

Modeling for Comment Trust Recommendation Based on Collaborative Filtering

Xinkao Liao

Department of Electronic and Information Engineering
Tongji University
Shanghai, 201804, China
E-mail: lxk861201@126.com

Lisheng Wang

Department of Electronic and Information Engineering
Tongji University
Shanghai, 201804, China, E-mail:
lishwang@tongji.edu.cn

Xiaojian Liu

Department of Electronic and Information Engineering
Tongji University
Shanghai, 201804, China
E-mail: 392501169@qq.com

Xiaojie Xu

Department of Electronic and Information Engineering
Tongji University
Shanghai, 201804, China
E-mail: 915600634@qq.com

Abstract— Asymmetry in the parties of the transaction leads to uncertainty in the transaction. Trust problem has been one of the bottlenecks restricting the development of e-commerce. For e-commerce product review issues, comment trust recommendation model was proposed based on comment credibility degree and user similarity, which combined with social networking trust mechanism and collaborative filtering to offer users with a more personalized trust recommendations. The experimental results demonstrate that the model can effectively improve the recommendation accuracy and solve cold startup problems in collaborative filtering.

Keywords— *E-commerce; Collaborative Filtering; PeopleRank; Trust Recommendation; Cold Startup Problem;*

I. INTRODUCTION

With the rapid development and improvement of living standards and e-commerce, online shopping has been one of the most popular online activities. Existing online shopping platform users can already feedback reviews after purchase. However, the evaluation of information still has the low credibility problem. In addition, each user feedback standards are quite different, and existing methods cannot provide personalized recommendation method for buyer.

Existing trusted computing focuses on the construction and analysis of influencing factors of trust models. Trust evaluation on online buyers and sellers is accomplished by eBay and Amazon online shopping websites through simple weighted average values with central establishment and maintenance of trust relationship. Some unreasonable problems, including irrationality of credit evaluation and credit deceit, eBay trust model were mainly analyzed in the literature[1]. Gan[2] proposed a trust computation method of multiple dimensions aiming at e-commerce social network. However, it is hard to guarantee the credibility and reliability of recommendation owing to the

lack of the comparison of similarity. Meanwhile, due to absent corresponding evaluation rewards and punishment mechanism, it fails to inhibit deceptive behaviors. Adding variables such as the number of trades and transaction value on the basis of Sporas model, Guo[3] put forward a modified computation model in terms of reputation value members in the C2C community but impact of factors, for example, time on reputation value was not yet considered in the paper. Liang[4] acquired the reputation value that users aggregate through simple accumulation of comments. It is easy to get the value via the algorithm. Wang[5] distinguished feedback ratings of familiar nodes from that of strange ones, proposing a semi-weighted reputation aggregation formula. However, it is difficult to find the familiar nodes due to the sparsity of virtual network. This is also the case for distinguishing familiar nodes from strange ones. Kamvar[6] raised a global trust management model EigenTrust which is used to simulate trust mechanism in the realistic society. However, being an iterative formula, it costs too much and fails to recognize malicious nodes effectively. Based on reputation calculation, Yu[7] trust model works out trust value for those who do not connect with others directly. Mui[8] set up a model to calculate trust and reputation from the perspective of sociology and evolutionism. Reputation value is aggregated by BETA density function in the Best algorithm in the literature[9]. A prediction technique of trust relationship for classification was put forward based on user similarity and globe reputation value according to sociological theory in the literature[10]. Malicious nodes can easily elevate its own global reputation value via the construction of anonymous neighbor nodes.

However, these trust computing methods all failed to deal with trust evaluation effectively. For different users, their comments on specific product will have certain impact on purchase. How to mine potential trusted feedback for target user in an ocean of comments on specific product? Comment trust recommendation model is

proposed combined with social network trust mechanism and collaborative filtering in the paper to provide more personalized recommendation of comments for users and help for purchase strategy of users finally.

II. DESCRIPTION OF COMMENTS TRUST RECOMMENDED PROBLEM

E-business environment mainly includes the following elements: user, product and comment. Trust relationships between users constitute the user trust network, as showed in Fig .1. Each node represents each user. And each edge is on behalf of the trust relationship between users. The weight of the edge is the strength of trust between users. Information of the commodity comment constitutes the network user comment network, as showed in Fig.2. Each node is a user or a product. Each edge is on behalf of user comment for the product. And the weight of the edge represents the score of the comment.

