User-credibility Based Service Reputation Management for Service Selection

Cao Jiuxin, Dong Yi, Qin Yi, Liu Bo, Dong Fang, Zhou Tao
School of Computer Science and Engineering
Key Laboratory of Computer Network and Information Integration of MoE of China under Grants No. 93K-9
Southeast University, Nanjing, 211189, China
{jx.cao, dongyi, qinyi, bliu, fdong, zhoutao}@seu.edu.cn

Abstract—**In composite services, the atomic service reputation is becoming important when many similar functional services could be provided for selecting under the heterogeneous and loose-coupled circumstance. However, existing reputation measurement methods pay little attention on user credibility which has a great influence on accuracy. In this paper, we define user credibility as the ability of honest users to provide rational feedbacks, and propose a novel service reputation management model based on service invocation information and user feedback. This method consists of three parts: malicious feedbacks distinguishing, user credibility evaluating and service reputation predicting. Firstly, a clustering algorithm is conceived to filter out malicious feedbacks. Then a feedback deviation based algorithm is proposed to evaluate the user credibility considering service QoS similarity. Finally, an algorithm on the basis of user credibility and honest feedbacks is applied to predict service reputation. Experiment proves that our model can effectively detect malicious feedbacks and precisely measure service reputation with low error.**

Keywords-service reputation; user credibility; malicious feedbacks;

I. INTRODUCTION

With the development of the web services technology and service-oriented architecture, a verity of web services offering the same functionalities are provided to improve the flexibility of service invocation on the Internet. In composite services, the atomic service reputation is becoming important when many services with similar function but different non-functional properties could be provided by different providers under the heterogeneous and loose-coupled circumstance. Reputation of web services which can be computed according to non-functional quality of service (QoS) has become a widely accepted service metric. QoS is a set of quality requirements, such as throughput, response time, reliability, and availability [1][2]. In recent years, reputation-based service selection mechanisms have received much attention, and much related research has been done. Many researchers have studied this problem from multiple perspectives and proposed different techniques to solve it. But some important details should be more deeply researched.

Firstly, the existence of malicious feedbacks plays a bad effect on the accuracy of predicting service reputation. Malicious feedbacks detection mechanism has only received limited attention, and it is still an important topic. For instance, water army is employed to defame a service in malicious business competition.

Secondly, because of the different user experience, even the honest feedback may be biased. In large distributed environments, it is difficult for users with different experience to provide same feedbacks. Such as a hotel booking service, users who use the service repeatedly can provide much more rational feedbacks than those who use it for the first time.

To address these problems, in this paper, a user-credibility based service reputation management model is presented with the following contributions:

1. Malicious feedbacks distinguishing: a clustering algorithm is given to discover various existing clusters of feedbacks and filter out malicious feedbacks by the average value of a cluster.
2. User credibility evaluating: the concept of user credibility is introduced and a feedback deviation based algorithm is also designed to measure the ability of providing rational feedbacks.
3. A novel method of predicting service reputation: in this model, different weights are given to various user feedbacks in predicting service reputation in term of their credibility. In addition, the timing of service assessment and the difference between deterioration and improvement are also taken into account as two important elements.

The rest of the paper is organized as follows. First, we provide an overview of related work in section II. Section III briefly presents our service reputation management model for service selection. Section IV describes the mechanism of distinguishing malicious feedbacks and the algorithm of evaluating user credibility in detail. Then, the novel method of predicting service reputation is shown in section V. Finally, we discuss various experimental results in section VI and conclude the paper in section VII.

II. RELATED WORK

Much previous research has been done on service reputation management for service selection. Service reputation models can be classified into three categories according to their rationale: models based on the direct user experience, models based on a Trusted Third Party who evaluates services truthfully and hybrid models merging the two former ones [3]. Although much research has been done in many ways, so far there is no uniform criterion of service reputation evaluation. Some researchers regard service reputation as a quality attribute. Zeng L et al. proposed a web service quality model where reputation is considered as one of five quality attributes [4]. On the other hand, service quality attributes and

