More Reputable Recommenders Give More Accurate Recommendations?

Weiwei Yuan[1,2]
[1]Harbin Engineering Univ.
[2]Kyung Hee Univ.
No. 145, Nantong Street
Harbin, China
+86-187-0450-3170
yuanweiwei00@gmail.com

Donghai Guan[1,2]
[1]Harbin Engineering Univ.
[2]Kyung Hee Univ.
Seocheon-dong, Giheung-gu
Yongin-si, Gyeonggi-do, Korea
+82-31-201-3467
donghai@oslab.khu.ac.kr

Yong-Koo Han
Kyung Hee Univ.
Seocheon-dong, Giheung-gu
Yongin-si, Gyeonggi-do, Korea
+82-31-201-2950
ykhan@khu.ac.kr

Sungyoung Lee
Kyung Hee Univ.
Seocheon-dong, Giheung-gu
Yongin-si, Gyeonggi-do, Korea
+82-31-201-2514
sylee@oslab.khu.ac.kr

Young-Koo Lee
Kyung Hee Univ.
Seocheon-dong, Giheung-gu
Yongin-si, Gyeonggi-do, Korea
+82-31-201-3732
yklee@khu.ac.kr

ABSTRACT

Existing models of the Trust-Aware Recommender System (TARS) build personalized trust networks for the active users to predict ratings. These models have reasonable rating prediction performances, while suffer from high computational complexity. One solution is to utilize the global rating prediction mechanism for TARS, in which an intuitive assumption is that more reputable recommenders give more accurate recommendations. In addition, due to the scale-freeness of the trust network, some users have and continuously have superior reputations than others. However, we show via comprehensive experiments on the real TARS data that the recommendations given by recommenders with higher reputations do not tend to be more accurate. Furthermore, even the recommendations given by the recommenders with superior high reputations do not tend to more accurate. Our experimental study provides promising directions for the future research on the rating prediction mechanism of TARS.

Categories and Subject Descriptors

F.2 [Theory of Computation]: Analysis of Algorithms and Problem Complexity.

General Terms

Algorithms, Experimentation

Keywords

Reputation, rating prediction accuracy, global rating prediction mechanism, trust-aware recommender system

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
ICUIMC(IMCOM)'13, January 17–19, 2013, Kota Kinabalu, Malaysia.
Copyright 2013 ACM 978-1-4503-1958-4…$15.00.

1. INTRODUCTION

The Trust-Aware Recommender System (TARS) suggests the worthwhile information to the users on the basis of trust. Though existing models of TARS have reasonable rating prediction performances [1-10], their computational cost is high: they search each user's trusted users, and build their personalized trust networks. The computational complexity is $O(k^d)$, where k is the average degree of the trust network and d is the trust propagation distance.

One way to reduce the computational complexity of the existing TARS models is to build the global rating prediction mechanism. The basis is to find the reputable recommenders in TARS, and use their recommendations to find the valuable information. A reputable recommender means this recommender has high reputation in the trust network, in which reputation is what is generally said or believed about a person's or thing's character or standing [11]. The global rating prediction mechanism is highly efficient: its computational complexity is only $O(C)$, where C is a constant. it is much more light weighted than the personalized rating prediction mechanism of TARS.

An intuitive assumption on building the global rating prediction mechanism is to assume that more reputable recommenders give more accurate recommendations. That is, the higher reputation a recommender has, the more accurate recommendations this recommender gives. These recommendations will be given higher weight when predicting the ratings on the target item. In this paper, a recommender's reputation is measured as the aggregation of the trust this recommender received from the trust network. So a recommender is more reputable if more users in the trust network trust this recommender.

An inspiring fact is that the trust network is the scale-free network [1-5], which means while most nodes have limited indegrees, some nodes have superior indegrees in the trust network. In addition, since the topology of the scale-free network enables its continuous scale-freeness [12], the nodes with superior indegrees continuously have much higher indegrees than other users of the trust network. Since a node's indegree in the trust network is the number of the trust relations pointing to this user, it measures the extent this user is trusted by other uses in the trust network. So we use the indegree of the user to evaluate its reputation. The scale-freeness of the trust network and the properties of the user's indegree indict that some users have the superior reputations in the trust network, and they continuously have much higher reputations than the other. It is very attractive to use the recommendations given by the recommenders with the superior high reputations in the global rating prediction mechanism since these users have very stable high reputations in the trust network.

