF such as data sparsity and cold start [4-8].Research has shown that people prefer recommendations from friends to those made by recommender systems. Massa and Avesani [6] analyze the drawbacks of conventional CF-based recommender systems, and elaborate the rationale why incorporating trust can mitigate those problems. They propose Mole Trust algorithm, which performs depth-first search, to propagate and infer trust in the trust networks. Guo et al. [4] proposed a novel method to incorporate trusted neighbors into traditional collaborative filtering techniques. They merge the ratings of trusted neighbors in order to form a new and more complete rating profile for the active users, aiming to resolve the cold start and data sparsity problems from CF suffer. On the other hand Ma et al. [9] propose a social trust ensemble method that linearly combines a social trust and a basic matrix factorization approach. This method is developed by Jamali and Ester [10] their approach focused on where trust propagation is enabled in the social networks. In conclusion, trust-aware recommendatìons can improve the performance of CF recommender systems; in this regard trust is able to provide an effective view of user preference in addition to similarity.

On the other hand, while confirmed to be efficient and scalable clustering-based approaches to large-scale data sets, but clustering-based recommendation have not been widely exploited in recommender systems.

Most previous researches focused on clustering users from the view of similarity. For example Sarwar et al. [11] base the neighborhood formation on the clustering the members by applying the bisecting k-means algorithm to cluster users in cluster-based recommendation. Finally, they find that in their method the accuracy is decreased around 5% in comparison with the KNN CF method. On the other hand Bellogìn and Parapar [12] show that the accuracy can be improved and even outperform the other Collaborative Filtering approaches by applying more advanced clustering method.

Therefore previous research shows that recommendation based on clustering approaches from similarity view, suffer from comparatively low accuracy; furthermore Coverage remains an unresolved issue too.

In this regard, this paper focuses on the development of a clustering-based approach based on both social trust relationships and similarity based on rating patterns. We adopt the trust as ''one's belief regarding the ability of other trustee users in providing valuable ratings''. Whereas, recommender system based on clustering suffer from relatively low accuracy and coverage. Therefore to cope with the above-mentioned issues, we develop a multi-view clustering approach by making use of both the view of similarity and the view of social trust based on Euclidean distance. In multi-view clustering approach user with different views integrated with each other in this regard more users' recommendation can be selected as an option for other users so directly coverage can be improved.

On the other hand to ameliorate accuracy prediction for cold-start users (users who cannot be clustered due to insufficient data), we proposed a novel trust-based approach by incorporating the explicit and implicit trusted neighbors.

In summary, the main contribution of this article is organized as follows: section 2 gives a brief overview of trust-aware recommender system. In section 3 we proposed a multi-view clustering method based on the views of trust and similarity. The proposed method, including aggregating trust neighbors and merging the rating of them, determining the similarity based on confidence, and rating prediction on a target item is present in detail in section 4. In section 5 we behave a series of experiments on film trust real-data set to

verify the impressiveness of the proposed method. Finally, some concluding comments are made in section 6.

2. Trust-Aware Recommender System

Recent research shows that incorporating trust information in recommender systems improving the quality of recommendation [4, 6, 7and 13]. Using trust can efficiently improve the accuracy prediction of recommender systems in comparison with traditional CF algorithms [6, 14]. Besides similarity, other factors like trust also play important role in providing high quality recommendations. Trust identify as efficient way to improve the recommendation quality and utilize the social network. In this regard various techniques proposed to employ trust information into the CF approaches which especially it called trust-aware recommender system [6, 10, 14, 15]. Trust-aware recommender system method are divided into two main approaches containing explicit and implicit [16, 17]. The explicit trust of the other users is used to calculate the direct trust value between each pair of users in social network [16]. The Implicit trust is calculated via implicit information obtained from a social network [17]. moreover, the implicit approach makes inferences from the trust statements between the users on the base of the item rating [17, 18]. An undefined trust value is roughly predict based on the supposition that users nearer in the trust network to the active user have higher trust value [14]. Finally, trust-aware recommender systems present opportunity for recommendations by utilizing users' trust statements especially for systems whose rating data is sparse [4, 19].