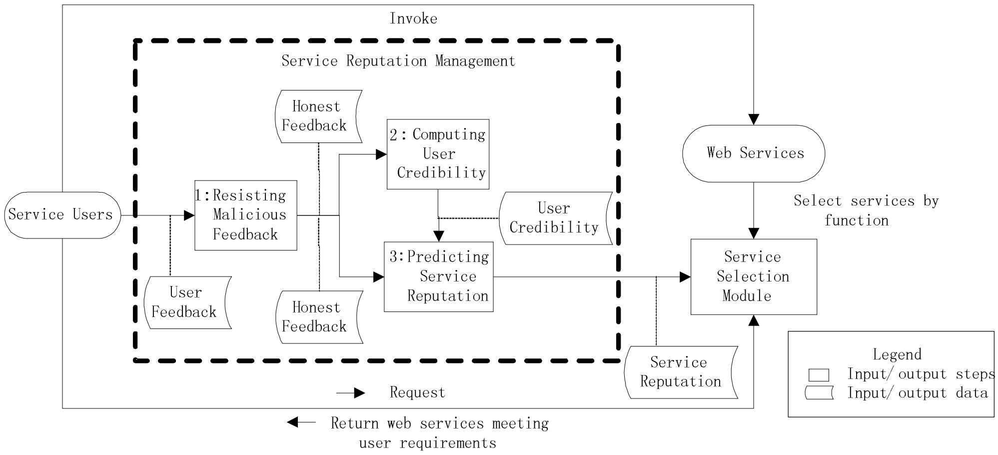


Figure 1. The reputation management model.

service reputation are treated separately in other research works [5, 6].

Conner W et al. present a trust-based reputation management framework which sets different service reputation functions for different services. That makes it flexible and effective to synthesize feedbacks on all kinds of services trust from many different entities [7]. A framework for reputation-aware service selection and rating automates the production of feedbacks to improve system efficiency [8]. A new QoS-based semantic web service selection and ranking solution proposed by Vu L H et al. aims at detecting and handling false ratings [9]. Maximilien E M et al. formally define service quality attributes by ontology, and propose a framework for dynamic web services selection based on a three-layer ontology structure [10, 11, 12]. Generally, same quality attributes may have different importance for different web service users in measuring service satisfaction [6, 13]. However, these solutions do not take into account user credibility which is very important to assure the accuracy of predicting service reputation. In addition, the service reputation has the incremental feature, and the fluctuation decays with time. The slot is regarded as the aging factor in the computing model [5]. Majithia S et al. introduce an attenuation function that reflects the weakening in the reputation value over time [14]. But these models do not pay enough attention to the timing of service assessment and the difference between deterioration and improvement.

III. A SERVICE REPUTATION MANAGEMENT MODEL

The service reputation management model is based on user direct experience, and the main idea is updating reputation of a service according to user feedbacks collected by service reputation management (SRM). Service reputation is a continuous variable based on time, and it is necessary and reasonable for us to adopt time discrete method which brings convenience to simplify computation complexity. In this paper, time series is divided equally into same slots, and Δt_j denotes the j_{th} time slot and t_j as the end of the j_{th} slot.

To illustrate our idea better, let us give the following fundamental definitions firstly:

Definition 1. Service Reputation Service reputation is a comprehensive assessment of a web service from the system, it is proportional to service quality. Let $R_{s_k}^{t_j}$ represent service reputation of service s_k at time t_j.

Definition 2. User Credibility User credibility is an objective evaluation of the ability to provide rational feedbacks, and only honest users is considered. Let $Cr(u_l, s_k)^{t_j}$ represent credibility of user u_l for service s_k at time t_j.

Definition 3. User Feedback User feedback is a subjective rating of a web service from a user, and expresses the level of satisfaction after invoking the web service. Here, $f_{u_i \to s_k}^{\Delta t_j}$ means the feedback on s_k from u_i in the slot Δt_j.

Definition 4. Similar Users Similar users are those who invoke same services and enjoy similar qualities.

Service reputation, user credibility and user feedback are all decimals ranging from 0 to 1, where 1 is assigned to the best one and 0 is assigned to the worst one.

Figure 1 gives the reputation management model, which mainly consists of two parts: Service Reputation Management (SRM) and Service Selection Module (SSM). Firstly, a service user sends SSM a request, including the functional requirements and a reputation threshold. Secondly, SSM selects objective service set meeting user functional requirements. Thirdly, the reputation is calculated by SRM for each service in the set. Finally, the services are selected by SSM, which have a higher reputation than the threshold from user requirement and are returned to users.

The most important part of the model is the design of the SRM module, whose performance determines the accuracy of the system. In SRM, a clustering algorithm is conceived to filter out malicious feedbacks, and similar users are discovered by calculating the cosine value of the angle between two QoS vectors from the same service, and a feedback deviation based algorithm is applied to get user credibility. After that, service reputation is updated on the basis of user credibility and honest feedbacks.

In the model, quality attributes are used to measure the similarity between users. We describe the service quality monitored by u_i with a vector $\overline{Q_i} = (Q_{i,1}, Q_{i,2}, \cdots Q_{i,m})$. There are two kinds of quality attributes. Some attributes are proportional to quality, and that means the higher the value, the higher the quality. Some attributes are opposite, and that means the higher the value, the lower the quality. Here, q^+ denotes the set of the former attributes, and q^- denotes the later. Thus, if a quality attribute is a member of q^+, we normalize $Q_{i,j}$ as formula (1). Otherwise, we normalize it as formula (2). Let $\overline{q_i} = (q_{i,1}, q_{i,2}, \cdots q_{i,m})$ represent the normalized QoS vector.

$$q_{i,j} = \frac{Q_{i,j} - \min_{\forall k} Q_{k,j}}{\max_{\forall k} Q_{k,j} - \min_{\forall k} Q_{k,j}} \tag{1}$$

$$q_{i,j} = \frac{\max_{\forall k} Q_{k,j} - Q_{i,j}}{\max_{\forall k} Q_{k,j} - \min_{\forall k} Q_{k,j}} \tag{2}$$

where $\min_{\forall k} Q_{k,j}$ is the minimum value and $\max_{\forall k} Q_{k,j}$ is the maximum value of the corresponding attribute. Besides, if $\max_{\forall k} Q_{k,j} = \min_{\forall k} Q_{k,j}$, we set $q_{i,j}$ to 1.