However, we show via comprehensive experiments that the reality is completely far from the wishful thinking: the recommenders with higher reputations do not tend to give more accurate recommendations on the items for the target users; furthermore, even the recommendations given by the users with superior high reputations do not tends to be more accurate than the recommendations given by other recommenders. These conclusions are achieved by examining the trust-aware recommender system on the open dataset of Epinions [13]. Regardless of the number of recommendations given by a user in the trust-aware recommender system, this user's reputation does not necessarily link to the accuracy of the recommendations given by him.

The rest of the paper is organized as follows: Section 2 summarizes the related works of TARS, Section 3 gives our experimental studying results on the global rating prediction mechanism of the trust-aware recommender systems, and Section 4 concludes this paper and points out the future directions.

2. RELATED WORKS

The architecture of TARS is shown in Fig. 1. The inputs are the trust matrix and the rating matrix. The output is the predicted ratings on the items for different users. The trust matrix is the collection of the trust relations between the users of the recommender system. Each element of the trust matrix describes the trust between two users. The rating matrix records the users' ratings on the items. Each element of the rating matrix is the rating given by a user on a particular item. There are two functional modules in the architecture of TARS: the trust estimation function module and the rating predictor module. Since it is the impossible for the users to state their trusts on all other users, the trust matrix is always sparseness, which means it has a lot of missing values. The trust estimation function module estimates each user's trust on every other user, which is recorded in the estimated trust matrix. The rating predictor module predicts the users' ratings on the items.

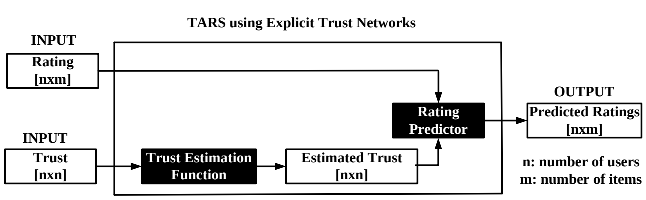


Figure1. Trust-aware recommender system architecture.

Table 1. The rating prediction mechanism of the conventional TARS model.

Input: $\mathbf{T}$ (trusts), $\mathbf{R}$ (ratings)

Parameter: a (active user), u (recommender), i (item), $d_{a,u}$ (a 's trust propagation distance to u), $d_{\max}$ (maximum allowable propagation distance), $w_{a,u}$ (a 's weight (trust) on u), m (number of recommenders), $r_{u,i}$ (u 's rating on i) $\overline{r_{.}}$ (average rating of .).

Output: $p_{a,i}$ (a 's predicted rating on i)

Phase 1: Recommender weighting (trust estimation):

$$w_{a,u} = \frac{d_{\max} - d_{a,u} + 1}{d_{\max}}$$

Phase 2: Rating calculation: $p_{a,i} = \overline{r_a} + \frac{\sum_{u=1}^{m} w_{a,u}(r_{u,i} - \overline{r_u})}{\sum_{u=1}^{m} w_{a,u}}$

The rating prediction mechanism of the conventional TARS model is similar as that of Collaborative Filtering (CF) [3-7]. The difference is that CF weights each recommendation based on the active user's similarity with the recommender, while TARS weights each recommendation based on the active user's trust on the recommender. The detailed rating prediction mechanism of the conventional TARS model is shown in Table 1. There are two phases: the recommender weighting phase and the rating calculation phase, where $d_{\max}$ is the maximum allowable propagation distance between users of the recommender system. Users that are not reachable within the maximum allowable propagation distance have no estimated trust value. $d_{a,u}$ is the active user a's trust propagation distance to the recommender u. In TARS, the trust propagation distance refers to the number of hops in the shortest trust propagation path from the trustor to the trustee.

3. EXPERIMENTAL STUDY ON THE GLOBAL RATING PREDICTION MECHANISM OF TARS

In this section, we use a set of experiments to examine the relationship between the user's reputation and the accuracy of their recommendations for the global rating prediction mechanism of TARS. The experiments are held on the real TARS dataset: Epinions [13]. Epinions is a product and shop review site. It has a pool of reviewers who can write reviews considering a set of aspects such as Ease of Use, On-Time Delivery etc. Other members can rate reviews as Not Helpful, Somewhat Helpful, Helpful, and Very Helpful. These ratings on the reviews are formatted to integers from 1 to 5 in the Epinions dataset. Epinions also has the Web of Trust which enables users to express their opinions on others: trust or block. A member's list of trusted members represents that member's personal Web of Trust. If an active user involve a target user in his web of trust, his trust on this target user is set to be 1 in the Epinions data set, otherwise, the value of trust is set to bet 1.

