Proceeding of the 2015 IEEE
International Conference on Information and Automation
Lijiang, China, August 2015

A Matrix Factorization Based Trust Factors Model *

Suhuan Sun, Changwei Zhao and Zhiyong Zhang
School of Information Engineering
Henan University of Science & Technology
Luoyang, China,471023
{sunsuhuan &_zhao_chw}@163.com, xidianzzy@126.com

***Abstract*: Considering the complexity of trust among users, we presented a new trust factor model. In this model, trust relationships among users are represented as an adjacent matrix, and trust matrix can be decomposed into two low rank dimensionality matrixes, which are trust factor matrix of users and trusted factor matrix by users, through matrix factorization in this model. User-to-user trust values are computed as inner products of corresponding vectors., Trust users can be identified according sorts of trust values. The model can give consideration to the global structure and finer-grained characteristics of trust. The experimental results show that this model has higher accuracy of trust identification and efficiency in prediction stage. The Trust factors model not only reflect users' characteristics of trust and trusted, but also can be used to explain trust propagation in trust chain.**

Index Terms – Matrix factorization; Trust model; Trust Chain.

I. INTRODUCTION

With the current popularity of online social networks, more and more users are joining and more and more information is shared through social network services [1]. In the process of information shared or acquired, users in online social networks want to transaction with trust ones and to refuse unreliable ones. Then trust management is very important for social network services. Trust plays a significant role on transaction among users, and it helps identify users who we can communicate with, share information with, and form friendships with and so on. There are three methods can be used to establish trust relationships among users: transaction trust, trust chain and recommended trust [2]. The first method makes trust relations according historical information of transaction. It is a directed method. However, it needs to remain transaction information and trust users are restricted to limited users who communicated with. The second method apply transitivity characteristic of trust [3, 4], and trust relationships is established through trust chain. In general, trust is not perfectly transitive, but decreases along a chain of networks[5]. the third method depends on recommend information from authority users. But it is difficult to select authority users.

What is the basis of trust among users? Which features are users making a trust decision? And which features are users trusted? In this paper, a new method of establishing trust among users is proposed, and it can give a preliminary answer to the above question. The method is based on matrix factorization method and it maps both trust factors of users and trusted factors by users to a joint latent factor space of low rank dimensionality, such that user-to-user trust values are modeled as inner products in that space. The latent vectors try to explain trust values by features both user trust and user trusted on factors which automatically inferred from user feedback of trust relations, and it need not exogenous information about users. After we obtain both trust factors of users and trusted factors by users, we can explain why trust can be propagated and decreased along a chain of networks.

This text is organized as follows: In section II the related work is briefly introduced. A model of trust factors is presented in section III. Trust chains are explained in section IV and empirical evaluation of our methods are in Section V. Finally, in section VI, we give our conclusions and future research about trust management.

II. RELATED WORKS

Trust management is a hot topic in social network services, and there are lots of researchers who have studied Identifying trusted people and preventing distrust people from obtaining illegal profits. Blaze etal. present a comprehensive approach to trust management, based on a simple language for specifying trust actions and trust relationship, and implement a trust management prototype called PolicyMaker which can be used in a wide range of network service [6]. [7] presented a reputation model that takes into account the social dimension of agents and a hierarchical ontology structure that allows to consider several types of reputation at the same time. For the mobile agent based e-commerce environment, [8] proposed a reputation based multi -dimensional trust(RMDT) algorithm which makes use of a self-confident coefficient to synthesize the directed and the reference trustworthiness to evaluate the node in the network. RMDT can uncover the influence on trust computation caused by the subjective factors, such as individual predilection and risk attitude. [2] proposed selecting transaction nodes and trust computing between them in social network. service transaction nodes acquire service opportunity and gain payment through biding method ，which can incentive enthusiasm of services transaction nodes and improve trust relations.

