Granularity-based Uncertain QoS Partitioning for Web Service Reliability

Guobing Zou[1,2], Zhimin Zhou[1], Mei Zhao[1], Sen Niu[1], Yanglan Gan[3]

[1] School of Computer Engineering and Science, Shanghai University, China
[2] Shanghai Institute for Advanced Communication and Data Science, Shanghai University, China
[3] School of Computer Science and Technology, Donghua University, China
{gbzou, zmzhou, mzhao, sniu}@shu.edu.cn, ylgan@dhu.edu.cn

Abstract

The performance of Web services may fluctuate due to their invocations in dynamic environment. Thus, quality of service (QoS) is inherently uncertain when reflecting non-functional features of Web services. QoS has been considered as a significant criterion for selecting those functionally similar services. In this paper, we proposed a novel granularity-based partitioning model (GPM) that is applied to evaluate the reliability of Web services with the consideration of uncertain QoS via cloud model and coefficient of variance. Extensive experiments with 1,558,224 service invocation records have been conducted to validate the effectiveness of our proposed approach. The results demonstrate that it outperforms the state-of-the-art traditional ones.

Keywords: Web services, Uncertain QoS, Granularity partitioning, Cloud model, Coefficient of variance

1 Introduction

Web services are self-describing software components that can be advertised, located and invoked across the Internet using a set of standards, including SOAP, WSDL and UDDI [1]. Due to the effective integration of distributed, heterogeneous, and autonomous applications, there has been a large flow of Web services deployed over the Web. It can be widely applied for cloud computing and its applications [2-4]. Therefore, as more and more Web services are registered and invoked on the Internet, many enterprises and organizations are willing to outsource part of their business processes. As a consequence, they only need to focus on their core activities. However, it may occur that there are multiple service providers competing to offer those Web services with the same functionality, while they share the different quality of service (QoS) [5].

Multi-dimensional service QoS model characterizes non-functional properties that are inherent to Web services, such as execution price, response time, reputation, reliability and availability [6]. It is often utilized to QoS-aware Web service composition, where a set of correlative services are integrated together to create a new value-added composite service to satisfy a complex request with many tasks. In most cases, a bunch of service candidates with the same functionality can be invoked for a specific task, although they have totally different QoS. Thus, at service run time, a concrete Web service is selected from service candidates by differentiating their QoS for each outsourced task. By doing so, we can set up a composite service with optimal aggregated QoS, while multiple global QoS constraints can be satisfied [7-8]. Aiming at selecting a qualified Web service, much research has been widely exerted for QoS-aware service selection and service composition [6-9]. In these works, composition optimization approaches assume that the quality delivered by service providers does not change over time, which is known as tentative QoS.

However, the network environment is dynamically changing for the invocation of Web services. QoS values of Web services may significantly vary due to the update of server hardware/software or workload changes. Moreover, some of the selected services may suddenly become unavailable at run-time, while new service candidates may be launched [10-11]. We observe that QoS parameters have played a major role in determining the success or failure of the composition applications [12]. For this reason, given a set of functionally similar Web services, the reliability of uncertain QoS of Web services has a strong impact on automatic service selection and composition. Recently, some efforts have been done to take the uncertainty of QoS into account [5, 12-13]. Benouaret et al. represented each QoS attribute of a Web service with uncertainty using a possibility distribution [5]. Wang et al. employed cloud model [12] to compute the QoS uncertainty. A novel approach is investigated for

*Corresponding Author: Yanglan Gan; E-mail: ylgan@dhu.edu.cn
DOI: 10.3966/160792642018081904007

computing the service skyline from uncertain QoWS [13]. Unfortunately, most of these works take entire transaction log as a union, while they do not make further analysis for each transaction. This leads to great deviation from its truth. Therefore, how to accurately and efficiently assess the reliability of uncertain QoS for a Web service has become a research challenge.

To solve above research challenge, in this paper we proposed a novel approach to evaluate the reliability of Web services. To the best of our knowledge, this is the first attempt to study the evaluation of Web service reliability with the consideration of QoS uncertainty using different partitioning granularities. The main contributions of our paper are threefold as below.

First, we propose a granularity-based partitioning model (GPM) for uncertain QoS, which divides uncertain QoS of a Web service into multiple independent regions to evaluate the reliability of Web services.

Second, two algorithms based on cloud model and coefficient of variance are respectively integrated into our model (GPM) to evaluate the reliability of each partitioned QoS region of a Web service.

Third, we have conducted extensive experiments to compare our approach with the state-of-the-art method [12] on a large-scale publicly realeased benchmarking dataset with more than 1500000 uncertain QoS transaction logs. The experimental results validate the effectiveness of our proposed approach for the evaluation of the reliability of QoS uncertainty of Web services.