The trust network of the Epinions dataset consists of 49288 users' 487183 directed trust relations. The rating matrix of the Epinions dataset records 40163 users' 664824 ratings on 139738 items. The distribution of the ratings in Epinions dataset is shown in Table 2. The indegree distribution of the trust network is shown in Fig. 2. It is obvious that the trust network's indegree distribution follows the power law. This means the trust network is the scale-free network [12]. Since a user's indegree in the trust network is the number of trusts pointing to this user, i.e., the number of users trust this user, we use the indegree of the node to represent its reputation in the trust network. In this paper, the term indegree and the term reputation are used alternatively, representing the same meaning.

Table 2. The distribution of the ratings in the rating matrix of the Epinions dataset.

	Probability
Rating=1	0.06502172003417446
Rating=2	0.07622769334440394
Rating=3	0.11360149453088336
Rating=4	0.2923179668604022
Rating=5	0.4528311252301361

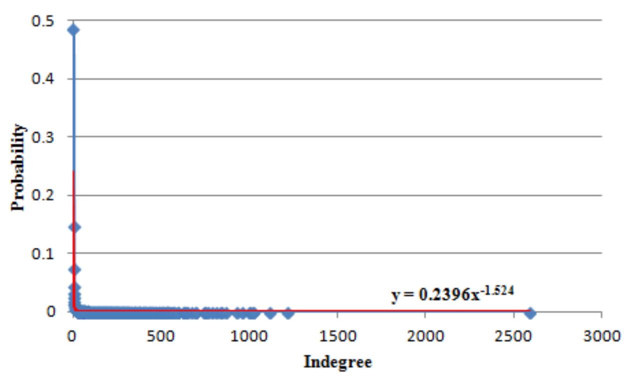


Figure 2. The indegree distribution of trust network in the Epinions dataset.

Since the scale of the Epinions dataset is large, we randomly choose around 1% of the nodes, i.e, around 500 users, from the Epinions dataset. We examine the relationship between these users, and investigative the relationship between their reputations and the accuracy of their recommendations. To ensure the statistical meaning of the experimental results, we repeat the experiments five times. Table 3 and Table 4 list the detailed information of the trust matrices and the rating matrices in the five sets of experiments respectively. The five experiments are represented by Experiment 1 to Experiment 5 respectively in this paper. And they are also represented by Ex 1 to Ex 5 in some figures for the conciseness.

Table 3. The detailed information of the trust matrices used in the five sets of experiments.

Trust matrix	Num of users	Num of trust relations
Experiment 1	500	23848
Experiment 2	456	7148
Experiment 3	443	3918
Experiment 4	426	2558
Experiment 5	387	1749

Table 4. The detailed information of the rating matrices used in the five sets of experiments.

Rating matrix	Num of users	Num of items	Num of ratings
Experiment 1	489	44288	66287
Experiment 2	492	26717	39467
Experiment 3	478	22534	36548
Experiment 4	489	19571	31510
Experiment 5	475	16624	24953

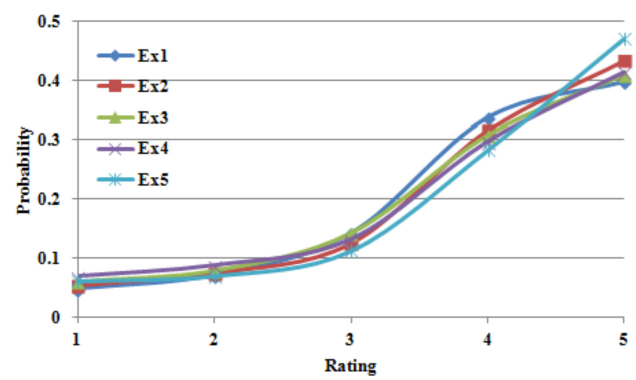


Figure 3. Distribution of the ratings in the rating matrices of our five experiments.

The distribution of the ratings in the five rating matrices mentioned in Table 4 is shown in Fig. 3. Their distributions are similar as the distributions of the overall ratings shown in Table 2. This indicates that our experiments on the five selected datasets can represent the actual situation of the original Epinions dataset. Our experiments are statistically meaningful.

The indegree distribution of the nodes in the five trust matrices mentioned in Table 3 is shown in Fig. 4. Similar as the indegree distribution of the original Epinions dataset, these five trust networks are all scale-free networks. This means some users of the trust networks have superior reputations than other users, while most users have low reputations in the trust network. In addition, the limited number of users with superior high reputations continuously has much higher reputations than other users of the trust network.

