Heterogeneous Trust-Aware Recommender Systems in Social Network

Na Wang, Xia Li, Zhaonan Chen
College of Information Engineering
Shenzhen University
Shenzhen, China
e-mail: wangna@szu.edu.cn
e-mail: lixia@cuhk.edu.cn
e-mail: 2150130423@email.szu.edu.cn

***Abstract*— Trust, as the basis of human interactions, has been playing an important role in addressing information sharing, experience communication, and public opinions. Trust-aware recommender systems are an effective solution to the information overload problem, especially in the online world where we are constantly faced with inordinately many choices. In this paper, to build a trust-aware recommender system with enhanced accuracy of recommendation, a novel approach is proposed which incorporates multi-faceted trust relationships between users into traditional rating prediction algorithms to reliably estimate users multi-faceted and asymmetry trust strengths. Experimental results on real-world data show that our work of discerning heterogeneous trust can be applied to improve the performance rating prediction and more robust to the cold-start problem.**

Keywords-component: Heterogeneous Trust; Trust-aware recommender systems; social media

I. INTRODUCTION

Popularized Online Social Network (OSN) applications, especially the Social Network Sites (SNSs), provide people with great convenience for information sharing, collaborating and interacting [1], [2]. The prevalent use of social media generates massive data at an unprecedented rate, which makes the information overload problem increasingly severe. Trust in online social networks plays a central role in exchanging relationships involving unknown risk, which provides information about with whom we should share information and from whom we should accept information [3], [4]. Due to the openness and anonymity of the internet, how to provide reasonable trust computing mechanism for users in online social networks is an urgent problem to solve.

In Epinions, for example (which is the focus of this paper), users can write and share their reviews about products, the rest of the user community comments and rates the reviews. Additionally, users can specify whom they trust. These trust connections constitute the user's trust network. To a certain extent, users are likely to have similar preferences with their trust networks. Furthermore, we are more willing to buy an item from a particular seller on E-Bay or Amazon, if either we or our friends had a positive experience with that seller in past. On the other hand, we are reluctant to engage in any relationship with strangers. According to a consumer of psychological tests, people are more likely to accept the recommendation of friends when choosing commodities. The trustworthiness of the users is often tantamount to the reliability of the information they provide, which is widely exploited in collaborative filtering [5], [6], intelligent recommender systems [7], review quality prediction [8], [9] and viral marketing [1], [10], to improve the accuracy.

Trust is especially critical in some online communities such as e-commerce sites and product review site, but most existing works on online trust assumes single and homogeneous trust relationships between users [6], [8], [9]. However, several theories in sociology show that the effect of the trust from different angles (topics) may be different [7], [11]. Trust takes many different meanings and highly depends on the context in which users interact with each other, indicating multiple and heterogeneous trust relationships between users. For example, Figure 1(a) demonstrates a single trust relationship between a real user from Epinions, represented by user 1, and her 5 representative friends. Figures 1(b) and 1(c) show their multi-faceted trust relationships in the categories of "computer hardware" and "family", respectively. The width of arcs in these figures indicates their trust strengths. A user may trust user 3 in computer hardware category while not trust him/her in family category. But when assuming a single trust relationship, user 2 is patently the most trustworthy person. This example suggests that treat trust relationships of different categories equally cannot capture the real relationship between users.

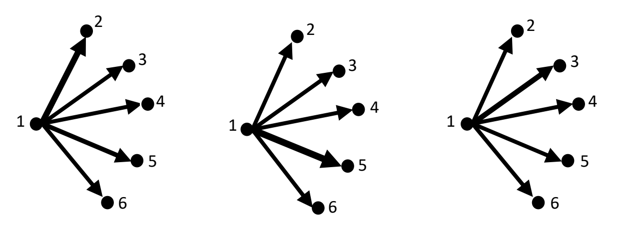


Figure 1. Single Trust and Multi-Faceted Trust Relationships of One User in Epinions

In this paper, we study the problem of trust prediction by exploiting discerning multifaceted trust relationships in the context of online product review sites. In essence, we extend the matrix representation to a tensor representation by adding an extra dimension faces, then calculate the strengths of multi-faceted and asymmetric trust relationships. we also address the cold-star problem when no historical rating on items or users are available. By incorporating these into rating prediction, performance improved significantly. Our empirical study on two real-world social media datasets demonstrates the effectiveness of our model.