Under the desired motivations, we have implemented a prototype system for Web service reliability with QoS uncertainty and performed extensive experiments from a dataset called WS-DREAM [15] with 1,558,224 service invocation records of uncertain QoS. Compared to existing traditional approach [12], the experimental results verify the effectiveness and efficiency of our proposed approach when evaluating the reliability of Web services under the characteristics of QoS uncertainty.

The rest of this paper is organized as follows. Section 2 reviews related work. Section 3 provides the preliminaries for computing the QoS uncertainty of Web services. Our approach is proposed for Web service reliability in Section 4. We show experimental results in Section 5. Finally, Section 6 concludes the paper.

2 Preliminaries

In this section, we provide fundamental knowledge on QoS uncertainty evaluation, including cloud model and coefficient of variance.

2.1 Cloud Model

Cloud model [14] is a widely used classic technique of uncertainty transition between a linguistic term of a qualitative concept and its numerical representation. A cloud model can be defined as below.

Definition 1 (Cloud model). Let U be the set as the universe of discourse, and C a qualitative concept associated with U. The membership degree of quantitative numerical representation x in U to the concept C, $\mu(x) \in [0,1]$, is a random number with a stable tendency, that is as in (1):

$$\mu : U \to [0,1], \forall x \in U, x \to \mu(x) \quad \textbf{(1)}$$

The distribution of x in the universe of discourse U is called cloud $C(X)$, and x is called a cloud drop. The overall characteristics of cloud model can be reflected by three numerical characteristics, i.e., expected value (E_X), entropy (E_n) and hyper-entropy (H_e). As shown in Figure 1, we illustrate the three numerical characteristics of cloud model by the number of 3000 cloud drops. With the computation, we have $E_x = 0$, $E_n = 1$, and $H_e = 0.1$.

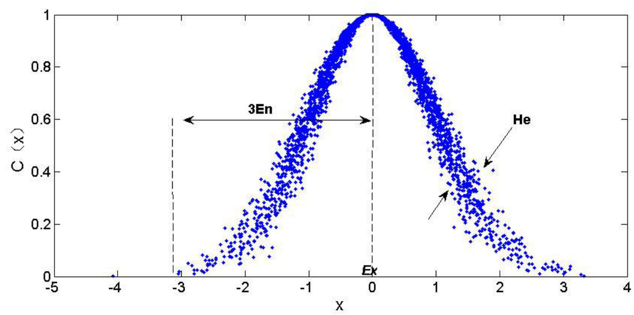


Figure 1. Three numerical characteristics of cloud model with 3000 cloud drops

In the discourse universe, E_x is the position corresponding to the center of the cloud gravity, whose elements are fully compatible with the linguistic concept. Moreover, E_n is a measure of the concept coverage, i.e., a measure of the fuzziness, which indicates how many elements could be accepted to the qualitative linguistic concept. Similarly, H_e is a measure of the dispersion on these cloud drops, which can also be considered as the entropy of E_n. With the combination of these three measure features, the vector $NC = \{E_x, E_n, H_e\}$ is called the eigenvector of cloud model.

2.2 Coefficient of Variance

Coefficient of variance is another measurement that is of highly correlative with our QoS uncertainty evaluation. It reflects the average ratio of standard deviation. Especially, when we compare the measures of dispersion of two datasets, if their measurement scopes have many differences or they have different dimensions, it is inappropriate to make comparisons with the direct use of standard deviation. To fairly eliminate the differences on measurement scale and dimension impact, the coefficient of variation (CV) can be effectively applied in such case.

Definition 2 (Coefficient of variance). Given a dataset with the size of numbers. the coefficient of variation is calculated by

$$CV = \sigma / \mu \tag{2}$$

$$\sigma = \sqrt{\sum (x_i - \mu)^2 / (n-1)} \tag{3}$$

$$\mu = \left(\sum x_i\right) / n \tag{4}$$

Where σ is the standard deviation and μ is the mean of the dataset respectively. Therefore, CV reflects the degree of dispersion of a given dataset. By means of this comparison, it not only computes the uncertainty by the size of its data value of the variable dispersion influence degree, but also it considers the size of the average value of the variable. In general, the larger the value of CV is, the more distributed the dataset holds.

3 Our Approach

In this section, we first focus on the understanding of computing QoS uncertainty of Web service by a set of formal definitions, and then we present a comprehensive framework that integrates the granularity partitioning model and two algorithms via cloud model and coefficient of variance for Web service reliability.

3.1 Problem Formulation

The uncertainty of a Web service can be observed from two aspects, including its functional and non-functional aspects. Here, we mainly focus on considering the latter one and it is defined as below.

