A Personalized Approach to Address Unfair Ratings in Multiagent Reputation Systems

Jie Zhang and **Robin Cohen**
David R. Cheriton School of Computer Science
University of Waterloo
Waterloo, ON, Canada N2L 3G1
{j44zhang, rcohen}@uwaterloo.ca

Abstract

In multiagent systems populated by self-interested agents, consumer agents would benefit by modeling the reputation of provider agents, in order to make effective decisions about which agents to trust. One method for representing reputation is to ask other agents in the system (called advisor agents) to provide ratings of the provider agents. The problem of unfair ratings exists in almost every reputation system, including both unfairly high and unfairly low ratings. We begin by surveying some existing approaches to this problem, characterizing their capabilities and categorizing them in terms of two main dimensions: public-private and global-local. The impact of reputation system architectures on approach selection is also discussed. Based on the study, we propose a novel personalized approach for effectively handling unfair ratings in an enhanced centralized reputation system. Experimental results demonstrate that the approach effectively adjusts the trustworthiness of advisor agents according to the percentages of unfair ratings provided by them. We then argue for the merits of our model as the basis for designing social networks to share reputation ratings of providers in multiagent systems.

Introduction

In many multiagent settings, agents are self-interested. They interact with each other to achieve their own goals. Provider agents provide services to consumer agents and try to maximize their profit. Consumer agents try to gain good services in terms of, for example, high quality and low prices. To ensure good interactions amongst agents, a reputation mechanism provides important social control in the multiagent system. In a reputation system, agents can rate each other. Agents estimate each other's reputation according to those ratings and choose the most reputable ones to interact with. However, a reputation system may be deceived by unfair ratings for an agent's personal gain. The problem of unfair ratings is fundamental and exists in almost every reputation system. A well-known example of this problem is that on the eBay system three men highly rate each other and later sell a fake painting for a very high price (Jøsang & Ismail 2002).

Dellarocas (Dellarocas 2000) distinguishes unfair ratings as unfairly high ratings and unfairly low ratings. Unfairly high ratings may be used to increase provider agents' reputations. They are often referred as "ballot stuffing". Unfairly low ratings of a provider agent may be provided by consumer agents who cooperate with other provider agents to drive the provider agent out of the system. They are often referred as "bad-mouthing". In bi-directional rating schemas where consumer and provider agents can rate each other, consumer agents may provide unfairly high ratings in hope of getting high ratings in return. This behavior is based on a social dictum, "be nice to others who are nice to you" (Mui, Mohtashemi, & Halberstadt 2002). Unfairly high ratings may also be given because agents tend to give high ratings as long as other agents pay for services or deliver the services requested (Chen & Singh 2001). Agents may also provide unfairly low ratings to retaliate against other agents for being rated low.

To ease the problem of unfair ratings, the eBay system only allows agents to provide ratings after their transactions have succeeded (Jøsang, Ismail, & Boyd 2005). It also adds a cost for each transaction. However, this approach still cannot stop fraudulent attempts.

Many researchers have proposed different theoretical approaches to handle unfair ratings. However, these approaches are only effective in limited situations. Dellarocas (Dellarocas 2000) proposes the *Cluster Filtering* approach to separate unfairly high ratings and fair ratings. This approach is unable to handle unfairly low ratings. Whitby et al. (Whitby, Jøsang, & Indulska 2005) extend the beta reputation system proposed by Jøsang and Ismail (Jøsang & Ismail 2002) to cope with unfair ratings by filtering out the ratings that are not in the majority amongst other ones. This approach is only effective when the majority of ratings are fair. The *GM-GC* approach is developed by Chen and Singh (Chen & Singh 2001) to compute reputations of agents based on their ratings. The computation of this approach is quite time consuming. Teacy et al. (Teacy *et al.* 2005) propose the *TRAVOS* model to cope with inaccurate reputation opinions. This model does not deal with changes of agents' behavior.

In this paper, we first survey different approaches for handling unfair ratings, and their advantages and disadvantages. We list the capabilities that an effective approach should have and compare these approaches based on their capabilities. We categorize these approaches in terms of two dimensions, a "public-private" dimension and a "global-local" dimension. We also discuss the impact of reputation system

architectures on the selection of approaches for handling unfair ratings.

Based on the study, we propose a personalized approach for effectively handling unfair ratings in enhanced centralized reputation systems. We consider the scenario where consumer agents elicit reputation ratings of provider agents from other consumer agents, known as advisor agents. The personalized approach first calculates what we refer to as the "private reputation" of an advisor agent, based on the consumer and advisor agents' ratings for commonly rated provider agents. When the consumer agent is not confident in its private reputation ratings it can also use what we refer to as the "public reputation" of the advisor agent. This public reputation is estimated based on the advisor agent's ratings for all provider agents in the system. The personalized approach ultimately computes a weighted average of private and public reputations to represent the trustworthiness of the advisor agent.

Experimental results demonstrate the effectiveness of the personalized approach in terms of adjusting advisor agents' trustworthiness based on the percentages of unfair ratings they provided. Our personalized model can therefore be seen a valuable approach to use when introducing social networks in order to model the reputations of providers in multiagent systems.

The rest of the paper is organized as follows. In the next section, we will survey different approaches for handling unfair ratings. We then propose a personalized approach for handling unfair ratings in an enhanced centralized reputation system. The following sections provide examples that go through each step of our approach and present our experimental results. Conclusions and future work are outlined in the last section.

Related Work

In this section, we summarize approaches used in different reputation systems, present our proposed categorization of reputation systems, and discuss the impact of reputation system architectures on the selection of approaches for handling unfair ratings.

Different Approaches

Different approaches have been proposed for handling unfair ratings. Those approaches are used in different reputation systems. We briefly summarize these systems and focus on their approaches for handling unfair ratings. Advantages and disadvantages of these approaches will be pointed out as well.