The rest of paper is organized as follows. The problem of trust is formally stated in section 2. Section 3 introduces a fine-grained representation for multi-faceted trust relationships and describes our method to estimate their heterogeneous and asymmetric trust strengths. Experimental results on real-world data are demonstrated to prove the advantages of this model in section 4. Finally, section 5 concludes this study with future work.

II. PROBLEM STATEMENT

A. *Problem Definition*

Before building the mathematical model, we would like to establish the notations that are used. A rating system of product review sites consists of two types of objects with two different actions, as shown in Fig 1(a). Let $U = \{u_1, u_2, ..., u_n\}$ be the set of users where n is the number of users. $I = \{i_1, i_2, ..., i_m\}$ is the set of items where m is the number of items. Users can perform two types of actions in the rating system: establishing trust relations and creating ratings. Let $X \in \mathbb{R}^{n \times n}$ denote the trust network where $X_{ij} = 1$ if u_i is trusted by u_j and zero otherwise. Let $R \in \mathbb{R}^{m \times m}$ represent the ratings and R_{ij} is used to represent the entity at i^{th} row and j^{th} column, i.e., the rating of I_j given by u_i.

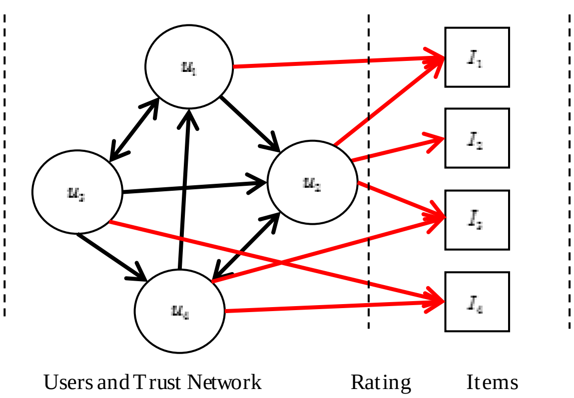


Figure 2. A Simple Example of Rating System

B. *Baseline Estimates*

Rating prediction is the task of predicting a given user's ratings for a given item based on past ratings or other information, regarded as one of the most important algorithms for recommendation system. Daily experience tells us that some users have the propensity to give a higher rating than others. Also, some items are more likely to receive a higher rating than others [12]. It is customary to adjust the data by accounting for these effects, which we encapsulate within the baseline estimates. A baseline model for rating prediction is based on user preferences and item characteristics. Thus, the unknown rating from u to i is shown below accounting for both the user and item effects.

$$R(u,i) = \mu + b(u) + c(i) \tag{1}$$

In order to estimate $b(u)$ and $c(i)$, the optimization formulation as follows:

$$\min_{b,c} \sum_{(u,i)\in O} (R(u,i) - \hat{R}(u,i))^2 + \lambda(\sum_u b^2(u) + \sum_i c^2(i)) \tag{2}$$

where $b(u)$ and $c(i)$ indicate the observed deviations of user u and item i, respectively, from the average, μ is the average rating for all products and $\lambda(\sum_u b^2(u) + \sum_i c^2(i))$ avoids overfitting by penalizing the magnitudes of the parameters.

It assumed that two users trusted will have similar preferences in terms of items. Another baseline method based on trust networks is the nearest-neighbor algorithm. The set of users who are trusted by user u and have rated product i is donated by $N \; u,i$. The predicted value of $R(u,i)$ is taken as a weighted average of the rating of neighboring items.

$$R(u,i) = \frac{\sum_{v\in N(u,i)} W(v,u) R(v,i)}{\sum_{v\in N(u,i)} W(v,u)} \tag{3}$$

Where $W(v,u)$ is the strength of single trust from u to v.