3. Multi-view Clustering in Recommender System

The clusters analysis is an important research field in the data mining, the main goal of these algorithms is to recognize natural groups among thousands of pattern. So clustering algorithm be able to group together the users or items with the same properties [15, 20]. The k-means and k-medoids algorithms are the most well-known partitioned clustering methods due to their effectiveness and simplicity [21-23]. Since the k-means produce a cluster center by averaging all the values of each characteristic, it will remove significant personal information for instance trusted neighbors. As a substitute, the k-medoids clustering algorithm chooses a real user as the centroid that minimizes the summation of pairwise distances inside a cluster [22, 24].

The multi-view clustering algorithm was introduced for the first time by [25] and this method develop through that rating pattern and social trust by [26]. The basic idea is to search for clustering in different subspaces of a user space. Users have two different kinds of information, namely trust information (social connections) and ratings information issued on items of interest for users. Hence, these types of information describe users from different views specifically, trust links and rating patterns (user behaviors) together. In this section, we intend to cluster users using both ratings and trust information. Mathematically, the objective function is given as follows [22]:

$$J = \min \sum_{c \in C} \sum_{u,v \in c} d(u,v) \qquad (1)$$

$$d^2 = d_s^2 + d_t^2 \qquad (2)$$

Where users u and v are members of cluster, C is a set of clusters, shows the distance of users u and v

based on Euclidean distance that calculate by equation(2). Then the k-medoids algorithm adopts in order to preserve individuals' ratings and trust information. User similarity is used as the distance metric to measure the closeness of two users; the Pearson correlation coefficient [1] adopted to compute user similarity:

$$S_{u,v} = \frac{\sum_{i \in I}(r_{u,i} - \bar{r}_u)(r_{v,i} - \bar{r}_v)}{\sqrt{\sum_{i \in I}(r_{u,i} - \bar{r}_u)^2}\sqrt{\sum_{i \in I}(r_{v,i} - \bar{r}_v)^2}} \qquad (3)$$

Where   denote the similarity between users' u and v; denote the set of items commonly that are rated by both users u and v. and finally, the average of ratings given by users u and v are   and , respectively. The user distance is thus computed by equation (4).

$$d_s = 1 - S_{u,v} \qquad (4)$$

Second, distance trust calculate using the trust information by Mole Trust[6]. The closer two users are located, the higher trustworthiness the users have.

$$t_{u,v} = \frac{1}{d_{u,v}} \qquad (5)$$

Where , is the dependability of user v relative to user u, and is the minimum distance between two users according to the trust network. The trust distance , is thus computed by equation (6):

$$d_t = 1 - t_{u,v} \qquad (6)$$

And similarity distance , is thus computed by equation (7):

$$d_s = 1 - s_{u,v} \qquad (7)$$

The pseudo code of our multi-view clustering algorithm is presented in Algorithms 1:

Algorithm1 multi-view clustering based on Euclidean distance

Input: *distance matrix D_s, D_t , Sort Trust ; cluster number k*

Output: *user cluster C*

1 $p \leftarrow 0$

2 *disvim* $\leftarrow \left(d_s^2 + d_t^2\right)$

3 *randomly select k medoids m_{medids} from dataset Sort Trust*

4 $C_{disvim} \leftarrow v$, *give* $\min\left(d_{disvim}(v, m_t)\right)$;

5 *while medoids change and < max iteration do*

6 $p \leftarrow p + 1$

7 *calculate* $sum(u) = \sum_v d_{disvim}(u, v), v \in C^{p-1}$

8 *if* $sum_{disvim}(u) < sum_{disvim}(m_{disvim})$ *then*

9 $m_{disvim} \leftarrow u$

10 $C_{disvim} \leftarrow u$ *for* $\forall v$ *find* m_{disvim} *s.t* $\min(d(d(v, m_{disvim})))$

11 $swap(m, u), u \in C^{p-1}$

In Algorithm 1, the rating distance and the trust distance are combination together via Euclidean distance then Euclidean distance as inputs to the multi-view clustering algorithm which outputs the clusters of users.

