A Two-Stage Clustering Approach to Filtering Unfair Testimonies for Reputation Systems

Siyuan Liu[1], Chunyan Miao[1], Yin-Leng Theng[2], and Alex C. Kot[3]

[1]Joint NTU-UBC Research Centre of Excellence in Active Living for the Elderly (LILY),
[2]Wee Kim Wee School of Communication and Information,
[3]School of Electrical and Electronic Engineering,
Nanyang Technological University, Singapore 639798

{syliu, ascymiao, tyltheng, eackot}@ntu.edu.sg

Abstract

Reputation systems have contributed much to the success of online service provision systems. By using reputation systems, a service consumer can benefit from modeling the behavior of a service provider. However, the problem of unfair testimonies makes it difficult and remains an open issue for the consumer to accurately evaluate the provider's reputation. To address this problem, various approaches have been proposed to mitigate the adverse effect of unfair testimonies. However, most existing works focus on countering unfair testimonies for reputation systems supporting binary ratings. To resolve this limitation, in this paper, we propose a two-stage clustering approach to filter unfair testimonies, especially for reputation systems using multi-nominal ratings. The proposed approach uses clustering technique to identify unfair testimonies and further contributes to providing the consumer with a more accurate reputation evaluation result regarding the target provider. The approach is evaluated within a simulated online service provision environment. Experimental findings have shown that the proposed approach is efficient in filtering out various types of unfair testimonies for reputation systems and outperforms the comparative approaches.

Keyword: Reputation system, Trust, Testimonies, Multi-nominal ratings, Clustering.

I. Introduction

With the development of Internet technology, online service provision systems are starting to permeate throughout our daily life in recent years. Various services can be provided to service consumers by different service providers in such a system. A service can be an e-commerce transaction (e.g., a seller selling products to a buyer in eBay) or a functional component implemented by Web service technologies. Due to the online nature of these systems, consumers and providers usually do not have the chance to meet face-to-face during a transaction process or to inspect the quality of a service before the transaction is complete. Therefore, it is crucial for online service provision systems to help consumers to make proper decisions on transaction partner selection by accurately evaluating the trustworthiness of potential service providers.

In order to achieve this goal, researchers in the agent community have been developing reputation systems [1] [2] [3] [4] for multiagent-based online service provision systems [5]. These reputation systems, which are referred to as *soft security* mechanisms, serve as a complement to traditional information security [6]. By using a reputation system, a consumer can rate a provider after the completion of a transaction. A reputation value can be derived through aggregating ratings with regard to the provider and made available to the general consumers. The derived reputation value can be further used to assist the consumers to evaluate the trustworthiness of the provider and decide whether to transact with him.

Although reputation systems have contributed much to the success of online service provision systems, their accuracy remains to be a big concern. One issue plaguing reputation systems is the problem of *unfair testimonies*. Suppose there is a consumer *U* and a provider *P*. *U* is now evaluating the reputation of *P* to decide whether to transact with *P*. To assist his evaluation, *U* requests ratings (called testimonies) from other consumers (called witnesses) who had transactions with *P* before. However, to mislead *U* into transacting with *P*, *P* might collude with some witnesses who only report positive ratings to *U* regarding *P* no matter what *P*'s real behavior is. It is possible that those unfairly positive ratings will lead to *U*'s inaccurate evaluation of *P*'s reputation. As a consequence, *U* might make a wrong decision to transact with *P*.

Various approaches [3] [7] [8] [9] [10] [11] have been proposed to cope with the problem of unfair testimonies in reputation systems. However, most of these approaches focus only on the reputation systems using binary ratings. But a reputation system using binary ratings excludes the possibility of providing ratings with finer granularity [12]. Therefore, to preserve consumers' opinions better, reputation systems using multi-nominal ratings are proposed to allow consumers to rate providers in a multi-nominal way [12]. But the approaches addressing the problem of unfair testimonies and specifically designed for the reputation systems accepting only binary ratings cannot be easily extended to reputation systems supporting multi-nominal rating levels. To resolve this limitation, in this paper we propose a two-stage clustering approach to filter unfair testimonies for reputation systems, especially for the ones using multi-nominal ratings. The proposed approach adopts the hierarchical clustering method and clusters similar testimonies together within two stages. Then the testimonies that are not in the same cluster as the consumer's personal ratings (if any) or as majority ratings are considered as unfair and discarded.

Experiments conducted in a simulated service provision environment where dishonest witnesses may provide different types of unfair testimonies were used to evaluate the proposed approach. The results show that the proposed approach is effective in filtering out unfair testimonies in many different scenarios, including provider behavior change and witness behavior change. The proposed approach also outperforms related work in the scenario where only binary testimonies are allowed. Thus, our approach is proven to improve the accuracy of reputation systems and contribute to the goal of developing reliable online service provision environment for users.

II. Related Work

Different approaches have been proposed to tackle the problem of unfair testimonies in reputation systems. According to their technical mechanisms, they can be classified into the following four categories.

A. Information Theory Based Approaches

Jøsang and Ismail proposed the Beta Reputation System (BRS) [1]. In BRS, ratings for a provider are expressed as either positive or negative, which can be considered as two events in the beta probability distribution [13]. The provider's reputation is calculated as the expected value of the positive rating happening in the future by substituting the numbers of the positive and negative ratings into the beta probability density function. Weng et al. proposed an entropy-based approach [9] to filter unfair testimonies for BRS. This approach calculates the quality values of the consumer's personal ratings and a particular witness's ratings by using an entropy based metric. If the difference between the two quality values exceeds a given threshold, the witness's ratings are considered as unfair and discarded. However, due to the use of entropy, the approach cannot distinguish the difference between the symmetry pairs of positive and negative testimonies (i.e., the number of a witness's positive ratings is the same as the number of the consumer's negative ratings, and the number of the witness's negative ratings is the same as the number of the consumer's positive ratings), which might lead to unfair testimonies not being accurately identified.

Yu and Singh proposed a provider reputation model [14] based on Dempster-Shafer theory [15]. This model divides a witness's ratings into three parts, then maps the three parts to a Dempster-Shafer basic belief assignment function [15]. A provider reputation value is derived based on the combination of all witnesses' belief assignment functions. The problem of unfair testimonies is addressed by comparing the provider reputation value calculated through the witness's testimonies only with that calculated through the consumer's personal ratings only

[7]. A metric based on the difference between the two reputation values is used to discount the witness's testimonies. This approach has the problem that it only compares the consumer and the witness's ratings to discount the witness's testimonies, leading to that the approach cannot work when the consumer's personal ratings regarding the target provider is missing. At the same time, it also has the problem that the witness cannot regain his trustworthiness back from the consumer once his testimonies are discounted.

Statistical Approaches

Whitby et al. proposed an iterated filtering approach [8] to filter out unfair testimonies for BRS. This approach filters out the unfair testimonies by calculating whether a particular witness's ratings are outside q or 1-q quantile of majority witnesses' ratings. This approach is simple to implement, but it has the disadvantage that its filtering accuracy decreases quickly with the increase of the portion of unfair testimonies.

A witness behavior model was proposed in [11]. In this model, a witness's testimonies are first filtered out if the average difference between his testimonies and the consumer's personal ratings for all commonly rated providers exceeds a given threshold. Then the approach calculates the similarity and tendency values of the remaining witnesses' testimonies by comparing the testimonies with the consumer's personal ratings. Finally, the approach models the witnesses' behaviors as optimistic or pessimistic by considering the similarity and the tendency values together. The approach has the advantage that it could differentiate the witnesses' behavior patterns. But it still considers binary ratings and assumes that the witnesses' behaviors are consistent over all providers.

An approach using clustering to cope with unfairly high ratings was proposed in [16]. This approach uses a divisive clustering algorithm to separate testimonies for a provider into two clusters: the one containing lower testimonies, and the one containing higher testimonies. The testimonies in the cluster including higher testimonies are considered as unfair. However, this approach cannot effectively handle unfairly low testimonies.

Probabilistic Approaches

Teacy et al. proposed the TRAVOS model [17] to evaluate the reputation of agents in agent-based virtual organizations. In TRAVOS, the problem of unfair testimonies is addressed by accomplishing two tasks. The first task is to estimate the probability that a witness's testimonies are fair by comparing the witness's ratings with the consumer's personal ratings regarding the commonly rated providers. The second task is to adjust the witnesses' testimonies according to the probability values obtained from the first task. However, TRAVOS is quite time-consuming as it will scan the witnesses' entire rating history for all providers every time when it estimates the probability that the witnesses' testimonies are fair.

Learning-Based Approaches

Regan et al. proposed the BLADE model [3] by using Bayesian learning to reinterpret a witness's ratings instead of filtering out the possible unfair testimonies. But the reinterpretation depends on the consistency of the witnesses' behaviors towards all providers. Otherwise, the reinterpretation of the witness' ratings may be incorrect.

In [18], the authors proposed a reinforcement learning based reputation model which adjusts the relative importance given to the testimonies from each witness based on the actual gain or loss derived from the actual transactions following their advice. The model provides a method for evaluating the impact of each witness's testimonies on the consumer's wellbeing and rewarding/penalizing them accordingly. But this approach is also limited to binary ratings.