In Peer-to-Peer (P2P) systems, Kamvar, Schlosser, and Garcia-Molina proposed an algorithm, called EigenTrust, which calculates unique global reputation values for each peer

* The work was sponsored by National Natural Science Foundation of China Grant No.61003234, Plan for Scientific Innovation Talent of Henan Province Grant No.134100510006.

in the system based on a peer's previous behavior [9]. Gong, Yang, Su, and Zhang presented a searching algorithm for a resource in an unstructured P2P system. It is fully based on social networks and makes use of trust relationships between peers [10]. In the multimedia social network environment, Zhang proposed a multimedia social network services trust management (MSNTM) based on small world theory. The model distinguishes direct trust and recommendation trust relationship between users, and proposed a direct trust computing window mechanism and trust recommend path finding algorithm according to the decay characteristics of trust with time and path, and comprehensively considering credible feedback of digital content, feedback weighting factor, user share similarity, and other share characterization factors [11].

Above all the models and methods, only link messages or local network structures be considered, but global network structures and all similarities among users are omit. For this reason, our paper represents a new trust factors model which considers the global structure information, and it is described as follows.

III. TRUST FACTORS MODEL

A. Notations

In order to give a formula describe, some basic notions are defined as follows.

U: a set of users. $|U|$ is number of users in set U. i, j, k represent different users, and we have $i, j, k \in U$.

T: trust matrix among users. T_{ij} is trust value of user i to user j. As only seldom trust values are known, matrix T is very sparse.

P: trust matrix of user. P_i is trust vector of user i. one item of P_i represent one trust factor of user i.

Q: trusted matrix by user, Q_i is i's trusted vector by other users. One item of Q_i represents one trusted factor of user i by other users.

$\hat{T}_{ij}$ represents predict trust value of user i to user j. For known P_i of user i and Q_j of user j, we have: $\hat{T}_{ij} = P_i \cdot Q_j^t$

B. The model of trust factors.

In the model of trust factors, trust relationships among users can be explained through trust factors. trust factors about one user are composed by two parts: user trust factors and user trusted factors. The process of building the trust model is getting factors of two parts, and methods of matrix factorization can be used to solve this question.

The core of trust matrix factorization is getting trust matrix of user P and trusted matrix by users Q through known trust values. To learn the matrix P and Q, the models minimizes the squared error on the set of known trust relationships. As known data of trust relationships is very sparse, regularized P_i and Q_j is necessary for preventing overfit. The object function described as follows:

$$\min \sum_{T_{ij} \in T} (T_{ij} - P_i \cdot Q_j^t)^2 + \lambda(\|P_i\|^2 + \|Q_j\|^2) \quad (1)$$

Where, T is the matrix of trust for which T_{ij} is known, and λ is regularized coefficient, which controls the extent of regularization, is usually determined through cross validation. The optimal object function is solved by either stochastic gradient descent or alternating least squares.

Alternating least squares techniques rotate between fixing the P_i to solve for the Q_j and fixing the Q_j to solve for the P_i. Notice that when one of these is taken as a constant, the optimization problem is quadratic and can be optimally solved by MSE algorithm.

An easy stochastic gradient descent optimization was popularized and successfully practiced by many others researches. The algorithm loops through all ratings in the known data. For each known T_{ij}, a prediction $\hat{T}_{ij}$ is made, and the associated prediction error $e_{ij} = T_{ij} - P_i \cdot Q_j^t$ is computed. For a given training trust case T_{ij}, we modify the parameters by moving in the opposite direction of the gradient, yielding:

$$P_i \leftarrow P_i + \gamma(e_{ij} \cdot Q_j - \lambda \cdot P_i) \quad (2)$$

$$Q_j \leftarrow Q_j + \gamma(e_{ij} \cdot P_i - \lambda \cdot Q_j) \quad (3)$$

where, γ is learning rate.

once optimal P_i and Q_j are get, we can predict final trust values $\hat{T}_{ij}$ applying formula(4) described as follows:

$$\hat{T}_{ij} = P_i \cdot Q_j^t \quad (4)$$

As trust values are predicted through simple inner products, so the method is very efficient in the stage of prediction trust.

IV. EXPLANATION OF TRUST CHAIN

A. Describing trust relationships

In social networks, trust chain is a basic concept about trust transitive. In general, trust can be transited, but decreases along a chain of networks. The mechanism of trust propagation and decay is not clear at present.