IV. User Credibility Evaluating

A. Malicious Feedbacks Distinguishing

In the heterogeneous and loose-coupled circumstance, most service-oriented computing systems are so large and open that consumers can arbitrarily evaluate services, such as eBay and Amazon. So there is a mass of malicious feedbacks which play a bad effect on the measurement of services reputation. In current research, there are two typical kinds of malicious feedbacks:

1) Malicious High Feedbacks: feedbacks are far higher than the actual performance of a service;
2) Malicious Low Feedbacks: feedbacks are far lower than the actual performance of a service.

In a slot, for the same service, honest users will provide feedbacks similar to the actual performance, but the situation is different of malicious users. We also divide malicious users into two groups:

1) General Malicious Users: users providing random feedbacks to services individually;
2) Collusive Malicious Users: a group of users providing malicious high or low feedbacks to services consistently.

Therefore, we apply data-mining techniques in this situation to discover malicious feedbacks. Firstly, algorithm 1 is conceived to discover various existing clusters of feedbacks on service s_k in the slot Δt_{j+1}. Time complexity of this algorithm in worst case is $O(\frac{sizeof(feedback)}{d})$, where $sizeof(feedback)$ denotes the number of feedbacks collected in current slot, and d is a distance threshold. The distance between a feedback and a cluster is defined as the difference between the feedback and the average value of feedbacks in the cluster. Let m_i represent the average value of feedbacks in the i_{th} cluster. After discovering clusters of feedbacks, we compare m_i with the reputation of service s_k at time t_j. $R_{s_k}^{t_j}$ is a function of user feedbacks, and the computing method will be introduced in section V. Given an explicit offset θ, we have reason to believe that all feedbacks in cluster Z_i are malicious if m_i is not within $[R_{s_k}^{t_j} - \theta, R_{s_k}^{t_j} + \theta]$. If $m_j > R_{s_k}^{t_j} + \theta$, these feedbacks are considered as malicious high feedbacks, and if $m_j < R_{s_k}^{t_j} - \theta$, they are considered as malicious low feedbacks. Then malicious feedbacks can be filtered out, and users who provide them are malicious.

Algorithm 1: Clusters of Feedbacks Discovery Algorithm, CFDA

Input: a set of feedbacks $F = \{f_{u_l \to s_k}^{\Delta t_j}\}$, d.

Output: clusters of feedbacks $Z = \{Z_1, Z_2, Z_3, ...\}$

initiate a cluster $Z_1 = \{F[1]\}$, $Z = \{Z_1\}$;
for i = 1 to $sizeof(F)$ **do** /* $sizeof(F)$ denotes the number of elements in F */
  for j = 1 to $sizeof(Z)$ **do** /* $sizeof(Z)$ denotes the number of elements in Z */
    if $dis(F[i], Z[j]) < d$ **then**
    /* $dis(F[i], Z[j])$ denotes the distance between $F[i]$ and $Z[j]$ */
      classify $F[i]$ as a member of cluster $Z[j]$;
      break;
    else if j == $sizeof(Z)$ **then**
      define a new cluster $Z[j+1]$ as a new member of Z;
      classify $F[i]$ as a member of cluster $Z[j+1]$;
      break;
    end if
    j++;
  end for
  i++;
end for
return $Z = \{Z_1, Z_2, Z_3, ...\}$;

B. Computing User Credibility

Even honest users cannot assure the rationality of their feedbacks. Different user feedbacks have different weights in predicting service reputation in term of their credibility. In this paper, user credibility is classified into two classes: local user credibility and global user credibility. The following are their definitions.

Definition 5. Local User Credibility local user credibility means the ability to provide rational feedbacks on a specific service. Let $Cr(u_l, s_k)_L^{t_j}$ represent the local credibility of u_l for s_k at time t_j.

Definition 6. Global User Credibility global user credibility means the ability to provide rational feedbacks on any web service. Let $Cr(u_l)_G^{t_j}$ signify the global credibility of u_l at time t_j.