Our work belongs to the category of statistical approaches. It differs from the existing approaches in the following ways. Firstly, as reviewed above, most of the existing approaches [3] [9] [11] [17] [18] to handling the problem of unfair testimonies are designed for reputation systems using binary ratings. In contrast, the proposed approach is applicable to reputation systems using multi-nominal ratings. Secondly, the proposed approach considers the consumer's personal ratings with a different importance from the witnesses' ratings [10]. Therefore, it has the advantage that it is able to work when majority witnesses are providing unfair testimonies. Lastly, as it focuses on the testimonies regarding the target provider, it is not influenced by the variations in witness behaviors regarding different providers (i.e., some witnesses may provide fair testimonies for some providers to build up their credibility, and then provide unfair testimonies for other providers).

III. Notations

Before getting into the details of the proposed approach, we first introduce the notations we used.

A. *Aggregating Consumers' Ratings*

Suppose that in a reputation system, there are M providers $\{P_1, P_2, \dots, P_M\}$, and N consumers $\{U_1, U_2, \dots, U_N\}$. After each transaction between a consumer U_n $(1 \le n \le N)$ and a provider P_m $(1 \le m \le M)$ is complete, U_n can rate P_m's behavior as a rating level from a set of predefined discrete rating levels. Suppose that there are L $(L \ge 2)$ different rating levels which are indexed by $1, 2, \dots, L$. If U_n rates P_m's behavior as rating level l, U_n's rating for P_m is represented as a row vector:

$$r_{U_n,P_m} = [r_{U_n,P_m}(1),\ r_{U_n,P_m}(2),\ \dots,\ r_{U_n,P_m}(L)],$$

where $r_{U_n,P_m}(l) = 1$ and $r_{U_n,P_m}(i) = 0$ $(1 \le i \le L,\ i \ne l)$. For example, suppose that $L = 5$ and U_n rates P_m's behavior as 4 after one transaction, then $r_{U_n,P_m} = [0, 0, 0, 1, 0]$. In a single time period t, the aggregated ratings from U_n to P_m can be represented as a row vector, expressed as:

$$R^t_{U_n,P_m} = [R^t_{U_n,P_m}(1),\ R^t_{U_n,P_m}(2),\ \dots,\ R^t_{U_n,P_m}(L)],$$

where $R^t_{U_n,P_m}(i)$ is the aggregated result of $r_{U_n,P_m}(i)$ $(i = 1, 2, \dots, L)$ in time period t. The updating of $R^t_{U_n,P_m}$ can be achieved by adding (i.e., matrix addition) the new rating vector r_{U_n,P_m} to the previous aggregated rating vector $R^t_{U_n,P_m}$.

B. *Aggregating Ratings with Temporal Decay*

Providers might change their behaviors over time. Therefore, U_n generally cares more about P_m's recent behavior and forgets P_m's old behavior by giving a relatively greater weight to more recent ratings, which can be achieved by introducing a forgetting factor λ ($\lambda \in [0,1]$) as proposed in [8] [12]. λ controls the rate at which the provider's old behavior is forgotten. When $\lambda = 0$, P_m's past behavior is completely forgotten after a single time period. When $\lambda = 1$, P_m's old behavior is never forgotten. Let P_m's accumulated ratings with temporal decay from U_n after time period *t* be denoted as:

$$A_{U_n,P_m}^t = [A_{U_n,P_m}^t(1),\ A_{U_n,P_m}^t(2),\ \ldots,\ A_{U_n,P_m}^t(L)],$$

then the accumulated rating vector with temporal decay after time period *t* can be calculated as:

$$A_{U_n,P_m}^t = \begin{cases} \vec{0} & (t = 0), \\ \lambda \times A_{U_n,P_m}^{t-1} + R_{U_n,P_m}^t & (t \in \aleph). \end{cases}$$

When U_n is evaluating P_m's reputation, he can collect rating vectors from other consumers to facilitate his evaluation. Then the set of the consumers who provide rating vectors to U_n regarding P_m are expressed as:

$$W_{U_n,P_m} = \{U_j | j \neq n \ and\ ||A_{U_j,P_m}^t|| \neq 0\}.$$

From U_n 's point of view, W_{U_n,P_m} is called the set of witnesses regarding P_m (each consumer in W_{U_n,P_m} is a witness), and the rating vector provided by each witness is called testimony from this witness. The set of the accumulated rating vectors regarding P_m after time period *t* is denoted as:

$$G_{U_n,P_m}^t = \{A_{U_j,P_{m_j}}^t | U_j \in W_{U_n,P_m} \cup \{U_n\}\}.$$

Then U_n can use G_{U_n,P_m}^t to estimate P_m's reputation by applying some existing reputation evaluation models, such as BRS [1] or DRS [12] [19]. But as mentioned previously, though U_n can use the testimonies to facilitate the evaluation regarding P_m's reputation, the testimonies may mislead U_n 's evaluation if the witnesses do not provide the testimonies in an honest way. This may even result in an opposite evaluation situation, e.g., where a very low reputation is estimated regarding a reputable provider. Therefore, we need to filter out the testimonies provided by dishonest witnesses before aggregating them together with the testimonies provided by honest witnesses.

IV. Using Clustering to Filter Unfair Testimonies

Before we present the proposed approach, we need to clarify what "*unfair testimonies*" mean in our context. According to the definition of trust -- the opinion (more technically, an evaluation) of an entity towards a person, a group of people, or an organization on a certain criterion [5], the trustworthiness of the target provider *P* (the provider whose reputation is under evaluation) is the opinion held by a consumer *U* towards *P*. Therefore, we consider that the likelihood that a witness is dishonest in reporting testimonies should also be the opinion held by *U* towards the witness. Intuitively, the higher the similarity between *U*'s past personal ratings regarding *P* and the witness's testimonies, the more likely the testimonies are fair. Hence, if we can group with *U* the witnesses that provide similar testimonies as *U*'s personal ratings, we can identify the honest witnesses and filter out the unfair testimonies provided by the dishonest witnesses. We need to emphasize that "*unfair testimonies*" do not always mean that the witnesses report ratings intentionally unfairly. The "*unfairness*" can be due to the subjective difference. In this work, we consider the different testimonies caused by subjective difference as unfair.

To group similar testimonies together, the technique of clustering is a good choice. Clustering is originally used to assign a set of observations into subsets (called clusters) so that the observations in the same cluster are similar to each other according to some criteria [20]. There are many clustering methods designed, such as *k*-means clustering method and hierarchical clustering method. For the proposed approach, we use hierarchical clustering method as it supports different clustering stopping criteria. For hierarchical clustering method, each observation is initially regarded as one cluster, then two clusters will be merged together according to some distance criteria (e.g., merge two clusters together which have shortest distance between each other). The merging process continues until the predefined clustering stopping criterion (e.g. predefined number of clusters) is met.

Here, we propose a two-stage clustering approach to filter out unfair testimonies. The basic idea of the proposed approach is as follows. Imagine the accumulated ratings from a particular witness or the consumer himself as a feature vector in an L-dimensional space (L is the number of rating levels used in the reputation system). The proposed approach tries to group the feature vectors with the same similarity into one cluster. After clustering, the testimonies will be discarded if they are not in the cluster including the consumer's personal rating vector (if any), or if they are not in the cluster including the majority witnesses' rating vectors.

More specially, the proposed approach includes two clustering stages. In the first stage, the proposed approach groups the rating vectors into a predefined number (K) of clusters. Initially, the rating vector from each witness is considered as a cluster, and two clusters with the shortest 2-norm distance are merged together to get a new cluster. Then the two clusters with the shortest 2-norm distance are selected from the new cluster and other remaining clusters. The two selected clusters are then merged together to get a new cluster again. The process continues until the predefined number of clusters K is met (we will show how K value will impact the accuracy of the proposed approach in next section. The primary aim of this stage is to merge the rating vectors with the smallest difference together.

However, if K is preset as a very large value, some similar vectors might not be clustered together in the first stage. Therefore, we need the second clustering stage, in which stage the merging process continues and stops when a different criterion is met. The proposed approach first calculates the furthest distance between any two clusters achieved from the first stage, then merge together the two clusters with the minimum furthest distance if the furthest distance of the two clusters is smaller than the predefined distance threshold D. The merging process continues until no furthest distance between any two clusters is smaller than D. The aim of this stage is to continue merging rating vectors with similarity together but ensure that the rating vectors with obvious difference are not merged by controlling D value (we will show how D value will impact the accuracy of the proposed approach in next section).The detailed algorithm is illustrated in Algorithm 1.

