A user reputation model for a user-interactive question answering system

Wei Chen1, Qingtian Zeng1,2, Liu Wenyin1,*† and Tianyong Hao1

1Department of Computer Science, City University of Hong Kong, 83 Tat Chee Avenue, Hong Kong, People’s Republic of China
2Department of Computer Science and Technology, Shandong University of Science and Technology, Qingdao 266510, People’s Republic of China

SUMMARY

In this paper, we propose a user reputation model and apply it to a user-interactive question answering system. It combines the social network analysis approach and the user rating approach. Social network analysis is applied to analyze the impact of participant users’ relations to their reputations. User rating is used to acquire direct judgment of a user’s reputation based on other users’ experiences with this user. Preliminary experiments show that the computed reputations based on our proposed reputation model can reflect the actual reputations of the simulated roles and therefore can fit in well with our user-interactive question answering system. Copyright © 2006 John Wiley & Sons, Ltd.

Received 1 September 2006; Accepted 2 September 2006

KEY WORDS: reputation system; question answering system; social network analysis; user rating; multi-agent system

1. INTRODUCTION

The Internet is playing an important role in people’s daily lives. However, participants in online transactions are often at risk because they have little information about their transaction partners. The difficulty of collecting evidence about unknown transaction partners makes it hard to distinguish...
between high- and low-quality service providers on the Internet [1]. Online reputation systems record and report an online trader’s reputation according to other traders’ feedback [2]. The information available from a reputation system helps to improve a prospective trader’s trust and conviction in conducting online transactions. It also reduces the chances of fraud because traders are typically expected to maintain good reputation records in order to maximize their profits. In this way, a reputation system may not only serve as a guide to otherwise clueless entrants but also help enhance the predictability of existing traders’ behavior and honesty [3]. The lack of this kind of information can be amended by trust and reputation. Therefore, the study of online reputation systems has increasingly attracted researchers’ attentions.

Traditional online reputation systems usually focus on online trade systems [4], for example eBay [5–7]. These systems are usually based on user ratings among the users, mainly because the relations between the participants are very simple in online trade systems as they are limited to selling and buying.

Our scenario is different. In this paper, we propose a reputation model for our user-interactive question answering (QA) system—CuteAid [8]. The reputation model also works for BuyAns [9] QA system. CuteAid is an experimental system for internal usage and BuyAns is a commercial system for public usage. In the QA system, users can ask questions and answer the questions of other users. If the answer of someone is chosen as the correct answer, he/she can obtain a payment for this answer from the asker. This kind of QA system is different from traditional online trade systems from the viewpoint of developing a reputation system. Relations between participants may not be limited to selling and buying. Instead, this kind of QA system has more in common with an online social community.

In this paper, we adopt the method of social network analysis to simulate the community in our QA system, and propose a method to compute users’ reputations. This method combines the user rating approach of traditional online trade systems and the approach based on social network analysis. Our simulation experiment with a multi-agent system on JADE [10] shows that the computed reputations based on our reputation model can reflect the actual reputations of participants.