Definition 3 (Uncertainty of service). A Web service s is 3-tuple s=<*I, O, Q*>, where <*I, O*> are the inputs and outputs for service functionality. Q represents the non-functional performance with uncertainty. $s.Q$ is denoted as the QoS of s.

Definition 4 (Uncertain service repository). An uncertain Web service repository S consists of a finite set of Web services, denoted as $S = \{s_1, s_2, \cdots, s_N\}$, where $\forall s_i \in S$ is a Web service with QoS uncertainty.

For the representation of non-functional performance of a Web service, we define the QoS criteria as below.

Definition 5 (QoS criteria). Given a Web service $s = \langle I, O, Q \rangle \in S$, its uncertain QoS $s.Q$ is aligned by a set of QoS attributes, $QS = \{q_1, q_2, \cdots, q_n\}$, where each q_i is used to represent one facet of non-functional values of s.

Multiple dimensional vector models the QoS criteria, such as execution price, response time, availability, etc. For a Web service s, an invocation results in a set of QoS values by $QS = \{q_1, q_2, \cdots, q_n\}$. Actually, it is partial of a service transaction log, which is defined as below.

Definition 6 (Transaction log). Given a service $s \in S$ and $QS = \{q_1, q_2, \cdots, q_n\}$, a transaction log t is a history record by the invocation of s for one time, denoted as $t = \{q_1(s), q_2(s), \cdots, q_n(s)\}$.

Note that in addition to regular QoS values, auxiliary information should also be a part of transaction log. Thus, we model service transaction log as a 5-tuple, $t = \{TID, IP, CountryCode, State, Q\}$, where *TID* is the transaction identifier of a Web service, *IP* represents the IP address of the user who invokes a service, *CountryCode* and *State* are the service requester's location information from two different granularities, and Q is the QoS values of a service invocation with $QS = \{q_1, q_2, \cdots, q_n\}$.

Example 1. Consider a Web service P that offers weather forecast service as shown in Table 1. The uncertain QoS performance of P is recorded by a series of transaction logs, which capture the actual QoS in practice. Transaction logs can be obtained from monitoring mechanisms. Here, we only consider 8 transaction logs $\{t_1, \cdots, t_8\}$, although the actual number should be much larger. Each transaction log t has its identifier, location information, IP, and response time.

Table 1. A set of transaction logs of a Web service P

Web Service: P (WSID16432)				
TID	Country Code	State	IP	RT (ms)
t1	US	Indiana	128.10.19.52	58
t2	US	Indiana	128.10.19.52	60
t3	US	Kansas	129.237.161.193	50
t4	US	Kansas	129.237.161.193	90
t5	BR	São Paulo	200.133.215.141	125
t6	BR	São Paulo	200.133.215.141	180
t7	KR	Qing zhou	210.125.84.15	75
t8	KR	Qing zhou	210.125.84.15	43

The dynamic environment causes the QoS uncertainty of its performance. This can be reflected by the fluctuation among different transaction logs. Thus, given a service $s \in S$ and a set of QoS criteria $QS = \{q_1, q_2, \cdots, q_n\}$, transaction logs of s consist of a finite set of Web services transactions, denoted as $T(s) = \{t_1, t_2, \cdots, t_m\}$, which indicates the QoS uncertainty of service s.

Definition 7 (Uncertain QoS of Web service). Given a Web service $s \in S$, a set of QoS criteria with n attributes $QS = \{q_1, q_2, \cdots, q_n\}$, and its transaction logs $T(s) = \{t_1, t_2, \cdots, t_m\}$, the QoS uncertainty of s can be formalized as a matrix M_{m*n}.

$$M(s)=\begin{bmatrix} t_1 \\ \vdots \\ t_m \end{bmatrix}=\left\{\begin{matrix} q_{11} & q_{12} & \cdots & \cdots & q_{1n} \\ \vdots & \ddots & & & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ q_{m1} & q_{m2} & \cdots & \cdots & q_{mn} \end{matrix}\right\} \quad \textbf{(5)}$$

Where m and n are the number of transaction logs of Web service and QoS criteria for each transaction log, respectively. That is, each row t_i stands for a transaction log and each column row q_j represents a set of values on the j^{th} QoS criterion across m transaction logs.

Example 2. Let's also take the Web service P as an example shown in Table 1, the uncertain QoS matrix of P is M with 8 rows and 5 columns. However, for the simplicity of further QoS uncertainty computation, we only show its response time by a dimension reduction. Thus it is $M(P)=[58,60,50,90,125,180,75,43]^T$.

Given a service s and its uncertain QoS matrix $M(s)$, we focus on how to accurately and efficiently compute the uncertainty of service s that leads to identify the reliability of the Web service.