Intuitively, local user credibility is a function of the difference between the user feedback and the actual service reputation. But the actual service reputation is difficult to acquire. So, we obtain user credibility by computing the deviation of his feedback from other similar users'. Here, the similarity of two users who enjoy from the same service is measured by the similarity of two normalized QoS vectors. The following formula is used to achieve a set of users who are similar to u_l.

$$sim(u_l) = \{u_i \mid \cos(\overline{q_l}, \overline{q_i}) < \sigma, l \neq i\} \tag{3}$$

where σ is a specific threshold, and the range of its value is 0 to 1.

After obtaining similar users, SRM computes the local credibility. Let $Cr(u_l, s_k)_L^{\Delta t_j}$ represent the local credibility of u_l for s_k in the slot Δt_j, and it can be calculated as formula (4).

$$Cr(u_l, s_l)_L^{\Delta t_j} = 1 - \sum_{i=1}^{n} \frac{(f_{u_i \to s_k}^{\Delta t_j} - f_{u_l \to s_k}^{\Delta t_j})^2}{n} \tag{4}$$

where n is the number of users similar to u_l, and u_i is a member of $sim(u_l)$.

Formula (4) just considers feedbacks in a slot, and user credibility is a continuous variable based on time. So we can obtain the local credibility at some point according to $Cr(u_l, s_k)_L^{\Delta t_j}$ in all previous periods. The following formula is used to obtain $Cr(u_l, s_k)_L^{t_j}$, which means the value at time t_j.

$$Cr(u_l, s_k)_L^{t_j} = \begin{cases} \gamma & j = 0 \\ Cr(u_l, s_k)_L^{t_{j-1}} + \alpha^j \times (Cr(u_l, s_k)_L^{\Delta t_j} - Cr(u_l, s_k)_L^{t_{j-1}}) & j > 0 \end{cases} \tag{5}$$

where γ is the initial local user credibility, j is an integer, α^j is an impact factor which decides how $Cr(u_l, s_k)_L^{\Delta t_j}$ affects overall local user credibility. Finally, if $Cr(u_l, s_k)_L^{t_j} < 0$, we set it to 0. If $Cr(u_l, s_k)_L^{t_j} > 1$, we set it to 1.

In formula (5), α^j is an important factor and the definition of α^j should take into account two aspects: the timing of user credibility evaluation and the difference between deterioration and improvement. So we let δ_1^j represent $Cr(u_l, s_k)_L^{\Delta t_j} - Cr(u_l, s_k)_L^{t_{j-1}}$, and define α^j as:

$$\alpha^j = F_1 + F_2 = \begin{cases} \dfrac{1}{a \times t_j + b} + \eta \times (e^{|\delta_1^j|} - 1) & \delta_1^j \leq 0 \\ \dfrac{1}{a \times t_j + b} + \eta \times \delta_1^j & \delta_1^j > 0 \end{cases} \tag{6}$$

where F_1 is a function of time, F_2 is a function of fluctuation, a, b and η are parameters.

As for the definition of global user credibility, we can achieve it based on the weight sum of the local credibility for every service, which is shown in formula (7).

$$Cr(u_l)_G^{t_j} = \sum_k \rho_k \times Cr(u_l, s_k)_L^{t_j} \tag{7}$$

where ρ_k is the weight of s_k and it is decided by the detail information of the specific service.

At last, local user credibility and global user credibility are used to achieve user credibility, which is illustrated as follow:

$$Cr(u_l, s_k)^{t_j} = \varphi \times Cr(u_l, s_k)_L^{t_j} + (1 - \varphi) \times Cr(u_l)_G^{t_j} \tag{8}$$

where φ is the weight of local user credibility. The more feedbacks on service s_k from user u_l is received before time t_j, the higher the weight of local user credibility is. So we define φ as:

$$\varphi = \begin{cases} 1 & N \geq N_{\min} \\ \dfrac{N}{N_{\min}} & N < N_{\min} \end{cases} \tag{9}$$

where N is the number of feedbacks on s_k from u_l received by SRM so far, $N_{\min}$ is a threshold which can be obtained according to Chernoff Boundaries Theorem [15].

Given confidence μ and acceptable calculation error ε, $N_{\min}$ can be calculated by using

$$N_{\min} = -\frac{1}{2\varepsilon^2} \ln \frac{1-\mu}{2} \tag{10}.$$

When the value of μ is closer to 1, the value of $N_{\min}$ is larger.

The procedure of evaluating user credibility is summarized to algorithm 2. Time complexity of this algorithm in worst case is $O(num(s) \times n)$, where $num(s)$ is the number of all web services.

