An evaluation approach based on word-of-mouth for trust models in recommendation systems

Feng Jiang[1,2], Min Gao[3,4,*], Hui Xia[3], Qingyu Xiong[3,4], Junhao Wen[3,4]

[1] *College of Civil Engineering, Chongqing University, Chongqing, 400044, China*

[2] *College of Construction Engineering, Chongqing Technology and Business Institute, Chongqing, China*

[3] *School of Software Engineering, Chongqing University, Chongqing 400044, China*

[4] *Key Laboratory of Dependable Service Computing in Cyber Physical Society, Ministry of Education, Chongqing 400044, China*

Received 1 June 2014, www.cmnt.lv

Abstract

Recommendation systems have been recognized as an effective approach to heavy information load. In recommendation systems, trust/reputation has attracted increasing attention because it helps to improve the precision of recommendation and the robustness of systems to shilling attacks. Recommendation system oriented trust models, mostly rating-based, used to build the reputation and trustiness among users. They are often evaluated in terms of how accurately they help to predict user ratings and how robustly they resist shilling attacks. However, those evaluation techniques disregard the trust values themselves: how accurately they calculate the trust values themselves is not measured. To solve the problem, in this work, we propose an approach to measure the trust values based on electronic word-of-mouth (eWoM) theory. The eWoM believes a user is reliability if he is of good public praise. In our approach, firstly, according to eWoM, the reliability value of a user can be judged by other users' votes - whether the user's ratings or feedbacks are positive or negative. Secondly, the trust values of users can be calculated by a trust model. Finally, we compare trust values and reliability values. As a case study, we propose a simple rating-based trust model and then evaluate the trust model based on the proposed evaluation approach and Amazon dataset.

Keywords: Recommendation Systems, Evaluation Approach, Trust Model

1 Introduction

Trust has been a spot in recommendation systems, because combining trust and recommendation systems can improve the accuracy and stability of recommendations and improve the user's experience [1]. Trust models have been used to recommend movies, songs, products in e-commercial websites, and friends in social networks.

Trust models, usually rating-based in recommendation systems, are often evaluated according to how accurately they help to predict these ratings [1, 2], or by measuring the stability of rating prediction and hit ratio [3, 8]. However, the set of evaluation techniques disregard the trust values themselves: how accurately they calculate the trust values themselves is not measured. To solve the problem, we proposed an evaluation approach to measure trust models according to eWoM theory [4, 5]. The approach can evaluate the accuracy of a trust model by users' assessment. Applied this evaluation approach, the well-performed trust model can be applied to other recommendation systems without user's assessment. For example, there is a trust model m1, according to the evaluation approach and datasets with user assessments (e.g. the datasets driving from Amazon.com), we can evaluate m1. If m1 is good enough, then it can be used to other recommendation systems without user assessments information, such as Movielens and Netflix.

In this work, we first provide an overview of trust model, evaluation approaches and their limitations (Section 2) to explain why a novel evaluation approach may be important in recommendation systems. We proposed a eWoM-based approach to evaluate trust models (Section 3): by proposing a eWoM-based algorithm for reliability values of users, and then comparing those reliability values and trust model based trust values. We finally to illuminate the evaluation approach using a case study (Section 4), comparing the trust values calculated by a trust model to the public reliability values on datasets deriving from Amazon. We draw conclusion in Section 5.

2 Related works and associated problems

With the advancement in networking and multimedia technologies enables the distribution and sharing of multimedia content widely. In the meantime, piracy becomes increasingly rampant as the customers can easily duplicate and redistribute the received multimedia content to a large audience.

Trust/reputation is what is generally said or believed about a person's or thing's character or standing [6, 7].

* ***Corresponding author*** e-mail: skycat1980@hotmail.com

Giorgos and Zacharia [8] proposed Sporas model based on eBay and consider the trustworthiness of the graded counterpart to mitigate influences of malicious behaviours; and then they propose a novel collaborative reputation mechanism for electronic marketplaces [9]. Guo, Jiang, and Cai [10] modified Sporas into E-Sporas by adding the factors of volume and number of trades to make Sporas better in C2C. Wang, Zhang, and Chen [11] proposed a cloud-based reputation reporting mechanism to solve the fake and joint cheat recommendations of malicious customers. Li, Liang, and Zhang [12] gave a systematic study on trustworthiness and reputation management model in e-commerce. Wu, Li, and Kuo [13] proposed several metrics for online auctions to evaluate the trust values of participants. Abdel-Hafez, Xu, and Tjondronegoro [14] provided a product reputation model for recommendation based on opinion mining techniques.