Algorithm 1: The Proposed Two-Stage Clustering Approach

Procedure: Two-Stage Clustering Filtering
Input : P, the provider whose reputation is being evaluated;
U, the consumer who is evaluating P's reputation;
K, threshold controlling stage 1 clustering stopping;
D: threshold controlling stage 2 clustering stopping
Output : a set of honest witnesses regarding P;

1 Find $W_{U,P}$ and collect rating vectors from $W_{U,P}$;
2 **if** $||A_{U,P}|| \neq 0$ **then**
3     $W'_{U,P} = W_{U,P} \cup \{U\}$;
4 **else**
5     $W'_{U,P} = W_{U,P}$;
6 Normalize each rating vector $A_{U_j,P}$ $(U_j \in W'_{U,P})$ as $A'_{U_j,P}$;
7 **foreach** $U_j \in W'_{U,P}$ **do**
8     $C_j = \{A'_{U_j,P}\}$;
9 $C = \bigcup\{C_j\}$;
10 **while** $|C| > K$ **do**
11   **foreach** $C_i \in C$, $C_j \in C$ $(i \neq j)$ **do**
12   **foreach** $A'_{U_{i'},P} \in C_i$, $A'_{U_{j'},P} \in C_j$ **do**
13             $d_s(i',j') = ||A'_{U_{i'},P} - A'_{U_{j'},P}||$;
14         $D_s(C_i,C_j) = \min(d_s(i',j'))$;
15     $p,q = \arg\{\min_{i,j}(D_s(C_i,C_j)\}$;
16   merge C_p and C_q together;
17 *stopcondition* = false;
18 **while** *stopcondition is false* **do**
19   **foreach** $C_i \in C$, $C_j \in C$ $(i \neq j)$ **do**
20   **foreach** $A'_{U_{i'},P} \in C_i$,$A'_{U_{j'},P} \in C_j$ **do**
21             $d_s(i',j') = ||A'_{U_{i'},P} - A'_{U_{j'},P}||$;
22         $D_f(C_i,C_j) = \max(d_s(i',j'))$;
23     $p,q = \arg\{\min_{i,j}(D_f(C_i,C_j)\}$;
24   **if** $D_f(C_p,C_q) \leq D$ **then**
25   merge C_p and C_q together;
26   **else**
27   *stopcondition* = true;
28 **if** $||A_{U,P}|| > 0$ **then**
29     $\exists u, A'_{U,P} \in C_u$;
30   return $W_T = \{U_j | A'_{U_j,P} \in C_u\}$ as honest witnesses;
31 **else**
32     $q = \arg\{\max_i(|C_i|)\}$
33   return $W_T = \{U_j | A'_{U_j,P} \in C_q\}$ as honest witnesses;

In Algorithm 1, the proposed approach first finds the witnesses and collects the testimonies regarding the target provider P (Lines 1-5). How to discover the distributed witnesses and testimonies is also an important issue [7] [17], but it is not the focus of our current work. Here, we assume that the consumer can find the witnesses and collect the testimonies in some way. For simplicity, we use $A_{U,P}$ to denote the accumulated rating vector with temporal decay from U regarding P till the moment when U is evaluating P's reputation. Before clustering, the collected rating vectors need to be normalized (Line 6). The normalization for each rating vector is achieved by having the value of each dimension divided by the sum of the value of each dimension (e.g., suppose that a rating vector is [0, 1, 4, 2, 1], then the normalized rating vector is [0, 0.125, 0.5, 0.25, 0.125]). After normalization, the clustering process begins. Lines 7-16 are for stage 1 clustering. Initially, each normalized rating vector is considered as a cluster (Lines 7-9). Then the proposed approach calculates the shortest distance between each two clusters as Lines 11-14 show. More specifically, if either or both clusters include more than one normalized rating vectors, the shortest distance is calculated as the minimum of the distance of any two rating vectors from each cluster. The two clusters with the minimum shortest distance is merged together (Lines 15-16). The process continues until the number of the clusters is smaller than or equal to *K*. Lines 17-27 are for stage 2 clustering. In this stage, the proposed approach first calculates the furthest distance between each two clusters achieved from the first stage (the similar method as in stage 1 clustering is adopted if either or both clusters include more than one normalized rating vectors) as Lines 19-22 show. Then the two clusters with the minimum furthest distance is merged together (Lines 23-25). The process continues until the furthest distance between any two clusters is larger than *D* (Lines 26-27). After getting the clustering results, there are two scenarios for consideration as shown below:

- If the consumer has personal ratings for the target provider, then the witnesses whose rating vectors are included in the cluster which contains the rating vector from the consumer are returned as honest witnesses (Lines 28-30).
- If the consumer does not have personal ratings for the target provider, then the witnesses whose rating vectors are included in the cluster which contains the greatest number of rating vectors are returned as honest witnesses (Lines 31-33). Here, we argue that there are other

options to select a particular cluster. For our current experiments, we follow the majority rule by selecting the cluster including the greatest number of rating vectors. Therefore, it will lead to inappropriate cluster selected in the scenario where unfair testimonies occupy the majority of the total testimonies. Other options can be, for example, that the consumer applies other information, such as context information, to make a decision.

It needs to point out that the proposed approach is for the purpose of filtering out unfair testimonies. To evaluate the provider's reputation, the proposed approach needs to be integrated with a reputation evaluation mechanism, e.g.,[1] [12] [19] to get the provider's reputation by passing the achieved honest witnesses' testimonies to the reputation system.

V. Experimental Studies

We carry out three sets of experiments to evaluate the accuracy and robustness of the proposed approach. The first set investigates the relationship between the threshold values and the accuracy of the proposed approach in filtering out unfair testimonies. The second set examines the robustness of the proposed approach in various scenarios where multi-nominal rating levels are adopted. The third set compares the proposed approach with other two representative approaches -- the iterated filtering approach [8] from the category of statistical approaches and TRAVOS [17] from the category of probabilistic approaches -- in terms of the accuracy of filtering out unfair testimonies and evaluating provider reputation when only binary ratings are accepted.

We measure the accuracy of an approach in filtering out unfair testimonies in two ways. One is its ability to detect dishonest witnesses, which can be measured by the false positive rate (FPR) and false negative rate (FNR), computed as:

$$FPR = \frac{f_p}{f_p + t_n}, FPR = \frac{f_n}{f_n + t_p},$$

where f_p, t_p, f_n, t_n represent the number of false positives, true positives, false negatives and true negatives, respectively. In our experiments, a true positive means that an honest witness is correctly detected as honest; a false positive means that a dishonest one is incorrectly detected as honest; a true negative means that a dishonest one is correctly filtered out as dishonest; a false negative means that an honest witness is incorrectly filtered out as dishonest. The lower values of FPR and FNR imply higher accuracy.

The other way is to use Matthew's correlation coefficient (MCC) [21] to generally measure the accuracy of an approach in filtering out unfair testimonies, which is computed as:

$$MCC = \frac{t_p \times t_n - f_p \times f_n}{(t_p + f_p) \times (t_p + f_n) \times (t_n + f_p) \times (t_n + f_n)}.$$

MCC value is between -1 and 1, where 1 represents a perfect filtering result, -1 represents an inverse filtering result, and 0 represents a random filtering result.

Our experiments involve two types of dishonest witnesses [16]: 1) *ballot-stuffing* witnesses report testimonies that the provider is trustworthy regardless of the true behavior of the provider; 2) *badmouthing* witnesses report testimonies that the provider is not trustworthy regardless of the true behavior of the provider.

II. Format Requirement

A. ***Threshold Value Exploration***

In this set of experiments, we explore how the threshold values (*K* and *D*) impact the accuracy of the proposed approach in filtering out unfair testimonies and try to establish an optimum threshold value range.

In current research regarding mitigating the adverse effects of unfair testimonies, it has always been a difficulty to obtain real-world data to conduct experiments due to the following two reasons. First, it is difficult to get the realistic data for an online transaction system. Second, even if we can get the data from the system (e.g., eBay), there is usually no ground truth on which ratings are unfair. Therefore, most of current research work uses simulated synthetic data to conduct experiments. We currently also use simulation to investigate the performance of the proposed approach and compare with other approaches.

In this set of experiments, we simulate an online service provision environment which includes 1 provider *P*, ω witnesses, and 1 consumer *U*. The provider *P* has an initial willingness (*iw*) value selected from the value set {0.1, 0.3, 0.5, 0.7, 0.9}. Each witness (or *U*) has *T* transactions with *P*. For each transaction, *P*'s rating is simulated by a willingness value which is generated from a normal distribution whose mean is equal to *iw*, and standard deviation is δ. The mapping between the generated willingness value and the rating level (5 rating levels are simulated) for *P* is shown in Table 1.

Willingness	(-∞,0.2]	(0.2,0.4]	(0.4,0.6]	(0.6, 0.8]	(0.8, ∞)
Rating	1	2	3	4	5

Table 1 Mapping from Wiliness Values to Rating Levels

Each dishonest witness will report ratings unfairly with a randomly generated probability value greater than *r*. The percentages of ballot-stuffing witnesses and badmouthing witnesses are p_h and p_l, respectively. We assume that when *P*'s *iw* value is very low (i.e., *iw*=0.1 or *iw*=0.3), there are no badmouthing witnesses. When *P*'s *iw* value is very high (i.e., *iw*=0.7 or *iw*=0.9), there are no ballot-stuffing witnesses.