3.2 The Framework of Our Approach

To accurately compute the QoS uncertainty, we observe that the idea of uncertain QoS partitioning in terms of the location information of their transaction logs could facilitate the evaluation of Web service reliability. Thus, we propose a novel approach that integrates a granularity-based partitioning model (GPM) to partition a QoS matrix into independent QoS transaction regions. By doing so, we further apply two uncertainty computation methods to make an evaluation on the reliability of a service. The overall framework of our approach is illustrated in Figure 2.

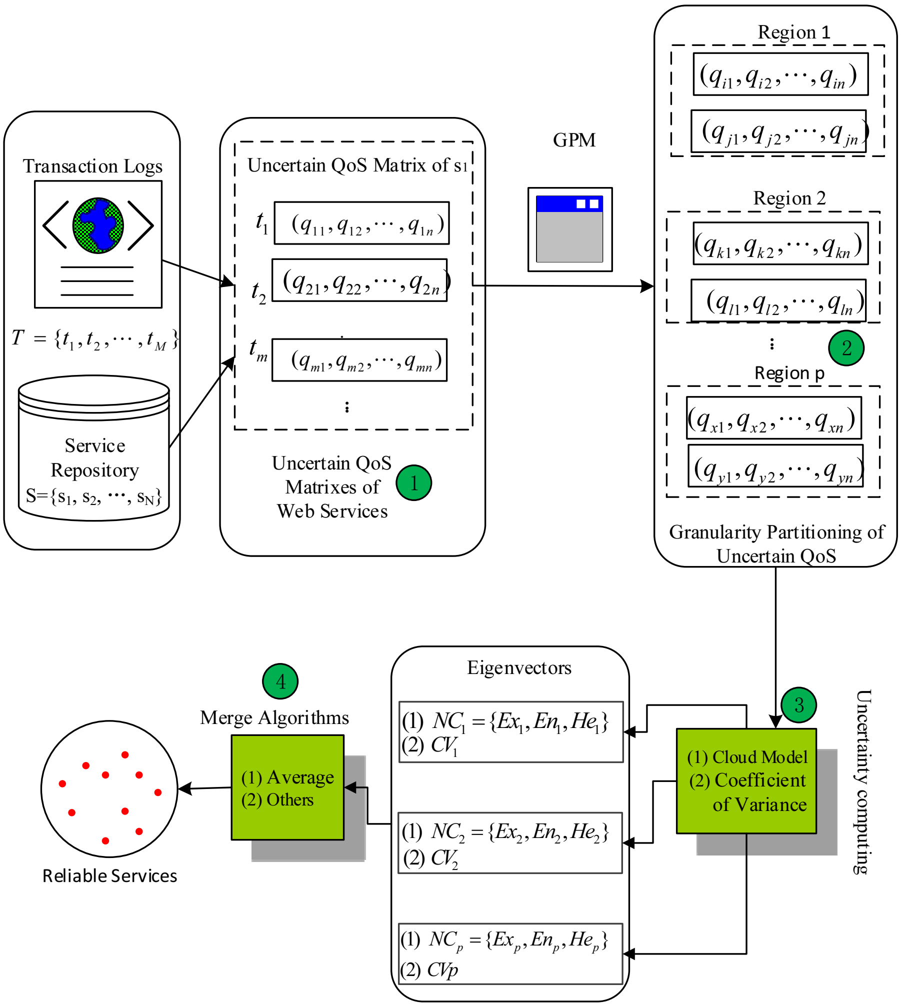


Figure 2. The framework of our approach for Web service reliability

Its input consists of N number of uncertain services in a service repository with their M number of associated transaction logs, and it outputs those reliable services of the service repository. More specifically, the framework goes through four crucial steps. First, we extract an uncertain QoS matrix M for each service in the Web service repository. Then, we make a partitioning on each service matrix M via GPM model to partition it into a set of independent QoS transaction regions $\{R_1, R_2, ..., R_p\}$. Subsequently, we apply two highly effective approaches (cloud model and coefficient of variance) for the uncertainty computation of each transaction region of each service. Finally, we identify the reliability of each Web service by merging the uncertainty results of p independent QoS transaction regions.

Note that, to achieve the goal of more accurately evaluating the reliabilty of Web services, the novelty of our approach is that we take uncertain QoS of every service transaction logs into consideration and partition them into serveral indepent regions from the perspective of granularity computation. Based on the above idea, a granularity-based partitioning model called GPM is presented and applied to partition the whole uncertain transaction logs of a Web service into different subsets with multiple uncertain QoS execution records, and then two kinds of uncertain computation strategies cloud model and coefficient of variance are integrated into our model to evaluate the reliability of each Web service. Finally, the experimental results validate the effectiveness of our proposed approach.