In recommendation systems, Golbeck et al. looked specifically at trust between people who have no directly links to another and proposed TidalTrust [15] mechanism to infer trust in continuous trust networks. Paolo Avesani et al proposed a time-efficient trust metric named Moletrust [16] and applied it in Moleskiing application to compute the trustworthiness of users. Levien et al. [17] investigated the role of trust metrics in attack-resistant public key certification. Kuter et al. described SUNNY [2], a new trust inference algorithm using a probabilistic sampling technique to estimate confidence. Cai-Nicolas Ziegler et al. provide a novel trust metric Appleseed [18] for local group trust of semantic web issues. Massa and Avesani [19, 20] proposed approaches to estimate all trust weights of users by propagate trust over a trust network. Verbiest et al. [21] proposed a new approach to calculate trust and distrust by introduce path length incorporating aggregation strategies. Victor et al. proposed bilattice-based aggregation approaches for gradual trust and distrust [22] and then incorporated the gradual trust and distrust in recommender systems [23].

Large numbers of literatures concern about trust and reputation recently. However, they are mainly proposed for constructing better recommendations, such as decreasing the accuracy (e.g. MAE, RMSE) of rating prediction, improving precision and recall of recommendations [9, 10, 14], and improving robustness [8, 11, 13] (especially stability ability under shilling attacks) and users' satisfactions [8-14]. However, the set of evaluation techniques disregard the trust values themselves: how accurately they calculate the trust values themselves is not measured. Thus, we propose a new evaluation approach to solve the problem.

3 An EWoM-based evaluation approach

Traditionally, WoM is to pass information from person to person by oral communication. When WoM is mediated through electronic means, the eWoM refers to any statement consumers share via the Internet about a product, service, brand, or company [4]. The eWoM proves to have more effectiveness in leading purchasing decisions, as the wide scale of online broadcasting and weak correlations between online customers and marketers [5]. The enormous impact of eWoM on consumer behaviour and product success thus attracts many researchers.

The steps for the evaluation approach are show as Figure 1. We can get reliability values from eWoM-based algorithm (subsections 3.1 and 3.2) and then compare those reliability values to trust values based on trust models (subsection 3.3).

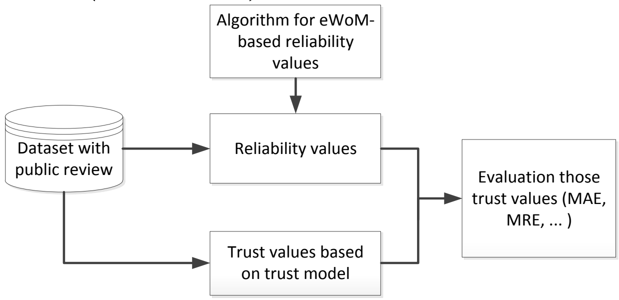


FIGURE 1 The steps for the evaluation approach

3.1 EWOM-BASED RELIABILITY VALUES

As known to us all that, people can rate their consumed product after each transaction. However, the reliability [7] of those consumers and their ratings is still a problem. Inspired by the effectiveness of eWoM, an empirically convincing evaluation approach under mass attitude is proposed to assess reputation of customer with their ratings.

Firstly, we collect support ϕ_j of the whole ratings from a customer u_j; at the same time, compute the weight φ_j of u_j's ratings over the whole system. Then the reliability value of u_j's from public views can be simply denoted as Equation (1).

$$Reli_j = \alpha f_j + \beta j_j, \tag{1}$$

where the weights of ϕ_j and φ_j are α and β, $\alpha + \beta = 1$; $\phi_j = \sum_{0 \le i \le k} h_{i,j} / v_{i,j}$, where $h_{i,j}$ is the helpful votes from u_j to item i, $v_{i,j}$ is the corresponding total votes; $\varphi_j = k_j / R$, $k_j = |r(u_j)|$, k_j is the number of u_j's all ratings; R is the number of all ratings in the dataset.

