Bayesian Memory-Based Reputation System

Weiwei Yuan, Donghai Guan, Sungyoung Lee, Young-Koo Lee
Dept. of Computer Engineering, Kyung Hee University
Seocheon-dong, Yongin-si, Gyeonggi-do, Korea
{weiwei, donghai, sylee} @oslab.khu.ac.kr, yklee@khu.ac.kr

Heejo Lee
Dept. of Computer Science and Engineering,
Korea University
Anam-dong, Seoul, Korea
heejo@korea.ac.kr

ABSTRACT
Reputation System provides a way to maintain trust through social control by utilizing feedbacks about the service providers' past behaviors. Conventional Memory-based Reputation System (MRS) is one of the most successful mechanisms in terms of accuracy. Though MRS performs well on giving predicted values for service providers offering averaging quality services, our experiments show that MRS performs poor on giving predicted values for service providers offering high and low quality services. We propose a Bayesian Memory-based Reputation System (BMRS) which uses Bayesian Theory to analyze the probability distribution of the predicted valued given by MRS and makes suitable adjustment. The simulation results, which are based on EachMovie dataset, show that our proposed BMRS has higher accuracy than MRS on giving predicted values for service providers offering high and low quality services.

Keywords
Reputation System, Bayesian Theory, Memory-based

1. INTRODUCTION
Reputation system is a way to maintain trust in ubiquitous environments where service requesters interact with service providers that (1) they might have never met, not even heard of, (2) or their own personal interaction experience is not enough to make the decision. This is achieved by the provision of information about the service providers' past performance [1], i.e. the reputation system is used to collect, distribute and aggregate feedbacks about the service providers' past behaviors.

The task of reputation system is to predict the utility of service providers to a particular user (called active user) based on other users' recommendations. Conventional Memory-based Reputation System (MRS) using Pearson Correlation Coefficient is one of the most successful mechanisms in terms of accuracy [2]. However, we found through experiments that MRS (using Pearson Correlation Coefficient) performs well on the Median Values, while performs poor on the Polar Values. For example, $R \in [1, 2, 3, 4, 5, 6]$, where R represents the rating, 1 represents the rating of the minimal trust on the service provider while 6 represents the maximal trust. Our experiments show that MRS has high accuracy when the real ratings given by the active user are 3 and 4. However, they have low accuracy when the real ratings are 1, 2, 5 and 6. The reason is that when evaluating the active user's rating on a certain service provider, MRS uses the active user's mean rating as the major part of the predicted values, and Pearson Correlation Coefficients based part, which is used to adjust the active user's mean rating, is always relatively small.

Compared with accuracy on Median Values, the accuracy on Polar Values is more important for the reputation system. The reason lies in the following two aspects: (1) Without the ability to distinguish service providers offering high quality services and average quality services, more and more service providers offering high quality services will leave since they can not effectively attract the usage of services. (2) Without the ability to distinguish service providers offering low quality services and average quality services, more and more service providers offering low quality services will join since they can attract the usage of services as effectively as others. Finally there are only service providers offering low quality services left and no user is willing to pursuit the services in this environment.

We propose a Bayesian Memory-based Reputation System (BMRS) to solve the above problem of MRS. Compared with conventional MRS, the main advantage of our method is that it can effectively improve the accuracy of the predicted values on service providers offering high and low quality services, i.e. with Polar Values. This is achieved by adjusting the predicted values given by MRS based on analyzing those values using Bayesian mechanism.

The rest of the paper is organized as follows. We briefly introduce related works in Section 2. And we present the proposed Bayesian Memory-based Reputation System as well as the simulation results in details in Section 3. Finally, conclusions and future work are presented in Section 4.