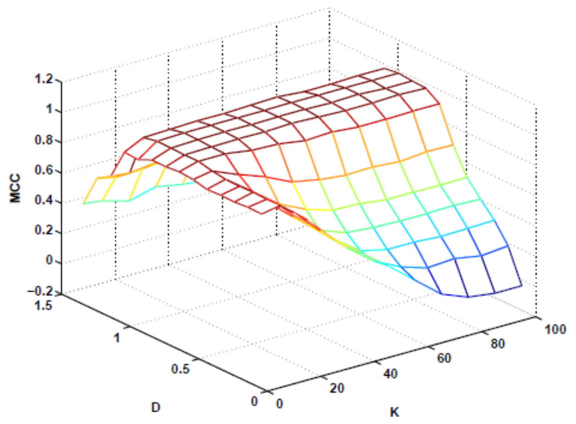
We explore how the threshold values will impact the accuracy of the proposed approach from two aspects -- scalability and stability. Here the scalability means the number of the witnesses reporting ratings to the consumer, and the stability means the number of the transactions happening between each witness (or *U*) and *P*. A high scalability means that there are a lot of witnesses. On the contrary, a low scalability means that there are few witnesses. A high stability means that there are a large number of transactions between *P* and the witnesses (or *U*), and a low stability means that there are only a few transactions between *P* and the witnesses (or *U*). We explore two levels of scalability and stability -- high (i.e., 100 witnesses and 100 transactions) and low (i.e., 10 witnesses and 10 transactions). Table 2 lists the parameter meanings and possible values adopted in our simulation.

Parameter	Meaning	Value
ω	The number of witnesses	{10,100}
iw	*P*'s initial willingness value	{0.1,0.3,0.5,0.7,0.9}
δ	The standard deviation of the normal distribution to simulate *P*'s behavior	{0.1,0.2}
T	The number of transactions between each witness or *U* and *P*	{10,100}
p_h	The percentage of ballot-stuffying witnesses	{0, 15%, 30%}
p_l	The percentage of badmouthing witness	{0, 15%, 30%}
r	The smallest probability value of the dishonest witnesses reporting ratings unfairly	0.5

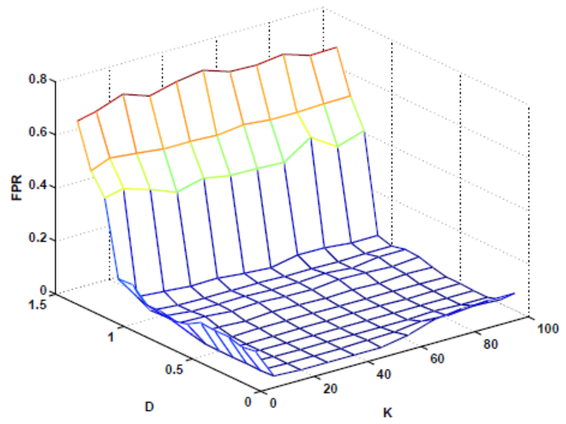
Table 2 Simulation Parameters, Values and Meanings

In our simulation, we set the forgetting factor as $\lambda = 0.9$. *K* will increase from 1 to ω with step 1. *D* will increase from 0.1 to 1.4 (the largest 2-norm distance value between two normalized rating vectors) with step 0.1. We totally have 40 parameter value combinations. For each parameter value combination, there are two scenarios: *U* has transactions with *P*, and *U* has no transactions with *P*. Therefore, we totally have 80 simulation scenarios. We run 100 rounds for each simulation scenario. As an illustration, we show the results of the MCC, FPR, and FNR value changes with *K* and *D* values in the scenario where $iw = 0.5$, $\delta = 0.2$, $p_h = 15\%$, and p_l =15% when scalability and stability are high or low, and *U* has no transactions with *P*. Then we have four sets of results as Figure 1 and 2 show.

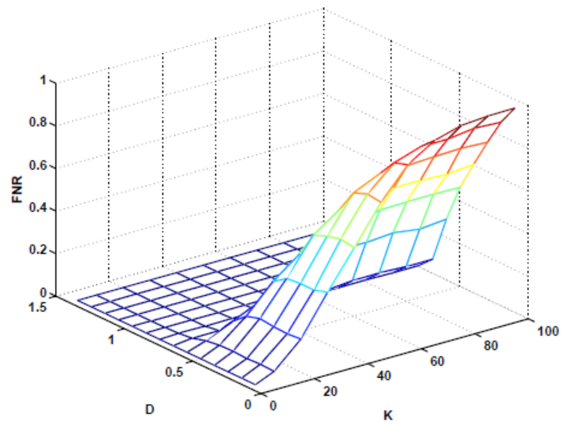
Figure 1 shows the MCC, FPR and FNR values when the scalability is high. According to the results, when the scalability is high (i.e, ω=100), MCC (FPR or FNR) value presents a similar changing trend when the stability is high or low. MCC value first increases with *D* value when $K > 10$, then it keeps stable for some *D* values, and then decreases until the largest *D* value is



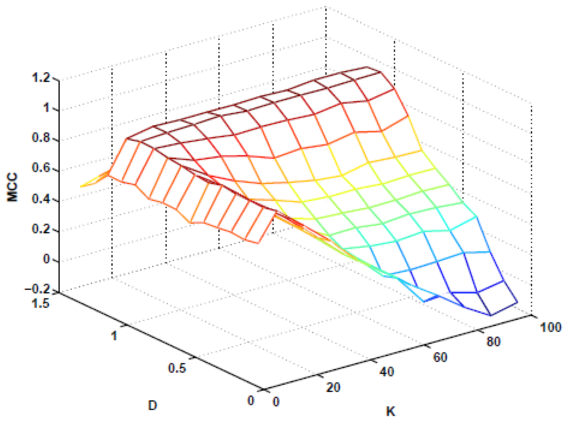
(a)



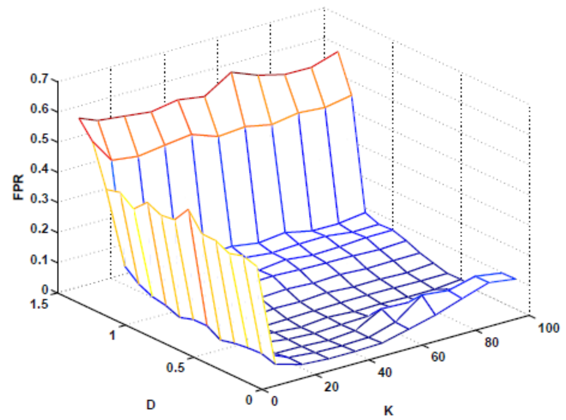
(b)



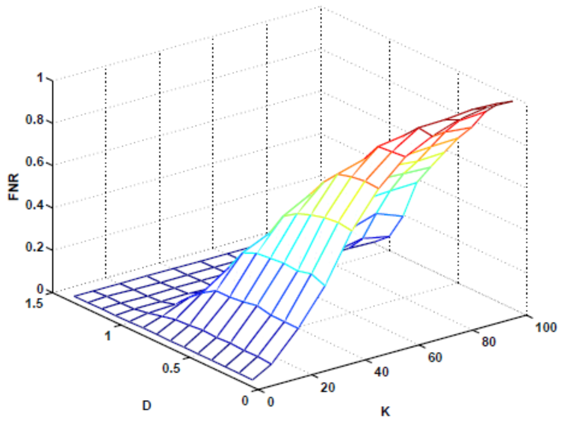
(c)



(d)



(e)



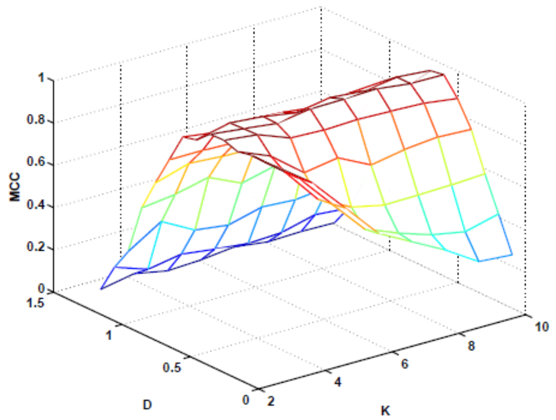
(f)

Figure 1 (a) MCC value when T = 100 and ω = 100; (b) FPR value when T = 100 and ω = 100; (c) FNR value when T = 100 and ω = 100; (d) MCC value when T = 10 and ω = 100; (e) FPR value when T = 10 and ω = 100; (f) FNR value when T = 10 and ω = 100 reached. When $K \leq 10$, it can be noticed that when $D < 1.1$, the proposed approach can achieve a good MCC value as shown in Figure 1 (a) and (d).

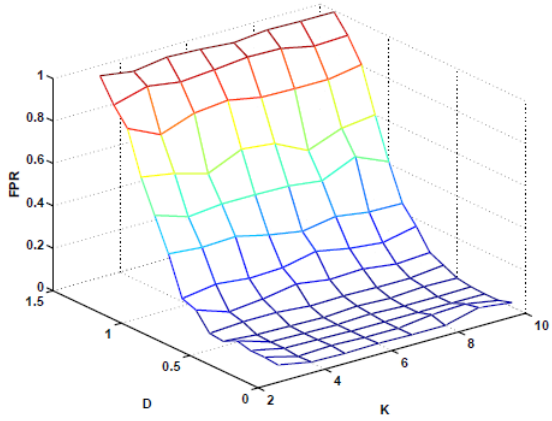
When D value is very large (i.e., 1.2-1.4), FPR is very high as shown in Figure 1 (b) and (e). As when D value is very large, the clusters achieved from the first stage will be wrongly merged together in the second stage. More specifically, the clusters including the dishonest witnesses are merged with the clusters including the honest witnesses, leading to a high FPR value. For FNR value, it can be noticed that FNR is very high when K value is large (i.e, 40-100) and D value is not large (i.e., 0.1-0.7) as shown in Figure 1 (c) and (f). The reason is that when K value is large, the proposed approach will finish the first stage clustering quite early, leaving some clusters including honest witnesses alone. But at the second stage, the clusters including the honest witnesses cannot be merged as D value is too small, finally leading to that FNR is very high. Generally speaking, when D value is too large, FPR will be very high, and when K value is too large and D value is too small, FNR will be very high. A noticed trend is that FNR value increases with the value of $K/_{\omega}$ increasing when D value is small.