3.3 Granularity Partitioning Model of Uncertain QoS

Definition 8 (Granularity set). A granularity set G consists of a finite set of partitioning granularities, denoted as $G=\{g_1,g_2,\cdots\}$, where each $g \in G$ is a minimum unit of partitioning for an uncertain QoS matrix *M*.

Taking the service *P* in Table 1 as an example, we have $G=\{g_1,g_2\}$, where g_1 and g_2 are country and state.

We observe that the QoS uncertainty of a Web service has regional property, because transaction logs *T* can divided into several independent subset by the information of a requester who invoked that Web service. Based on this regional property, an uncertain QoS matrix of a service can be partitioned as follows.

Definition 9 (QoS granularity partitioning). Given an uncertain QoS matrix $M=\left[t_1,t_2,\cdots t_m\right]^T$ of a Web service $s \in S$, where each t_i $(1 \le i \le \mathrm{m})$ represents a transaction log, and a $g_j \in G=\{g_1,g_2,...\}$ is designated as partitioning granarity, then we partition M into p QoS independent transaction regions, $R=\{R_1,R_2,\cdots,R_p\}$, where $\bigcup_{i=1}^{k} R_i = M$ is satisfiable.

Form the above definition, we can observe that local information similarity ensures that their QoS values of those transaction logs within the same region leads to more accurate uncertainty evaluation because of fewer gaps during service invocations.

Example 3. Let us take the uncertain QoS matrix of Web service *P* in example 2 as an example, its matrix is $M(P)=\left[58,60,50,90,125,180,75,43\right]^T$ and we set the partitioning granularity as $g = CountryCode$. After the QoS partitioning, $M(P)$ is divided into three independent QoS regions, which is illustrated at the second level in Figure 3. More specifically, we partition it into $M(P)=\{R_1,R_2,R_3\}$, where $R_1 = \{58,60,50,90\}^T$, $R_2=\{125,180\}^T$, and $R_3=\{75,43\}^T$.

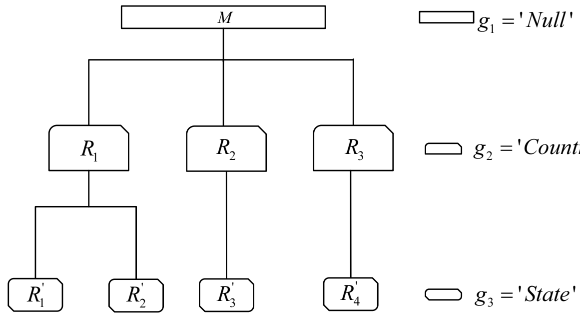


Figure 3. Granularity partitioning results of uncertain QoS of Web service *P*

In each QoS region, we can see that QoS values have less fluctuation. That means QoS values to some extent are more stable in a single QoS partitioning region, even though there are many differences between each other. However, another observation is that within the same QoS region, QoS values can be further partitioned into more stable independent regions, such as $R_1=\{58,60,50,90\}^T$.

Example 4. Based on the results of QoS regions in Example 3, we make a further partitioning with a smaller granularity $g = State$. After the partitioning, the matrix is divided into four QoS regions, $M(P)=\{R_1',R_2',R_3',R_4'\}$, where $R_1'=\{58,60\}^T$, $R_2'=\{50,90\}^T$, $R_3'=\{125,180\}^T$, and $R_4'=\{75,43\}^T$. We illustrate the partitioning results at the third level in Figure 3.

At the first level, without any granularity is taken to partition QoS matrix *M*, thus we calculate its QoS uncertainty by putting the entire transaction logs as a whole. As for the application of GPM model, we can further disperse its entire transaction logs into a set of independent QoS regions by using different granularity levels. We can conclude that the smaller the granularity we apply, the greater likelihood transaction logs within a partitioned QoS region is. Thus, it will result in more accurate for service reliability evaluation.

Here, we further explain why our model could be effectively applied to facilitate the computation of uncertain QoS that can better the evaluation of Web service reliability. We find that network distance of Web service between service provider and service requester is the crucial factor, which affects the non-functional performance when invoking a service on the Internet. Under this assumption, those requesters at the same region have the same network distance from the same Web service. Thus, the QoS measurements for these service requesters should be affected in the same condition. As a result, if a Web service can be reliably performed, the QoS values of transaction logs for service requesters from the same region should have less fluctuation.

Here, we apply the geographic location as a partitioning criterion. In realistic applications, however, our proposed QoS partitioning model can be applicable to any scenarios where the transaction logs can be divided into several independent transaction regions using QoS properties, such as CIDR aggregation of IP address.

3.4 Algorithms

In this section, we present two algorithms based on QoS granularity partitioning for the evaluation of Web service reliability via cloud model and coefficient of variance, respectively.