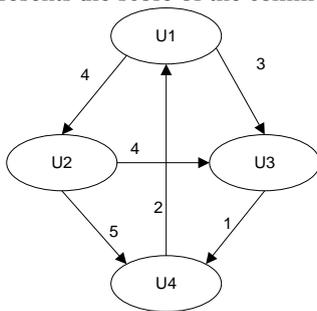


Figure 1. The user trust network

Each user can evaluate the product. These comments directly affect the views of other users who are going to buy this product. However, that excessive comment information causing the user cannot identify which comment is authentic. As show in Fig .2, users {U1,U2,U3} have published the comment on product I1. If the user U4 wishes to purchase the product I1, then the comment will be more credible. Therefore, it is necessary to find effective ways to tap the potential of credible comment information, in order to provide users with a more reliable product comments recommended.

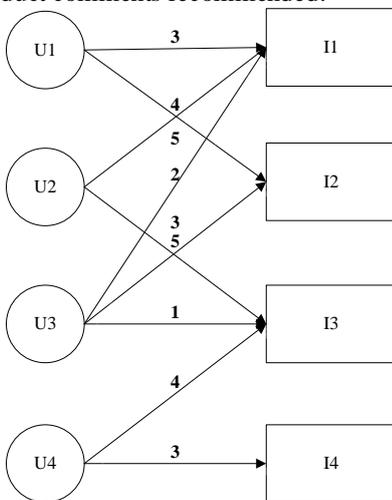


Figure 2. The user comment network

III. PRELIMINARY KNOWLEDGE

A. PageRank and PeopleRank algorithm

PageRank is used to measure the relative importance of a particular page in a search engine. Now it is being widely used by the Google Web search engine. It was invented by the late Larry Page and Sergey Brin in the 1990s. It sorts the pages by the relative importance in the network. The mathematical expression of the equation as follows:

$$PR(P_i) = \frac{1-d}{n} + d \sum_{P_j \in M(i)} \frac{PR(P_j)}{L(j)} \quad (1)$$

{P1; P2; P3...Pn } represents different pages, M(i) is a collection of page which links to pi, L(j) is the number of external links on the page pj , d(0 < d < 1) is the damping factor.

According to the idea of PageRank, researchers can define the PeopleRank algorithm in the social network. The basic idea of PeopleRank algorithm that if the number of friends are more, the more important the user will be in the social network, and then the trust level should be higher. The mathematical expression of the equation as follows:

$$PeR(N_i) = \frac{1-d}{n} + d * \sum_{N_j \in F(i)} \frac{PeR(N_j)}{F(j)} \quad (2)$$

PeR(Ni) represents the reputation value of user Ni, {N1;N2;N3...Nn} is on behalf of social network users, F(i) represents the set of users that have socialrelationships with Ni, d(0 < d < 1) is a damping factor.

B. Collaborative Filtering

Collaborative filtering is the most successful algorithm for recommendation. It can be divided into item-based collaborative filtering and user-based collaborative filtering. The user-based collaborative filtering first found the set of similar neighbor user, and then recommended interesting items in the neighbor set to the target user. The item-based collaborative filtering algorithm first finds similar neighbor set, and then these neighbor set is recommended to the target user. Computing similarity mainly includes the following methods: Euclidean distance, Pearson correlation coefficient, Cosine distance and so on. Collaborative filtering algorithm has an assumption that ratings reflect the preferences and interest of the user. But collaborative filtering algorithm exist data sparse and cold start problems. Literature[11] proposed that can effectively solve the problem of data sparsity by using trust mechanism to find the neighbor users.

IV. COMMENT TRUST RECOMMENDATION MODEL

Here establish the trust relationship information on user reviews generally consider the following factors:

- Comment credibility: If the reputation of the reviewer is higher, then the credibility he commented information is also higher.
- User similarity: In two user history evaluation, if the same goods they have similar evaluation, it can be speculated that these two users on the same goods are more likely to have a similar rating. So if two users have a similar evaluation, they will trust each other more.