Figure 2 shows the MCC, FPR and FNR values when the scalability is low (i.e., $\omega = 10$). It can be noticed that MCC (FPR or FNR) values show a similar changing trend as those shown in Figure 1 when the stability is high or low.

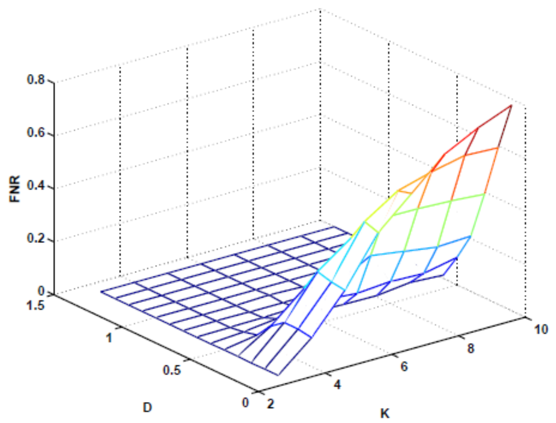
As a summary, a larger D value will cause a higher FPR no matter the scalability is high or low. And a smaller D value will cause a higher FNR when $K/_{\omega}$ is large. Therefore, before choosing the appropriate threshold values, the consumer better evaluates the scalability and stability of the collected testimonies. If the consumer more concerns with FPR, then a smaller D value is a good choice though this may lead to a higher FNR. If the consumer more concerns with FNR, then a smaller K value is a better choice. As a balance, it is a good practice to adopt a relatively small K value ($\leq$ 10) and a medium D value (0.7-0.9). In our following experiments, we generally explore the accuracy of the proposed approach in the scenarios



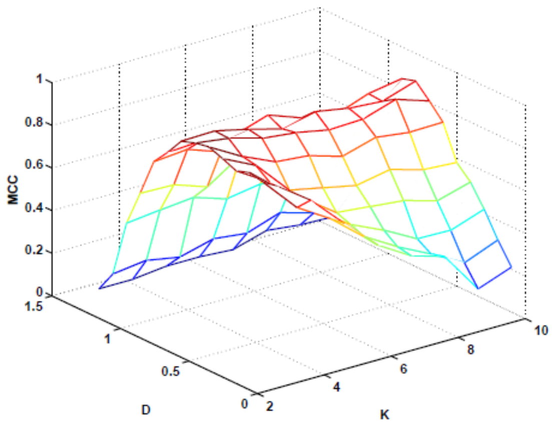
(a)



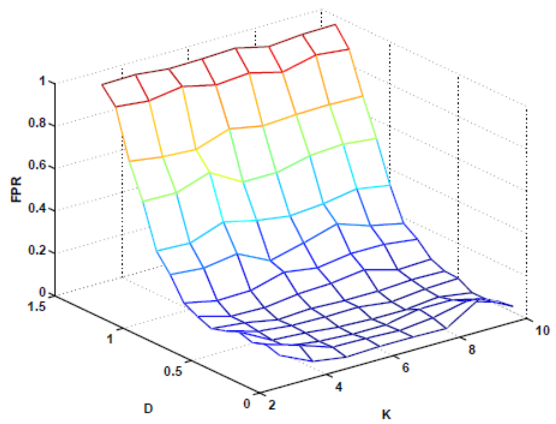
(b)



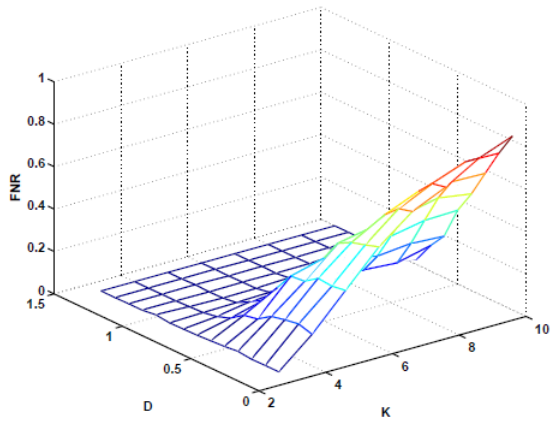
(c)



(d)



(e)



(f)

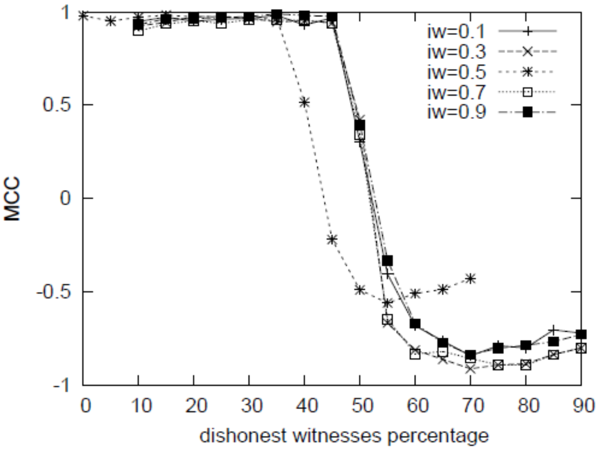
Figure 2 (a) MCC value when T = 100 and ω = 10; (b) FPR value when T = 100 and ω = 10; (c) FNR value when T = 100 and ω = 10; (d) MCC value when T = 10 and ω = 10; (e) FPR value when T = 10 and ω = 10; (f) FNR value when T = 10 and ω = 10 when the scalability is medium (i.e., ω = 20) and stability is randomly generated. The threshold values we use in the following experiments are K = 10 and D = 0.7.

B. *Robustness Investigation*

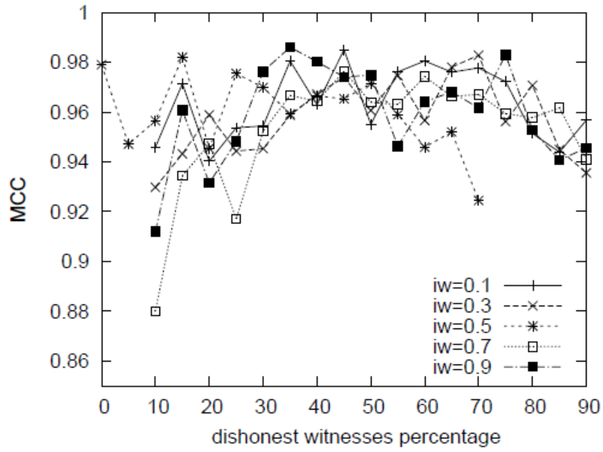
The aim of this set of experiments is to investigate the robustness of the proposed approach in various scenarios when appropriate threshold values are set. In this set of experiments, we simulate a similar service provision environment as that in the first set of experiments, including 1 provider P, 20 witnesses and 1 consumer U. The value of the standard deviation of the normal distribution to simulate the provider's behavior is 0.2.

In this experiment, we first explore the accuracy of the proposed approach when the percentage of dishonest witnesses changes. Figure 3 (a) and (b) show the MCC value changes with the increase of the percentage of dishonest witnesses. When P's iw value is 0.1 or 0.3, p_h = 0, and p_l increases from 10% to 90%. When P's iw value is 0.7 or 0.9, p_l, and p_h increases from 10% to 90%. When P's iw value is 0.5, p_h = 20%, and p_l increases from 0 to 70%. Each witness (or U) will have a randomly generated number of transactions (in the range of [10,100]) with P.

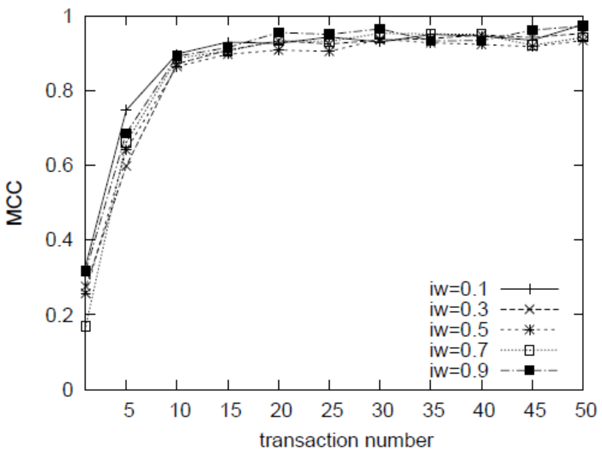
According to the results, when U has no transactions, MCC value is close to 1 until the dishonest witnesses are the majority. For example, when P's iw value is 0.5 and p_l is smaller than 40%, MCC value is close to 1. As the percentage of honest witnesses is greater than 40% which is greater than p_l or p_h, the proposed approach will follow the majority rule and a correct filtering result can be achieved. But when p_l is greater than 40%, the percentage of honest witnesses is smaller than p_l. In such a case, the proposed approach will return the badmouthing witnesses as honest witnesses by following the majority rule. When U has transactions with P, however, the proposed approach can work properly even when the dishonest witnesses are the majority according to the results shown in Figure 3 (b) (note that the y-axis value starts from 0.85 in Figure 2 (b)). MCC value is close to 1 even when the percentage of dishonest witnesses in the population is 90%. As when U has transactions with P, the proposed approach will return the witnesses that have similar rating vectors as U's personal rating vector as honest witnesses. Therefore, a correct filtering result can be achieved even when the dishonest witnesses are the majority.