4. Proposed Method

In this paper, clustering base trust-aware recommender system by using a multi-view clustering algorithm is proposed. The proposed method consist of four phases which shown in fig 1. In the first phase, trust matrix is applied to collection the trust user. This phase consists of two steps that include: finding implicit trust also calculates explicit trust by mole trust. Then obtain the trust distance by both of trust neighbors. In the second phase, rating matrix is

applied to earn similarity between users, the Pearson correlation coefficient adopted is one of the most well known techniques to compute user similarity, and therefore we used it. The similarity distance is computed. Then in the three phases, a multi-view clustering based on Euclidean distance algorithm is applied to group the similar user in several clusters by trust distance and similarity distance. Based on this new distance, the k-medoids algorithm can be applied for clustering users that have higher similarity. Finally, in four phases, for each unseen item, a rate is predicted for recommendation to the active user. Prediction is based on the cluster's member rating and by similarity based on confidence.
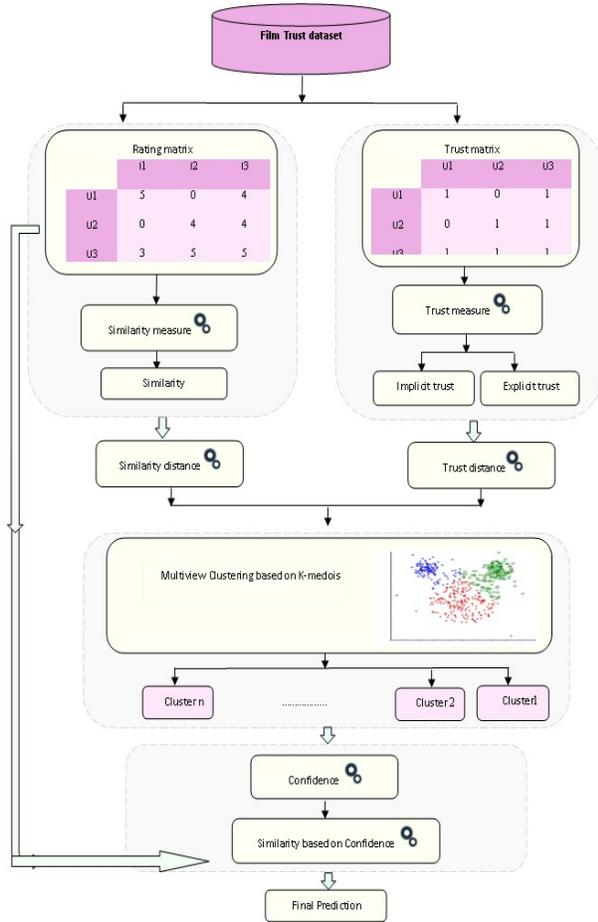
Fig. 1. Overview of proposed method

4.1. Prediction

After determination the trustworthy neighbors, all the ratings of trusted neighbors merged into a single rating based on the weights of trusted neighbors [4] is defined as follow:

$$\tilde{r}_{u,j} = \frac{\sum_{v \in TN_u} W_{u,v} \cdot r_{v,j}}{\sum_{v \in TN_u} |W_{u,v}|} \tag{8}$$

That is, $\tilde{r}_{u,j}$ is the merged value for user u on item j based on the ratings of all the trusted neighbors, and $W_{u,v}$ represent the importance weight of user v's ratings relation to the user u [4]. The weight $W_{u,v}$ is consisting of three parts: trust value $t_{u,v}$, rating similarity $S_{u,v}$ and social similarity $j_{u,v}$ so, $W_{u,v}$ is computed by equation (9):

$$W_{u,v} = \alpha S_{u,v} + \beta t_{u,v} + (1 - \alpha - \beta).j_{u,v} \tag{9}$$

The social similarity is computed by the Jaccard Index and specified as the ratio of common trusted neighbors over all the trusted neighbors, by equation (10):

$$j_{u,v} = \frac{|TN_u \cap TN_v|}{|TN_u \cup TN_v|} \tag{10}$$

Where TN_u and TN_v are the trusted neighbors user u and v respectively and $j_{u,v} \in (0,1]$.

However, the quality or usefulness of the merged ratings is unknown. Therefore determining the confidence [27] of the merged ratings by equation (11):

$$C_{u,j} = C(p_{u,j}, n_{u,j}) = \frac{1}{2} \int_0^1 \left| \frac{x^{p_{u,j}}(1-x)^{n_{v,j}}}{\int_0^1 x^{p_{u,j}}(1-x)^{n_{u,j}} dx} - 1 \right| dx \tag{11}$$

Where, $C_{u,j} \in (0,1]$ is rating confidence $p_{u,i}$ and $n_{u,i}$ referring to the number of positive and negative rating which provided by the trusted users.