2. RELATED WORK

Reputation is an important organization asset, particularly in the era of e-commerce [3]. In an online consumer-to-consumer (C2C) auction market, a trader’s reputation sends an important signal to his/her trading partners in their respective decisions on C2C transactions, owing to the nature of the anonymous transaction process [3]. The online reputation system has attracted researchers from the fields of behavioral science and economics to investigate the new issues of reputation in the e-market, such as the effects of online reputation on a trader’s trust (e.g. [4]) and auction price (e.g. [6]). Dellarocas introduces a model for analyzing marketplaces that rely on binary reputation mechanisms for quality signaling and quality control [7]. In this model, sellers keep their actual quality private and choose what quality to advertise. The reputation mechanism is primarily used to determine whether sellers advertise truthfully. Buyers may exercise some leniency when rating sellers, which needs to be compensated for by a corresponding strictness when judging sellers’ feedback profiles [7]. Xiong and Liu [11] present a reputation-based trust supporting framework, which includes a coherent adaptive trust model for quantifying and comparing the trustworthiness of peers based on a transaction-based feedback system, and a decentralized implementation of such a model over a structured peer-to-peer (P2P) network.
Their model has two main features. First, they introduce three basic trust parameters and two adaptive factors in computing the trustworthiness of peers, namely, the feedback a peer receives from other peers, the total number of transactions a peer performs, the credibility of the feedback sources, the transaction context factor, and the community context factor. Second, they define a general trust metric to combine these parameters. The reputation system in eBay is described in detail by Resnick and Zeckhauser [12]. A reputation computation engine is used to compute users’ reputations after user ratings are collected. Several kinds of principles are used to compute reputations based on user ratings, such as the simple summation model [5] and the Bayesian model [13]. In order to avoid the unfair factors caused by subjective ratings, Dellarocas proposed a method of immunizing against unfair ratings and discriminatory behavior [14]. However, the basic idea of these current reputation systems, and especially in many current online trade systems, is to let participants rate each other. Reputation scores are computed based on these ratings using aggregative approaches. Meanwhile, social network analysis has been applied by Sabater and Sierra [15] in the reputation system of a multi-agent system.

The CuteAid QA system can be used as a collaborative approach to knowledge acquisition. As more knowledge is accumulated and well represented, more accurate automatic QA is possible. Another research effort of collaborative knowledge accumulation and management is the Knowledge Grid [16]. The Knowledge Grid is an intelligent, sustainable Internet application environment that enables people or virtual roles to effectively capture, publish, share, and manage explicit knowledge resources [17]. In the environment of Knowledge Grid, a QA system can act as a recommender system to provide contents from users to users based on the users’ models [18], in which reputation is an important aspect. However, the use of reputation systems in current online user-interactive QA systems is not extensive. Google Answers [19], which is a well-known example of this type of system, has introduced a simple answerer evaluating mechanism relying on user ratings. Other systems, such as Sina iAsk [20], Baidu Zhidao [21], and Yahoo! Answers [22], have employed similar mechanisms.

In this paper, we propose a reputation model and system for an online user-interactive QA system. In contrast to the work in [3,6,7,12,15,19], our system has the following features.

- Our model combines the user rating approach in traditional online trade systems [12] and the approach based on social network analysis [15].
- The reputation model proposed is used for a Web-based user-interactive QA system, not the traditional online trade systems [3,6,7].
- A simulation experiment with a multi-agent system is presented to show the computed reputations based on our reputation model. The experiment results indicate that the reputation model can reflect the actual reputations of participants.

3. INTERACTIONS BETWEEN USERS IN CUTEAID

The user-interactive QA system, CuteAid, is developed as a platform for users to ask and answer questions. Figure 1 shows all kinds of interactions between users in CuteAid. The main interactions are as follows.

(1) A user can post his/her questions with the intention of offering an amount of money for the correct answers.
(2) Other participants can post answers to those questions they think they have answers for. All the answers in the system are visible to the public.

(3) An answer may be supported by others if they think it is correct. The number of supports of an answer is also publicly visible.

(4) When the asker of a question thinks there is a correct answer to the question, he/she should mark the correct answer and close the question in time. Only the earliest correct answer should be marked.

(5) If an answer is marked as correct by the asker, the answerer can get the money offered by the asker.

(6) People are also encouraged to complain about misbehaviors in the system, which include marking correct answers improperly.

The actions of choosing the correct answer and an arbitrator’s processing of complaints will control the flow of money, as shown in Figure 1 with thick arrowheads.

We can define a directed graph $G = (V, E)$ to represent the relations between the participants in CuteAid, where $V = \{v_1, v_2, \ldots, v_m\}$ is a set of nodes of the graph $G$, each of which represents a participant, and $E = \{e_1, e_2, \ldots, e_n\}$ is a set of edges of $G$, each of which represents a relation between two participants. Each edge has a type attribute, which is defined as follows:

$$\text{typeof}(e_i) \in E \cdot \text{TYPE} = \{\text{Answer. Question}, \text{Support. Answer}, \text{Choose. As. Correct. Answer}, \text{Personal. Message}, \text{Complain}\}, \quad i = 1, 2, \ldots, n$$  (1)
Table I. Types of relations between participants and their impact weights.