The weights α and β can be set to suitable values by experiments. We will illuminate how to find the suitable values in Section 3.2.

3.2 SET THE WEIGHTS USING AMAZON DATASET

The dataset is Amazon product co-purchasing network metadata* collected in summer 2006. The dataset contains 548,552 different items (Books, music CDs, DVDs, and VHS video tapes). Each item contains reviews from customers, including ratings, the number of helpful votes for ratings, and total number of votes.

* http://snap.stanford.edu/data/amazon-meta.html

To get suitable values of the weights, we divided the datasets to four subsets G1, G2, G3, and G4 according to the number of all ratings of a user k (k>0, k>5, k>10, k>20). We calculated the average reliability value avg_R_{ci}, mean absolute error (MAE) MAE_{ci}, and mean relative error (MRE) MRE_{ci} (see Equations 2, 3, and 4), using ten combinations of α and β, C = (0.0, 1.0; 0.1, 0.9; 0.2, 0.8; 0.3, 0.7; 0.4, 0.6; 0.5, 0.5; 0.6, 0.4; 0.7, 0.3; 0.8, 0.2; 0.9, 0.1; 1.0, 0.0), marked by c_1, c_2, ..., and c_n.

$$avg\_Reli_{c_i} = \sum_{i=1}^{n} \left| Reli_{u_i}(c_i) \right| \tag{2}$$

$$MAE_{c_i} = \frac{\sum_{i=1}^{n} \left| Reli_{u_i}(c_i) - R_{u_i} / all\_avg\_I * avg\_Reli_{c_i} \right|}{n} \tag{3}$$

$$MRE_{c_i} = \frac{\sum_{i=1}^{n} \left(\left| Reli_{u_i}(c_i) - R_{u_i} / all\_avg\_I * avg\_Reli(c_i) \right| / Reli_{u_i} \right)}{n}, \tag{4}$$

where n means the total number of users in group; $Reli_{ui}(c_i)$ can be calculated by Equation (1); Ru_i is a set of u_j's all ratings; all_ave_I is the average of all items average rating scores.

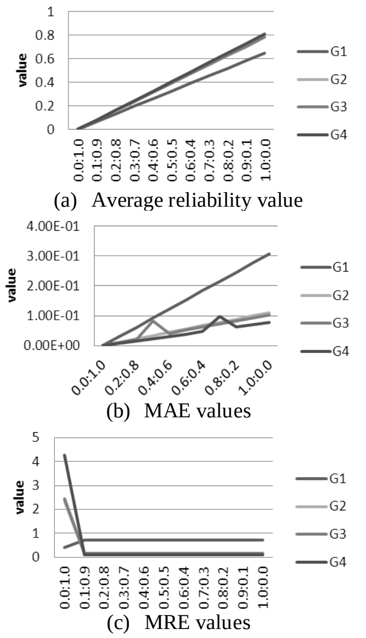


FIGURE 2 Four groups of average E, absolute error, and relative error

The avg_Reli_{ci} results are shown as Figure 2 (a). It is simply that 4 curves grow linear with the value of α increase, though β decreases. Actually, k / all_num is normally too small to have nearly no impacts on the results.

From sampling ratings from Amazon dataset, the value of all_avg_I is 4.18, thus we could get the absolute and relative error values according to Equations (3) and (4) respectively, and then have the corresponding curves as Figure 2 (b) and (c) denotes.

In Figure 2 (b), all groups' absolute errors are very small even though the proportion of α and β are different, which seem to remind us that proportion of α and β takes no effect. In Figure 2 (c), we can find a phenomenon that there are 4 outlier points where the value of (α, β) is (0.0, 1.0). The phenomenon gives us a further authentication that the first part of the equation (1) plays an important role in computing the value of $Reli_{ui}$ in another direction. Another obvious phenomenon in (c) is that all curves rise and fall slightly around certain numbers, which look like 4 straight lines no matter what happens to the value of (α, β) later. From the above two figures, we could make sure that any proportion of α and β can work except α equals 0.0. Thus, in the next experiments, we keep (α, β) is (0.8, 0.2).