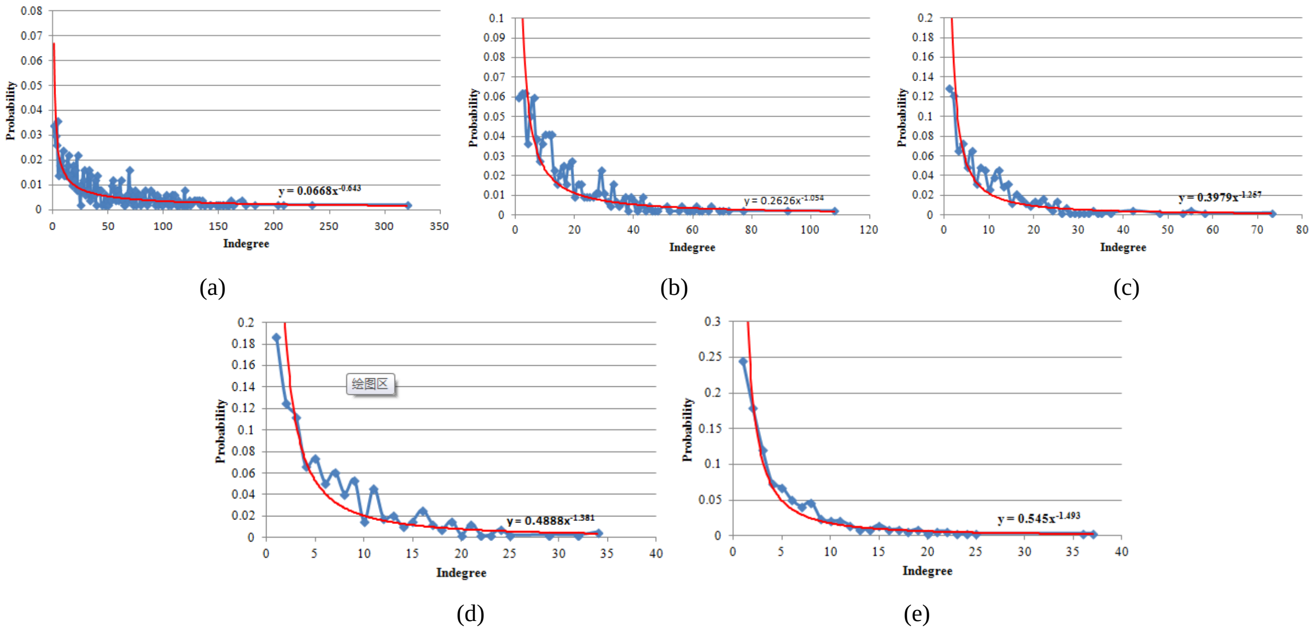


Figure 4: The indegree distribution of the nodes in the trust matrix of (a) Experiment 1, (b) Experiment 2, (c) Experiment 3, (d) Experiment 4, and (e) Experiment 5.

Table 5. The detailed information of the trust networks in the test datasets of the five experiments held in this work.

Test dataset	Num of target users	Num of target items	Num of ratings
Experiment 1	47	3854	4062
Experiment 2	48	2128	2325
Experiment 3	49	2938	3408
Experiment 4	48	2159	2306
Experiment 5	49	4264	4830

Table 6. The detailed information of the recommendations related to the test datasets in the five experiments held in this work.

Test dataset	Num of recommenders	Num of recommendations
Experiment 1	433	8636
Experiment 2	408	6250
Experiment 3	386	10543
Experiment 4	393	5032
Experiment 5	405	10703

We randomly choose around 10% users from the trust networks of the five experimental datasets to compose the test datasets. These users are regarded as the target users of the rating prediction. The ratings given by these users are regarded as the target of the rating prediction. All other users which have given ratings on the target items are regarded as the recommenders, and their ratings on the target items are regarded as the recommendations. The reputations of the target users are their indegrees in their corresponding trust networks. The detailed information of the test datasets in the five experiments is given in Table 5 and Table 6. The distribution of the real ratings in the test datasets and the distribution of the recommendations on the target items are given in Fig. 5.