Algorithm 2: User Credibility Evaluation Algorithm, UCEA

Input: $\{f_{u_i \to s}^{\Delta t_{j+1}}\}$, $\{\overline{q_i}\}$, $Cr(u_l, s_k)_L^{\Delta t_j}$, $Cr(u_l)_G^{t_j}$ and other parameters.
Output: $Cr(u_l, s_k)^{t_{j+1}}$, $Cr(u_l, s_k)_L^{\Delta t_{j+1}}$, $Cr(u_l)_G^{\Delta t_{j+1}}$

discover a set of users similar to u_l according to formula (3);
for k = 1 to $num(s)$ **do** /* $num(s)$ denotes the number of all web services*/
  $Cr(u_l, s_k)_L^{\Delta t_{j+1}} = 1 - \sum_{i=1}^{n} \frac{(f_{u_i \to s_k}^{\Delta t_{j+1}} - f_{u_l \to s_k}^{\Delta t_{j+1}})^2}{n}$;
  $\delta_1^j = Cr(u_l, s_k)_L^{\Delta t_j} - Cr(u_l, s_k)_L^{t_{j-1}}$;
  if $\delta_1^j \leq 0$ **then**
    $\alpha^{j+1} = \frac{1}{a \times t_{j+1} + b} + \eta \times (e^{|\delta_1^{j+1}|} - 1)$;
  else then
    $\alpha^{j+1} = \frac{1}{a \times t_{j+1} + b} + \eta \times \delta_1^{j+1}$;
  end if
  $Cr(u_l, s_k)_L^{t_{j+1}} = Cr(u_l, s_k)_L^{t_j} + \alpha^{j+1} \times \delta_1^{j+1}$;
  if $Cr(u_l, s_k)_L^{t_{j+1}} < 0$ **then**
    $Cr(u_l, s_k)_L^{t_{j+1}} = 0$;
  else if $Cr(u_l, s_k)_L^{t_{j+1}} > 1$ **then**
    $Cr(u_l, s_k)_L^{t_{j+1}} = 1$;
  end if
end for
$Cr(u_l)_G^{t_{j+1}} = \sum_k \rho_k \times Cr(u_l, s_k)_L^{t_{j+1}}$;
if $N \geq N_{\min}$ **then**

$\quad\varphi = 1$;
else then
$\quad\varphi = \dfrac{N}{N_{\min}}$;
end if
$Cr(u_l, s_k)^{t_{j+1}} = \varphi \times Cr(u_l, s_k)_L^{t_{j+1}} + (1-\varphi) \times Cr(u_l)_G^{t_{j+1}}$;
return $Cr(u_l, s_k)^{t_{j+1}}$, $Cr(u_l, s_k)_L^{\Delta t_{j+1}}$, $Cr(u_l)_G^{\Delta t_{j+1}}$;

V. SERVICE REPUTATION PREDICTING

Service reputation is considered as the only non-functional criterion for selecting satisfactory items. In order to simplify the model but without losing accuracy, if a user provides more than one feedback on a service in the same time slot, only the latest feedback is taken into account. In this model, service reputation is measured by the weighted average of honest feedbacks, and the weight is transformed from user credibility. Formula (11) is used to compute $R_{s_k}^{\Delta t_j}$ that means the reputation of service s_k in the slot Δt_j.

$$R_{s_k}^{\Delta t_j} = \sum_i w_{u_i \to s_k}^{\Delta t_j} f_{u_i \to s_k}^{\Delta t_j}, \; w_{u_i \to s_k}^{\Delta t_j} = \frac{Cr(u_i, s_k)^{t_j}}{\sum_m Cr(u_m, s_k)^{t_j}} \tag{11}$$

where $w_{u_l \to s_k}^{\Delta t_j}$ is the weight of the feedback on s_k from u_l in the slot Δt_j.

Like user credibility, service reputation is also a continuous variable. Thus, overall service reputation can be obtained according to $R_{s_k}^{\Delta t_j}$ in previous periods. We compute service reputation at time t_j as follow.

$$R_{s_k}^{t_j} = \begin{cases} \tau & j = 0 \\ R_{s_k}^{t_{j-1}} + \beta^j \times (R_{s_k}^{\Delta t_j} - R_{s_k}^{t_{j-1}}) & j > 0 \end{cases} \tag{12}$$

where τ is the initial reputation of service s_k, β^j is an impact factor which decides how $R_{s_k}^{\Delta t_j}$ affects overall service reputation. Finally, if $R_{s_k}^{t_j} < 0$, we set it to 0. If $R_{s_k}^{t_j} > 1$, we set it to 1.