3.3 EVALUATION APPROACH

To evaluate a trust model, we calculate *wMAE* (weighted MAE) and *wMRE* (weighted MRE) for the trust model based trust values and eWoM-based reliability values (See Equations (5) and (6)).

$$wMAE = \sum_{j=1}^{n} (T_j * \overline{Reli} / \overline{T} - Reli_j) / n, \tag{5}$$

$$wMRE = \sum_{i=1}^{n} ((T_j * \overline{Reli} / \overline{T} - Reli_j) / Reli_j) / n. \tag{6}$$

In the Equations,

$$\overline{T} = \sum_{j=1}^{n} T_j / n, \tag{7}$$

$$\overline{Reli} = \sum_{j=1}^{n} Reli_j / n. \tag{8}$$

T_j is the trust value for user j. $\overline{T}$ is the average value of T_j. $Reli_j$ is the reliability value for user j.

4 Promoting the proposed evaluation approach

To illuminate the steps for applying the evaluation approach, we will use a simple trust model first as an example, then evaluate the model using the proposed approach.

4.1 A SIMPLE TRUST MODEL – AS A CASE STUDY

Here we have a simple but effective trust model for customer according to a basic truth, that is, each rating has a potential bias since each rating is affected by users' sentimental factor. Thus we should use correct function (5) to objectify the biased rating r_{ij}.

$$r_{ij}^{'} = r_{ij} * \frac{\overline{r}}{\overline{r_i}}, \tag{9}$$

where $\overline{r}$ is the average rating of the whole system, while $\overline{r_i}$ is the average rating of item i. The objective part of r_{ij} is $r_{ij}^{'}$.

After the correction, we could modify the existed trust value of u_j iteratively (See Equation 10).

$$T_j = \frac{T_j^{'} * k + r_{ij}^{'}}{k+1}\ , \tag{10}$$

where $T_j^{'}$ is the previous trust value before u_j rated item i.

4.2 THE COMPARISON AND ANALYSIS

We derived two random selected groups from Amazon dataset to validate the tendency: the consistency - how the average trust values and average reliability values looked like. Experimental results are shown in Figure 3.

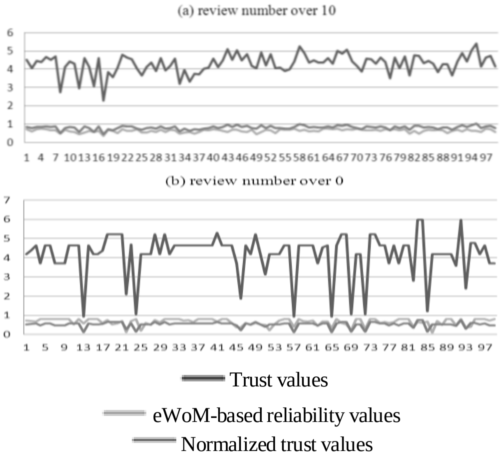
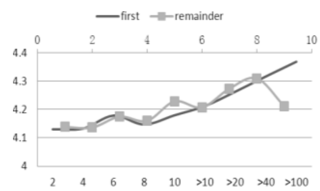
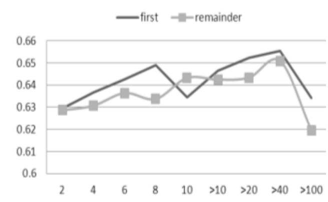


FIGURE 3 Reputation trends in review number (a) over 10, (b) over 0, with initial trust values, the normalized trust values from equation and reliability values from equation (1) with α=0.8 and β=0.2

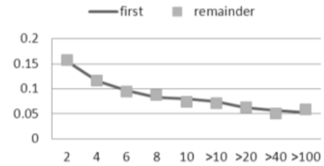
We then derived from selected sample data with 40000 items. We divided the data to two groups according to the number of users' total reviews k. One group was the *first* 20000 sample items, and the other one was the *remaining* 20000 sample items. We firstly grouped 9 classes to perform experiments where k=2, k=4, k=6, k=8, k=10, k>10, k>20, k>40, and k>100 respectively, and then calculated the average trust values (T), average reliability values (*ave_Reli*), *wMAE*, and *wMRE* according to Equation (5) - (8). The results are shown in Figure 4.