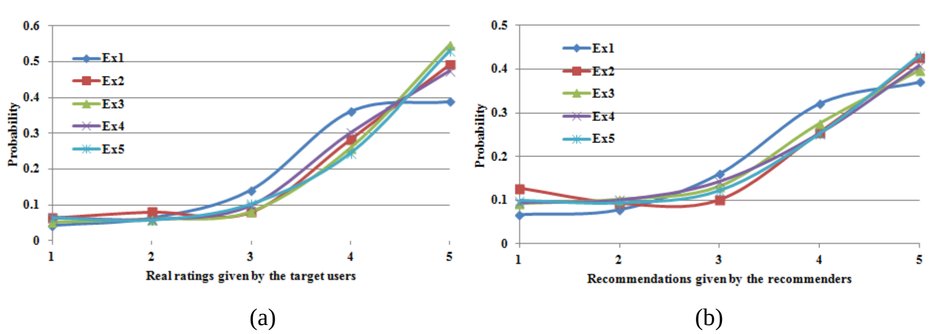
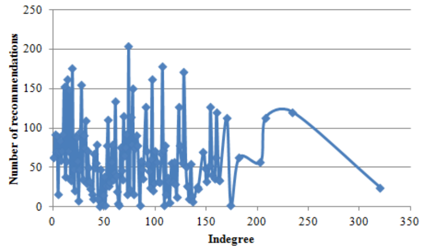
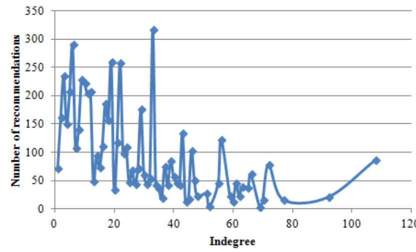


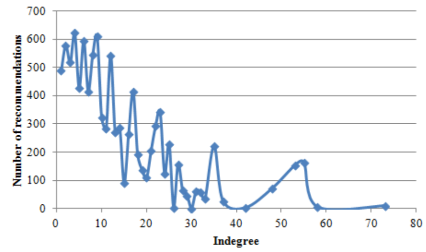
Figure 5. The distribution of (a) the real ratings given by the target users, and (b) the recommendations given by the recommenders.



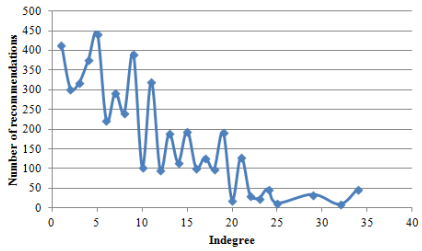
(a)



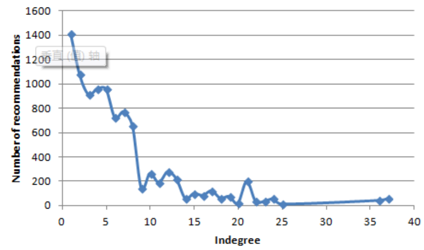
(b)



(c)

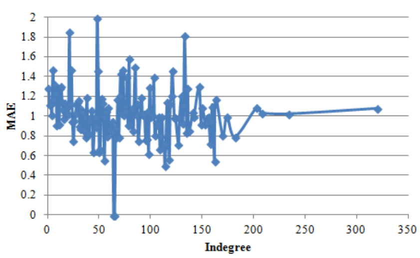


(d)

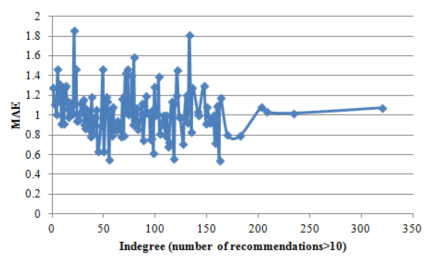


(e)

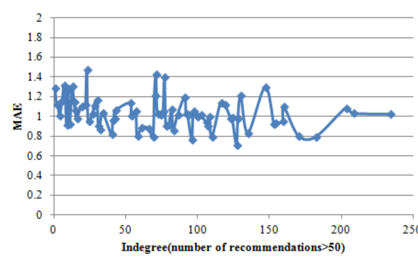
Figure 6. Number of recommendations given by recommenders with differenct repuations of (a) Experiment 1, (b) Experiment 2, (c) Experiment 3, (d) Experiment 4, and (e) Experiment 5.