2. RELATED WORKS
A number of reputation systems have been proposed in previous literatures, in which some of them have already been used to commercial applications. The simplest reputation model is to compute the ratee's reputation by summing all the positive ratings and negative ratings. A famous example is eBay's reputation forum [3]. Some reputation systems are based on Bayesian Theory, for example [4, 5, 6, 7]. These models get a posteriori (i.e. the

Mobimedia'07, Month 8, 2007, Nafpaktos, Aitolokarnania, Greece.

updated) reputation from the computing of combining the priori (i.e. previous) reputation with the new ratings. To use the Bayesian reputation systems, we need to get enough training data to get the priori knowledge. There are also some reputation systems based on Dempster-Shafter Theory (belief model) [8, 9]. Dempster-Shafter Theory is a generalization of Bayesian theory of subjective probability. Some reputation systems are based on flow models. These systems calculate reputation by transitive iteration through looped or arbitrarily long chains [10]. The ratee's reputation increases as a function of income flow and decreases as a function of outgoing flow [11]. A famous example is Google's PageRank [12]. Discrete reputation systems are proposed based on the fact that humans are often better able to rate performance in the form of discrete variables instead of continuous means, e.g. [13, 14, 15, 16]. There are also some reputation systems based on the fuzzy models, e.g. [17, 18, 19]. In fuzzy reputation systems, reputations are expressed as linguistically fuzzy concepts in which membership functions describe to what degree an agent can be described [20].

3. Our Proposed Reputation System

3.1 A Brief Introduction to Memory-Based Reputation System

Memory-based Reputation System (MRS) motivates from the observation that people usually trust the recommendations from like-minded friends. MRS applies a nearest neighbor-like scheme to predict a user's ratings based on the ratings given by like-mined users. We use $r_{i,j}$ to represent user i 's rating on service provider j. SP_i is used to represent the set of service providers on which user i has given ratings. The mean rating for user i is defined as:

$$\bar{r}_i = \frac{1}{|SP_i|} \sum_{j \in SP_i} r_{i,j} \tag{1}$$

We use $p_{a,j}$ to represent the predicted rating value given by the active user (indicated with a subscript a) on service provider j. Using MRS, $p_{a,j}$ is calculated as:

$$p_{a,j} = \bar{r}_a + k \sum_{i \in 1}^{n} w(a,i)(r_{i,j} - \bar{r}_i) \tag{2}$$

where $w(a,i)$ is the weight which reflect distance, correlation, or similarity between each user i and the active user a; n is number of users who gave rating on service provider j; k is the normalizing factor such that the absolute values of $w(a,i)$ sum to unity.

Pearson Correlation Coefficient is one of the most effective methods to calculate $w(a,i)$. Using Pearson Correlation Coefficient:

$$w(a,i) = \frac{\sum_q (r_{a,q} - \bar{r}_a)(r_{i,q} - \bar{r}_i)}{\sqrt{\sum_q (r_{a,q} - \bar{r}_a)^2 \sum_q (r_{i,q} - \bar{r}_i)^2}} \tag{3}$$

where $q = SP_a \cap SP_i$.

3.2 Limitation of MRS

We use the following experiment to analyze the limitation of the conventional MRS.

The dataset we used for analysis is EachMovie Dataset, which is collected by DEC (now Compaq) research. It consists of 72916 users, 1628 movies and 2811983 movie votes. Each rating is one number of [0, 0.2, 0.4, 0.6, 0.8, 1]. To make the analysis result more distinct, we amplify the rating as shown in Table 1.

Table 1. Amplification on the rating

Original Rating	0	0.2	0.4	0.6	0.8	1
Our Representation	1	2	3	4	5	6

The experiment steps are:

1. Randomly choose one user from EachMovie Dataset as the active user.

All the ratings given by this active user on different movies are the objects of analysis. The active user we chose had voted for 418 movies.

2. Split ratings given by the active user into two parts: training dataset and test dataset.

In this experiment, we randomly choose 100 ratings to act as test dataset, the left ratings on 318 different movies are used as training dataset. That is, for training dataset, $|SP_i|$=318; for test dataset $|SP_i|$=100. Thus we get two vectors, training dataset TR and test dataset TS, with length of 100 and 318 respectively.

3. Randomly choose 500 users which are different from the active user from the EachMovie Dataset. Ratings given by these users are used to calculate $w(a,i)$ with the TR in formula (3), and calculate $p_{a,j}$ with TS in formula (2).