In formula (12), the definition of β^j is similar to α^j. The definition of β^j should consider two aspects: the timing of service assessment and the difference between deterioration and improvement. So we let δ_2^j represent $R_{s_k}^{\Delta t_j} - R_{s_k}^{t_{j-1}}$, and define β^j as:

$$\beta^j = \lambda \times (F_1' + F_2') = \begin{cases} \lambda \times (\dfrac{1}{a' \times t_j + b'} + \eta' \times (e^{|\delta_2^j|} - 1)) & \delta_2^j \le 0 \\ \lambda \times (\dfrac{1}{a' \times t_j + b'} + \eta' \times \delta_2^j) & \delta_2^j > 0 \end{cases} \tag{13}$$

where F_1' is a function of time, F_2' is a function of fluctuation, λ means the importance of $(F_1' + F_2') \times \Delta_2^j$ in overall service reputation, a', b' and η' are parameters. In addition, the definition of λ is similar to that of φ. The more honest feedbacks on s_k is received by SRM in the slot Δt_j, the higher the value of λ is. So we define λ as:

$$\lambda = \begin{cases} 1 & N' \ge N_{\min}' \\ \dfrac{N'}{N_{\min}'} & N' < N_{\min}' \end{cases} \tag{14}$$

where N' is the number of honest feedbacks on s_k received by SRM, $N_{\min}'$ is a threshold which can also be computed according to Chernoff Boundaries Theorem.

The procedure of predicting service reputation is summarized to algorithm 3. Time complexity of this algorithm in worst case is $O(num(u))$, where $num(u)$ is the number of all users.

Algorithm 3: Service Reputation Prediction Algorithm, SRPA

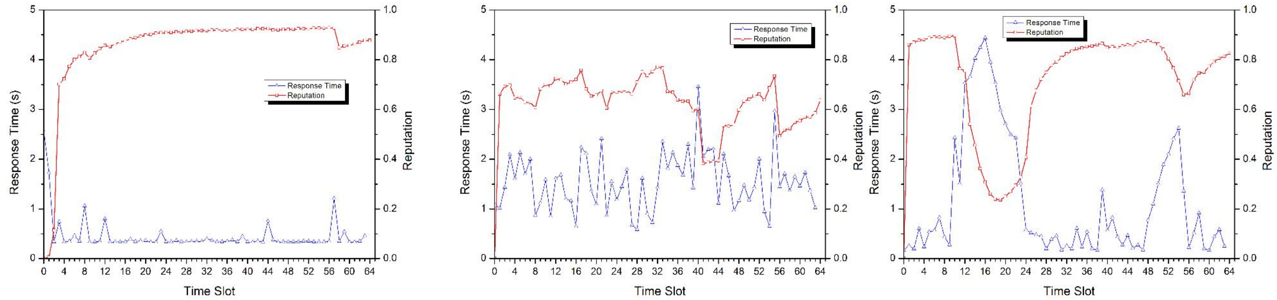
Input: $\{f_{u_i \to s}^{\Delta t_{j+1}}\}$, $\{Cr(u_i, s_k)^{t_{j+1}}\}$, $R_{s_k}^{t_j}$ and other parameters.
Output: $R_{s_k}^{t_{j+1}}$

$R_{s_k}^{\Delta t_{j+1}} = \sum_i \dfrac{Cr(u_i, s_k)^{t_{j+1}}}{\sum_m Cr(u_m, s_k)^{t_{j+1}}} f_{u_i \to s_k}^{\Delta t_{j+1}}$;
$\delta_2^{j+1} = R_{s_k}^{\Delta t_{j+1}} - R_{s_k}^{t_j}$;
if $N' \ge N_{\min}'$ **then**
$\quad\lambda = 1$;
else then
$\quad\lambda = \dfrac{N'}{N_{\min}'}$;
end if
if $\delta_2^{j+1} \le 0$ **then**
$\quad\beta^{j+1} = \lambda \times (\dfrac{1}{a' \times t_{j+1} + b'} + \eta' \times (e^{|\delta_2^{j+1}|} - 1))$;
else then
$\quad\beta^{j+1} = \lambda \times (\dfrac{1}{a' \times t_{j+1} + b'} + \eta' \times \delta_2^{j+1})$;
end if
$R_{s_k}^{t_{j+1}} = R_{s_k}^{t_j} + \beta^{j+1} \times \delta_2^{j+1}$;
return $R_{s_k}^{t_{j+1}}$;

VI. EXPERIMENTAL RESULTS AND DISCUSSIONS

To evaluate the performance of our user-credibility based service reputation management model, several experiments are conducted with WS-DREAM dataset provided by Zhang Y, Zheng Z and Lyu M R [16]. The dataset is stringency and authority, and includes 4,532 actual web services and QoS data monitored by 142 users in 64 time slots. In addition, we assume that users can be classified into honest users, general malicious users and collusive malicious users. Feedbacks of them are randomly generated based on following different strategies:

1) Feedbacks from honest users are similar to the actual service reputation;
2) Feedbacks from general malicious users are generated randomly;



(a) (b) (c)

Figure 2. The comparison between response time and service reputation of the three services: (a) a typical service with stable data; (b) a typical service with data which has great volatility in several consecutive slots; (c) a typical service with data which is changing all the time.

3) Feedbacks from a group of collusive malicious users are far lower or higher than the actual service reputation consistently.

The experiment is comprised of two parts: the accuracy of predicting service reputation and the effectiveness of distinguishing malicious feedbacks. Table 1 shows the description and value of every parameter in our model, and some parameters can change in different environment.