(a)

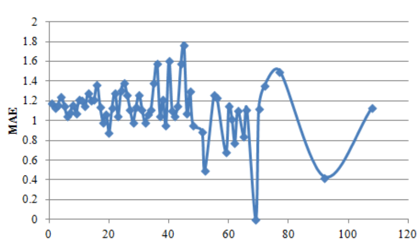


(b)

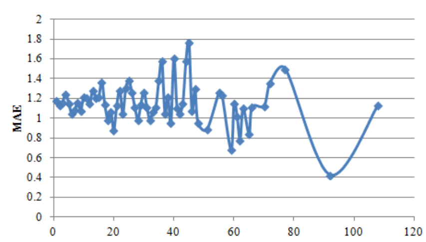


(c)

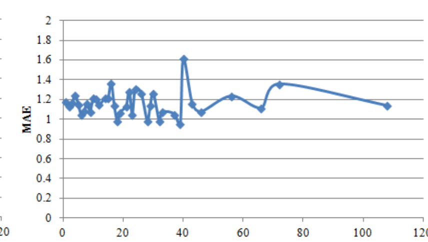
Figure 7. The relationship between the repuations of the recommenders and the accrucy of their recommendations in Experiment 1, in which (a) is for all recommenders, (b) is for recommenders with the same reputations totally given more than 10 recommendations, and (c) is for recommenders with the same reputations totally given more than 50 recommendations.



(a)

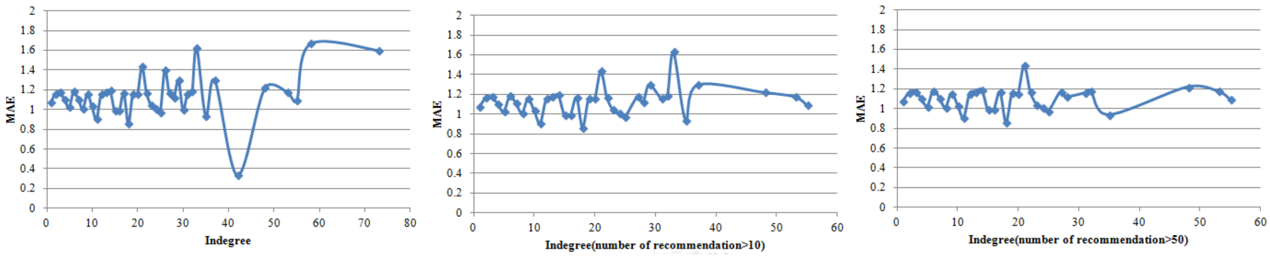


(b)



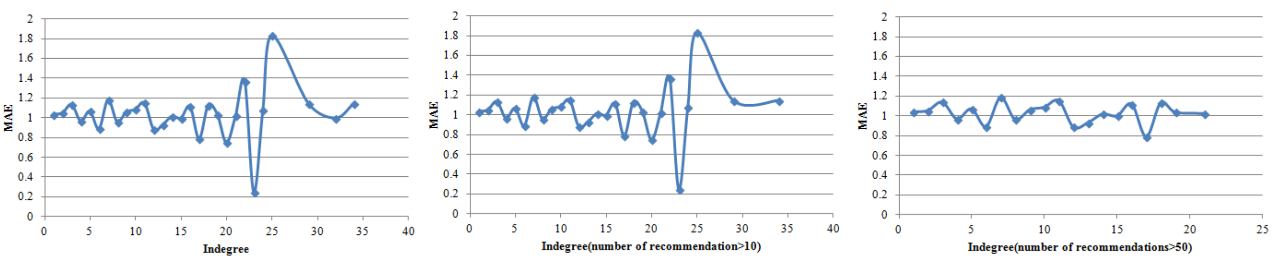
(c)

Figure 8. The relationship between the repuations of the recommenders and the accrucy of their recommendations in Experiment 2, in which (a) is for all recommenders, (b) is for recommenders with the same reputations totally given more than 10 recommendations, and (c) is for recommenders with the same reputations totally given more than 50 recommendations.



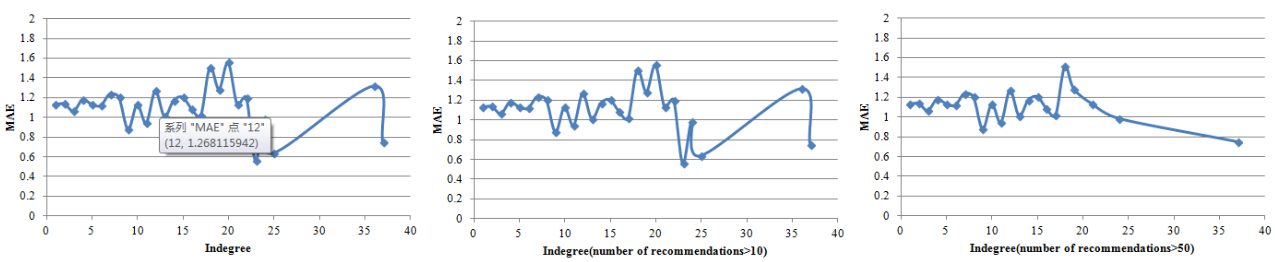
(a) (b) (c)