Dellarocas (Dellarocas 2000) simplifies the problem of unfair ratings by introducing the mechanism of controlled anonymity to avoid unfairly low ratings and negative discrimination. To reduce the effect of unfairly high ratings and positive discrimination, Dellarocas first uses collaborative filtering techniques to identify the nearest neighbors of a consumer agent based on their preference similarity with the consumer agent on commonly rated provider agents. He then proposes the *Cluster Filtering* approach to filter out unfairly high ratings provided by those neighbors. The idea of this approach is to apply a divisive clustering algorithm to separate the neighbors' ratings into two clusters, the lower rating cluster and the higher rating cluster. Ratings in the lower rating cluster are considered as fair ratings. Ratings in the higher rating cluster are considered as unfairly high ratings, and therefore are excluded or discounted. To deal with the situation where ratings vary over time, the *Cluster Filtering* approach considers only the ratings within the most recent time window whose width is influenced by the frequency of fair ratings. The *Cluster Filtering* approach copes with unfairly high ratings, takes into account preference similarity between consumer agents and advisor agents, and deals with changes of agents' ratings. One problem about this approach is that it does not handle unfairly low ratings. Dellarocas points out that the mechanism of controlled anonymity cannot avoid unfairly high ratings and positive discrimination because of identity signals between consumer and provider agents, for instance, provider agents may use a particular pattern in the amounts of their services. Identity signaling may not be able to avoid unfairly low ratings as well because consumer agents may rate against all other provider agents except their partners. In addition, controlled anonymity may only work in a sufficiently large system. In many smaller systems, however, it cannot be used due to the fact that agents may easily locate their conspirators' identity signals.

The beta reputation system (BRS) proposed by Jøsang and Ismail (Jøsang & Ismail 2002) estimates reputations of provider agents using a probabilistic model. This model is based on the beta probability density function, which can be used to represent probability distributions of binary events. This model is able to estimate the reputation of a provider agent by propagating ratings provided by multiple advisor agents. Ratings are combined by simply accumulating the amount of ratings supporting good reputation and the amount of ratings supporting bad reputation. To handle unfair feedback provided by advisor agents, Whitby et al. (Whitby, Jøsang, & Indulska 2005) extend the BRS to filter out those ratings that are not in the majority amongst other ones by using the *Iterated Filtering* approach. More specifically, feedback provided by each advisor agent consists of ratings supporting both good reputation and bad reputation of a provider agent, and is represented by a beta distribution. If the cumulated reputation of the provider agent falls between the lower and upper boundaries of feedback, this feedback will be considered as fair feedback. However, the *Iterated Filtering* approach is only effective when the significant majority of ratings are fair. This approach also does not consider consumer agents' personal experience with advisor agents' feedback.

Chen and Singh (Chen & Singh 2001) develop a general method, *GM-GC*, to automatically compute reputations for raters based on all the ratings given to each object. More specifically, the *GM-GC* approach computes a rater's reputation through three steps. The first step is to compute quality and confidence values of each of the rater's ratings for each object in a category. The quality value, called local match (LM) is calculated based on the frequency distribution of all ratings given to the same object. The confidence

level, called local confidence (LC) is determined by a piecewise function. LC is the same for all ratings for the same object. The second step is to compute the cumulated quality and confidence values of all ratings for each category of objects, which are called global match (GM) and global confidence (GC) respectively. GM and GC are computed by combining LM and LC for each object in the category. Finally, the *GM-GC* approach computes the rater's reputation based on the rater's GM and GC for each category. The *GM-GC* approach is different from filtering approaches. It explicitly computes reputations for raters to cope with unfair ratings. Ratings from less reputed raters will carry less weight and have less impact on accumulated reputations of provider agents. However, the computation of *GM-GC* is quite time consuming.

Teacy et al. (Teacy *et al.* 2005) propose the *TRAVOS* model, which is a trust and reputation model for agent-based virtual organizations. This model copes with inaccurate reputation advice by accomplishing two tasks. The first task is to estimate the accuracy of the current reputation advice based on the amount of accurate and inaccurate previous advice which is similar to that advice. The second task is to adjust reputation advice according to its accuracy. The aim of this task is to reduce the effect of inaccurate advice. This task is necessary because it can deal with the situation where an advisor agent untruthfully rates a provider agent a large number of times, also known as the problem of advisors "flooding" the system (Dellarocas 2000). Experimental results show that the *TRAVOS* model outperforms the *Iterated Filtering* approach. However, this model also has some problems. It assumes that provider agents act consistently. This assumption might not be true in many cases. The second problem is that this model repeatedly goes over an advisor agent's past advice at each time when estimating accuracy of this advisor agent's current advice. This could be a problem when the number of advisor agents is large and/or the amount of past advice provided by each advisor agent is large.

Wang and Vassileva (Wang & Vassileva 2003) propose a *Bayesian network*-based trust model in a peer-to-peer file sharing system. In this system, file providers' capabilities are evaluated by different aspects, including download speed, file quality, and file type. A naïve Bayesian network is constructed to represent conditional dependencies between the trustworthiness of file providers and the aspects. Each user holds a naïve Bayesian network for each file provider. If a user has no personal experience with a file provider, it may ask other users (advisors) for recommendations. A recommendation provided by an advisor will be considered by the user according to the trust value it has of the advisor. The trust value is updated by a reinforcement learning formula. More specifically, it will be increased/decreased after each comparison between the naïve Bayesian networks held by the user and the advisor for the file provider. The *Bayesian network*-based trust model takes into account preference similarity between users and advisors. However, this approach assumes that the aspects of file providers' capabilities are conditionally independent. This assumption is unrealistic in many systems. For instance, users may prefer high quality video and picture files, but do not care much about the quality of text files.