3.4.1 GPM-based Service Reliability by Cloud Model

Given a set of transaction logs T of a Web service, Algorithm 1 extracts the QoS matrix and partitions it into a set of independent QoS regions, and then applies cloud model to compute the QoS uncertainty. The pseudo code of Algorithm 1 is shown as below.

Example 5. Let's denote En and He of a Web service s as $NC(s)=\{En(s),He(s)\}$. Taking the QoS matrix M in Table 1 as an example, we respectively compute the reliability of service P by using three different partitioning granularities. When the granularity is set as NULL, $NC(s)=\{239.875,187.5271\}$. If we set the granularity as CountryCode, the QoS matrix M is partitioned into $R=\{R_1,R_2,R_3\}$, where $R_1=[58,60,50,90]^T$, $R_2=[125,80]^T$, and $R_3=[75,43]^T$. Thus, we have $NC(R_1)=\{15.9798,7.2328\}$, $NC(R_2)=\{34.4661,18.0163\}$, and $NC(R_3)=\{20.0530,10.4822\}$. With the combination of these results, we calculate the eigenvector of service s as $NC(s)=\{23.4993,11.9104\}$. If the parameters are set as λ= 25 and h=15, we draw a conclusion that applying pure cloud model the service s is identified as unreliable, while it is evaluated as a reliable one, if we make a QoS granularity partitioning.

3.4.2 GPM-based Service Reliability by Coefficient of Variance

As the reliability evaluation of Web service by cloud model, after the QoS partitioning to a QoS matrix, we apply coefficient of variance to compute the QoS uncertainty. The pseudo code is shown in Algorithm 2.

Algorithm 1. GPM-based uncertainty computation via cloud model

Input: Transaction logs $T(s)=\{t_1,t_2,\cdots,t_m\}$ of service s; a partitioning granularity $g\in G=\{g_1,g_2,...\}$; λ and h as the thresholds of E_n and H_e;

Output: Reliability of service s;

Step 1: Given transaction logs $T(s)$, we extract uncertain QoS matrix $M(s)$ as

$$M(s)=\begin{bmatrix} t_1 \\ \vdots \\ t_m \end{bmatrix}=\left\{\begin{matrix} x_{11} & x_{12} & \cdots & \cdots & x_{1n} \\ \vdots & \ddots & & & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & \cdots & x_{mn} \end{matrix}\right\}$$

Where, m and n are the number of the transaction logs and QoS criteria.

Step 2: Apply $g\in G=\{g_1,g_2,...\}$ to QoS granularity partitioning model and partition the $M(s)$ into p independent QoS regions $R=\{R_1,R_2,\cdots,R_p\}$, such that $\bigcup_{i=1}^{k} R_i = M$.

Step 3: Foreach $R_i \in M$ do

①L is denoted as the number of R_i, $L=|R_i|$;

②Considering the only j^{th} QoS criterion, compute the sample mean of R_i as, $\overline{X}=\frac{1}{L}\sum_{l=1}^{L} x_{lj}$, and the sample variance $S^2=\frac{1}{L-1}\sum_{l=1}^{L}(x_{lj}-\overline{X})^2$;

③Calculate the Expected value of R_i by $Ex(R_i)=\overline{X}$;

④The Entropy of region R_i is calculated by $En(R_i)=\frac{\sqrt{\pi/2}}{L}\sum_{l=1}^{L}|x_{lj}-Ex(R_i)|$;

⑤Compute The Hyper-Entropy of region R_i by $He(R_i)=\sqrt{S^2-En(R_i)^2}$;

Step 4: With the combination of p independent QoS regions, we calculate the Ex,En,He of Web service s by the arithmetic average, thus we have $Ex(s)=\frac{\sum_{k=1}^{p}Ex(R_k)}{p}$, $En(s)=\frac{\sum_{k=1}^{p}En(R_k)}{p}$, and $He(s)=\frac{\sum_{k=1}^{p}He(R_k)}{p}$;

Step 5: If $En(s)\le\lambda$ and $He(s)\le h$, then Web service s is reliable; otherwise, s is filtered out from the reliable services;

Algorithm 2. GPM-based Uncertainty Reliability Computation via Coefficient of Variance

Input: A set of QoS transaction logs of Web service s, $T(s)=\{t_1,t_2,\cdots,t_m\}$; a partitioning granularity $g\in G=\{g_1,g_2,...\}$; α as the threshold of coefficient of variance;
Output: Reliability of Web service s;
Step 1: Given $T(s)$, extract uncertain QoS matrix $M(s)$;
Step 2: Partition matrix $M(s)$ into p QoS regions $R=\{R_1,R_2,\cdots,R_p\}$;
Step 3: Foreach $R_i\in M$, calculate its coefficient of variance as $CV(R_i)$;
Step 4: Combine the results of p coefficient of variances as $CV(s)=\frac{\sum_{k=1}^{p}CV(R_k)}{p}$;
Step 5: If $CV(s)\le\alpha$, the s is identified as reliable service; otherwise, it is filtered out from reliable services;

4 Experimental Evaluation

4.1 Experiment Setup and Datasets

To demonstrate the feasibility and effectiveness of our proposed approach for the reliability of Web services, we implemented a prototype system that integrates the QoS partitioning model and all the algorithms in Java. All the experiments are run on a PC with Intel Core 2.0 GHz processor, 2G RAM in Windows 7.