Figure 9. The relationship between the repuations of the recommenders and the accrucy of their recommendations in Experiment 3, in which (a) is for all recommenders, (b) is for recommenders with the same reputations totally given more than 10 recommendations, and (c) is for recommenders with the same reputations totally given more than 50 recommendations.



(a) (b) (c)

Figure 10. The relationship between the repuations of the recommenders and the accrucy of their recommendations in Experiment 4, in which (a) is for all recommenders, (b) is for recommenders with the same reputations totally given more than 10 recommendations, and (c) is for recommenders with the same reputations totally given more than 50 recommendations.



(a) (b) (c)

Figure 11. The relationship between the repuations of the recommenders and the accrucy of their recommendations in Experiment 5, in which (a) is for all recommenders, (b) is for recommenders with the same reputations totally given more than 10 recommendations, and (c) is for recommenders with the same reputations totally given more than 50 recommendations..

To predict ratings for the target users, we first examine the number of recommendations given by the recommenders. The experimental results are given in Fig. 6. This step is held since some users may give limited number of recommendations. To insure that the experimental results are statistically meaningful, we shall not consider the relationship between the reputation and the accuracy of the recommendations for this kind of recommenders.

The simulation results on the five sets of experiments are shown in Fig. 7, Fig. 8, Fig. 9, Fig. 10 and Fig. 11 resptectivly. These results illsturate the relationship between the recommenders' reputations and the accuracy of their recommendations. MAE, which is the vertical axis of each figure mentioned above, means the Mean Absolute Error. MAE is used to measure the error of the predicted ratings since it is very appropriate and useful for evaluating prediction accuracy in offline tests [3]. To calculate MAE, the predicted rating is compared with the real rating and the difference (in absolute value) is the prediction error, this error is then averaged over all predictions to obtain the overall MAE. The predicted rating is the mean of all recommendations with the given reputations in the experiments held in this work. Based on the simulation results shown in Fig. 6, we also examine the relationship between recommenders' reputations and the accuracy of their recommendations on different type of user: all recommenders and recommenders given a number of recommendations. The simulation results for all users are given in (a) subgraph of each figure, and the simulation results for the recommenders given a number of recommendations are given in (b) subgraph and (c) subgraph of each figure.

The simulation results of the five sets of experiments all illustrate that there is no direct relationship between the recommenders' reputations and the accuracy of their recommendations: the recommendations given by the recommenders with higher reputations do no contribute to better rating prediction accuracy, i.e, the recommendations given by these recommenders do not tend to be more accurate. Furthermore, the trust network is the scale-free network, as shown in the degree distributions of the trust network in Fig. 4, so some users in the trust network have superior indgrees. However, it is shown in the experimental results that the recommenders with superior high reputation in the trust network also do not contribute to better rating prediction accuracy. This means the recommendations given by the recommenders with superior reputations also do not tend to be more accurate. The simulation results, as given in (b) and (c) subgraphs of Fig. 7 - Fig. 11, further show that even for the recommenders with sufficient number of recommendations, i.e., the recommenders actively participated in, even if they have higher reputations, their recommendations do not tend to be more accurate.

The experiments held in this work show that the recommendations given by more reputable recommenders do not tend to be more accurate, and these recommendations are not more valuable for the rating prediction. This is far from our expectation. It has been verified in our previous work [1-5] that if a recommender has shorter trust propagation distance from an active user, i.e., this recommender is more trustful to the active user, the recommendations given by this recommender tend to be more accurate for the rating prediction of this active user. A user's reputation is the aggregation of the trusts obtained from other users in the trust network. If a user has higher reputation, it means this user is trusted by more users in the trust network of TARS. However, it is supervising that the recommendations given by the recommenders trusted by many other users in the trust network do not provide more accurate recommendations for the active users. The reason may lie in that: though trust and reputation are closely related concepts, they have their own characteristics. Reputation measures the reliability of a user from the overall point of view, while trust measures the reliability of a user from each user's own personalized point of view. Since the ratings are personalized opinions representing each user's personal tastes, the personalized measure trust plays an active role in the rating prediction, while the global measure reputation does not have necessary relationship with the accuracy of the recommendations.