Buchegger and Boudec (Buchegger & Boudec 2003) propose a robust reputation system for mobile Ad-hoc networks (*RRSMAN*). *RRSMAN* is a fully distributed reputation system that can cope with false disseminated information. In *RRSMAN*, every node in the network maintains a reputation rating and a trust rating about every node else that it cares about. The trust rating for a node represents how likely the node will provide true advice. The reputation rating for a node represents how correctly the node participates with the node holding the rating. A modified Bayesian approach is developed to update both the reputation rating and the trust rating that node i holds for node j based on evidence collected in the past. Evidence is weighted according to its order of being collected. To detect and avoid false reports, *RRSMAN* updates the reputation rating held by node i for node j according to the advice provided by node k only if node k is trustworthy or the advice is compatible with the reputation rating held by node i. The advice is considered as compatible if its difference with the reputation rating held by node i is less than a deviation threshold, which is a positive constant. Three problems exist in the *RRSMAN* approach. Evidence collected by a node is weighted only according to its order of being observed. Therefore, the weights of two pieces of evidence collected on one month ago and on one year ago have no much difference as long as they have been collected one after another. Another problem is that this approach determines the preference similarity between two nodes based on only their current reputation ratings to one other node, which is certainly insufficient. The third problem concerns its way of integrating advice. The *RRSMAN* approach updates the reputation rating of a node by considering other nodes' advice. Pieces of advice provided by other nodes are considered equally as long as these nodes are trustworthy or each piece of advice is compatible.

Capabilities

To compare the above approaches, we analyze the capabilities they have based on their summaries. We list the following four capabilities that an effective approach should have.

- **Preference:** Agents may have different preferences. When a consumer agent estimates the reputation of a provider agent from advice provided by advisor agents, advisor agents with different preferences may have different opinions about the provider agent's reputation. Therefore, an effective approach should be able to take into account preference similarity between consumer and advisor agents when it copes with unfair ratings. For example, the *Cluster Filtering* approach (Dellarocas 2000) uses collaborative filtering techniques to identify nearest neighbors of a consumer agent;
- **High:** The approach should be able to handle unfairly high ratings;
- **Low:** The approach should be able to handle unfairly low ratings;
- **Varying:** The approach should be able to deal with changes of provider agents' behavior. Because of changes

Table 1: Capabilities of Approaches for Handling Unfair Ratings

Approaches	Preference	High	Low	Varying
Iterated Filtering		$\surd$	$\surd$	$\surd$
TRAVOS	$\surd$	$\surd$	$\surd$	
Cluster Filtering	$\surd$	$\surd$		$\surd$
GM-GC		$\surd$	$\surd$	
Bayesian Network	$\surd$	$\surd$	$\surd$	
RRSMAN	$\approx \surd$	$\surd$	$\surd$	$\approx \surd$

Table 2: Categorization of Approaches for Handling Unfair Ratings

Categories	Public	Private
Global	GM-GC	TRAVOS, RRSMAN Bayesian Network
Local	Iterated Filtering, Cluster Filtering	

of provider agents' behavior, agents may provide different ratings for the same provider agent. Even though two ratings provided within different periods of time are different, it does not necessarily mean that one of them must be unfair. Different ways are proposed to deal with this situation. BRS (Whitby, Jøsang, & Indulska 2005) uses a forgetting factor to dampen ratings according to the time when they are provided. The older ratings are dampened more heavily than the more recent ones.

Table 1 lists capabilities of the approaches summarized in the previous section. In this table, the mark "$\surd$" indicates that an approach has the capability. For example, the *Iterated Filtering* approach is capable of handling unfairly high and low ratings, and dealing with changes of agents' behavior. The mark "$\approx \surd$" indicates that an approach has the capability, but in a limited manner. For example, the *RRSMAN* approach determines the similarity between a consumer agent and an advisor agent based on only their current opinions on one provider agent, which is certainly insufficient. In addition, it deals with changes of agents' behavior by dampening advisor agents' ratings according to only their orders of being provided.

Categories

We have summarized different approaches proposed to handle unfair ratings, including *Cluster Filtering*, *Iterated Filtering*, *TRAVOS*, *GM-GC*, *Bayesian Network*, and *RRSMAN*. These approaches can be categorized in terms of two dimensions, a "public-private" dimension and a "global-local" dimension.

Public versus Private: When a consumer agent lacks personal experience with a provider agent, it can estimate the reputation of the provider agent based on collected ratings of the provider agent provided by advisor agents. Ratings will be considered differently according to trustworthiness of advisor agents. Ratings provided by more trustworthy advisor agents will be considered more heavily. An approach of handling unfair ratings is *private* if the consumer agent estimates the trustworthiness of an advisor agent based on only its personal experience with previous ratings provided by the advisor agent. The current rating provided by the advisor agent is likely to be fair if the advisor agent's past ratings are also fair. For example, the *TRAVOS* model (Teacy *et al.* 2005) estimates the accuracy of the advisor agent's current rating based on the amount of fair and unfair previous ratings provided by it that are similar to its current rating. An approach of handling unfair ratings is *public* if the consumer agent estimates trustworthiness of the advisor agent based on all the ratings it has supplied for any of the provider agents in the system. A rating is likely to be reliable if it is the same as/similar to most of the other ratings to same provider agents. For example, the *Iterated Filtering* approach (Whitby, Jøsang, & Indulska 2005) filters out unfair ratings that are not majority amongst others.

Global versus Local: An approach is *local* if it filters out unfair ratings based on only the ratings for the current provider agent. The *Cluster Filtering* approach (Dellarocas 2000) applies a divisive clustering algorithm to separate the ratings to a provider agent into two clusters, the lower rating cluster and the higher rating cluster. The ratings in the higher rating cluster are then considered as unfair ratings. An approach of handling unfair ratings is considered as *global* if it estimates the trustworthiness of an advisor agent based on ratings for all the provider agents that the advisor agent has rated. The *GM-GC* proposed in (Chen & Singh 2001) is a *global* approach.