(a)



(b)



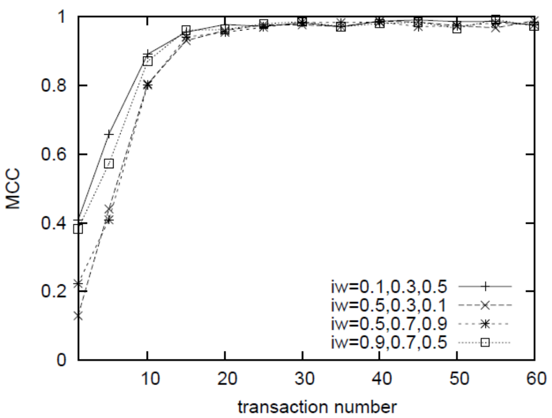
(c)

Figure 3 (a) *U* has no transactions; (b) *U* has transactions; (c) MCC value changes with the increase of the number of transactions

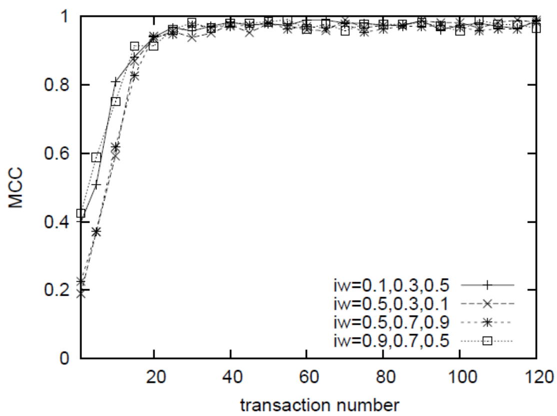
Secondly, we explore the accuracy of the proposed approach changes with the number of transactions between each witness and *P* (for simplicity, *U* has no transactions with *P* because when the dishonest witnesses are not the majority, the filtering result when *U* has no transactions is similar to that when *U* has transactions). When *P*'s *iw* value is 0.1 or 0.3, $p_l = 0$ and $p_h = 30\%$. When *P*'s *iw* value is 0.7 or 0.9, $p_l = 30\%$ and $p_h = 0$. When *P*'s *iw* value is 0.5,

p_l = 15% and p_h =15%. The number of transactions between each witness and *P* increases from 1 to 50. The results are shown in Figure 3 (c). We can see that MCC value increases with the number of transactions. Initially, MCC value is very low. And after about 15 transactions, MCC value approximates to 1, meaning that the proposed approach performs more and more stably.

Thirdly, we investigate the robustness of the proposed approach in the scenario where *P*'s behavior changes over time. As *P*'s behavior is simulated by the normal distribution, we change the mean of the normal distribution to simulate his behavior change. In this experiment, we simulate four scenarios: *P*'s *iw* value increases from 0.1 to 0.3 then to 0.5; *P*'s *iw* value decreases from 0.5 to 0.3 then to 0.1; *P*'s *iw* value decreases from 0.9 to 0.7 then to 0.5; *P*'s *iw* value increases from 0.5 to 0.7 then to 0.9. In the first two scenarios, p_l = 30% and p_h = 0. In the last two scenarios, p_h = 30% and p_l = 0. For simplicity, *U* has no transactions with *P*. Figure 4 shows the MCC value changes with the number of transactions.



(a)

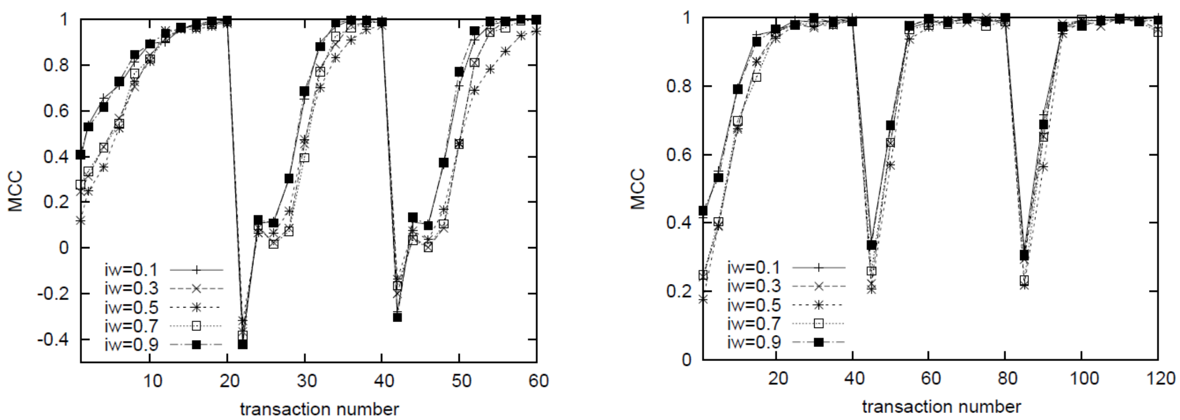


(b)

Figure 4 (a) *P* changes his behavior after each 20 transactions; (b) *P* changes his behavior after each 40 transactions

According to the results, MCC value is initially very low as the number of transactions is small. Then after about 15 transactions, the proposed approach performs more stably and MCC value approaches to 1. Though after each 20 or 40 transactions, the provider changes its behavior, the proposed approach still can work properly and stably.

Finally, we explore the robustness of the proposed approach in the scenarios where the witnesses' behaviors change over time. We use the following settings. When *P*'s *iw* value is 0.1 or 0.3, p_h = 30% and p_l = 0. When *P*'s *iw* value is 0.7 or 0.9, p_h = 30% and p_h = 0. When *P*'s *iw* value is 0.5, p_l = 15% and p_h = 15%. For simplicity, *U* has no transactions with *P*. Initially, the last 30% witnesses are selected as dishonest witnesses, then after a time period (i.e., 20 or 40 transactions), the first 30% witnesses are selected as dishonest witnesses. After the same length of time period again, the simulation is set as the initial setting. In such a way, the experiment simulates the witnesses' behavior changes -- from honest to dishonest, or from dishonest to honest over time. Figure 5 shows the MCC value changes with witness behavior variation over time.



(a) (b)

Figure 5 (a) Witnesses change their behaviors after each 20 transactions; (b) Witnesses change their behaviors after each 40 transactions

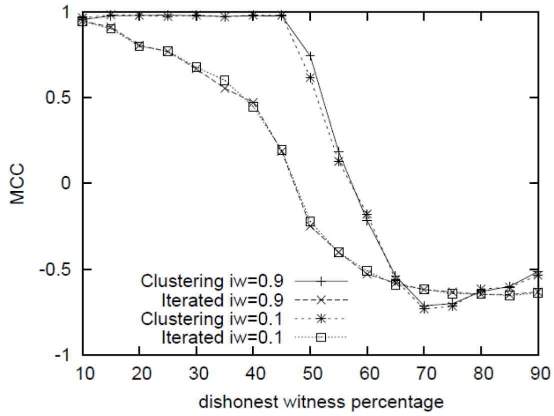
As Figure 5 (a) shows, in the first 20 transactions, MCC value is initially very low as the witnesses only have a few transactions with *P*. After about 15 transactions, MCC value is close to 1. Then in the time window of 20-40 transactions and 40-60 transactions, MCC value first suddenly decreases to a very low value and then increases continuously and approaches to 1 again after about 15 transactions. The results in Figure 5 (b) present the similar trend as those in Figure 5 (a). When the witnesses' behaviors change, MCC value has a sharp decrease and then increases continuously and approaches to 1 after about 15 transactions. The results suggest that the proposed approach needs a period to learn the witnesses' behaviors, which is about 15 transactions. The results are also consistent with the results shown in Figure 3 (c) and Figure 4.

C. *Comparative Experiments*

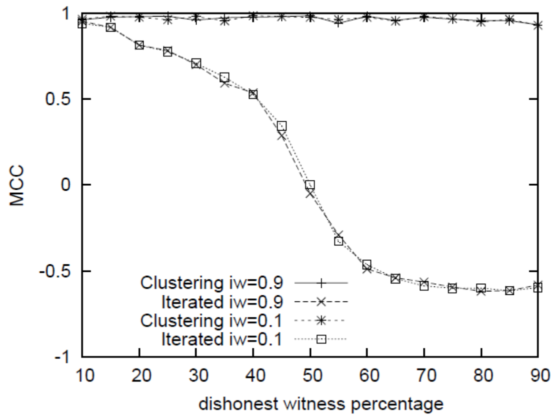
We compare the proposed approach with other approaches to addressing the problem of unfair testimonies from two aspects -- the accuracy of filtering out unfair testimonies and evaluating provider reputation. Though the motivation for the proposed approach is to address the problem of unfair testimonies for reputation systems supporting multi-nominal rating levels, the proposed approach can also work for reputation systems accepting binary rating levels. In this experiment, we compare the proposed approach with the iterated filtering approach [8] in terms of filtering accuracy. We also compare the proposed approach with both the iterated filtering approach and TRAVOS [17] in terms of the accuracy of evaluating provider reputation. Note that in the experiments, after filtering out unfair testimonies by the proposed approach and the iterated filtering approach, achieved fair testimonies will be aggregated to evaluate provider reputation using BRS [1].