For a simple trust chain composed by users of i, j, k, Arrows indicate the direction of trust, "+" represent as trust and "-" represent as untrust. The figure of trust relationships can be described as Fig.1

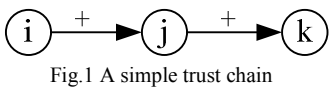

Fig.1 A simple trust chain

According the mechanism of trust propagation, generally, we have

$$T_{ik} = T_{ij} \cdot w_{jk} \cdot T_{jk} \quad (5)$$

Where, w_{jk} is weight of trust transitive or decay factor. At present, there is no direct theory for assigning the weights of trust transitive. The weights depend on user-specific character of trust. But users are generally not motivated to provide it. Therefore, some other ways of assigning weights include measuring the importance of a node based on out- and in-degrees or measuring the similarities between two nodes are presented. More generally, the weight is assigned subjectively the number which ranges from 0 to 1. In our method, the weights can be calculated through users self-trust value.

Trust chain method often meets with failure in computing trust value between users. Considering the chain composed by users of i, j, k, i trust j and j trust k, we have two types trust relationships that are i trust k, which we called positive trust, as Fig.2(a), and i untrust k, which we called negative trust, as Fig.2(b).

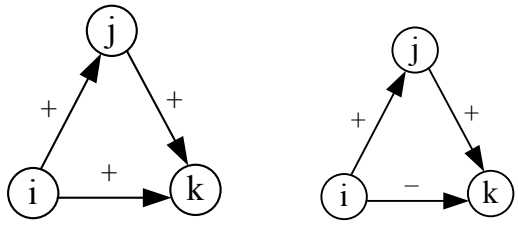

Fig.2(a) Positive trust Fig.2(b) Negative trust

In a trust relationship triangle, we call user i as the initial trust node, j as the transfer node and k as the end trust node. For positive trust, it can be easily explained by trust chain. But for negative trust, it is difficult. However, when we use trust factors model, two situations can be explained.

B. Explanation of trust chain

For users i, j, k, its trust and trusted latent factors are P_i, P_j, P_k and Q_i, Q_j, Q_k differently. the trust value of user i to j can be expressed as $\hat{T}_{ij} = P_i \cdot Q_j^t$ and user j to k as $\hat{T}_{jk} = P_j \cdot Q_k^t$, user i to k as $\hat{T}_{ik} = P_i \cdot Q_k^t$. Simply, if we view P_i, P_j, P_k and Q_i, Q_j, Q_k as unit vectors, the trust among users equal Cosine similarity of correlation vectors.

If we use cosine similarity for trust, the prediction trusts of user i to j can be expressed as:

$$\hat{T}_{ij} = \cos < P_i, Q_j > \approx 1 \quad (6)$$

and user j to k expressed as:

$$\hat{T}_{jk} = \cos < P_j, Q_k > \approx 1 \quad (7)$$

If the angles of P_i, Q_j and P_j, Q_k is very little, $\hat{T}_{ij}$ and $\hat{T}_{jk}$ are equal to 1 approximately. The relations of vectors can be described as Fig.3:

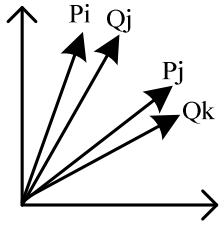

Fig.3 The relations of latent factors vectors

The trust value of user i to k can be expressed as:

$$\hat{T}_{ik} = \cos < P_i, Q_k > \approx \cos < P_j, Q_j > \quad (8)$$

It is inner product of P_i, Q_j, which we called as self-trust of user j

Thus, for short trust chain composed by three user nodes, the trust values from initial nodes to end nodes are related with self-trust of transfer nodes. The more the self-trust of transfer node is, the better the trust propagation performance is.

V. RESULTS

A. Data

To evaluate our presented method, we experimented with a directed, signed trust network of Epinions.com2[12]. Users in Epinions can indicate whether reviewers are trusted or distrusted. The original dataset contained 131,828 nodes (users) and 841,372 edges (trust or distrust). From this dataset, we selected nodes that have more than 50 out-neighbors in original dataset. Consequently the dataset used in our experiments has 3623 users and 285303 trust/distrust connections that link pairs of users, and data sparsity is 97.82164

Since the dataset is labelled, we randomly divided dataset into 5 equal segments and select one segment as test set, the other four segments as training set. 5-fold cross-validation method is used to test performance of our model.