The categorization of approaches for handling unfair ratings is summarized in Table 2. Note that there is no approach falling in the category of "private and local". It is simply because that there is a conflict in this category. A consumer agent asks advice about a provider agent from an advisor agent only when it lacks personal experience with the provider agent. An approach belonging to the "private and local" category will evaluate the trustworthiness of the advisor agent based only on the consumer agent's ratings and the advisor agent's ratings for the provider agent currently being evaluated as a possible partner (referred to as the current provider agent). The consumer agent's limited experience with the current provider agent is certainly not sufficient for determining the trustworthiness of the advisor agent.

Impact of Reputation System Architectures

Reputation system architectures have an impact on the selection of approaches for handling unfair ratings. There are basically two types of reputation systems in terms of their different architectures, centralized reputation systems and distributed reputation systems (Jøsang, Ismail, & Boyd 2005).

In centralized reputation systems, central servers collect ratings for each provider agent from consumer agents after transactions between them have succeeded. Central servers do not record all of the ratings of each individual consumer agent. Therefore, approaches used in these systems cannot consider consumer agents' personal experience with advisor agents' advice. The approaches used in centralized reputation systems, such as *Iterated Filtering*, *Cluster Filtering* and *GM-GC*, are based on all ratings of provider agents and belong to the "public" category. Results from those approaches do not differ for different consumer agents.

In distributed reputation systems, there is no central location for submitting ratings or obtaining advisor agents' ratings. A consumer agent should simply request advice about a provider agent from advisor agents. Even though some of distributed reputation systems have distributed stores for collecting ratings, it is still costly to obtain all ratings for the provider agent. Therefore, approaches used in these systems cannot consider all agents' ratings for the provider agent. The approaches used in distributed reputation systems, such as *TRAVOS*, *Bayesian Network* and *RRSMAN*, handle unfair ratings by estimating the trustworthiness of an advisor agent based on each individual consumer agent's personal experience with the advisor agent's advice. These approaches belong to the "private" category.

A Personalized Approach

As discussed in the previous section, approaches for handling unfair ratings are limited by reputation system architectures. Specifically, the approaches used in centralized reputation systems, such as *Iterated Filtering*, *Cluster Filtering* and *GM-GC*, cannot consider consumer agents' personal experience with advice provided by advisor agents. However, consumer agents' personal experience with advisor agents' advice is very important for estimating trustworthiness of advisor agents because agents tend to trust their own experience more than others' opinions. Furthermore, the *Iterated Filtering* approach is only effective when the significant majority of ratings are fair, the *Cluster Filtering* approach cannot handle unfairly low ratings, and the *GM-GC* approach is computationally intractable.

Inspired by the approaches used in distributed reputation systems, we propose a personalized approach for an enhanced centralized reputation system. This system creates a profile for each consumer agent to record its ratings for each provider agent it has rated. The personalized approach essentially combines advantages of both approaches used in centralized and distributed reputation systems. It allows a consumer agent to estimate the reputation (referred to as private reputation) of an advisor agent based on their ratings for commonly rated provider agents. [1] In this case, agents' preferences are also taken into account. If an advisor agent is trustworthy and has similar preferences with the consumer agent, the consumer and advisor agents will likely have many ratings in common. When the consumer agent has limited private knowledge of the advisor agent, the reputation (referred to as public reputation) of the advisor agent will also be considered. [2] The public reputation is estimated based on all ratings for the provider agents ever rated by the advisor agent. Finally, the trustworthiness of the advisor agent will be modeled by combining the weighted private and public reputations. The weights of them are determined based on the estimated reliability of the private reputation.

Private Reputation

Our approach allows a consumer agent C to evaluate the private reputation of an advisor agent A by comparing their ratings for commonly rated provider agents $\{P_1, P_2, ..., P_m\}$. For one of the commonly rated providers P_i ($1 \leq i \leq m$ and $m \geq 1$), A has the rating vector R_{A,P_i} and C has the rating vector R_{C,P_i}. A rating for P_i from C and A is binary ("1" or "0", for example), in which "1" means that P_i is reputable and "0" means that P_i is not reputable. [3] The ratings in R_{A,P_i} and R_{C,P_i} are ordered according to the time when they are provided. The ratings are then partitioned into different elemental time windows. The length of an elemental time window may be fixed (e.g. one day) or adapted by the frequency of the ratings to the provider P_i, similar to the way proposed in (Dellarocas 2000). It should also be considerably small so that there is no need to worry about the changes of providers' behavior within each elemental time window. We define a pair of ratings (r_{A,P_i}, r_{C,P_i}), such that r_{A,P_i} is one of the ratings of R_{A,P_i}, r_{C,P_i} is one of the ratings of R_{C,P_i}, and r_{A,P_i} corresponds to r_{C,P_i}. The two ratings, r_{A,P_i} and r_{C,P_i}, are correspondent only if they are in the same elemental time window, the rating r_{C,P_i} is the most recent rating in its time window, and the rating r_{A,P_i} is the closest and prior to the rating r_{C,P_i}. [4] We then count the number of such pairs for P_i, N_{P_i}. The total number of rating pairs for all commonly rated providers, N_{all} will be calculated by summing up the number of rating pairs for each

[1] We call this type of reputation private reputation because it is based on the consumer agent's own experience with the advisor agent's advice, and is not shared with the public. The private reputation value of the advisor agent may vary for different consumer agents.

[2] We call this type of reputation public reputation because it is based the public's opinions about the advisor agent's advice, and it is shared by all of the public. The public reputation value of the advisor agent is the same for every consumer agent.

[3] For the purpose of simplicity, we assume ratings for providers are binary. Possible ways of extending our approach to accept ratings in different ranges will be investigated as future work. Further discussion can be found in the future work section.