III. RECOMMENDATION BASED ON MULTI-FACETED TRUST RELATIONSHIPS

A. *Multi-Faceted trust relationships*

Trust has a strong correlation with user preference similarity in rating systems, reflected in their rating information [7], [13]. In fact, conventional algorithms measure the similarity that two user's favor of any types of items through the single rating similarity. However, daily experience tells us that people usually have a different degree on their favor of different types of objects. Trust is context dependent. This means that trusting someone on one topic does not guarantee to trust him in other topics. For example, a user who is trustworthy in technology might not be also trustworthy in astronomy, as well, so trust is multi-faceted. In this work, we explore the user multi-preferences to model multi-faces trust in rating prediction.

Due to the strong correlation between user similarity and trust, multi-faceted trust relationships between users can be indicated by facet similarities. Thus we define $s(u,v,k) \in R$

as the multi-preference similarity vector between u_u and u_v in the kth facet, based on their average preferences in kth facet $R_u^{\bar{k}}$ and $R_v^{\bar{k}}$, for example,

$$s(u,v,k)=\frac{\sum_{i\in I_{u,v}^k}(R(u,i)-R_u^{\bar{k}})(R(v,i)-R_v^{\bar{k}})}{\sqrt{\sum_{i\in I_{u,v}^k}(R(u,i)-R_u^{\bar{k}})^2}\sqrt{\sum_{i\in I_{u,v}^k}(R(v,i)-R_v^{\bar{k}})}} \tag{4}$$

However, that similarity based on the similarity in product ratings is just one of a number of possible factors that might be used to influence recommendation and prediction. The most difference between rating similarity and trust value is that trust is a subjective and personal relation between users. Therefore, it creates a directed relation in social networks. In other words, if w_{uv} represents the value of trust from user u to user v, it might not necessarily be equal to the value of trust from user v to user u. Based on this attribute, user u trusting user v does not guarantee that user v also trusts user u to the same extent. We further assume a linear relation between trust strength $w(u,v,k)$ and $s(u,v,k)$, formally stated as Eq. 5,

$$w(u,v,k)=f(w^{\mathrm{T}}s(u,v,k)+b_i) \tag{5}$$

where f: R → [0, 1] is an activation function such as sigmoid function, in which the trust strength is mapped into (0, 1) and b_i is a user specific bias of u_u.

B. Incorporating trust in recommender systems

Trust is a concept with many facets and dimensions, which can be used to further improve the performance or rating prediction. We know that no user has the same bias to all products. Instead, users might have a different bias toward different facets. Also, the average ratings for different facets are different. Thus, the function is defined as Eq.6,

$$\hat{R}(u,i)=c(i)+\frac{\sum_k PF(i,k)(\mu(k)+B(u,k))}{\sum_k PF(i,k)} \tag{6}$$

where $u(k)$ is the average rating for $f(k)$, $B(u,k)$ is the bias from u toward $f(k)$ and $PF(i,k)\in\mathbb{R}^{m\times k}$ that we refer to as the item-facet distribution with $PF(i,k)=1$ indicating the item i belongs to the facet k.

Intuitively, a user may trust different people in different domains. When faced with difference situations, people will seek advice from difference trusted sources to make decisions. If u strongly trusts v in f_k, $R(u,i)$ should be similar to $R(v,i)$. Multi-faceted trust relationships calculated in section 3.1, can be easily incorporated into this method as follows:

$$\hat{R}(u,i)=\frac{\sum_{k=1}^{K}\sum_{v\in N(u,i)}PF(i,k)w(v,u,k)R(v,i)}{\sum_{k=1}^{K}\sum_{v\in N(u,i)}PF(i,k)w(v,u,k)} \tag{7}$$

This model only considers the single factor, we believe that rating is not only determined by the bias of u toward f_k and the characteristics of i, but also influenced by the trust network of u, thus the rating prediction algorithm to estimate strengths of multi-faceted trust are shown below:

$$\begin{aligned}\hat{R}(u,i)=\alpha(c(i)+\frac{\sum_k PF(i,k)(\mu(k)+B(u,k))}{\sum_k PF(i,k)})+\\(1-\alpha)\frac{\sum_{k=1}^{K}\sum_{v\in N(u,i)}PF(i,k)w(v,u,k)R(v,i)}{\sum_{k=1}^{K}\sum_{v\in N(u,i)}PF(i,k)w(v,u,k)}\end{aligned} \tag{8}$$

The parameter α controls the contributions of these two parts. This is our final prediction rule, which allows fast online prediction.