In this experiment, we simulate a similar service provision environment as that in the second set of experiments, which includes 1 provider *P*, 20 witnesses, and 1 consumer *U*. *P* has an initial willingness (*iw*) value which is from the value set {0.1, 0.2, 0.3, 0.7, 0.8, 0.9}. For each transaction, one willingness value is randomly generated to indicate the probability that the provider is rated as positive for this transaction. The willingness value for the first transaction is equal to the initial willingness value. The willingness values for the subsequent transactions are generated through one of the following three strategies -- the willingness value of last transaction adding 0.02, equal to the willingness value of last transaction, or the willingness value of last transaction subtracting 0.02. The three strategies are randomly selected with equal probability. The willingness value for each transaction is also limited in the range of [*iw*-0.1, *iw*+0.1]. In this experiment, we study one type of dishonest witnesses, who will report the opposite of the actual ratings as testimonies. When comparing with TRAVOS, *U* will have transactions with *P* as TRAVOS needs the consumer's personal transactions to work.

Figure 6 shows the MCC value changes with the increase of the percentage of dishonest witnesses. The number of transactions between each witness (or *U*) and *P* is 20. As Figure 6 (a) shows, the proposed approach (i.e., labelled as clustering) can filter out the dishonest witnesses if they are not the majority when *U* has no transactions. When *U* has transactions, the proposed approach can filter out the unfair testimonies no matter what the percentage of dishonest witnesses is. The MCC value using the iterated filtering approach continuously decreases with the increase of the percentage of dishonest witnesses as Figure 6 (b) shows. It can be noticed that when there are 30% dishonest witnesses, MCC using the iterated filtering approach is close to 0.5, implying that using the iterated filtering approach is close to random guessing.



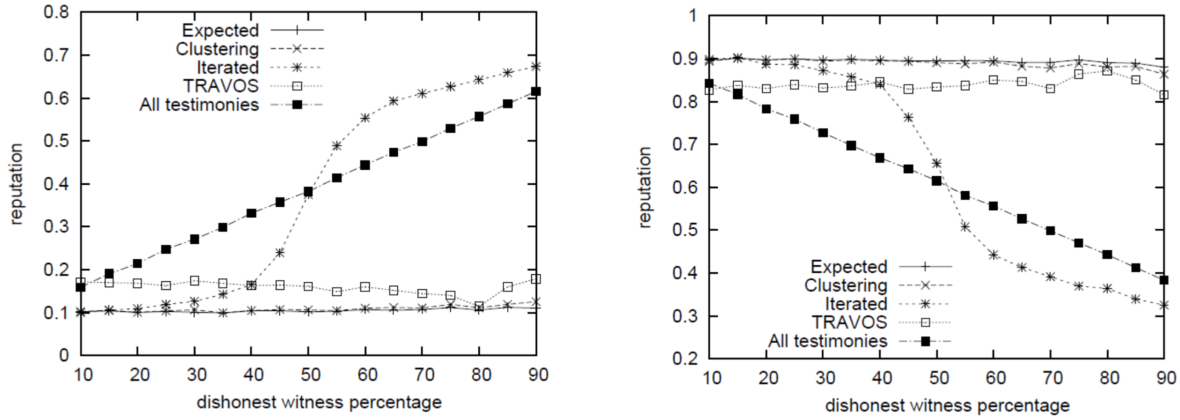
(a)



(b)

Figure 6 (a) *U* has no transactions; (b) *U* has transactions

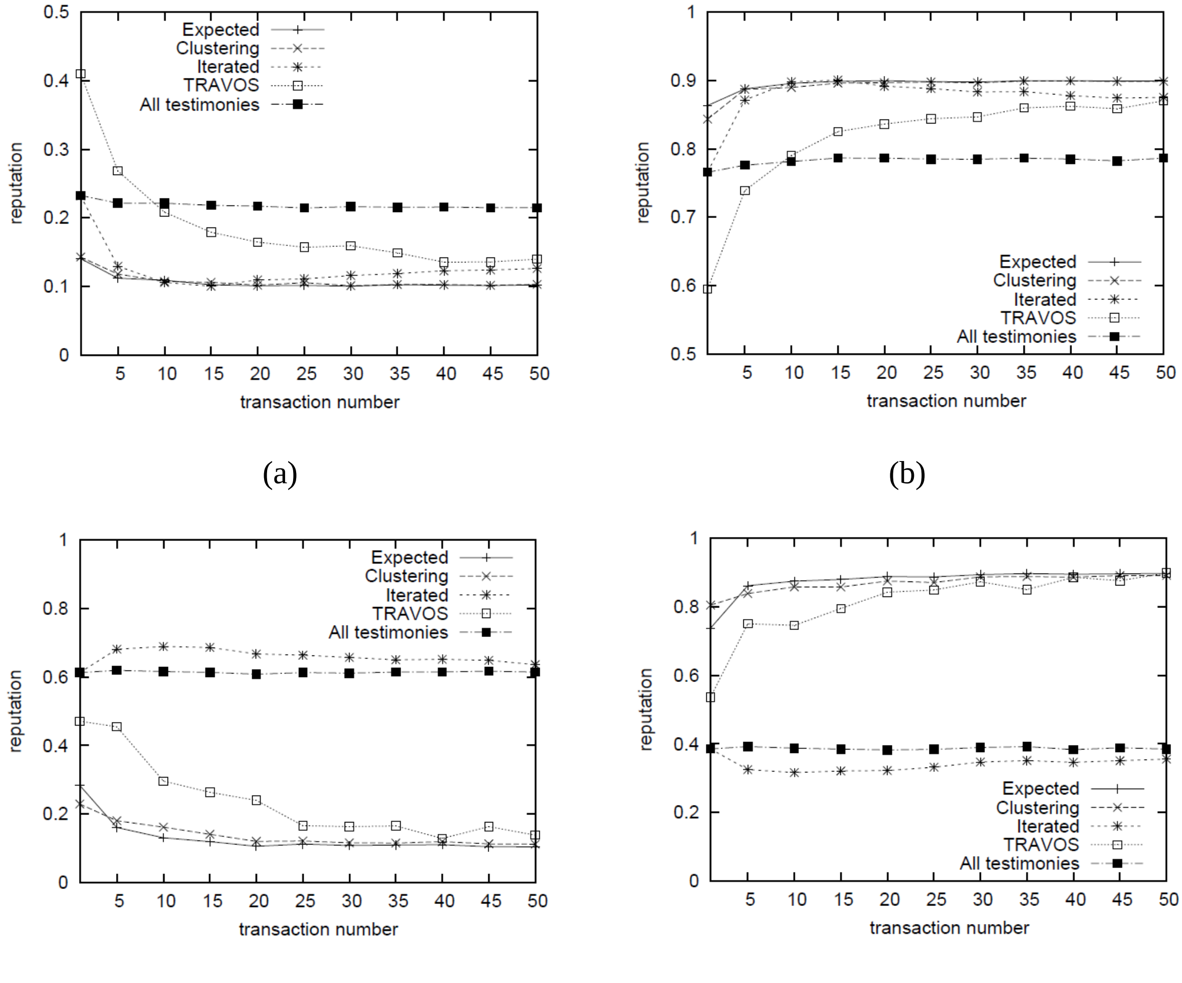
Figure 7 shows the reputation value after using different approaches. According to the results, the reputation value by using the proposed approach is very close to the expected reputation value, which is calculated using the actually fair testimonies only. The reputation value after using the iterated filtering approach is initially very close to the expected reputation value, then continuously deviates from the expected reputation value with the increase of the percentage of dishonest witnesses. TRAVOS works stably but the reputation value produced by it is not as close to the expected reputation value as the proposed approach.



(a) (b)

Figure 7 (a) *iw* = 0.1; (b) *iw*=0.9

Figure 8 shows the evaluated provider reputation value changes with the increase of the number of transactions after using different approaches. In Figure 8 (a) and (b), the percentage of dishonest witnesses is 20%, and *iw* value is 0.1 and 0.9, respectively. In Figure 8 (c) and (d), the percentage of dishonest witnesses is 90%, and *iw* value is 0.1 and 0.9, respectively. According to the results, when the percentage of dishonest witnesses is 20%, the reputation value after using the proposed approach is very close to the expected reputation value after about 10 transactions. The reputation value after using the iterated filtering approach is close to the expected reputation value after about 10 transactions, but then deviates from the expected reputation value after about 30 transactions. The TRAVOS reputation value continuously approximates to the expected reputation value with the increase of the number of transactions but it needs more transactions to approach to the expected reputation value. When the percentage of dishonest witnesses is 90%, the reputation value after using the proposed approach is still very close to the expected reputation value. But the reputation value after using the iterated filtering approach seriously deviates from the expected reputation value, and is even worse than the result without using any approach. The TRAVOS reputation value continuously approaches to the expected reputation value but needs a longer time.