[4] We consider ratings provided by C after those by A in the same time window, in order to incorporate into C's rating anything learned from A during that time window, before taking an action. According to the solution proposed by Zacharia et al. (Zacharia, Moukas, & Maes 1999), by keeping only the most recent ratings, we can avoid the issue of advisors "flooding" the system.

commonly rated provider agent as follows:

$$N_{all} = \sum_{i=1}^{m} N_{P_i}$$

The private reputation of the advisor agent is estimated by examining rating pairs for all commonly rated providers. We define a rating pair (r_{A,P_i}, r_{C,P_i}) as a positive pair if r_{A,P_i} is the same value as r_{C,P_i}. Otherwise, the pair is a negative pair. Suppose there are N_f number of positive pairs. The number of negative pairs will be $N_{all} - N_f$. The private reputation of the advisor A is estimated as the probability that A will provide reliable ratings to C. Because there is only incomplete information about the advisor, the best way of estimating the probability is to use the expected value of the probability. The expected value of a continuous random variable is dependent on a probability density function, which is used to model the probability that a variable will have a certain value. The beta family of probability density functions is commonly used to represent probability distributions of binary events. Therefore, the private reputation of A can be calculated as follows:

$$\alpha = N_f + 1, \beta = N_{all} - N_f + 1$$

$$R_{pri}(A) = E(Pr(A)) = \frac{\alpha}{\alpha + \beta},$$

where $Pr(A)$ is the probability that A will provide fair ratings to C, and $E(Pr(A))$ is the expected value of the probability.

Public Reputation

When there are not enough rating pairs (discussed in the next section), the consumer agent C will also consider the advisor agent A's public reputation. The public reputation of A is estimated based on its ratings and other ratings for the providers rated by A. Each time A provides a rating $r_{A,P}$, the rating will be judged centrally as a fair or unfair rating. We define a rating for a provider agent as a fair rating if it is consistent with the majority of ratings to the provider up to the moment when the rating is provided. [5] As before, we consider only the ratings within a time window prior to the moment when the rating $r_{A,P}$ is provided, and we only consider the most recent rating from each advisor. In so doing, as providers change their behavior and become more or less reputable to each advisor, the majority of ratings will be able to change.

Suppose that the advisor agent A totally provides N'_{all} ratings. If there are N'_f number of fair ratings, the number of unfair ratings provided by A will be $N'_{all} - N'_f$. In the same way as estimating the private reputation, the public reputation of the advisor A is estimated as the probability that A will provide fair ratings. It can be calculated as follows:

$$\alpha' = N'_f + 1, \beta' = N'_{all} - N'_f + 1$$

$$R_{pub}(A) = \frac{\alpha'}{\alpha' + \beta'},$$

which also indicates that the more the percentage of fair ratings advisor A provides, the more reputable it will be.

Trustworthiness of Advisors

To estimate the trustworthiness of advisor agent A, we combine the private reputation and public reputation values together. The private reputation and public reputation values are assigned different weights. The weights are determined by the reliability of the estimated private reputation value.

We first determine the minimum number of rating pairs needed for C to be confident about the private reputation value it has of A. Based on the Chernoff Bound theorem (Mui, Mohtashemi, & Halberstadt 2002), the minimum number of rating pairs can be determined by an acceptable level of error and a confidence measurement as follows:

$$N_{min} = -\frac{1}{2\varepsilon^2} ln \frac{1-\gamma}{2},$$

where ε is the maximal level of error that can be accepted by C, and γ is the confidence measure. If the total number of pairs N_{all} is larger than or equal to N_{min}, consumer C will be confident about the private reputation value estimated based on its ratings and the advisor A's ratings for all commonly rated providers. Otherwise, there are not enough rating pairs, the consumer agent will not be confident about the private reputation value, and it will then also consider public reputation. The reliability of the private reputation value can be measured as follows:

$$w = \begin{cases} \frac{N_{all}}{N_{min}} & \text{if } N_{all} < N_{min}; \\ 1 & \text{otherwise.} \end{cases}$$

The trust value of A will be calculated by combining the weighted private reputation and public reputation values as follows:

$$Tr(A) = wR_{pri}(A) + (1-w)R_{pub}(A)$$

It is obvious that the consumer will consider less the public reputation value when the private reputation value is more reliable. Note that when $w = 1$, the consumer relies only on private reputation.

Examples

To illustrate how our approach models trustworthiness of advisors, this section provides examples that go through each step of the approach. Examples are also provided to demonstrate how trust values different consumer agents have of same advisors may vary, and to show the effectiveness of our approach even when the majority of ratings are unfair.

In a multiagent reputation system, a consumer agent C needs to make a decision on whether to interact with a provider agent P_0, which depends on how much C trusts P_0. To model the reputation of the provider P_0 when the consumer has had no or only limited experience with P_0, C seeks advice from three advisor agents A_x, A_y and A_z who have had experience with P_0. The advice about P_0 from A_x, A_y and A_z are ratings representing the reputation of P_0. Before aggregating the ratings provided by A_x, A_y and A_z, the

[5] Determining consistency with the majority of ratings can be achieved in a variety of ways, for instance averaging all the ratings and seeing if that is close to the advisor's rating.

Table 3: Ratings of Providers Provided by Advisors

A_j	A_x					A_y					A_z				
T	T_1	T_2	T_3	T_4	T_5	T_1	T_2	T_3	T_4	T_5	T_1	T_2	T_3	T_4	T_5
P_1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0
P_2	1	1	1	1	1	0	1	0	1	1	0	0	0	0	0
P_3	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0
P_4	1	1	1	1	1	1	0	0	1	0	0	0	0	0	0
P_5	1	1	1	1	1	1	1	0	0	1	0	0	0	0	0

consumer agent C needs to evaluate the reliability of those ratings, which depends on the trustworthiness of the advisors A_x, A_y and A_z. Our approach effectively models the trustworthiness of advisors based on how reliable the previous ratings provided by them are.

Consider the case where the advisors A_x, A_y and A_z each has rated only the five provider agents (P_1, P_2, P_3, P_4, and P_5). In this case, we may assume P_0 and P_5 refer to the same provider agent. Table 3 lists the ratings provided by A_x, A_y and A_z for the five providers. The symbol "T" represents a sequence of time windows, in which T_1 is the most recent time window. To simplify the demonstration, we assume that each advisor agent provides at most one rating within each time window. We also assume that those are the only ratings provided by them.