We might be introduced this situation when one user buys an item that no one in his trust network buys it before. A challenge for our model is how to address the cold-start problem where no historical ratings on items or users are available. Homophily [13] is employed to deal with the cold-start problem: similar users are more likely to trust each other. Therefore, for a new user, we first find his top-ℓ similar users based on their profiles to establish his trust network and then use the average user preferences from his trust network to estimate his preferences.

More computational work is needed at a pre-processing stage where parameters are estimated. Thus, model parameters B and c are estimated by solving the following optimization problem:

$$\min_{b,c}\sum_{(u,i)\in O}E^2(u,i)+\lambda(\sum_{u,k}B^2(u,k)+\sum_i c^2(i)) \tag{9}$$

where $E(u,i)=R(u,i)-\hat{R}(u,i)$.

Unlike most work use least square solvers to obtain convex problem, we have found that the following simple gradient descent solver works much faster. For a given training case $R(u,i)$, we modify the parameters by moving in the opposite direction of the gradient, yielding:

$$B(u,k)\leftarrow B(u,k)-\beta(E(u,i)-\xi B(u,k)) \tag{10}$$

$$c(i)\leftarrow c(i)-\beta(E(u,i)-\xi c(i)) \tag{11}$$

The meta-parameters β (step size) and ξ are determined by cross-validation. We used $\beta=0.005$ and $\xi=0.002$ for the social network data.

IV. EXPERIMENTS

In this section, we conduct experiments to evaluate the proposed framework. We first introduce the details about the dataset used in the experiments. Next describe the findings and results of applying our model to online applications, i.e., rating prediction.

A. Datasets

For the purpose of this study, we crawled two datasets from two popular product review sites Epinions and Ciao in the month of May 2011. On both sites, people not only write critical reviews for various products but also read and rate the reviews written by others. Furthermore, people can add members to their trust networks or "Circle of Trust", if they find their reviews consistently interesting and helpful. For each user, we collected information about profiles, trust networks, and product rating entries. For each product rating entry, we collected date, product name, categories of a product and its ratings. Users with fewer than 5 reviews are pruned. Some statistics of the datasets are shown in Table 1: Epinions has a much larger trust network while Ciao has more close-knit trust relationships, indicated by its higher clustering coefficient and network density.

TABLE I. STATISTICS OF THE DATASETS

Items	Epinions	Ciao
#of Users	22166	12375
#of Products	296277	106797
#of Categories	27	28
#of Rating	922267	484086
#of Link	355813	237350
Ave Rating	4.05	4.21
Trust Network Density	0.0014	0.0031
Clustering Coefficient	0.1518	0.1969

We choose six representative facets (categories) from Epinions and Ciao to study multi-faceted trust relationships. The statistical information for these chosen facets is shown in Table 2. On average, each product receives more than two ratings. The average ratings for different facets are different and users have different preferences for different facets.

TABLE II. STATISTICS OF CHOSEN CATEGORIES

Epinions				
Order	*Name*	*Products*	*Ratings*	*Ave Score*
1	Electronics	9594	23571	3.96
2	Home & Garden	16029	28759	4.12
3	Computer Hardware	7584	17532	4.02
4	Hotel & Travel	11723	33410	4.15
5	Restaurants	14368	32477	3.91
6	Kids & Family	20926	50783	4.11

Ciao				
Order	*Name*	*Products*	*Ratings*	*Ave Score*
1	Entertainment	5059	14349	4.13
2	House & Garden	5804	15398	4.37
3	Household app	5334	14478	4.34
4	Electronics	4802	12426	4.41
5	Family	6227	18923	4.08
6	Games	6237	18491	4.31