A. Comments Credibility

Comments credibility primarily considers the reviewer. User trust created by reputation is a global trust. In PeopleRank algorithm, trust is segmented by the number of neighboring nodes. However, in social networks, trust weights in the collection are not completely equivalent. Therefore, the value of each node to other nodes reputation contribution should be distributed according to the weight ratio rather than simply the number of neighbor nodes. Improved PeopleRank mathematical expression is as follows:

$$\text{Re } p(N_u) = \frac{1-d}{n} + d \sum_{N_v \in F(u)} \frac{W_{v,u} * \text{Re } p(N_v)}{\sum W(u)} \quad (3)$$

PeR(Nu) represents the reputation value of user Nu, {N1;N2;N3...Nn} on behalf of social network users, F(u) Represents a collection that has social relationships with user Nu. W(v,u) indicates that the strength of value that user v trust user u. In the trust network architecture means that weights between node u and node v. $\sum F(u)$ represents the sum of the trust weight to user u.

However, in a single global reputation that do not involve trading related information, customer value influence on comments credibility cannot be reflected. In CRM Customer Classification, researchers usually measure customer behavior characteristics and analyze customer intrinsic value using the RFM model[12]. RFM model is used to calculate the measurement of buyers during a period of time. R represents recency, which is used to measure conditions of customers from the latest deal to current time. R is bigger when the time is shorter. F represents frequency, which refers to the number of purchase during a period of time. F is bigger with more trades. M represents monetary, which is the amount of purchase during the period of time. M increases with bigger amount[13]. The formula as follows[14] :

$$S_{RFM}(i) = 0.2(W_R S_R(i) + W_F S_F(i) + W_M S_M(i)) \quad (4)$$

{SR(Nu);SF(Nu);SM(Nu)} denote the Recency, Frequency and Monetary value of the user Nu, {WR;WF ;WM } are the weight for the three indicators, this paper were taken 0.22,0.35,0.43. Combined RFM customer model, the mathematical expression evaluates as follow:

$$\text{Trust}(N_u) = \text{Re } p(N_u) * S_{RFM}(N_u) \quad (5)$$

B. User Similarity

Similar trust is based on the idea of collaborative filtering of users. Grades of users on products constitute an item score matrix when one item represents one product. When researchers predict the score a target user gives one product, researchers firstly work out score information of similar users through calculation. Finally, researchers give the target user feedback of similar users. The similar users are calculated by cosine similarity the range of which is [0,1]. The comment on specific product of one user reflects to some extent his interest and taste. In this text, the product comment matrix of users is chosen as the calculation basis of similarity between users. It is expected two users are more likely to give similar comments on the same product in the future if there are similar comments on the same product in the past. That is to say, they will trust each other in terms of comments if they have similar

comments on the same product. The calculation formula is as follows:

$$\text{Sim}(u,v) = \frac{\sum_{i=1}^n u_i v_i}{\sqrt{\sum_{i=1}^n u_i^2 \sum_{i=1}^n v_i^2}} * \frac{\|U\|}{\|V\|} \quad (6)$$

Among them, n is the number of products the user v and user u both commented on. u_i and v_i are the scores of the product i that the user u and user v gives. The similarity between two users is higher if the value is closer to 1. On the contrary, the differences between them is bigger when the value is smaller. And U and V represent the number of commented products for the user u and user v. The number of comments on products will be one of the influencing factors on user similarity.

Based on the above analysis, trust recommend of comment users are influenced by comment credibility and similar trust. Hence, the formula about comment trust recommendation for a target user v is as following:

$$\text{Trust}(N_u, N_v) = \alpha * \text{Trust}(N_u) + \beta * \text{Sim}(N_u, N_v) \quad (7)$$

Among them, α and β are experienced parameters that control the influence degree of comment credibility and recommendation trust with similar credibility generates. $\alpha + \beta = 1$.