B. Evaluation indexes

Considering the facts that there are two types of relationships among users, which are trust relationships and untrust relationships, we divided the performance evaluation into two perspectives：trust relationship identification and Error-Hit untrusted relationship identification.

In order to measure the accuracy of the trust user identification, two types of evaluation metrics were adopted: precision at top-n and recall at top-n.

$$P@n = \frac{1}{|U|} \sum_{i=1}^{|U|} \frac{|Test^+(i) \cap T_i|}{|T_i|} \quad (9)$$

$$R@n = \frac{1}{|U|} \sum_{i=1}^{|U|} \frac{|Test^+(i) \cap T_i|}{|Test^+(i)|} \quad (10)$$

where $Test^+(i)$ is the set of trusted users positively linked from user i in the test set and T_i is a top-n identified list with regard to user i .

For untrust relationships，we use the Error-Hit at top-n，which measures how often a model under examination includes untrustworthy users in a list of trusted users.

$$EH@n = \frac{1}{|U|}\sum_{i=1}^{|U|}\frac{|Test^{-}(i) \cap T_i|}{|T_i|} \quad (11)$$

where $Test^{-}(i)$ is the set of users distrusted by user i. The lower EH@n value, the more effectively an algorithm refuses distrusted users.

C. Baseline methods for trust link identification

For performance comparisons, we conducted experiments with several different algorithms designed for link prediction in social networks, including methods based on node neighbourhoods and methods based on the ensemble of all paths [13]. We evaluated the performance of our model in comparison with the following baseline methods: Common neighbours approach, Jaccard coefficient approach, Extended Advogato trust metric [5], Personalized PageRank approach etc.

D. Effect of parameters

In this section, we present empirical results that show the effects of the parameter adjustments. Our model has two parameters that can be set to a range of values: the number of trust factors and the value of regularization.

We first examined the performance with respect to number of factors that controls the complexity of models. Generally，the more complex the models，the more details can be showed through it. But, with the increase of complexity，the efficiency of the algorithm will decrease，and longer training and predicting time will be cost.

We compared the results obtained from setting different factors. when regularization= 0.015，Fig.4 shows recall, precision, and error-hit at top-10 according to different factors, and Fig.5 shows recall, precision, and error-hit at top-5 according to different factors.

From Fig.4 and Fig.5, it can be seen that：When the number of trust factors is less than a certain value, the precision and recall will increase with the number of factors. When the number of trust factors exceeds a certain value, the precision and recall will increase slowly, and may even reduce. The experiments results show that trust is complex. It is inappropriate only using one-dimensional numerical for representation trust characters. but trust metric dimension is not always better . When the trust dimension is too large, there may suffer problems of overfit or introducing noise, and it can cause performance degradation. On the other hand, with the increase of dimension confidence factor, the computational complexity will increase. Therefore, in the process of trust identification, selecting appropriate dimension of trust factors is necessary.

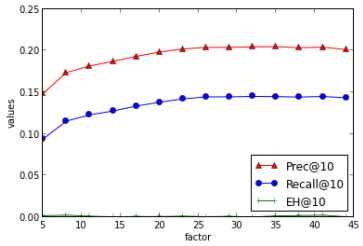

Fig.4 Recall, precision, and error-hit values at top-10 according to different factors

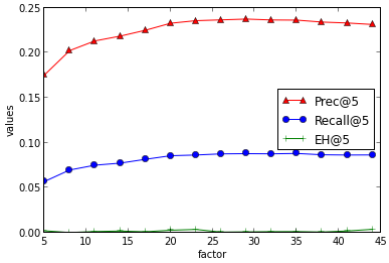

Fig.5 Recall, precision, and error-hit values at top-10 according to different factors

We continued to look into the effect of change regularization on the performance of trust identification, and results are shown as Fig.6.