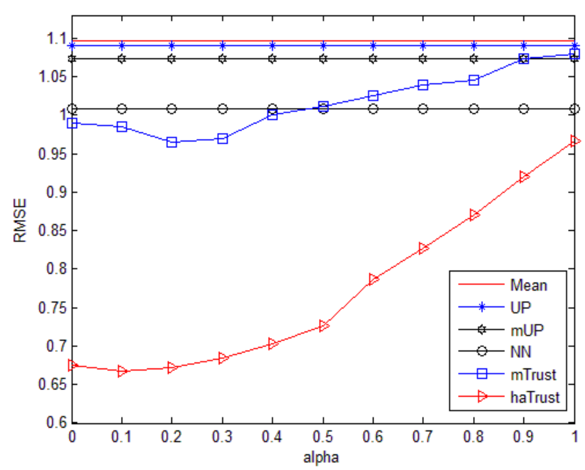
B. Experiment and evaluation

We choose the top 70% of the datasets as training data and the rest as testing data. The estimated parameters, i.e., $\{B, c\}$ can be used to predict unknown ratings. Our model can be incorporated into other rating prediction methods thus our objective is not to compare rating prediction methods. Instead, we want to verify that heterogeneous and asymmetry trust relationships between users allow us to improve the performance of prediction. A common metric, Root Mean Squared Error(RMSE) are used to evaluate prediction accuracy.

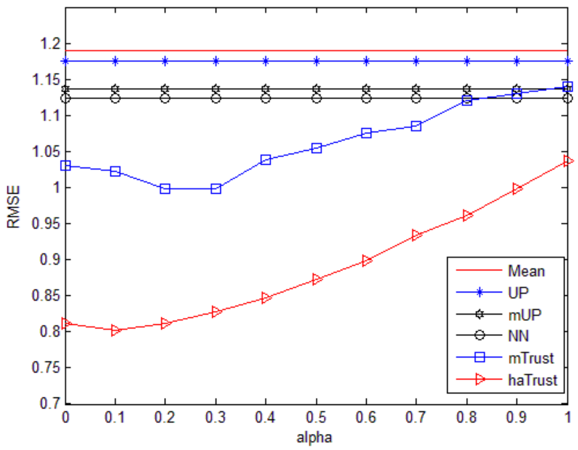
$$RMSE(u) = \sqrt{\frac{\sum_{(u,i)\in U} (R(u,i) - \hat{R}(u,i))^2}{|U|}} \tag{12}$$

Where $|U|$ is the size of testing data U. Through cross validation, we set $\lambda = 0.05$. The parameter controls the contributions of two factors of our formulation. We test different values of α to investigate the importance of these two factors in our study datasets. Results are shown in Figure 3. The methods mentioned in the figure are defined as follows:

- **Mean**: the rating of a product is predicted by the mean of known ratings of the product.
- **UP**: the rating of a product is predicted by Eq. (2) and this model assumes that one user has the same bias to all products.
- **mUP**: the model is a variant of UP, which incorporates the facet differences such as average facet ratings differences and users' facet bias differences.
- **NN**: it refers to the nearest neighbor algorithm, as shown in Eq. (3) NN assumes single trust relationships between users.
- **mTrust**: the rating of a product is predicted by the combination of mUP and mNN.
- **haTrust:** exploiting the heterogeneous and asymmetry trust relations and considering cold-start.



(a) Ciao



(b) Epinions

Figure3. Performance of Rating Prediction in Epinions and Ciao

We notice that RMSE values for Mean and UP are very close, because most of the time, the majority of user's actual ratings are close to the average and by closely examining the dataset, we find that more than 70% users give a score of 4 or 5. When facet differences are incorporated into UP, the performance of mUP is apparently improved. This directly supports the existence of facet differences, i.e. average facet rating differences and facet bias differences. Users can be influenced by their trustworthy friends, and are more likely to accept recommendations made by their trusted friends than recommendations from strangers. NN is the most common approach in rating prediction, but NN treats all trust relations equally. However, people may trust a part of their trust networks more strongly than others. mTrust that combination of mUP and mNN, considers multi-faceted trust relations and obtains better performance, but we will note that all baseline methods cannot tackle the cold-start problem. In addition, the fact that interpolation weights in mTrust sum to one forces the method to fully rely on the neighbors even in cases where neighborhood information is absent (i.e., user u did not rate items similar to i), and it would be preferable to rely on baseline estimates.