From the probability prospect of view, if users gave ratings on only several movies, we will get $q = \varnothing$ in formula (3) and it is meaningless to use MRS. So in the selected 500 users, we filter out the users whose voted movies are below a small number. In this paper, we set this small number to 5. And we finally get 431 users qualified to be used in the calculation of MRS. These 431 users totally gave ratings on 899 movies. Thus we get a 431×899 matrix M.

4. Calculate $w(a,i)$ for the selected 431 users in M. This is achieved by compare TR and M using formula (3).

5. Give predicted values on the movies rated by the active user in TS. This is achieved by calculating $p_{a,j}$ using formula (2) and $w(a,i)$ we got in step 4. Thus we get a predicted value vector P whose length is 100.

6. Compare TS with P.

We get the distribution of *P* as shown in Fig. 1. Compare with the distribution of *TS* shown in Fig. 2, we find that: though *TS* takes values in whole interval of [1 6], $p_{a,j}$ is always a Middle Value using MRS, which means that MRS are not able to give proper prediction on service providers offering high and low quality services.

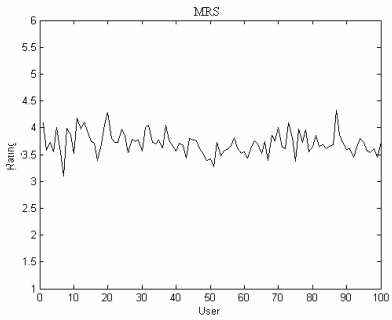


Figure 1. Distribution of *P* using MRS.

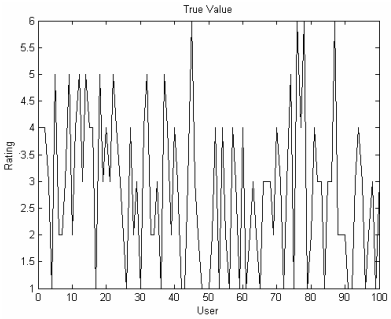


Figure 2. Distribution of *TS* .

The reason for the above observation is that the evaluation of $p_{a,j}$ uses $\overline{r_a}$ as the major part, and $k\sum_{i\in 1}^{n} w(a,i)(r_{i,j} - \overline{r_i})$ is used to adjust $\overline{r_a}$ (as shown in formula (2)). From the probability aspect of view, the active user's mean rating is always a value in Median Values ($\overline{r_a}$ =3.6667 in our selected *TR*). At the same time, the interval of the adjustments ($k\sum_{i\in 1}^{n} w(a,i)(r_{i,j} - \overline{r_i})$) is relatively small compared with $\overline{r_a}$. Fig. 3 shows the adjustment for $\overline{r_a}$ in formula (2) in our experiment. The adjustment belongs to the interval of [-0.6 0.8], which is small compared with $\overline{r_a}$. So MRS failed to perform well on the Polar Values.

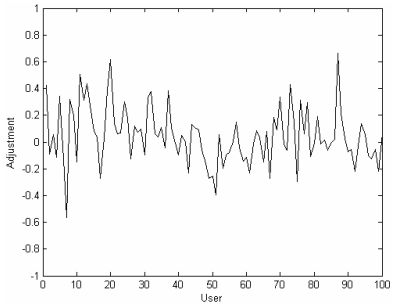


Figure 3. Adjustment for $\overline{r_a}$ in formula (2).

3.3 Our Proposed Bayesian Memory-Based Reputation System

To deal with the limitation of MBR as shown in Section 3.2, we propose a Bayesian Memory-based Reputation System (BMRS). Our main motivation is to give ratings on service providers offering high and low quality services as accurate as on those offering average quality services.

The key idea of BMRS is that: instead of directly using formula (2) to calculate $p_{a,j}$ for *TS* , as did by MRS, we adjust *P* by analyzing $p_{a,j}$'s statistics distribution. The adjustment is based the usage of Bayesian Theorem. Bayesian theorem is a mathematical theorem that follows very quickly from the axioms of probability theory. In practice, it is used to calculate the updated probability of some target phenomenon or hypothesis given new empirical data and the prior probability. In our experiment, we calculate $p_{a,j}$ for *TR* , and compare $p_{a,j}$ with the real rating values in *TR* given by active user. The comparison results are used as the prior probability.