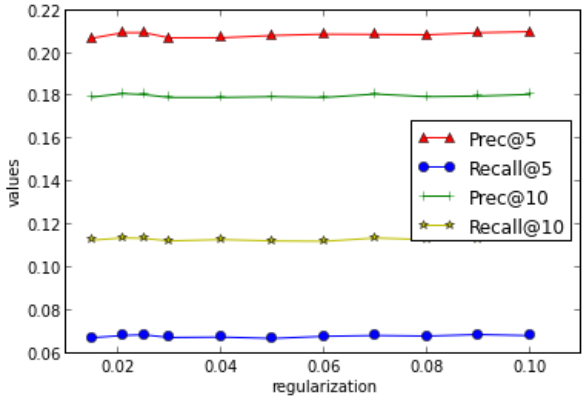

Fig.6The performance of trust identification according to different regularization

Generally, Regularization can avoid overfitting problem in the method of sparse matrix factorization. But, in this experiment, there is little affected on the performance of trust identification about regularization parameter adjustment. One possible cause could be that we only retained users which out-degree is greater than 50 in data selection process , and the degree of users is close.

E. Comparison results

The following experiment compares our matrix factorization method (denoted MF) to other baseline methods Common neighbors approach，Jaccard coefficient approach，Extended Advogato trust metric[5]，Personalized PageRank approach. We selectively varied the number of n about top-n from 5 to 30 with an increment of 5. Each result is plotted as data points on the graph curves. for MF. trust factors was set to10,and regularization was set to 0.015. For

ExAdvogato, the decay factor was set to 0.5, and power capacity was set to 6. Fig. 7 shows the precision-recall curves for each method.

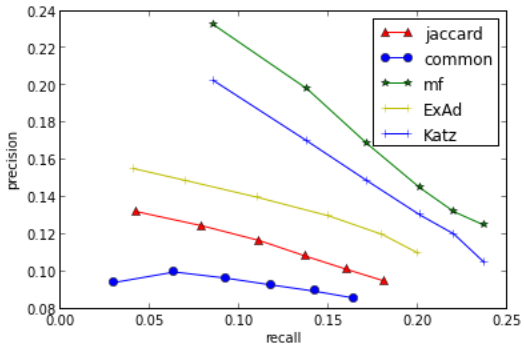

Fig. 7 Comparison of precision and recall values at different top-n.

The graph shows that matrix factorization achieved the best performance. An important reason is the method considers the global link structure. Compared with Katz, which considers the ensemble of all links, matrix factorization improved precision by 3%. Compared with other method, which only consider the neighbour, our method has certain advantages in trust identification.

F. Explanation of trust chain

In the trust propagation process, it is generally believed that trust will be decay with the length of trust chain increases. But there is no explanation about it. On the other hand, negative propagation often appears as in Figure 2. (b). For the condition of trust propagated, the preliminary experiments was carried out in this paper. In our experiment, we considered trust relationships which composed by three users as Figure 2. (a), which we called as trust triangle, and Figure 2. (b), which we called as distrust triangle. In our data set, there are 7,189,669 trust triangle and 109406 distrust triangle.

For each transfer user j, we calculated trust triangle numbers and distrust triangle numbers. The ratio reflects in part trust propagation probability. Fig.8 shows self-trust vs ratio about each transfer users, which ratio is less than 10. Fig.9 shows self-trust vs ratio about each transfer users, which ratio is greater than 10.

Fig.9 shows: for most transfer users, when self-trust about ones is little, the probability of trust propagation is low. but, some portion transfer users self-trust is big. Reasons for this phenomenon may be only three users trust chain considered and a number of chains greater than 3 are ignored.

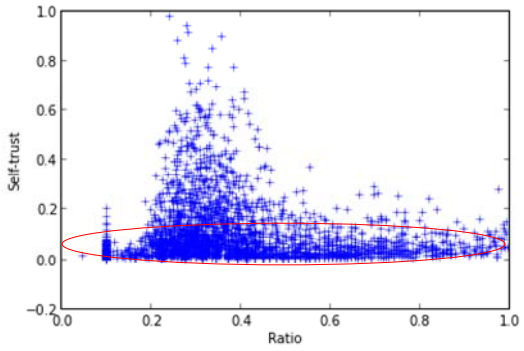

Fig.9 self-trust vs ratio (ratio<10)