(a) (b) (c) (d)

Figure 8 (a) *iw* = 0.1 and 20% dishonest witnesses; (b) *iw* = 0.9 and 20% dishonest witnesses; (c) *iw* = 0.1 and 90% dishonest witnesses; (d) *iw* = 0.9 and 90% dishonest witnesses

D. Discussion

We have conducted three sets of experiments. In the first set of experiments, we explore the influence of threshold values on the accuracy of the proposed approach when different levels of scalability and stability are set. According to the results, a larger *D* value will lead to a higher FPR, a smaller *D* value and a larger $K/_{\omega}$ value will lead to a higher FNR. Therefore, a medium *D* value and a smaller $K/_{\omega}$ value are more preferred. For different scenarios, it is better to consider current scalability and stability levels and set the threshold values according to the FPR and FNR concerns.

In the second set of experiments, we investigate the robustness of the proposed approach in various scenarios when multi-nominal rating levels are adopted. When the witnesses or the consumer has more than 15 transactions with the provider, the proposed approach can work more stably and accurately in filtering out unfair testimonies. Therefore, it also suggests that when the consumer is collecting testimonies from the witnesses, it is better to consider the testimonies from the witnesses that have more than 15 transactions with the provider. When the consumer has no transactions with the provider, the proposed approach can work accurately until the dishonest witnesses are the majority. When the consumer has some transactions with the provider, the proposed approach can work properly no matter what the percentage of dishonest witnesses is.

In the scenario where the provider's behavior varies over time, the variation has no obvious influence on the accuracy of the proposed approach working. The proposed approach still works more stably and accurately after about 15 transactions. In the scenario where the witnesses' behaviors change over time, the proposed approach cannot work at the moment when the witnesses change their behaviors suddenly. As the results show, the proposed approach needs a period to learn the witnesses' behaviors, which is about 15 transactions. Therefore, if the witnesses change their behaviors frequently, the proposed approach may not identify the dishonest witnesses. But for such scenarios, all the witnesses can be treated as dishonest from a global point of view.

In the third set of experiments, we compare the proposed approach with the iterated filtering approach and TRAVOS from two aspects -- the accuracy of filtering out unfair testimonies and evaluating provider reputation. Compared with the iterated filtering approach, the proposed approach performs better on the accuracy of filtering out unfair testimonies, and hence achieves a more accurate result on evaluating provider reputation. Compared with TRAVOS, the proposed approach can accurately evaluate the provider reputation at a faster speed.

V. Conclusion

Reputation systems have contributed much to the success of online service provision systems. However, the reliability of reputation systems can easily deteriorate due to the existence of unfair testimonies. To cope with the problem of unfair testimonies, we propose a two-stage clustering approach to filter unfair testimonies for reputation systems using multi-nominal ratings. As the experimental results show, the proposed approach can effectively filter out unfair testimonies in various scenarios when appropriate threshold values are set.

References

[1] A. Jøsang, "The Beta reputation system," in *Proceedings of the 15th Bled Conference on Electronic Commerce*, 2002, pp. 50-59.

[2] L. Xiong and L. Liu, "Peertrust: supporting reputation-based trust for peer-to-peer electronic communities," *IEEE Transactions on Knowledge and Data Engineering*, vol. 16, no. 7, pp. 843-857, 2004.

[3] K. Regan, P. Poupart, and R. Cohen, "Bayesian reputation modeling in e-marketplaces sensitive to subjectivity, deception and change," in *Proceedings of the 21st National Conference on Artificial Intelligence, 2006*, pp. 206-212.

[4] W. Teacy, N. R. Jennings, A. Rogers, and M. Luck, "A hierarchical bayesian trust model based on reputation and group behaviour," in *the 6th European Workshop on Multi-Agent Systems*, 2008, pp. 206-212.

[5] A. Jøsang, R. Ismail, and C. Boyd, "A survey of trust and reputation systems for online service provision," in *Decision Support System*, vol. 43, no. 2, pp. 618-644, 2007.

[6] L. Rasmusson and S. Janssen, "Simulated social control for secure internet commerce," in *1996 New Security Paradigms Workshop*, 1996.

[7] B. Yu and M. Singh, "Detecting deception in reputation management," in *Proceedings of 2nd International Joint Conference on Autonomous Agents and Multiagent Systems*, 2003, pp. 73-80.

[8] A. Whitby, A. Jøsang, and J. Indulska, "Filtering out unfair ratings in bayesian reputation systems," in *ICFAIN J. Management Res.*, vol. 4, no. 2, pp. 48-64, 2005.

[9] J. Weng, C. Miao, and A. Goh, "An entropy-based approach to protecting rating systems from unfair testimonies," in *IEICE Trans. INF. and SYST.*, vol. E89-D, no.9, pp. 2502-2511, September 2006.

[10] J. Zhang and R. Cohen, "Evaluating the trustworthiness of advice about seller agents in e-marketplaces: a personalized approach," in *Electronic commerce Research and Applications*, vol. 7, pp. 330-340, 2008.

[11] Z. Noorian, S. Marsh, and M. Fleming, "Multi-layer cognitive filtering by behavioral modeling," in *Proceedings of the 10th International Conference on Autonomous Agents and Multiagent Systems*, 2011, pp. 871-878.

[12] A. Jøsang and J. Haller, "Dirichlet reputation systems," in *Proceedings of the 2nd International Conference on Availability, Reliability and Security*, April 2007, pp. 112-119.

[13] A. Gelman, *Bayesian Data Analysis*. Florida, USA: Chapman & Hall/CRC, 2004.

[14] B. Yu and M. Singh, "An evidential model of distributed reputation management," in *Proceedings of the 1st International Joint Conference on Autonomous Agents and Multiagent Systems*, 2002, pp. 294-301.

[15] G. Shafter, *A Mathematical Theory of Evidence*. Princeton University Press, 1976.

[16] C. Dellarocas, "Immunizing online reputation reporting systems against unfair ratings and discriminatory behavior," in *Proceedings of the 2nd ACM Conference on Electronic Commerce*, 2000, pp. 150-157.

[17] W. Teacy, J. Patel, N. R. Jennings, and M. Luck, "TRAVOS: trust and reputation in the context of inaccurate information sources," in *Autonomous Agents and Multi-Agent Systems*, vol. 12, no. 2, pp. 183-198, 2006.

[18] H. Yu, Z. Shen, and B. An, "An adaptive witness selection method for reputation-based trust models," in *Proceedings of the 15th International Conference on Principles and Practice of Multi-Agent Systems (PRIMA)*, 2012, pp. 184-198.

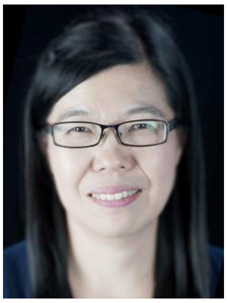
[19] C. J. Fung, J. Zhang, I. Aib, and R. Boutaba, "Dirichlet-based trust management for effective collaborative intrusion detection networks," in *IEEE Transactions on Network and Service Management*, vol. 8, no. 2, pp. 79-91, 2011.

[20] R. O. Duba, P. E. Hart, and D. G. Stork, *Pattern Classification*. Wiley-Interscience, 2001.

[21] B. Matthews, "Comparison of the predicted and observed secondary structure of t4 phage lysozyme," in *Biochim. Biophys. Acta*, vol. 405, pp. 442-451, 1975.

Siyuan Liu got her B.Sc and M.Sc degrees in computer science from Peking University, China in 2002 and 2005, respectively. She obtained her PhD degree from Nanyang Technological University (NTU), Singapore in 2013. Currently she is a Research Fellow in Joint NTU-UBC Research Centre of Excellence in Active Living for the Elderly (LILY). Her research interest includes trust and reputation in multi-agent system and incentive mechanism design in crowdsourcing.

Chunyan Miao received the B.S. degree in computer science from Shandong University in 1988 and the M.Eng. and Ph.D. degrees in computer engineering from Nanyang Technological University (NTU), Singapore in 1996 and 2001, respectively. She has been an Assistant Professor at the School of Computer Engineering (SCE) in NTU since 2003. She is currently an Associate Professor in NTU and the Director of the Joint NTU-UBC Research Centre of Excellence in Active Living for the Elderly (LILY). Her main research focuses on the study of human factors of agents, social-ecological models of multiagent systems, and their applications in real world systems.



Yin-Leng Theng got her B.Sc from National University of Singapore in 1983, M.Sc from Manchester Institute of Science & Technology in 1995, and PhD from Middlesex University in 1997. She is currently the Associate Chair (Research) and an Associate Professor with the Division of Information Studies, Wee Kim Wee School of Communication and Information at Nanyang Technological University, Singapore. Her research interests are mainly in user-centered design, interaction design and usability engineering, and digital libraries.

Alex C. Kot has been with the Nanyang Technological University, Singapore since 1991. He headed the Division of Information Engineering at the School of Electrical and Electronic Engineering for eight years and served as Associate Chair/ Research and Vice Dean Research for the School of Electrical and Electronic Engineering. He is currently Professor and Associate Dean for College of Engineering in NTU. He is the Director of Rapid (Rich) Object SEearch Lab (ROSE) in NTU. He has published extensively in the areas of signal processing for communication, biometrics, data-hiding, image forensics and information security.