Formula (4) gives the well-known Bayesian theorem:

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)} \tag{4}$$

where $P(H|E)$ is the posteriori probability, which is a measure of belief about a hypothesis *H* updated in response to evidence E ; $P(E|H)$ is the conditional probability, which is the probability of E given H ; $P(H)$ is the prior probability, which is the belief about *H* in absence of evidence; $P(E)$ is the probability of E .

By using the Bayesian theorem, we give the prediction mechanism of BMRS based on the conventional MRS as follows:

1. Calculate $p_{a,j}$ for *TR* and compare with *TR* to get the prior knowledge.

(1) Calculate P_R , which is used to represent the vector of $p_{a,j}$ for *TR* , use formula (2).

(2) Analyze P_R and divide the interval of $p_{a,j}$ into m suitable categories.

$$Cat_{a,j} = f(p_{a,j}) \tag{5}$$

where $Cat_{a,j}$ is the category of $p_{a,j}$. $Cat_{a,j}$ is a function of $p_{a,j}$, and there are totally m categories for P_R. $p_{a,j} \in P_R$. The function f is decided by the analysis of the probability distribution of $p_{a,j}$, and a concrete example will be given in Section 3.4.

(3) Calculate the probability of each divided category (i.e. $P(Cat_{a,j})$) by analyzing P_R based on formula (5).

(4) Calculate the probability of each possible rating value i.e. $P(TrueValue = i)$. For EachMovie Dataset, $i \in$ [0, 0.2, 0.4, 0.6, 0.8, 1]. Since we amplify the ratings as shown in Tab.1, $i \in$ [1, 2, 3, 4, 5, 6] in our paper.

(5) Calculate the conditional probability $P(Cat_{a,j} \mid TrueValue = i)$, i.e. the probability of each category given the real predicted value equals to i. This is based on the comparison of P_R, TR and the calculation of formula (5).

2. For TS, calculate posteriori probability for each $p_{a,j} \in P$.

(1) Calculate P using formula (2).

(2) Map each $p_{a,j} \in P$ into a category using formula (5).

(3) Using formula (6), calculate the posteriori probability $P(TrueValue = i \mid Cat_{a,j})$.

$$P(TrueValue = i \mid Cat_{a,j}) = \frac{P(Cat_{a,j} \mid TrueValue = i) P(TrueValue = i)}{P(Cat_{a,j})} \tag{6}$$

The calculation of the right side of formula (6) is based on the prior knowledge gotten in step 1.

3. Give the predicted value for each item $j \in TS$ using formula (7).

$$p^{B}_{a,j} = \sum_{i=1}^{n} P(TrueValue = i \mid Cat_{a,j}) * i \tag{7}$$

where $p^{B}_{a,j}$ is the predicted value for active user a on item j; i is the possible real rating value; n is total number of i.

3.4 Simulation Results of BMRS

This section is an extension of the experiment in Section 3.2 by using the mechanism of BMRS shown in Section 3.3. The simulation results in this section give the comparison of accuracy on predicted values given by MRS and our proposed BMRS.

The simulation steps are shown as follows:

(1) Calculate P_R and divide the interval of $p_{a,j}$ into m suitable categories.

For the TR we chose, we get $p_{a,j} \in [3.076\ 4.3332]$. We divide the interval of $p_{a,j}$ into 5 categories as shown in Table 2. Fig. 4 gives the probability of each category.

Table 2. Five Categories.

Category Name	Interval
A	[3.0 3.6]
B	[3.6 3.7]
C	[3.7 3.8]
D	[3.8 3.9]
E	[3.9 4.4]

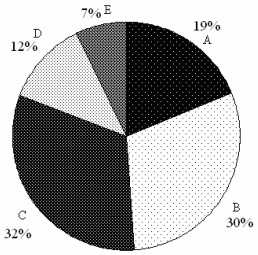


Figure 4. Probability distribution of each category.