As can be seen from Table 4, the consumer agent C has also provided some ratings for the five providers. The consumer agent C might have not provided any rating for some providers within some time window. For example, it has provided only one rating for the provider P_5, which is in the time window T_1. We assume that the ratings provided by C are after those provided by A_x, A_y and A_z if they are within the same time window.

Table 4: Ratings Provided by the Consumer Agent C

T	T_1	T_2	T_3	T_4	T_5
P_1	1	1	1	1	1
P_2	1	1	1	1	-
P_3	1	1	1	-	-
P_4	1	1	-	-	-
P_5	1	-	-	-	-

We compare the ratings provided by A_x, A_y and A_z in Table 3 and ratings provided by C in Table 4. The consumer agent C has the same number of rating pairs with each advisor agent ($N_{all}(A_j) = 15$ and $j \in \{x, y, z\}$). However, C has different numbers of positive rating pairs with A_x, A_y and A_z, which are listed in Table 5. Accordingly, as can be seen from Table 5, the private reputation values of A_x, A_y and A_z are different, in which the private reputation value of A_x is the highest and that of A_x is the lowest. It indicates that the advisor agent A_x is most likely to provide fair ratings and have similar preferences with the consumer agent C, whereas A_z most likely will lie and have different preferences with C.

According to Table 3, the total number of ratings provided by each advisor agent is the same ($N'_{all}(A_j) = 25$). We also count the number of fair ratings each advisor agent provides. A rating here is considered as a fair rating when it is consistent with the majority of ratings for the provider agent within a same time window. Consider the case where all of the five provider agents are reputable and the majority of ratings are fair. In this case, a rating of 1 provided by an advisor agent will be considered as a fair rating, whereas a rating of 0 will be considered as an unfair rating. From the advisor agents' ratings listed in Table 3, we can see that ratings provided by the advisor agent A_x are all fair, the advisor agent A_z always lies, and some of the ratings provided by the advisor agent A_y are unfair. Table 5 lists the number of fair ratings provided by each advisor agent and the corresponding public reputation value of it. From Table 5, it is clear that the advisor agent A_x is most likely to provide fair ratings, and the advisor A_z most likely will lie.

Table 5: Private and Public Reputations of Advisors

A_j	A_x	A_y	A_z
$N_f(A_i)$	15	8	0
α	16	9	1
β	1	8	16
$R_{pri}(A_j)$	0.94	0.53	0.06
$N'_f(A_j)$	25	12	0
α'	26	13	1
β'	1	14	26
$R_{pub}(A_j)$	0.96	0.48	0.04

Table 6: Trustworthiness of Advisor Agents

ε	0.1	0.15	0.2
N_{min}	115	51	29
w	0.13	0.29	0.52
$Tr(A_x)$	0.957	0.954	0.950
$Tr(A_y)$	0.487	0.495	0.506
$Tr(A_z)$	0.043	0.046	0.05

To combine private reputation and public reputation, the weight w should be determined. The value of w depends on the values of ε and γ, and the number of rating pairs $N_{all}(A_j)$, which is the same for every advisor agent in our example. Suppose we have a fixed value, 0.8 for γ, which means that the confidence value should be no less than 0.8 in order for the consumer agent to be confident with the private reputation values of advisor agents. In this case, the more errors it can accept, the more confident it is with the private

reputation values of advisor agents, which also means that the more weight it will put on the private reputation values. Table 6 lists different acceptable levels of errors, their correspondent weights of private reputation values, and different results of trust values. It clearly indicates that A_x is the most trustworthy, and A_y is more trustworthy than A_z. As a result, the consumer agent C will place more trust in the advice provided by A_x. It will consider the advice provided by A_x more heavily when aggregating the advice provided by A_x, A_y and A_z for modeling the reputation of the provider agent P_0. Discussion of possible aggregation functions is out of the scope of this paper. Our framework serves the purpose of representing the trustworthiness of advisors, so that this may be taken into account, when determining how heavily to rely on their advice.

Table 7: Ratings Provided by the Consumer Agent C'

T	T_1	T_2	T_3	T_4	T_5
P_1	1	1	-	-	1
P_2	1	-	-	1	-
P_3	1	1	-	-	-
P_4	1	1	-	-	-
P_5	1	-	-	-	-

Table 8: Trust Values C' Has of Advisors

A_j	A_x	A_y	A_z
$R_{pri}(A_j)$	0.92	0.58	0.08
$R_{pub}(A_j)$	0.96	0.48	0.04
$Tr(A_j)$	0.947	0.514	0.054

To demonstrate how the trust values different consumer agents have of the same advisors may vary, we consider another consumer agent C', that also needs to make a decision on whether to interact with a provider agent P_0' (P_0' may differ from P_0). We may assume P_0' and P_4 refer to the same provider agent. The ratings provided by C' for the five provider agents are listed in Table 7. By going through the same process as above, we can calculate the trust values the consumer agent C' has of A_x, A_y and A_z, when $\varepsilon = 0.2$ and $\gamma = 0.8$. The results are presented in Table 8. Comparing Table 8 with Tables 5 and 6, we can see that the private reputations the consumer agent C' has of advisors are different from those the consumer agent C has. Although the public reputations of advisors that the consumers have are the same, the trust values that the consumers have are still different.