Extensive experiments have been conducted on real world large-scale dataset WS-DREAM, which is shown in Table 2. It can be found and downloaded from [15]. WS-DREAM dataset contains the number of 1,558,214 service invocation records (i.e., transaction logs), which are invoked by 150 service requesters from 27 countries all over the world. All the QoS values of transaction logs are collected from these 150 service requesters on 100 Web services.

Three experiments have been done for comparing our approach with the existing approch for Web service reliability. (1) GPM-based QoS uncertainty computation via cloud model and coefficient of variance, respectively; (2) Computational time comparisons and analyses; (3) The influences of parameters λ, h and α. We also summarize the comparsions with these approaches and analyze the merits of our proposed one for service reliability.

4.2 Experimental Results of Web Service Reliability via Cloud Model

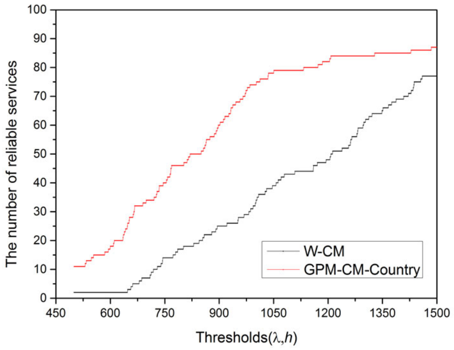
In this experiment, our GPM-based proposed approach for Web service reliability is compared with the existing state-of-the-art approach for the evaluation of uncertain QoS of Web services. The experimental data is based on the uncertain QoS logs in Table 2.

Table 2. The experimetnal datasets with the number of 1,558,214 transaction logs

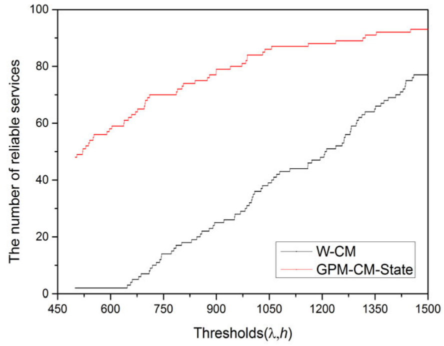
CountryCode	Trans. logs	CountryCode	Trans. logs	CountryCode	Trans. logs
IE	10,380	HR	20,711	SI	10,404
AT(EU)	20,602	CA	46,574	UY	10,334
BR	30,769	CZ	20,587	ES	20,906
BE	10,436	US	742,964	GR	31,116
PR	10,337	NO	10,512	HU	10,399
PL	20,895	PT	10,393	IL(EU)	10,274
DE	218,250	JP	62,871	IT(EU)	41,773
FR	62,538	CH(EU)	51,507	GB	41,613
FI	10,359	CY	10,314	TW	10,306

We set λ and h with different scopes that range from 450 to 2500. Then, we compare the results of the reliability of Web services among these three approaches, including traditional Web service reliability approach based on cloud model (called W-CM) proposed by Wang et al in [12], our GPM-based Web service reliability approach via cloud model under the granularity of CountryCode level (called GPM-CM-Country), and our GPM-based Web service reliability approach via cloud model under the granularity of State (called GPM-CM-State). Different from W-CM that directly considers the entire uncertain QoS matrix as a whole to evaluate the reliability of a Web service, while in terms of different partitioning granularites our approach applies GPM model to divide the whole uncertain QoS matrix into several independent regions. By doing so, we can more accurately evaluate the reliability of a Web service. The experimental results are illustrated in Figure 4(a), (b), (c) and (d).

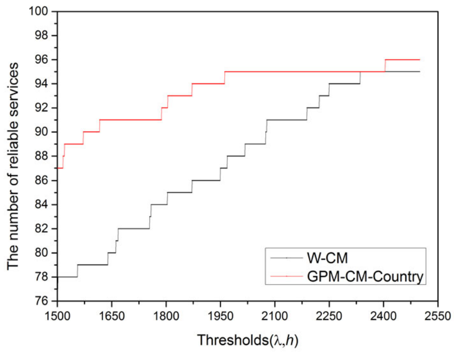
The results from Figure 4 demonstrate that the number of reliable services increases for all three approaches, when the threshold parameters λ and h become larger. However, it will keep stable as thresholds reach a certain value. As expected, the smaller the granularity applies, the larger the number of reliable services increases.