Thus we get the following function for formula (5) in our simulation:

$$Cat_{a,j} = f(p_{a,j}) = \begin{cases} A & p_{a,j} < 3.6 \\ B & 3.6 \le p_{a,j} < 3.7 \\ C & 3.7 \le p_{a,j} < 3.8 \\ D & 3.8 \le p_{a,j} < 3.9 \\ E & p_{a,j} \ge 3.9 \end{cases}$$

Based on the above classification of categories, we use formula (6) (7) to calculate $p^{B}_{a,j}$. And $p^{B}_{a,j}$ is used to make a comparison with P. To make the comparison results more distinct, we divide TS into six Sub-Test Datasets, where each Sub-Test Dataset consists of the items on which the active user gave the same rating values (1, 2, 3, 4, 5, 6 respectively). E.g., all the real rating values on the items in the first Sub-Test Dataset are equal to 1. We give the comparison results on the six Sub-Test Datasets in Fig. 5, Fig. 6, Fig. 7, Fig. 8, Fig. 9, Fig. 10 respectively, where 'TRUE' means the real rating value given by the active user on the item;

'MRS' means the predicted value calculated by MRS; 'BMRS' means the predicted value calculated by our proposed BMRS.

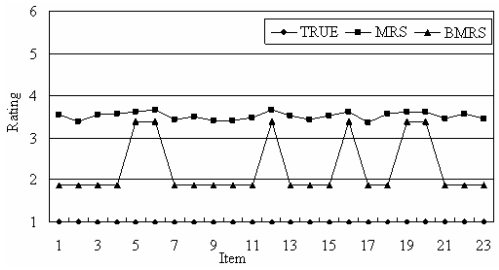


Figure 5. Comparison on items with real ratings equal to 1.

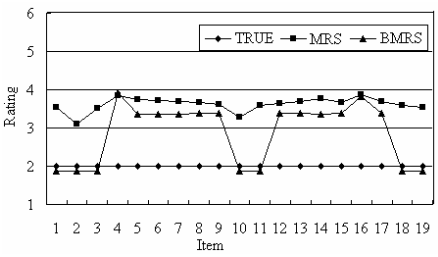


Figure 6. Comparison on items with real ratings equal to 2.

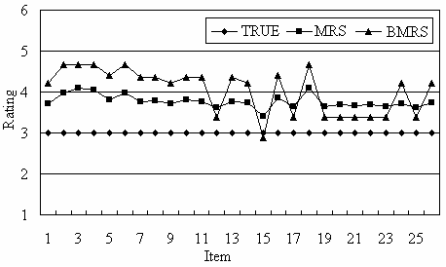


Figure 7. Comparison on items with real ratings equal to 3.

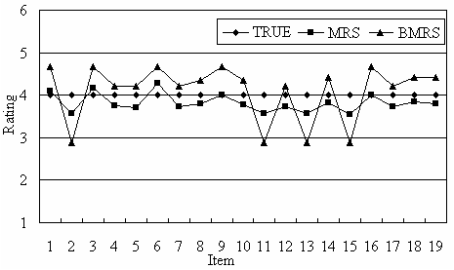


Figure 8. Comparison on items with real ratings equal to 4.

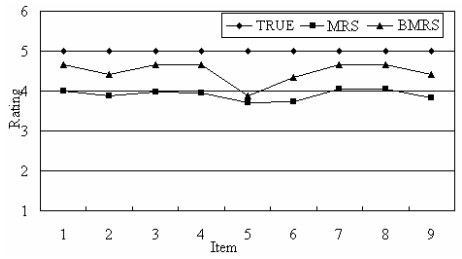


Figure 9. Comparison on items with real ratings equal to 5.