In addition to the merged ratings, the confidence is also important to show the quality of the merged ratings of trust neighbor. Therefore introduce a confidence-aware PCC [4] to compute user similarity in finally prediction, denoted by $S'_{u,v}$:

$$S'_{u,v} = \frac{\sum_{i \in I_{u,v}} c_{u,i} (\tilde{r}_{u,i} - \bar{r}_u)(r_{v,i} - \bar{r}_v)}{\sqrt{\sum_{i \in I_{u,v}} c_{u,i}^2 (\tilde{r}_{u,i} - \bar{r}_u)^2} \sqrt{\sum_{i \in I_{u,v}} (r_{v,i} - \bar{r}_v)^2}} \qquad (12)$$

Where $I_{u,v} = \tilde{I}_u \cap I_v$ refer to the set of items is rated by both users' u and v after the merging process, and $\tilde{I}_u$ consists of the items that rated by at least one trusted neighbor, $\bar{r}_u$ and $\bar{r}_v$ are the average ratings for users' u and v respectively, and finally $c_{u,i}$ is the confidence measurement.

Finally, all the rating's users which pertain to the multiview clusters are accumulating to produce a final prediction on a target item j that the active user u has not rated. So Prediction [4] calculates the average value of all ratings provided by the nearest neighbors v weighted via similarity based on confidence $S'_{u,v}$, and formally it computed by:

$$\hat{r}_{u,j} = \frac{\sum_{v \in MV} S'_{u,v} \cdot r_{vj}}{\sum_{v \in MV} |S'_{u,v}|} \qquad (13)$$

5. Experiments

In order to verify the performance of the proposed method, various experiments were performed with pure multi-view clustering, in addition to a number of trust-based method in recommender system. Specifically, the proposed method was compared to the multi-view by regarding both of trust and similarity view separately, that the basic model of multi-view clustering according to similarity and trust [26], the pure Collaborative Filtering (CF)[3], the basic model of trust-aware recommender system based on Mole Trust [6], and the merge method [4].

5.1. Dataset

In this paper, real-world FilmTrust dataset (trust.mindswap.org/FilmTrust/)   is   used   in experiments. That consists of 1642 users, 2071 movies. This dataset is a trust-based social site that users can rate the interested movies besides this site allows users to share movie ratings and explicitly specify to other users. The users of FilmTrust website will be able to review items and assign them numeric ratings values in the range 0.5 (min) to 4 (max) with step 0.5. Therefore FilmTrust is a dataset that contain both the user-item ratings and data of explicit trust too.

5.2. Evaluation Measures

The leave-one-out method mostly is used to compare two recommendation systems [6].In this method, in each step user rating is hidden from the dataset and then its value will be predicted by applying a definite method up to all the testing ratings are covered, finally main value is compared with the predicted rate obtained by the recommendation method. In this paper, the evaluation is former by applying this method on the two data views including cold start and all user. Then the experimental results are analyzed according to the accuracy and coverage measures by Mean Absolute Error (MAE) [12] and Rating Coverage (RC) [12] respectively. The evaluation metrics are described as follows:

$$MAE = \frac{\sum_{i=1}^{n} |r_i - r_p|}{N} \qquad (14)$$

Where, is real rating and is predicated rating of the item, i for the active user. Also, N is the total number of testing ratings.

In addition, the rating coverage defined as number of predictable ratings (M) over the all testing ratings (N), it given by equation (15):

$$RC = \frac{M}{N} \qquad (15)$$

5.3. Result and Analysis

In this section, we present a series of experiments on FilmTrust dataset to demonstrate the effectiveness of our approach. Finally results for both view all user and cold user are reported in the table1.

Table 1

The Mae And Rc Result Over The Filmtrust Dataset

View	Approach measured by MAE,RC				
	Measure	*CF*	*Mole Trust*	*Merge*	*MV based on Euclidean distance*
All Users	MAE	**0.703**	0.771	0.708	0.7795
	RC (%)	%93.84	%30.38	**%95.06**	72.17%
Cold user	MAE	0.744	0.819	0.768	**0.7001**
	RC (%)	%39.64	%23.19	%54.28	**%65.75**

From the table 1 results, it can be conclude that for cold user the proposed method outperformed than the other methods. On the other hand, the result shows that this method obtained the lower values in MAE for all user view; moreover result also show that merge method obtained the higher rate coverage from this view.

5.4 performances for cold users

Whereas the main objective of this research overcomes the accuracy and coverage of cold users, therefore in this section we compare the proposed method with merged method.