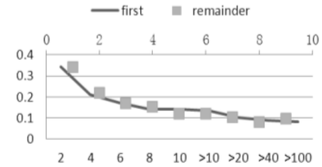
(a) Average trust values



(b) Average reliability values



(c) Weighted *MAE*



(d) Weighted *MRE*

FIGURE 4 The comparisons of two sets between (a) Average trust values (*T*), (b) average reliability values (*ave_Reli*), (c) *wMAE*, (d) *wMRE*

From Figure 4 (a) and 4 (b), each two lines have small fluctuations respectively, but still can fit to others to some extent, which confirms that both of the two samples can reflect the overall dataset. From Figure 4 (c) and 4 (d), we can see that both two sets' curves of *wMAE* and *wMRE* firstly decrease quickly before k = 8 or so, then slowly and have signs of tending towards stability after k = 8.

The largest *wMAE* in the first group is 0.156 at the first point, and falls down to 0.1 at the fifth point, and preserves stable around 0.08 after 8. It is almost the same trend to the second group's curve.

Therefore, we can conclude that, most *wMAE* is smaller than 0.1, and larger value of k means lower absolutely error. Similarly, we can analyse that the range of relative error is 0.0 - 0.2, and larger k means lower relative error.

5 Conclusions and future work

In this paper, we have analysed up-to-date trust/reputation models especially in recommendation systems. We have introduced evaluation approaches of the models and their problem: disregard the accuracy of trust values themselves. To solve the problem, we have proposed an evaluation approach based on WoM theory and given a case study and experiments on Amazon dataset to illuminate how to apply and analyse the proposed evaluation approach.

We have just used a trust model to show how to apply the proposed approach in the paper. We will try to apply the evaluation approach to more trust models and further analyse and improve the approach.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (71102065), the China Postdoctoral Science Foundation (2012M521680), the Fundamental Research Funds for the Central Universities (CDJZR12090001), the Scientific and Technological Research Program of Chongqing Municipal Education Commission (KJ121607 and KJ131603), and the Frontier and Application Foundation Research Program of CQ CSTC (cstc2014jcyjA40037).

References

[1] Golbeck J 2009 Tutorial on using social trust for recommender systems. *Proc. Conf. on Recommender systems* ACM New York USA 425-6

[2] Kuter U, Golbeck J 2007 Sunny: A new algorithm for trustence in social networks using probabilistic confidence models. *Proc. Conf. on Artificial Intelligence* AAAI Vancouver British Columbia Canada 1-6

[3] Gao M, Wu Z, Jiang F 2011 Userrank for item-based collaborative filtering recommendation *Information Processing Letters* **111**(9) 440-6

[4] Kietzmann J H, Canhoto A 2013 Bittersweet! Understanding and Managing Electronic Word of Mouth *Journal of Public Affairs* **13**(2) 146–59.

[5] He C 2009 *Research of Source Credibility of Online Review Based on Relationship Stages, Master Thesis* Dalian University of Technology

[6] McLeod Carolyn 1999 Trust *The Stanford Encyclopedia of Philosophy* 2011 Edition Edward N Zalta (ed.) URL=http://plato.stanford.edu/archives/spr2011/entries/trust

[7] Jøsang A, Ismail R, Boyd C 2007 A Survey of Trust and Reputation Systems for Online Service Provision *Decision Support Systems* **43**(2) 618-44

[8] Zacharia G 1999 *Collaborative Reputation Mechanisms for Online Communities* Massachusetts Institute of Technology USA

[9] Zacharia G, Moukas A, Maes P 2000 Collaborative Reputation Mechanisms in Electronic Market places *Decision Support Systems* **29**(4) 371-88.