<table>
<thead>
<tr>
<th>Relation</th>
<th>Weight of relation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Answer question</td>
<td>+0.5</td>
</tr>
<tr>
<td>Support answer</td>
<td>+2</td>
</tr>
<tr>
<td>Choose as correct answer</td>
<td>+3</td>
</tr>
<tr>
<td>Personal message</td>
<td>+1</td>
</tr>
<tr>
<td>Complain</td>
<td>−3</td>
</tr>
</tbody>
</table>

where $E_{TYPE}$ is the set of all possible types of relation that may affect participants’ reputations in our QA system.

A new edge is drawn in graph $G$ when a new relation action is formed between two participants. A type value is also assigned to the new edge to represent the type of the relation, such as AnswerQuestion. When a new relation is formed based on a user’s behavior, the graph will be updated from $G_i$ to $G_{i+1}$ by adding a new edge into $G_i$.

4. USER REPUTATION MODEL

A reputation model is used to evaluate the users’ reputations in CuteAid. Reputation is what is generally said or believed about a person’s or thing’s character or standing. Reputation is also used to judge the trustworthiness of a user in a transaction.

In CuteAid, the user reputation model is a combination of two approaches, social network analysis and user rating. We will explain these approaches in detail in the following sections.

4.1. Social network and user relation analysis

Social network analysis is based on an assumption of the importance of relationships among interacting units [23]. As CuteAid is similar to an online virtual community, it is natural for us to use a social network to analyze users’ reputations.

In social network analysis, relational data are represented using graphs called sociograms, which are directed and weighted graphs [15]. Each participant is represented as a node and each relation is represented as an edge. The value of a node represents the importance of the node, which is an important factor that forms the corresponding user’s reputation.

There may be many kinds of relations in CuteAid. Here we choose the main relations which are relevant to the users’ reputations. Table I lists the types of relation between participants and the weights that can affect their reputations. Different kinds of relations affect users’ reputations differently. The weights in the table are experimental values because they mainly depend on the characteristics of the QA system. For example, Choose as Correct Answer is given a higher positive value than Support Answer because a correct answer chosen by the asker is considered more important to form the
answerer’s reputation than an anonymous support, and Complain is given a high negative value because a complaint is a strong negative evidence of a person’s reputation. These weight values are used in our current system, which may be changed by the system administrator based on more experiments.

All users’ reputations are updated periodically in fixed time intervals. As we can see from (5), the change is the sum of the increments caused by all relations occurring during this interval. The increments are calculated separately for each individual type of relation. For a particular type (τ) of relation between participants, its initial importance value of each node is assigned to zero. We first use δ(v → b) to define the influence of a single relation v → b that occurs between time $t_i$ and $t_{i+1}$ on the reputation of user b:

$$\delta(v \rightarrow b) = \arctan \frac{R_t(v) - R_0}{\pi} + \frac{1}{2}$$  \hspace{1cm} (2)

where $R_t(v)$ is the reputation of v at time $t_i$, and $R_0$ is a parameter determined by system administrators to identify the reputations of average people.

Let $\Phi(b, \tau)$ represent the set of all relations $v \rightarrow b$ that occur between time $t_i$ and $t_{i+1}$ with a specified type $\tau$,

$$\Phi(b, \tau) = \{v \rightarrow b \mid v \in V \land (v \rightarrow b) \in E \land \text{typeof}(v \rightarrow b) = \tau \land t_i < \text{timeof}(v \rightarrow b) < t_{i+1}\}$$  \hspace{1cm} (3)

where timeof $(v \rightarrow b)$ is the occurring time of $v \rightarrow b$.

The increment of the importance of node b due to all relations with type $\tau$ from time $t_i$ to $t_{i+1}$ can be represented as

$$\Delta I^\tau_{t_{i+1}}(b) = \sum_{v \rightarrow b \in \Phi(b, \tau)} \delta(v \rightarrow b)$$  \hspace{1cm} (4)

where $\tau$ is a specified type of relation between participants.