Table 9: Public Reputations of Advisors When Majority of Ratings are Unfair

A_j	A_x	A_y	A_z
$N_f'(A_j)$	0	13	25
α'	1	14	26
β'	26	13	1
$R_{pub}(A_j)$	0.04	0.52	0.96

To show the robustness of our model, now consider a case where the majority of ratings provided by advisor agents are unfair. Adjusting our earlier example, a rating of 1 provided by an advisor agent for any provider agent will now be considered as an unfair rating, whereas a rating of 0 will be considered as a fair rating. As a result, the public reputations that the consumer C has of the advisor agents A_x, A_y and A_z will be different, which can be seen from Table 9. We model the trust values the consumer agent C has of the advisors A_x, A_y and A_z, when C's acceptable levels of errors of private reputation values are different. Results are presented in Table 10. From this table, we can see that our approach can still correctly represent the trustworthiness of advisor agents by making adjustments to rely more heavily on the private reputations.

Table 10: Trustworthiness of Advisors When Majority of Ratings are Unfair

ε	0.1	0.2	0.25
N_{min}	115	29	19
w	0.13	0.52	0.79
$Tr(A_x)$	0.157	0.508	0.751
$Tr(A_y)$	0.521	0.525	0.528
$Tr(A_z)$	0.843	0.492	0.249

Experimental Results

Our approach models the trustworthiness of advisors according to the reliability of the ratings provided by them. To demonstrate the effectiveness of the approach, we carry out experiments involving advisors that provide different percentages of unfair ratings. The expectation is that trustworthy advisors will be less likely to provide unfair ratings, and vice versa. We also examine how large numbers of dishonest advisors will affect the estimation of advisors' trustworthiness. Results indicate that our approach is still effective by making adjustments to rely more heavily on private reputations of advisors, in this case.

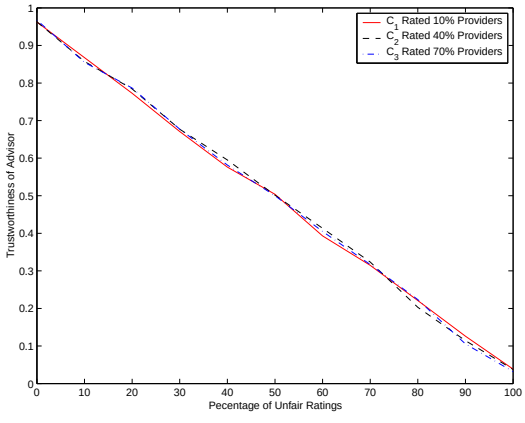

Figure 1: Trustworthiness of Advisor

The first experiment involves 100 providers, 3 consumers, and one advisor. The 3 consumers, C_1, C_2 and C_3, rate 10,

40 and 70 randomly selected providers, respectively. The advisor totally rates 40 randomly selected providers. [6] We examine how the trust values the consumers have of the advisor change when different percentages (from 0% to 100%) of its ratings are unfair. As illustrated in Figure 1, the trust values the consumers have of the advisor decrease when more percentages of the advisor's ratings are unfair. From this figure, we can also see that our approach is still effective when the consumer C_1 does not have much experience with providers, in the sense that C_1 can still reduce the reputation of the advisor when it provides more unfair ratings.

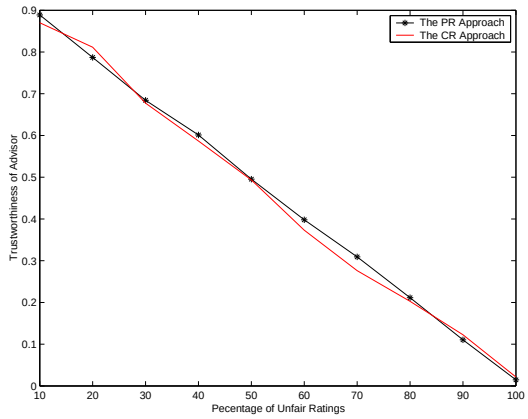

Figure 2: Trustworthiness of A When Majority of Advisors are Honest

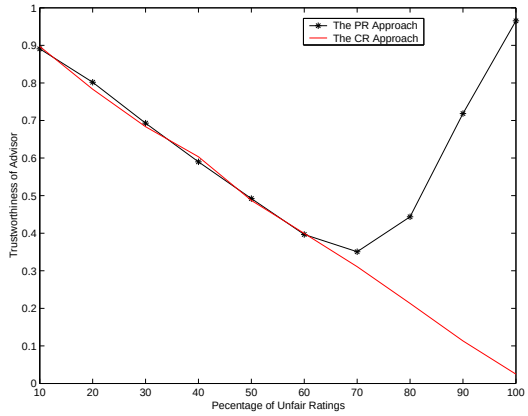

Figure 3: Comparison of the CR and PR Approaches

The second experiment involves 100 providers, 80 advisors, and one consumer. The consumer and each advisor rate 80 of the randomly selected providers. We model the trust value the consumer has of one of the advisors, A. The trustworthiness of the advisor will be modeled as the combination of its private and public reputations (referred to as the CR approach) and as only its public reputation (referred to as the PR approach), respectively. The advisor A will provide different percentages (from 10% to 100%) of unfair ratings. Figure 2 illustrates the trustworthiness of A when 24 (30% of all) advisors are dishonest. Those dishonest advisors provide the same percentage of unfair ratings as the advisor A does. Results indicate that the trustworthiness of A modeled by using the CR and PR approaches decreases when more percentages of ratings provided by A are unfair. Therefore, these two approaches are not affected when only a small number of advisors are dishonest. Figure 3 represents the trustworthiness of A when 48 (60% of all) advisors are dishonest. In this figure, the trustworthiness of A modeled by using the CR approach still decreases when more percentages of ratings provided by A are unfair, which indicates that our approach is still effective when the majority of advisors provide large numbers of unfair ratings. In contrast, the trustworthiness modeled by using the PR approach increases when more than 60% of ratings provided by the dishonest advisors are unfair, which indicates that the PR approach is only effective when the majority of ratings are fair.

[6]Note that we simplify the experiments by limiting each consumer or advisor to provide at most one rating for each provider.