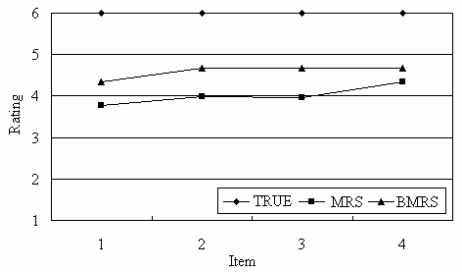


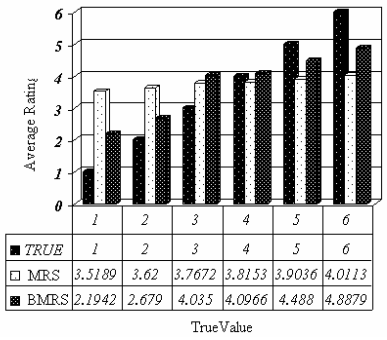
Figure 10. Comparison on items with real ratings equal to 5.

The simulation results in the above 6 figures show that:

1. When the real rating values given by the active user are 1, 2, 5 and 6 (Fig. 5, Fig. 6, Fig. 9 and Fig. 10), almost all the predicted values given by BMRS are closer to the real rating values than MRS, which means BMRS has higher accuracy.

2. When the real rating values given by the active user are 3 and 4 (Fig. 7 and Fig. 8), most of the predicted values given by MRS are closer to the real rating values than our proposed BMRS, which means MRS has higher accuracy in this case.

For the predicted values in each Sub-Test Dataset using MRS and BMRS, we calculate their average value and give the results in Fig. 11. And it is easy to observe that when the real rating value equals to 1, 2, 5 and 6, MRS is far from enough to give the correct prediction. However, our BMRS has better performance at those situations. When the real rating value equals to 3, MRS has higher accuracy than ours. And when the real rating value equals to 4, though the average rating value predicted by BMRS is closer to the 4, MRS is better than BMRS since it has smaller variance refers to Fig. 8.



	1	2	3	4	5	6
TRUE	1	2	3	4	5	6
MRS	3.5189	3.62	3.7672	3.8153	3.9036	4.0113
BMRS	2.1942	2.679	4.035	4.0966	4.488	4.8879

Figure 11. The comparison of the average predicted values given by MRS and BMRS.

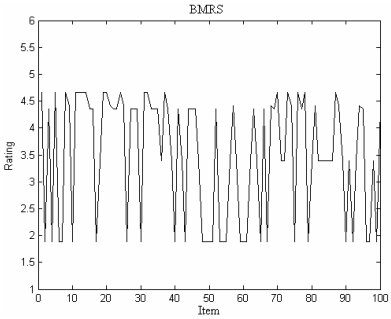


Figure 12. Distribution of the predicted values using BMRS.

Fig. 12 gives the distribution of the predicted values given by our BMRS on the test dataset. Recall the predicted values given by MRS in Fig. 1 and real rating values in Fig. 2, it is easy to notice that the interval of the predicted values given by BMRS has distinctly extended the interval of those given by MRS, and is closer to the interval of the real ratings.

4. Conclusions and Future Work

Reputation Systems provide a way for building trust through social control by utilizing community based feedback about past experiences of peers to help making recommendation and judgment on quality and reliability of the transaction [21]. MRS using Pearson Correlation Coefficient is one of the most effective methods. However, we found though experiments that MRS has low accuracy when make prediction on Polar Values, i.e. ratings on services providers give high and low quality service. We propose a Baysian Memory-based Reputation System (BMRS) which uses Bayesian Theory to analyze the probability distribution of the predicted values given by MRS and make suitable adjustment. Simulation results show that our proposed BMRS has higher accuracy than MRS on Polar Values.

In the future, we plan to focus on how to filter out unfair ratings in our reputation system. And to filter out the unfair ratings is one of the basic requirements to build up a robust reputation system. Based on our comparison between BMRS and MRS, we believe that the usage of BMRS in dynamic environments presents a promising path for the future research.

5. ACKNOWLEDGMENTS

This research was supported by the MIC (Ministry of Information and Communication), Korea, Under the ITFSIP (IT Foreign Specialist Inviting Program) supervised by the IITA (Institute of Information Technology Advancement).