In a social network, a node that connects to a large number of nodes is considered as being important in the network. In (4), when a new relation $v \rightarrow b$ is formed, it means that v communicates with b. In other words, v contributes to the importance of b in the network. Therefore, we add an edge from v to b in the graph to represent this contribution and mark the edge with its type.

For example, when user v supports an answer of user b, it means that v contributes to the importance of b. Hence, we add an edge in the graph that points to b from v and mark the type of this edge with SupportAnswer.

In fact, the weight of the edge does not necessarily contribute to b fully. It also depends on the reputation of v at time $t_i$. Suppose that a bad user v with a very low reputation complains to user b; it is reasonable to believe that maybe v is just playing tricks on b, and the bad influence should not be fully charged on b. Therefore, we calculate the influence of a relation also based on the reputation of v.

Our basic idea is that people who have a higher reputation should affect other people’s reputations to a greater extent than someone with a lower reputation. The parameter $R_0$ in (2) is used to adjust how much a participant should affect others, and it is set by the administrators to represent the reputations of average people. In (2), it is assumed that the newly added relation is of type $\tau$. If the reputation of the user v is much higher than $R_0$, the result $\delta(v \rightarrow b)$ is nearly 1, which indicates that the influence caused by relation $v \rightarrow b$ is nearly fully charged on b. If the reputation of v gets smaller and is close to $R_0$, the result of $\delta(v \rightarrow b)$ is reduced rapidly. When the reputation of v is equal to $R_0$, which means v is a person with average reputation, the result falls to 0.5, which shows that user b is only affected by half of the relation $v \rightarrow b$. When the reputation of v is much smaller than $R_0$, the result is nearly zero, which indicates that participants with very low reputations can hardly affect others by a relation.
The increment of the reputation of a user $b$ coming from social network analysis from time $t_i$ to $t_{i+1}$ can be calculated by making a weighted sum of all new relations

$$\Delta R_{t_{i+1}}^N (b) = \sum_{\forall \tau \in E_{\text{TYPE}}} (\Delta I_{t_{i+1}}^\tau (b) \times W_{\tau})$$

where $E_{\text{TYPE}}$, defined in (1), is the set of all types of relations that may affect the reputation of $b$, $I_{t_{i+1}}^\tau (b)$, defined in (4), is the increment of importance of $b$ due to all the relations of a particular type $\tau$ from time $t_i$ to $t_{i+1}$, and $W_{\tau}$ is the weight of relation type $\tau$.

### 4.2. User rating

It is a common method to compute users’ reputations by using ratings in the reputation systems of current online trade systems. Because of the nature of online trade systems, they usually let the buyer and seller in a transaction rate each other. The systems collect these kinds of ratings to compute the reputations for each user.

We adopt the idea of user rating and customize it to fit CuteAid. In our system, there are two kinds of ratings that are relevant to a user’s reputation.

First, after a question is posted, there may be many answers to it. All users in the system can choose the action support to support the answers that they think are correct. A large number of supports to a certain answer can show not only that this answer is probably correct but also that the answerer may be particularly knowledgeable on the subject. People who often give correct answers and receive many supports may be an expert and should therefore be given a high reputation value. Hence, the number of supports of an answer can be used as a kind of rating to the answerer.

Second, in our system, a question should be closed by the asker when he/she finds the correct answer. When the question is closed, the asker should choose one or several answers as the best (correct) answers and give the relevant credit to them. Similarly, choosing the best (correct) answer can also be considered as a rating of the answerer.

In the section on social network analysis, we have mentioned these kinds of users’ behaviors such as answer supporting and answer selection. In this section, their usage is different. In social network analysis, we focus specifically on the user’s behavior. We consider it as a kind of relation between the two users and use it to form the social network. However, in user rating, we focus on the score derived from the user’s behavior. For example, we will give the rated user a score based on the rating.

Different kinds of ratings contribute differently to a participant’s reputation. Table II shows the weight of each kind of rating in the model used in our current system. As for Table I, Choose as Correct Answer is given a higher value than Support Answer. These weights in the table are experimental values and may be changed by the system administrator based on more experiments.