4. CONCLUSIONS AND FUTURE WORK

We use five sets of experiments to examine the relationship between the recommender's reputation and the accuracy of its recommendations. The experiments are held on the global rating prediction mechanism of the trust-aware recommender system, where reputation is achieved by aggregating the user trust in the trust network. Since the trust network is the scale-free network, some users have superior reputations than other users, and these users continuously have high reputations. The experimental results clearly show that more reputable recommenders do not tend to give more accurate recommendations. Not only the recommenders with superior reputations, but also the recommenders actively participant in TARS, do not tend to gave more accurate recommendations than other recommenders. The experimental results imply that when utilizing the global rating prediction mechanism, the performance of TARS will not be improved by applying the weighted methodology, at least no better than the simple calculation of the mean value of the recommendations.

Besides applying the global rating prediction mechanism, there are some other ways to reduce the computational complexity of the existing TARS models. For example, we can deeply mining the trust relationship between users, use graph mining or other related methods to reduce the scale of each user's trust network, e.g., use the periodic pattern recognition finding the users in the same user group, and use the co-current pattern recognition finding the users in the similar user groups, assuming the users in the same user group and similar group have similar tastes, then build up a core trust network recording the most trusted and valuable users to each active user. The effective global rating prediction mechanism and the efficient personalized rating prediction mechanism are our future directions on the work of the trust-aware recommender systems.

5. ACKNOWLEDGMENTS

This research was supported by the MKE (The Ministry of Knowledge Economy), Korea, under the ITRC (Information Technology Research Center) support program supervised by the NIPA (National IT Industry Promotion Agency)" (NIPA-2012-(H0301-12-2001). Weiwei Yuan and Donghai Guan's work were also supported by National Natural Science Foundation of China (Grant No. 61100007, 61100081), and Fundamental Research Funds for the Central Universities (HEUCFR1120, HEUCF041204).

6. REFERENCES

[1] Yuan, W., Han, Y., Guan, D., Lee, Y.K., and Lee, S. Efficient routing on finding recommenders for trust-aware recommender systems. *Proc. of the 6th Int. Conf. on Ubiquitous Information Management and Communication (ICUIMC '12)*, 2012, Article No. 29.

[2] Yuan, W., Guan, D., Shu, L, and Niu, J. Efficient Searching Mechanism for Trust-Aware Recommender Systems Based on Scale-Freeness of Trust Networks. *Proc. of IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*, 2012, pp. 1819-1823.

[3] Yuan, W., Guan, D., Lee, Y.K., Lee, S., and Hur, S. J. Improved trust-aware recommender system using small-worldness of trust networks. *Knowledge-Based Systems* 23 (2010) 232-238.

[4] Yuan, W., Guan, D., Lee, Y.K., and Lee, S. The Small-World Trust Network. *Applied Intelligence* (2010): 1-12, April 27, 2010.

[5] Yuan, W., Guan, D., Lee, Y.K., and Lee, S. iTARS: Trust-Aware Recommender System using Implicit Trust Networks. *IET Communications* 14 (2010) 1709 - 1721.

[6] Massa, P., and Avesani, P.. Trust-aware Collaborative Filtering for Recommender Systems. *Proc. Of Federated Int. Conf. on the Move to Meaningful Internet*, 2004, pp. 492-508.

[7] Massa, P., and Avesani, P.. Trust Metrics in Recommender Systems. *Proc. of Computing With Social Trust*, 2009, pp. 259-285.

[8] Li, Y. and Kao, C.,.TREPPS: A Trust-based Recommender System for Peer Production Services. *Expert Systems with Applications*. 36 (2009) 3263-3277.

[9] Walter, F., Battistion, S. and Schweitzer, F.. A model of a

trust-based recommendation on a social network. *Autonomous Agents and Multi-Agent System*. 16 (2008) 57-74.

[10] Massa, P., and Avesani, P.. Trust-aware recommender systems. *Proc. Of the 2007 ACM Conference on Recommender Systems*, 2007, pp. 121-126.

[11] Jøsang, A., Ismail R., and Boyd C.. A Survey of trust and reputation systems for online service provision. *Decision support systems*, Vol. 43, Is. 2, 2007, pp. 618–644.

[12] Watts, D. and Strogatz, S.. Collective dynamics of 'small-world' networks. Nature, 1998, 393, pp. 440-442.

[13] http://www.trustlet.org/wiki/Epinions_dataset