Conclusions and Future Work

In this paper, we first survey different approaches for handling unfair ratings, and their advantages and disadvantages. We list the capabilities that an approach should have. Approaches for handling unfair ratings should be able to take into account the preference similarity between consumer agents and advisor agents. They should be able to handle both unfairly high and low ratings. They should also be able to deal with changes of agents' behavior over time. We compare those existing approaches based on the four capabilities. We then categorize these approaches in terms of two dimensions, a "public-private" dimension and a "global-local" dimension. We also discuss the impact of reputation system architectures on the selection of approaches for handling unfair ratings. Approaches used in centralized reputation systems belong to the "public" category and cannot consider consumer agents' personal experience with advisor agents' advice (ratings), whereas approaches used in distributed reputation systems belong to the "private" category and cannot consider all ratings for provider agents. This categorization of the different approaches provides a valuable perspective on the key challenges faced in designing an effective reputation system that makes use of advice from other agents, but takes care to consider the trustworthiness of those ratings.

Based on the study of these approaches, we propose a personalized approach for effectively handling unfair ratings in enhanced centralized reputation systems. The personalized approach has all four of the capabilities. It also has the advantages of both approaches used in centralized reputation systems and approaches used in distributed reputation systems. It allows a consumer agent to estimate the private reputation of an advisor agent based on their ratings for commonly rated provider agents. When the consumer agent is not confident with the private reputation value, it can also use the public reputation of the advisor agent. The public reputation of the advisor agent is evaluated based on all ratings for the provider agents rated by the advisor agent. Experimental results demonstrate the effectiveness of the per-

sonalized approach in terms of adjusting agents' trustworthiness based on the percentages of unfair ratings they provided. Trustworthiness of advisor agents will be decreased more/less if advisor agents provide more/fewer unfair ratings. Our approach can effectively model the trustworthiness of advisors even when consumer agents do not have much experience with provider agents. Furthermore, our approach is still effective when the majority of advisor agents provide large numbers of unfair ratings, by adjusting to rely more heavily on private reputations of advisor agents.

In future work, the personalized approach will be implemented and embedded in a simulated trust and reputation model. Experiments will be carried out to compare the performance of the personalized approach with the performance of other existing approaches, such as the *Iterated Filter* approach and the *TRAVOS* model. The performance could be evaluated, for instance, based on average estimation error, which is the average difference between provider agents' actual reputation values and estimated reputation values. We could also conduct experiments to specifically explore the benefit of our use of a time window for the private reputation model, in comparison with models like *TRAVOS* that also try to determine the reputability of advisor agents. [7]

Another avenue for future work is to make adjustments to the current model, to broaden its applicability. For example, we could move beyond binary ratings for provider agents to accept ratings in different ranges. In this case, we could begin with a modest set of possible values, each with a qualitative interpretation (e.g. very reputable, neutral, not reputable, etc.) as in (Chen & Singh 2001). Another possible extension is to allow advisors and consumer agents to represent the reputation of a provider agent not as a single rating but as a rating of different dimensions of trustworthiness. We could, for example examine different aspects (e.g. delivery time, quality and prices) of providers' services similar as used by Wang and Vassileva (Wang & Vassileva 2003), but take into account relationships among those aspects by using for example, a quality of service ontology used by Maximilien and Singh (Maximilien & Singh 2005).

Another potential future work is to distinguish ratings for the current provider agent from ratings for other provider agents. As stated earlier in the related work section, there is no approach belonging to the "private and local" category because consumer agents' limited experience with the current provider agent is insufficient to estimate trustworthiness of advisor agents. However, we believe that ratings for the current provider agent should influence consumer agents' decisions more heavily, and therefore should gain more weight when estimating trustworthiness of advisor agents.

[7] Relying on private reputation alone is a feature as well of collaborative filtering recommender systems, but these systems tend to focus on how best to select like-minded agents in order to acquire advice and are less concerned with judging the reputability of the advice being provided.

References

Buchegger, S., and Boudec, J.-Y. L. 2003. A robust reputation system for mobile ad-hoc networks. *Technical Report IC/2003/50, EPFL-IC-LCA*.

Chen, M., and Singh, J. P. 2001. Computing and using reputations for internet ratings. In *Proceedings of the 3nd ACM Conference on Electronic Commerce*.

Dellarocas, C. 2000. Immunizing online reputation reporting systems against unfair ratings and discriminatory behavior. In *Proceedings of the 2nd ACM Conference on Electronic Commerce*.

Jøsang, A., and Ismail, R. 2002. The beta reputation system. In *Proceedings of the 15th Bled Electronic Commerce Conference*.

Jøsang, A.; Ismail, R.; and Boyd, C. 2005. A survey of trust and reputation systems for online service provision. *(to appear). Decision Support Systems*.

Maximilien, E. M., and Singh, M. P. 2005. Agent-based trust model using multiple qualities. In *Proceedings of 4th International Autonomous Agents and Multi Agent Systems (AAMAS 2005)*.

Mui, L.; Mohtashemi, M.; and Halberstadt, A. 2002. A computational model of trust and reputation. In *Proceedings of the 35th Hawaii International Conference on System Science (HICSS)*.

Teacy, W. T. L.; Patel, J.; Jennings, N. R.; and Luck, M. 2005. Coping with inaccurate reputation sources: Experimental analysis of a probabilistic trust model. In *Proceedings of 4th International Autonomous Agents and Multi Agent Systems (AAMAS 2005)*.

Wang, Y., and Vassileva, J. 2003. Bayesian network-based trust model. In *Proceedings of the 6th International Workshop on Trust, Privacy, Deception and Fraud in Agent Systems*.

Whitby, A.; Jøsang, A.; and Indulska, J. 2005. Filtering out unfair ratings in bayesian reputation systems. *The Icfain Journal of Management Research* 48–64.

Zacharia, G.; Moukas, A.; and Maes, P. 1999. Collaborative reputation mechanisms in electronic marketplaces. In *Proceedings of the 32nd Hawaii International Conference on System Sciences (HICSS-32)*.