Each user has an initial reputation (coming from user rating) of zero. The change of a user’s reputation (that comes from user rating) from time $t_i$ to $t_{i+1}$ is defined using the following formula:

$$\Delta R_{t_{i+1}}^R (b) = \sum_{t \in \{t \mid t_i < t < t_{i+1} \}} W_t$$

where $W_t$ is the weight of the rating at time $t$ if there is a rating to the user $b$ at time $t$ or 0 if there is none at that time.
Table II. Weights of different kinds of ratings.

<table>
<thead>
<tr>
<th>Type of rating</th>
<th>Weight of rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Support answer</td>
<td>+1</td>
</tr>
<tr>
<td>Choose as correct answer</td>
<td>+2</td>
</tr>
</tbody>
</table>

4.3. Reputation computation

Combining the methods presented in Sections 4.1 and 4.2, the following formula is used to compute a participant’s reputation:

\[ R_{t_i+1}(b) = \beta R_{t_i}(b) + \mu_1 \Delta R^N_{t_i+1}(b) + \mu_2 \Delta R^R_{t_i+1}(b) \]  

(7)

where \( R_{t_i+1}(b) \) represents the reputation of user \( b \) at time \( t_{i+1} \), \( \Delta R^N_{t_i+1}(b) \) is the increment of the reputation of \( b \) computed based on social network analysis from time \( t_i \) to \( t_{i+1} \), \( \Delta R^R_{t_i+1}(b) \) is the increment of the reputation of \( b \) computed based on user rating from time \( t_i \) to \( t_{i+1} \), \( \mu_1 \) and \( \mu_2 \) are empirical parameters set by administrators, and \( \beta \) is an attenuation factor.

In human society, the recent behavior of a person usually affects his/her reputation more heavily than events that happened a long time ago, since the past events may gradually be forgotten. In order to reflect this characteristic of human society, an attenuation factor \( \beta \) is introduced into the user reputation model. The reputation value of every person is weakened by this factor periodically when users’ reputations update, meaning that past behavior contributes an increasingly smaller weight to a person’s current reputation.

Our objective of computing users’ reputations is to show them to the public so that one can take them into account when considering transactions with another user. In order to show an intuitive number to the public to represent each person’s reputation, the reputation value should be normalized before being shown to the public.

5. PRELIMINARY EXPERIMENTS

A set of experiments have been performed to improve our user reputation model and also to prove its accuracy and stability. As this work was performed while our user-interactive QA system was in its initial stages of development, there were insufficient historical data of users’ behaviors to evaluate our user reputation model.

Therefore, we have developed a multi-agent system on the JADE platform to simulate the interactions between users in our QA system. There are two kinds of agents in the multi-agent system. A server agent is deployed to act as the QA system and 100 user agents are used to act as 100 personal users in the system. The user agents are designed in a role-playing method. Some user agents act as if they are good people and some as if they are bad. Each user agent is assigned an internal reputation value at initialization that represents its actual reputation.

We analyze each kind of users’ behavior characteristics in CuteAid and define their behaviors accordingly. A user agent’s behaviors in the multi-agent system are designed such that the user agent
Figure 2. The changes of the average difference between the computed and actual reputations of 100 user agents.

communicates with the server agent to ask and answer questions and performs other actions according to its actual reputation.

The server agent simulates the function of our QA system and traces the user agents’ actions, and computes their reputations using our reputation model‡. We then compare the difference between their computed reputations and their actual reputations determined by their roles so as to verify the model.

Figure 2 shows the changes of the average difference between the computed reputations and the actual reputations of the 100 user agents in the system as time goes. Their reputations are normalized to a standard normal distribution when displayed.