TABLE I. THE DESCRIPTION AND VALUE OF EVERY PARAMETER

Parameter	*Description*	*Value*
σ	a threshold to compute sim(u)	0.1
γ	the initial local credibility value	0.8
a	a parameter to compute α^j	0.8
b	a parameter to compute α^j	10
η	a parameter to compute α^j	0.5
τ	the initial reputation value	0.5
a’	a parameter to compute β^j	0.8
b’	a parameter to compute β^j	10
η’	a parameter to compute β^j	0.5
ε	acceptable calculation error to compute $N_{\min}$	0.2
μ	confidence value to compute $N_{\min}$	0.9
θ	a threshold to distinguish malicious feedbacks	0.2
d	a distance threshold used in the algorithm 1	0.1

A. *The Accuracy of Predicting Service Reputation*

After analyzing the data of quality attributes of every service in the dataset, services can be classified into three categories in term of data stability: services with stable data; services with data which has great volatility in several consecutive slots; services with data which is changing all the time. To evaluate the reasonableness of our model, we observe whether service reputation can faithfully reflect the performance. Three typical services are respectively chosen from above three categories. The comparison between response time and service reputation of the three services in every slot is made in this experiment.

Figure 2 shows the consistency of the fluctuation of respond time and service reputation. In the three categories, service reputation always changes with respond time. When quality performance bottleneck appears in a service, reputation drops rapidly so that the system filters out this service due to the low reputation. So it indicates that our model can precisely and timely reflect the fluctuation of service performance.

To evaluate the accuracy of predicting service reputation, we compare our model and models proposed in [7] and [17] (TMS and CUSUM). Mean absolute percentage error is the evaluation criterion in this scenario. Following is the definition of it:

Definition 7. Mean Absolute Percentage Error (MAPE) MAPE is defined as:

$$MAPE = \frac{\frac{1}{m} \times \sum_{k=1}^{m} R_{s_k}^{\Delta t_j} - F^{\Delta t_j}}{F^{\Delta t_j}} \times 100\% \qquad (15)$$

where m is the number of all web services, $F^{\Delta t_j}$ is the ideal average reputation of all services equaling to 0.5. The actual reputation of every service can be used to calculate the error precisely, but it is impossible to acquire. So the difference between the ideal average reputation and predicted reputation is used to evaluate the accuracy of different methods, and the ideal average reputation is set to a middle value [17].

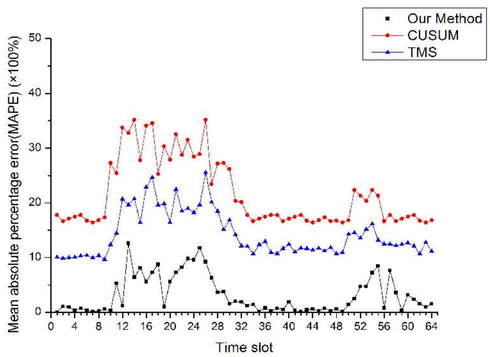


Figure 3. Evaluating the accuracy of three methods.

On the basis of services whose QoS data has great volatility in several consecutive slots, we compute MAPE of the three methods: Our method, CUSUM and TMS. The result of this experiment is shown in Figure 3. Obviously, MAPE computed with our method are lower than others, so the accuracy of our method is higher. However, when quality performance changes quickly and suddenly, MAPE still increase quickly. So the problem is an issue in our future work. On the whole, our method of predicting service reputation is better than other approaches.

B. The Effectiveness of Distinguishing Malicious Feedbacks

In this experiment, we assume three different distributions of general malicious users and collusive malicious users as following: the percentage of former is 30% and the percentage of latter is 10%, the percentage of former is 20% and the percentage of latter is 20%, the percentage of former is 10% and the percentage of latter is 30%. At the same time, the initial credibility of them is set to 0.8 so that malicious users can only be detected by the mechanism of distinguishing malicious feedbacks. We note down the quantity of general malicious users and collusive malicious users detected by the system in every slot in the three cases. The result is shown in Figure 4.

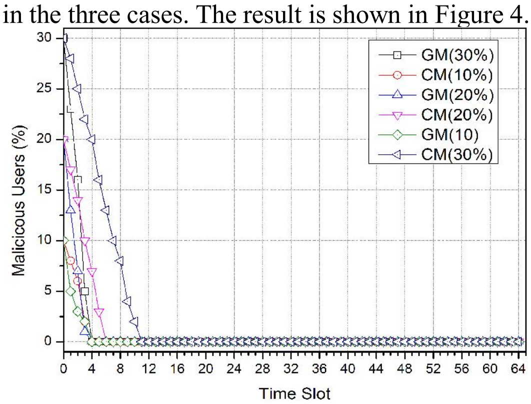


Figure 4. Results of distinguishing malicious feedbacks.