V. EXPERIMENT AND ANALYSIS

In order to verify the validity and accuracy of the proposed method, this paper selects online social network datasets. researchers randomly selected 2000 product nodes and user nodes to build user trust network and user comment network. Comment reliability was computed by trusted computing network and customer RFM value model, and the product reviews evaluated the similarity of network computing users trust value. During the experiment, mainly consider the following MAE(Mean Absolute Error) indicator to judge the accuracy of experimental results.

$$\text{MAE} = \frac{\sum |T_{i,j} - P_{i,j}|}{n} \quad (9)$$

Here n represents the total number of data, $T_{i,j}$ represents the actual value, $P_{i,j}$ indicate the predicted value.

A. Parameter Experimental Analysis

Parameters in comment trust recommendation model affects the proper weight distribution between comment credibility and user similarity. The introduction of the trust mechanism improved cold-start problem to a certain extent. Thus, different parameters have different effects on different types of users. So two types of users that ordinary users and cold-start users were chosen. The results were shown in Fig .3 and Fig .4. For the ordinary user, it could be observed in the figure, with the value of parameter increment. MAE value was bottoming out slowly at _rst and then go pick up the next. At (0.42, 0.58), MAE reached the optimal situation. For cold start user, with the value of parameter increment, MAE value was also lower bottom out slowly at first and then picked up, but the decrease was significantly slower than the ordinary user, and the recovery rate was also steeper than the ordinary user. At (0.67, 0.33), MAE reached the optimal situation.

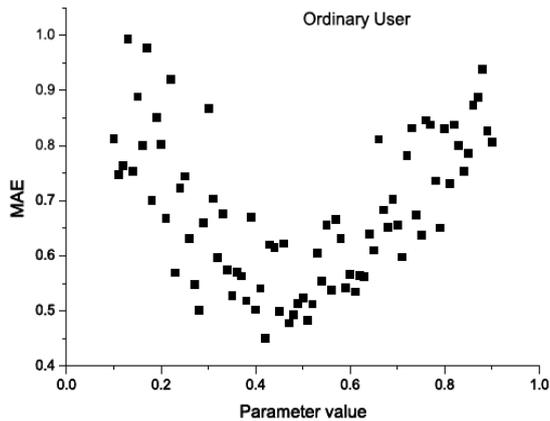


Figure 3. The ordinary user result

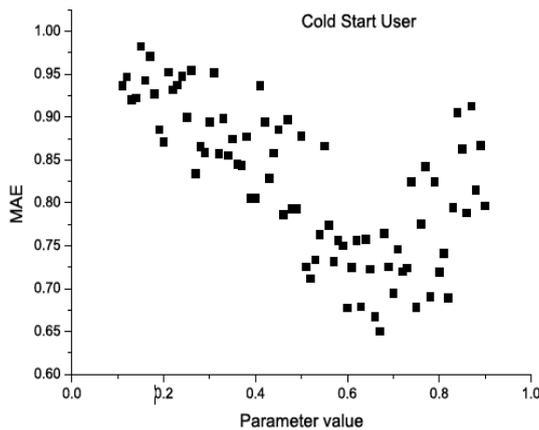


Figure 4. The cold start user result

B. Comparative Experimental Analysis

In the comparative experimental analysis, the sample data were divided into two parts, 80 percent of the data as training set and 20 percent of data as a test set to examine the results of this experiment. Finally, the experimental results were compared with other methods. The results were shown in Table 1. From Table 1, researchers could see that the CTRM algorithm improved the accuracy of the recommendation to a great extent compared with the other method. For the ordinary user, collaborative filtering MAE is 0.6423, while PeopleRank algorithm is slightly higher. However, PeopleRank was better than collaborative filtering algorithms for cold start user. Compared with PeopleRank and collaborative filtering, the literature[10] also increased the recommended effective to some extent. CTRM algorithm introduced trust mechanism into collaborative filtering and effectively improves the cold start problem. Therefore, compared to literature[10], the CTRM algorithm improved the accuracy of recommendation to a certain extent.

TABLE I. TABLE I MAE RESULT COMPARE WITH OTHER ALGORITHM

Algorithm	Ordinary User	Cold Start User
Collaborative Filtering	0.6423	0.9342
PeopleRank	0.7376	0.8028
Literature[10]	0.4917	0.7348
CTRM	0.4559	0.6501

VI. CONCLUSION

This paper proposes a more reasonable comment trust recommendation model based on improved PeopleRank algorithm and collaborative filtering. The simulation results show that this model can effectively improve the accuracy of prediction of recommendation. The reputation value will be more accurate to reflect real credibility for the buyer. However, the CTRM model still has some shortcomings, such as how to reduce the complexity.