From Figure 2 we can see that the average difference mentioned above is large at the start time of the system. At this point, the social network has just been formed and there are no sufficient data for the system to model and compute the users’ reputations. The average difference reduces rapidly as time goes by, which shows that the reputation system is collecting information about users as more actions occur between the users, and the computed reputations are closer to the actual values. After sufficient information is collected, the average difference reduces more slowly than before and keeps at a low

‡We set $\beta = 0.9999898$, $\mu_1 = \mu_2 = 1$ in (7) in the experiments. These are experimental values.
level for a long time. This shows that the user reputation model converges quickly and the results tend to be stable when sufficient information is collected.

Figure 2 only shows the change of the average difference between the actual reputations and the computed values of the whole system. Figure 3 presents six typical cases of differences between the computed reputations and the actual ones. Figure 3(a) shows the best results of the user reputation model. We can see that there is little difference between the computed and actual reputations after a short running time. Figure 3(b) shows the average results, which illustrate a small difference between the computed and actual reputations. However, Figure 3(c) shows the worst-case results where the differences between the computed and actual reputations are still significant after a long time.

Table III shows the actual and computed reputations of the user agents. Limited by the space, we only present 10 of them here in the table. All test data are available at http://www.cs.cityu.edu.hk/~wchen2/repexp06/.

Figure 4 shows the distribution of the users at each difference level after the system is stable.
Table III. The actual and computed reputations of the first 10 user agents.

<table>
<thead>
<tr>
<th>Number</th>
<th>Actual reputation</th>
<th>Computed reputation</th>
<th>Number</th>
<th>Actual reputation</th>
<th>Computed reputation</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.620918</td>
<td>1.000990</td>
<td>1</td>
<td>0.561792</td>
<td>0.394693</td>
</tr>
<tr>
<td>2</td>
<td>0.263612</td>
<td>-0.086973</td>
<td>3</td>
<td>-2.057856</td>
<td>-2.330079</td>
</tr>
<tr>
<td>4</td>
<td>-1.232341</td>
<td>-1.180947</td>
<td>5</td>
<td>0.813657</td>
<td>0.591097</td>
</tr>
<tr>
<td>6</td>
<td>-0.922178</td>
<td>-1.232341</td>
<td>7</td>
<td>1.086568</td>
<td>0.421668</td>
</tr>
<tr>
<td>8</td>
<td>2.057856</td>
<td>2.057856</td>
<td>9</td>
<td>0.289397</td>
<td>0.289397</td>
</tr>
</tbody>
</table>

Figure 4. Distribution of the user agents at each level of difference between the user agents’ computed and the actual reputations after the system is stable.

6. CONCLUSION AND FUTURE WORK

This paper presents a user reputation model that is used in our user-interactive QA system. The main merit of the user reputation model is that it adopts and combines two kinds of approaches, social network analysis and user rating. Social network analysis computes a participant’s reputation by analyzing the importance of the node based on the relations between nodes in a social network of the virtual community. The importance of a participant in the social network can reflect positively
on the participant’s reputation. The approach of user rating computes a participant’s reputation using subjective user ratings to the participant given by others in the community. These ratings can also reflect the participant’s reputation, although they are subjective. Using the combination of the two approaches leads to a good performance of our user reputation model, as we can see from our preliminary experiments.

Our user-interactive QA system is a kind of online virtual community. Its characteristics are different from other online trade systems that restrict participants’ behaviors to buying or selling goods. Participants in our system can perform many other actions, such as supporting answers and choosing correct answers, etc. The relations of participants in our system are sufficiently complex such that social network analysis can be applied. In addition, we also retain the traditional user rating approach. Their combination makes the user reputation model fit our user-interactive QA system well.

We have used a multi-agent system to simulate the users’ behaviors in our QA system for the evaluation of the reputation model. Although our preliminary evaluation to our user reputation model demonstrates its stability and flexibility, we hope that we can collect more real data from our CuteAid system and use them to validate our reputation model. We also hope that our reputation model can fit common online virtual communities, which have similar structures to our user-interactive QA system. However, more experiments will be needed to verify its applicability.

ACKNOWLEDGEMENTS

The work described in this paper was fully supported by a grant from City University of Hong Kong (Project No. 7001842) and the China Semantic Grid Research Plan (National Grand Fundamental Research 973 Program, Project No. 2003CB317002).

REFERENCES