[10] Guo H, Jiang J, Cai H 2009 Modelling for Reputation Computing in C2C Communities *Chinese Journal of Management* **6**(8) 1056-60

[11] Wang P, Zhang S, Chen X 2011 A Novel Reputation Reporting Mechanism Based on Cloud Model and Grey System Theory *International Journal of Advancements in Computing Technology* **3**(10) 75-84

[12] Li D, Liang Y, Zhang W 2010 Survey on trust management for e-commerce systems, *Application Research of Computers* **27**(4) 1208-11

[13] Wu F, Li H H, and Kuo Y H 2011 Reputation evaluation for choosing a trustworthy counterparty in C2C e-commerce. *Electronic Commerce Research and Applications* **10**(4) 428-36

[14] Abdel-Hafez A, Xu Y, Tjondronegoro D 2012 Product Reputation Model: An Opinion Mining Based Approach SDAD *Proc. Conf. on Sentiment Discovery from Affective Data* Bristol UK 16-27

[15] Golbeck J 2005 *Computing and Applying Trust in Web-based Social Networks* PhD thesis, University of Maryland, College Park, MD, USA

[16] Avesani P, Massa P, Tiella R 2005 Moleskiing.it: a trust-aware recommender system for ski mountaineering *International Journal for Infonomics* **20**(1) 1-19

[17] Levien R, Aiken A 1998 Attack-resistant trust metrics for public key certification *The 7th USENIX Security Symposium* San Antonio, Texas 229-42

[18] Ziegler C-N, Lausen G 2004 Spreading activation models for trust propagation *Proc. IEEE Conf. on e-Technology, e-Commerce, and e-Service* Taipei, Taiwan 83-97

[19] Massa P, Avesani P 2007 Trust-aware recommender systems *Proc. Conf. on Recommender Systems* Minnesota, USA 17-24

[20] Massa P, Avesani P 2009 Trust metrics in recommender systems *Computing with Social Trust* Springer: London

[21] Verbiest N, Cornelis C, Victor P, Herrera-Viedma E 2012 Trust and distrust aggregation enhanced with path length incorporation *Fuzzy Sets and Systems* **202** 161-74

[22] Victor P, Cornelis C, Cock M De, Herrera-Viedma E 2011 Practical aggregation operators for gradual trust and distrust *Fuzzy Sets and Systems* **184**(1) 126-47

[23] Victor P, Cornelis C, Cock M De, Pinheiro da Silva P 2009 Gradual trust and distrust in recommender systems *Fuzzy Sets and Systems* **160**(10) 1367-82

Authors

Feng Jiang, born on February 13, 1979, China

Current position, grades: researcher at Chongqing Technology and Business Institute, China.
University studies: MS degree in Civil Engineering from Chongqing University, China in 2006. Currently, he is a PhD candidate in the College of Civil Engineering, Chongqing University.
Scientific interests: information processing and Construction Engineering.
Publications: more than 10 refereed journal and conference papers in these areas.

Min Gao, born on April 08, 1980, China

Current position, grades: professor at the School of Software Engineering, Chongqing University.
University studies: MS and PhD degrees in computer science from Chongqing University in 2005 and 2010 respectively.
Scientific interests: recommendation system, service computing, and data mining.
Publications: more than 30 refereed journal and conference papers in these areas.
Experience: visiting researcher at the School of Business, University of Reading. Grants from the National Natural Science Foundation of China, the China Postdoctoral Science Foundation, and the China Fundamental Research Funds for the Central Universities.

Hui Xia, born on September 24, 1989, China

University studies: PhD candidate in the College of Computer Science, Chongqing University.
Scientific interests: recommendation system and computer networking.
Experience: participated in 5 industrial projects and developed many commercial systems and software tools.

Qingyu Xiong, born on October 03, 1965, China

Current position, grades: dean of the School of Software Engineering, Chongqing University.
University studies: BS and MS degrees from the School of Automation, Chongqing University in 1986 and 1991, respectively, and the PhD degree from Kyushu University of Japan in 2002.
Scientific interests: neural networks and their applications.
Publications: more than 100 journal and conference papers in these areas.
Experience: more than 20 research and applied grants.

Junhao Wen, bor on May 01, 1969, China

Current position, grades: deans of the School of Software Engineering, Chongqing University. Professor in the School of Software Engineering.
University studies: MS and PhD degrees in computer science from Chongqing University in 1999 and 2008 respectively.
Scientific interests: computing and recommendation systems.
Experience: more than 80 journal and conference papers in these areas.