Dr. Young-Koo Lee is the corresponding author.

6. REFERENCES

[1] Andrews P. Reputation management: Building trust among virtual strangers. IBM Global Business Services. Sept. 2006.

[2] Yu K,Schwaighofer A,Tresp V,et al.Probabilistic memory-based collaborative filtering[J].IEEE Transaction on Knowledge and Data Engineering,2004,16(1):56-69.

[3] Andrews P, “Reputation management: Building trust among virtual strangers”, IBM Global Business Services, Sept. 2006.

[4] A. Abdul-Rahman and S. Hailes, “Supporting trust in virtual communities”, Proc of the 33rd Hawaii International Conference on Systems Science,2000..

[5] Cahill, V. and Gray, E. et al, “Using trust for secure collaboration in uncertain environments”, IEEE Pervasive Computing, 2/3. pp. 52-61.

[6] Carbone, M., Nielsen, M. and Sassone, V., “A formal model for trust in dynamic networks”, In Proceedings of Int. Conference on Software Engineering and Formal Methods, SEFM 2003., pp. 54-61.

[7] Manchala D W, “Trust metrics, models and protocols for electronic commerce transactions”, In Proceedings of 18th International Conference on Distributed Computing Systems. Amsterdam:IEEE Computer Society, 1998. 312-321.

[8] Sepandar D.K., Mario T.S. and Hector G.M., “The EigenTrust algorithm for reputation management in P2P networks”, In Proceedings of the 12th Int'l Conf. on World Wide Web. ACM Press, 2003. 640-651.

[9] R. Levien, Attack Resistant Trust Metrics. PhD thesis, University of California at Berkeley.

[10] C.N. Ziegler and G. Lausen, “Spreading activation models for trust propagation”, In Proceedings of IEEE. International Conference on e-Technology, e-Commerce, and e-Service (EEE ’04), 2004.

[11] A. Jøsang, R. Ismail and C. Boyd, “A survey of trust and reputation systems for online service provision”, Decision Support Systems, Volume 43 , Issue 2 (March 2007), Pages: 618-644.

[12] S. Song, K. Hwang, R. Zhou, and YK Kwok, "Trusted P2P Transactions with Fuzzy Reputation Aggregation," Internet Computing, IEEE, Vol. 9, No. 6. (2005), pp. 24-34.

[13] P. Resnick and R. Zeckhauser, “Trust among strangers in internet transactions: empirical analysis of eBay’s reputation system”, The Economics of the Internet and E-Commerce, 2002.

[14] A. Jøsang and J. Haller, “Dirichlet Reputation Systems” (to appear). Proceedings of the Second International Conference on Availability, Reliability and Security (ARES 2007), Vienna, April 2007

[15] A. Jøsang and R. Ismail, “The Beta Reputation System”. Proceedings of the 15th Bled Conference on Electronic Commerce, Bled, Slovenia, 17-19 June 2002.

[16] Mui, L., Mohtashemi, M., Halberstadt, A., “A computational model of trust and reputation”, In Proceedings of 35th Hawaii International Conference on System Science. (2002) 280-287

[17] Withby A, Jφsang A and Indulska J., “Filtering Out Unfair Ratings in Bayesian Reputation Systems”, Journal of Management Research, 2005, 4(2):48～64.

[18] Jφsang A., Pope S., “Semantic constraints for trust transitivity”, In:Proceedings of the 2nd Asia-Pacific Conference on Conceptual Modelling(APCCM2005), 59～68

[19] Yu B . and Munindar P . , “An evidential model of distributed reputation management” . In proceeding of AAMAS02. 2002-07 : l5～l9

[20] Brin S., Page L. Motwami R. and Winograd T., “The PageRank Citation Ranking: Bringing Order to The Web”, Technical Report. Standford Digital Library Technologies Project. Stanford: Stanford University, CA, 1998

[21] Xiong L, Liu L.PeerTrust:Supporting Reputation-Based Trust for Peer-to-Peer Electronic Communities.IEEE transactions on knowledge and data engineering,2004,16 (7)