ACKNOWLEDGMENT

Acknowledge National High Technology Research and Development Program 863 (No. 2013AA040302)..

REFERENCES

- [1] CHEN Xicun, ZHOU Huan, PIAO Wen. Analysis of eBay's virtual terminal network[J]. Journal of Liaoning Technical University(Social Science Edition),2010(12).
- [2] GAN Zao-Bin, DING Qian, LI Kai, XIAO Guo-Qiang. Reputation-Based Multi-Dimensional Trust Algorithm[J]. Journal of Software,2011,22(10):24012411.
- [3] GUO Honghai JIANG Jinhua CAI Han. Modeling for Reputation Computing in C2C communities. Chinese Journal of Management.2009.
- [4] Liang Z,shi W. Enforcing cooperative resource sharing in untrusted peer-to-peer environment[J]. ACM Journal of Mobile Networks and Applications(MONET),2005,10(6):771-783.
- [5] Wang Y,Vassileva J. Bayesian network-based trust model[C]//Web Intelligence,2003. WI2003. Proceedings. IEEE, 2003: 372-378.
- [6] Kamvar et. The Eigentrust Algorithm for Reputation Management in P2P Network, Proceedings of the 12th International World Wide Web Conference, Budapest, Hungary: ACM Press, 2003:640-651.
- [7] Yu B. Singh M P. A social mechanism of reputation management in electronic communities. In: Proceeding: fmg Cooperative Information Agents, 7th International conference,2000: 154-165P.
- [8] Mui L. Computational Models of Trust and Reputation [D]. PhD thesis, MIT, 2002.
- [9] Josang A, Ismail R. The beta reputation system[C]//Proceedings of the 15th blede electronic commerce conference. 2002: 41-55.
- [10] D O'Doherty, S Jouili, P Van Roy. Towards trust inference from bipartite social networks[C]// Proceedings of the 2nd ACM SIGMOD Workshop on Databases and Social Networks, Scottsdale, AZ, USA, 2012:13-18.
- [11] Paolo Massa, Bobby Bhattacharjee. Using Trust in Recommender Systems: An Experimental Analysis. Second International Conference on Trust Management, 2004.
- [12] LIN Sheng, XIAO Xu. A method of telecom consumer market segmentation based on the RFM model[J]. Journal of Harbin Institute of Technology, 2006,38(5):758-760.
- [13] Miglatsch J. Thoughts on RFM scoring[J]. The Journal of Database Marketing, 2000,8(1):35-43
- [14] XIONG Jianying. Study on Trust Model of C2C E-Commerce Based on Feedback[D]. PhD thesis, Finance and Economics University of Jiangxi, 2013.

- [15] Golbeck J. Trust and nuanced Profile similarity in online social networks. *ACM Trans. Web*, 2009, 3(4): 1-33
- [16] Elizabeth BLA, R., Kiruthika, P., Shrada, M.N., et al. Bayesian based confidence model for trust inference in MANETs. *Recent Trends in Information Technology (ICRTIT)*, 2011 International Conference on, 2011. 402-406.
- [17] Xie L, Jia X, Zhou K. QoS Multicast Routing in Cognitive Radio Ad Hoc Networks. *International Journal of Communication Systems*, 2012, 25(1): 30-46.
- [18] Hoque M, Hong X. BioStaR: A Bio-inspired Stable Routing for Cognitive Radio Networks. *Proceedings of 2012 International Conference on Computing, Networking and Communications*. Washington: IEEE Computer Society, 2012. 402-406.
- [19] Hofmann T. Latent Semantic Models for Collaborative Filtering [J]. *ACM Transaction Information Systems*, 2004, 22(1): 89-1
- [20] Balabanovic M, Shoham Y. Fab: Content-based Collaborative Recommendation [J]. *Communications of the ACM*, 1997, 40(3): 66-72.