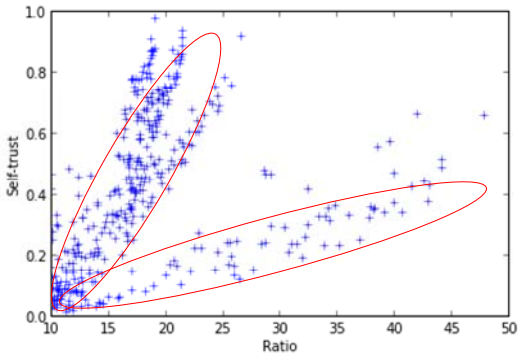

Fig.10 self-trust vs ratio (ratio>10)

Fig.10 shows that self-trust of transfer users is the key of trust propagation in trust chain. when the probability of trust propagation is big, self-trust of transfer nodes will increase with the probability of trust propagation. In trust chain, the larger self-trust of transfer user is, the more probability of transfer trust is. The figure reflects rules of trust propagation in some extent. Trust propagation is a complex issue. The data can be divided into two categories according distribution about values of transfer users' self-trust. The possible reason for this phenomenon is different user link structure types.

VI. Conclusion

In this paper , we present a new trust model based on matrix factorization method. The model maps both trust factors of users and trusted factors by users, which can reflect trust characters of users, to a joint latent factor space of low rank dimensionality , such that user-to-user trust values are modeled as inner products in that space. The latent space tries to explain trust values by characterizing both user trust and user trusted on factors which automatically inferred from user feedback of trust relations, and it need not exogenous information about users. This model has higher accuracy of trust identification and it's efficiency in prediction stage. We also analysed trust chain using trust factors model. In trust chain, the larger self-trust of transfer user is , the more probability of transfer trust is. In our model, we only researched building trust and transfer trust in static social networks. Considering the fact that building relationships among users is a dynamic process , next step , we will future research the rule of trust evolution and building dynamic model for trust。

References

[1] Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. Journal of Computer- Mediated Communication, 13(1), 210-230.

[2] Wang, G., & Gui, X. L. (2013). Selecting and trust computing for transaction nodes in online social networks. Jisuanji Xuebao(Chinese Journal of Computers), 36(2), 368-383.

[3] Golbeck, J. A. (2005). Computing and applying trust in web-based social networks.

[4] Golbeck, J., & Hendler, J. (2006). Inferring binary trust relationships in web-based social networks. ACM Transactions on Internet Technology (TOIT),6(4), 497-529.
[5] Al-Oufi, S., Kim, H. N., & El Saddik, A. (2012). A group trust metric for identifying people of trust in online social networks. Expert Systems with Applications, 39(18), 13173-13181.
[6] Blaze, M., Feigenbaum, J., & Lacy, J. (1996, May). Decentralized trust management. In Security and Privacy, 1996. Proceedings., 1996 IEEE Symposium on (pp. 164-173). IEEE.
[7] Sabater, J., & Sierra, C. (2001, May). Regret: A reputation model for gregarious societies. In Fourth workshop on deception fraud and trust in agent societies (Vol. 70).
[8] Gan, Z. B., Ding, Q., Li, K., & Xiao, G. Q. (2011). Reputation-based multi-dimensional trust algorithm. Ruanjian Xuebao/Journal of Software, 22(10), 2401-2411.
[9] Kamvar, S. D., Schlosser, M. T., & Garcia-Molina, H. (2003, May). The eigentrust algorithm for reputation management in p2p networks. InProceedings of the 12th international conference on World Wide Web (pp. 640-651). ACM.
[10] Gong, Y., Yang, F., Su, S., & Zhang, G. (2009, December). Improve peer cooperation using social peer-to-peer networks. In Information Science and Engineering (ICISE), 2009 1st International Conference on (pp. 253-257). IEEE.
[11] Zhang, Z., & Wang, K. (2012). A trust model for multimedia social networks.Social Network Analysis and Mining, 1-11.
[12] Leskovec, J., Huttenlocher, D., & Kleinberg, J. (2010, April). Predicting positive and negative links in online social networks. In Proceedings of the 19th international conference on World wide web (pp. 641-650). ACM.
[13] Liben- Nowell, D., & Kleinberg, J. (2007). The link- prediction problem for social networks. Journal of the American society for information science and technology, 58(7), 1019-1031.