The parameter θ is an important parameter that is used as the threshold value and this parameter shows of the proposed reliability measure. Therefore In this paper we used varies the threshold θ from 0.0 to 0.9 with step 0.1 to test the performance of the proposed method based on 3, 5 and 10 cluster in compare merge method for the FilmTrust dataset. Fig. 2 and Fig. 3 reports the results of different θ values on the proposed method over MAE and RC measure respectively.
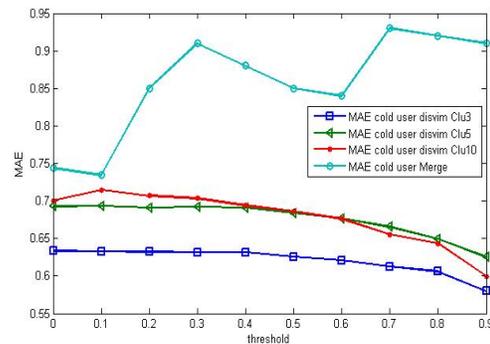
Fig. 2. the MAE in the view of Cold Users

The result according to fig.2 shows that MAE measure depend on the threshold value and based on the proposed method according to 3, 5 and 10 clustering, performance to be better than merge method. So multi-view clustering can be improved the accuracy of cold user in cluster-based recommendation.
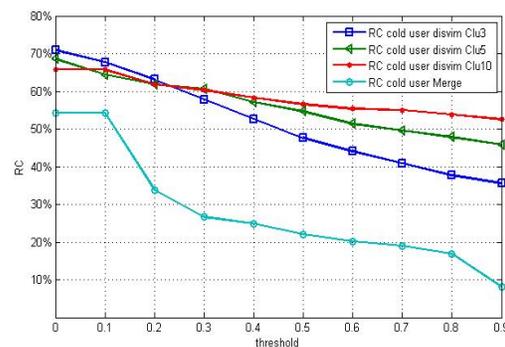
Fig. 3. the RC in the view of Cold Users

The results coming from Fig. 3 show that when the value of θ increases, the value of the rating coverage will be reduced. So it can be concluded that the lower

value of θ improving the coverage measure. On the other hand, the higher value of threshold results reducing the rating coverage (RC). Furthermore, in comparison with merge method we obtain that proposed method achieve best result; it confirms that multi-view clustering to improve the coverage cold start user.

6.   Conclusion

Traditional user-based recommender systems like CF- based recommender systems are proposed the likeminded user for providing the recommendation based on similarity. Cold start is one of the main problems of these systems. Trust is a concept that has recently takes much attention to improve the cold start problem. This paper proposed a novel method for trust-aware recommender system by multi-view clustering based on Euclidean distance to alleviate the cluster-based recommender system problems including: the low accuracy and coverage. In the proposed method, According to Euclidean distance, similarity-based distances and trust-based distances combined together. Then User Clustered by applying a k-medoids clustering method based on this new distance. In addition to accommodate the accuracy prediction's cold start users (the users who have rated less than five items) in this clustering method we employed both of the implicit trust and explicit trust for trust statement.

The experimental results on FilmTrust dataset showed that (1) the combination of similarity pattern and trust information were the useful in determining a real prediction for active user; (2) the multi-view clustering method outperformed the accuracy and coverage and it can effectively handle especially for cold start problem.(3) our proposed method worked better than other user-based recommender system

approaches. To sum up, the proposed method effectively increase clustering-based methods by asset of the combination view of trust information and similarity.

References

[1] G. Adomavicius, A. Tuzhilin, toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions, IEEE Trans.Knowl. Data Eng. (TKDE) 17 (2005) 734–749.

[2] D. Jannach, M. Zanker, M. Ge, M. Gröning, Recommender systems in computer science and information systems a landscape of research, E-Commerce Web Technol. (2012) 76–87.

[3] J. Zhang, Q. Peng, S. Sun, C. Liu, Collaborative filtering recommendation algorithm based on user preference derived from item domain features, Physica A 396 (2014) 66–76.

[4] G. Guo, J. Zhang, D. Thalmann, Merging trust in collaborative filtering to alleviate data sparsity and cold start, Knowl.-Based Syst. (KBS) 57 (2014) 57–68.