As we can see in Figure 4, relative to the same percentage of collusive malicious users, general malicious users are discovered more quickly. For example, general malicious users that account for 30 percent of the total only need 4 slots to be detected, but collusive malicious users need 11 slots. The reason is that the randomness and the individuality are the two basic features of general malicious users, so the difference between their feedbacks and the actual reputation value is relatively obvious. But the feedback behaviors of collusive malicious users are relatively hard to be detected, and the results show the efficiency of our method is higher.

VII. CONCLUSION

In this paper, we propose a user-credibility based reputation management model for service selection. The clustering algorithm is first conceived to filter out malicious feedbacks. Then, the concept of user credibility is introduced, and a feedback deviation based algorithm is proposed to evaluate user credibility considering service QoS similarity. Finally, a service reputation prediction algorithm on the basis of user credibility and honest feedbacks is applied in the model. The accuracy and efficiency is evaluated and validated by simulation experiments. However, there are some issues we have not taken into account in our model. The mechanism of distinguishing collusive malicious feedbacks needs to be better improved.

ACKNOWLEDGMENT

This work is supported by National Natural Science Foundation of China under Grants 61272531, 61070161, 61070158, 61003257 and 90912002, National Basic Research Program of China under Grants No.2010CB328104, National Key Technology R&D Program under Grants No.2010BAI88B03, Ph.D. Programs Foundation of Ministry of Education of China under Grants No.200802860031, Natural Science Foundation of Jiangsu Province under Grants No.BK2008030, National Science and Technology Major Project of China under Grants No.2009ZX 03004-004-04, Program of Jiangsu Province Key Laboratory of Network and Information Security under Grants No.BM2003201.

REFERENCE

[1] Ran S. A model for web services discovery with QoS[J]. ACM Sigecom exchanges, 2003, 4(1): 1-10.

[2] Lee K, Jeon J, Lee W, et al. Qos for web services: Requirements and possible approaches[J]. W3C Working Group Note, 2003, 25: 1-9.

[3] Dragoni N. A survey on trust-based web service provision approaches[C]//Dependability (DEPEND), 2010 Third International Conference on. IEEE, 2010: 83-91.

[4] Zeng L, Benatallah B, Dumas M, et al. Quality-driven web services composition[J]. 2003.

[5] Xu Z, Martin P, Powley W, et al. Reputation-enhanced qos-based web services discovery[C]//Web Services, 2007. ICWS 2007. IEEE International Conference on. IEEE, 2007: 249-256.

[6] Yau S S, Yin Y. Qos-based service ranking and selection for service-based systems[C]//Services Computing (SCC), 2011 IEEE International Conference on. IEEE, 2011: 56-63.

[7] Conner W, Iyengar A, Mikalsen T, et al. A trust management framework for service-oriented environments[C]//Proceedings of the 18th international conference on World wide web. ACM, 2009: 891-900.

[8] Limam N, Boutaba R. Assessing software service quality and trustworthiness at selection time[J]. Software Engineering, IEEE Transactions on, 2010, 36(4): 559-574.

[9] Vu L H, Hauswirth M, Aberer K. QoS-based service selection and ranking with trust and reputation management[J]. On the Move to Meaningful Internet Systems 2005: CoopIS, DOA, and ODBASE, 2005: 466-483.

[10] Maximilien E M, Singh M P. An ontology for Web service ratings and reputations[C]//3rd Workshop on Ontologies in Agent Systems. 2003: 25.

[11] Maximilien E M, Singh M P. A framework and ontology for dynamic web services selection[J]. Internet Computing, IEEE, 2004, 8(5): 84-93.

[12] Maximilien E M, Singh M P. Toward autonomic web services trust and selection[C]//Proceedings of the 2nd international conference on Service oriented computing. ACM, 2004: 212-221.

[13] Wang S, Sun Q, Zou H, et al. Reputation measure approach of web service for service selection[J]. Software, IET, 2011, 5(5): 466-473.

[14] Majithia S, Ali A S, Rana O F, et al. Reputation-based semantic service discovery[C]//Enabling Technologies: Infrastructure for Collaborative Enterprises, 2004. WET ICE 2004. 13th IEEE International Workshops on. IEEE, 2004: 297-302.

[15] Shamarova E. Chernoff's theorem for evolution families[J]. arXiv preprint arXiv:0706.4079, 2007.

[16] Zhang Y, Zheng Z, Lyu M R. WSPred: A Time-Aware Personalized QoS Prediction Framework for Web Services[C]//Software Reliability Engineering (ISSRE), 2011 IEEE 22nd International Symposium on. IEEE, 2011: 210-219.

[17] Wang S, Zheng Z, Sun Q, et al. Evaluating Feedback Ratings for Measuring Reputation of Web Services[C]//Services Computing (SCC), 2011 IEEE International Conference on. IEEE, 2011: 192-199.