Our model haTrust consistently outperforms all other methods and we believe that this improvement is contributed by exploiting the heterogeneous and asymmetry trust relations and considering cold-start. The algorithm achieves the peak performance with $\alpha = 0.1$ for both datasets. A small α means that a higher weight would be put on user trust network. This suggests the importance of trust networks in rating prediction, and that an appropriate combination of these two factors is crucial to achieving better performance.

V. CONCLUSION

In this paper, we study heterogeneous trust relationships between users in the domain of product review sites Epinions and Ciao. A fine-grained approach, haTrust, is proposed to capture multi-faceted and asymmetry trust relationships. Experiments on two real-world datasets show that haTrustcan effectively improve the performance of rating prediction and more robust to the cold-start problem.

There are new research directions to be investigated. First, haTrust does not consider temporal information related to trust networks and product ratings. When a longer time span is studied (e.g., a year), it would be wise to include temporal effects on trust between users. Second, haTrust only uses users' explicit feedback i.e. user rating history, ignore user implicit feedback, i.e. rating helpless.

ACKNOWLEDGMENT

This work is supported by the National Science Foundation of China under Grant No.60902069, No.61171124, supported by Science Technology Planning Project of Guangdong (Grant No.2013B021500017). Supported by Science Technology Planning Project of Shenzhen (Grant No.JCYJ20160422112909302, GJHS20160328145558586).

REFERENCES

[1] Elaheh Momeni, Claire Cardie, and Nicholas Diakopoulos. A Survey on Assessment and Ranking Methodologies for User-Generated Content on the Web. *ACM Comput. Surv.* 48, 3, Article 41 (December 2015), 49 pages. doi: http://dx.doi.org/10.1145/2811282.

[2] Jiliang Tang, Suhang Wang, Xia Hu, Dawei Yin, Yingzhou Bi, Yi Chang, and Huan Liu. Recommendation with Social Dimensions. The Thirtieth AAAI Conference (AAAI 2016), February 12-17, 2016. Phoenix, Arizona.

[3] Aggarwal C C. Social and Trust-Centric Recommender Systems[M]. Recommender Systems. Springer International Publishing, 2016: 345-384.

[4] Kidus Dejen Yirgu. Propagation and Aggregation in Trust Enhanced Recommender Systems[D]. PhD thesis, University of Fribourg, Switzerland, May 2016.

[5] Gao, Peixin, et al. STAR: Semiring Trust Inference for Trust-Aware Social Recommenders. *Proc. of the 10th ACM Conference on Recommender Systems*. ACM, 2016.

[6] Ting Yuan, Jian Cheng, Xi Zhang, Qingshan Liu, and Hanqing Lu. How friends affect user behaviors? An exploration of social relation analysis for recommendation. *Know.-Based Syst.* 88, C (November 2015), 70-84. doi=http://dx.doi.org/10.1016/j.knosys.2015.08.005.

[7] Golbeck J. Trust and nuanced profile similarity in online social networks[J]. ACM Transactions on the Web (TWEB, 2009), 3(4), pp:1-33.

[8] Zolfaghar K, Aghaie A. A syntactical approach for interpersonal trust prediction in social web applications: Combining contextual and structural data[J]. Knowledge-Based Systems, (2012), 26, pp:93-102.

[9] Zhang J, Cohen R. A framework for trust modeling in multi-agent electronic marketplaces with buying advisors to consider varying seller behavior and the limiting of seller bids[J]. ACM Transactions on Intelligent Systems and Technology (TIST, 2013), 4(2): 1-24.

[10] Abdul-Rahman A, Hailes S. Supporting trust in virtual communities[C]. Proc. of the 33rd Annual Hawaii International Conference on System Sciences, (IEEE 2000), pp:1-9.

[11] Tang J, Gao H, Liu H. mTrust: discerning multi-faceted trust in a connected world[C]. Proc. of the fifth ACM international conference on Web search and data mining. (ACM, 2012), pp:93-102.

[12] Mishra A, Bhattacharya A. Finding the bias and prestige of nodes in networks based on trust scores[C]. Proc. of the 20th international conference on World wide web. (ACM, 2011), pp: 567-576.

[13] Tang J, Gao H, Hu X, et al. Exploiting homophily effect for trust prediction[C]. Proc. of the sixth ACM international conference on Web search and data mining. (ACM, 2013), pp:53-62.