[5] S. Lee, J. Yang, S. Y. Park, Discovery of hidden similarity on collaborative filtering to overcome sparsity problem. In Proceedings of discovery science Berlin, Germany: Heidelberg: Springer-Verlag. (2004) 396–402.

[6] P. Massa, P. Avesani, Trust-aware collaborative filtering for recommender systems, in: Federated Int. Conf on the Move to Meaningful Internet, 2004.

[7] N. Lathia, S. Hailes, L. Capra, Trust-based collaborative filtering, in: Proceedings of Trust Management II, Springer, US, (2008) 119–134.

[8] C.C. Chen, Y.H. Wan, M.C. Chung, Y.C. Sun, an effective recommendation method for cold start new users using trust and distrust networks, Inform. Sci. 224 (2013) 19–36.

[9] M. Hao, D. Zhou, C. Liu, M. R. Lyu, Irwin King. "Recommender systems with social regularization." In Proceedings of the fourth ACM international conference on Web search and data mining. ACM, (2011) 287-296.

[10] J. Mohsen, and M. Ester. "A matrix factorization technique with trust propagation for recommendation in social networks." In Proceedings of the fourth ACM conference on Recommender systems, ACM, (2010) 135-142.

[11] M. Sarwar Badrul, G. Karypis, J. Konstan, and J. Riedl. "Recommender systems for large-scale e-commerce:

Scalable neighborhood formation using clustering." In Proceedings of the fifth international conference on computer and information technology, vol. 1. 2002.

[12] B. Jesús, F. Ortega, Antonio Hernando, and Abraham Gutiérrez. "Recommender systems survey." Knowledge-Based Systems 46 (2013) 109-132.

[13] P. Bedi, P. Vashisth, Empowering recommender systems using trust and argumentation, Inform. Sci. 279 (2014) 569–586.

[14] J. Golbeck, Generating predictive movie recommendations from trust in social networks, in: 4th International Conference on Trust Management Pisa, Italy, 2006.

[15] T. DuBois, J. Golbeck, J. Kleint, A. Srinivasan, Improving recommendation accuracy by clustering social networks with trust, in: ACM RecSys Workshop Recommender Systems and the Social Web, 2009.

[16] Liu, F., Lee, H. Use of social network information to enhance collaborative filtering performance. Expert Systems with Applications, 37, (2010) 4772–4778.

[17] Guo, G., Zhang, J., Smith, N.Y. TrustSVD: Collaborative filtering with both the explicit and implicit influence of user trust and of item ratings. In: Association for the advancement of artificial intelligence (AAAI) (2015b).

[18] Li, Y. M., Kao, C. P. TREPPS: A trust-based recommender system for peer production services. Expert Systems with Applications, 36, (2009) 3263–3277.

[19] Z. Fuzhi, H. Wang, Yi. Huawei, "An adaptive recommendation method based on small-world implicit trust network." Journal of Computers 9, no. 3 (2014) 618-625.

[20] Q. Wang, W. Cao, Y. Liu, A novel clustering based collaborative filtering recommendation system algorithm, Lect. Notes Electr. Eng. 260 (2014) 673–680.

[21] J. Kelleher, D. Bridge, Rectree centroid: An accurate, scalable collaborative recommender, in: Proceedings of the Fourteenth Irish Conference on Artificial Intelligence and Cognitive Science, Citeseer, (2003)89–94.

[22] G. Xue, C. Lin, Q. Yang, W. Xi, H. Zeng, Y. Yu, Z. Chen, Scalable collaborative filtering using cluster-based smoothing, in: Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), (2005) 114–121.

[23] C. Bishop, et al., Pattern Recognition and Machine Learning, vol. 4, 2006.

[24] A. Bellogín, J. Parapar, Using graph partitioning techniques for neighbor selection in user-based collaborative filtering, in: Proceedings of the 6th ACM Conference on Recommender Systems (RecSys), (2012) 213–216.

[25] S. Bickel, T. Scheffer, Multi-view clustering, in: Proceedings of the IEEE International Conference on Data Mining (ICDM), 2004.

[26] G. Guibing, J. Zhang, N. Yorke-Smith. "Leveraging multiviews of trust and similarity to enhance clustering-based recommender systems." Knowledge-Based Systems 74 (2015) 14-27.

[27] Y. Wang, and P. S. Munindar "Formal Trust Model for Multiagent Systems." In IJCAI, vol. 7, (2007) 1551-1556.