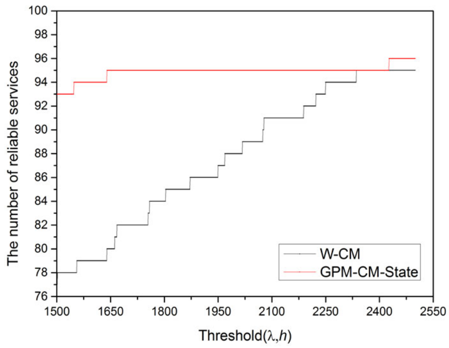
(a) The results on threshold from 450 to 1500 between W-CM and GPM-CM-Country



(b) The results on threshold from 450 to 1500 between W-CM and GPM-CM-State



(c) The results on threshold from 1500 to 2550 between W-CM and GPM-CM-Country

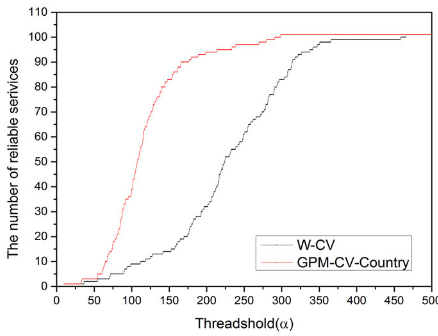


(d) The results on threshold from 1500 to 2550 between W-CM and GPM-CM-State

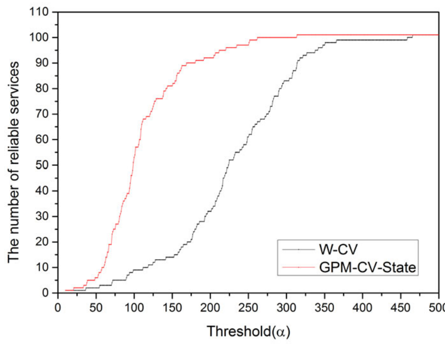
Figure 4. The Experimental results among three approaches for Web service reliability via cloud model

4.3 Experimental Results of Web Service Reliability via Coefficient of Variance

As the experiment of Web service reliability via cloud model, we set α ranging from 0 to 500. Comparisons on the number of reliable Web services are received on three approaches via coefficient of variance, which are called W-CV, GPM-CV-Country, and GPM-CV-State, respectively. Here, W-CV represents the approach [12] where the reliability evaluation of Web services is performed by the whole uncertian QoS matrix with coefficient of variance. As for GPM-CV-Country and GPM-CV-State, they are our proposed coefficient of variance-based approaches where the reliability evaluation of Web services is performed by the partitioning of uncertain QoS transaction logs from the whole QoS matrix. These indepedent QoS transaction logs are partitioned by two kinds of different granularities via GMP model. The experimental results are shown in Figure 5.



(a) The results on threshold from 0 to 500 between W-CV and GPM-CV-Country



(b) The results on threshold from 0 to 500 between W-CV and GPM-CV-State

Figure 5. Experimental results among three approaches of Web service reliability via coefficient of variance

From the experimental results in Figure 5, we can observe that the number of reliable Web services increases among all of the three compared approaches, when α becomes larger. Similarly, it tends to be unchanged as α arrives at a certain level around 450.

4.4 The Time Computational Results

We compare the time computation among three approaches via cloud model and coefficient of variance. The experimental results are shown in Figure 6.

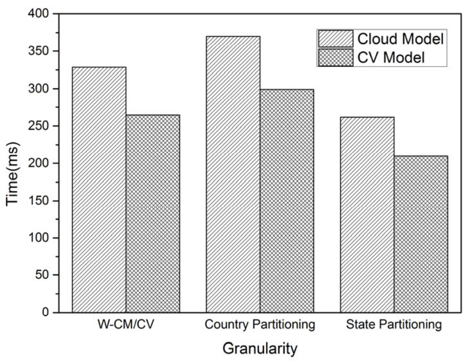
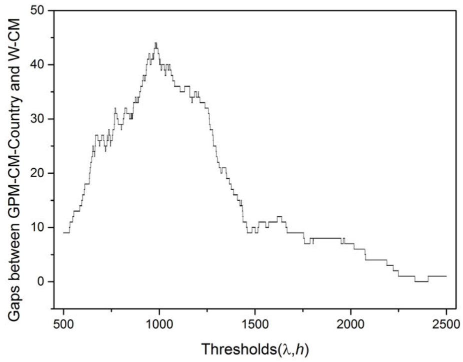


Figure 6. Comparisons of time computation among three approaches via cloud model and coefficient of variance

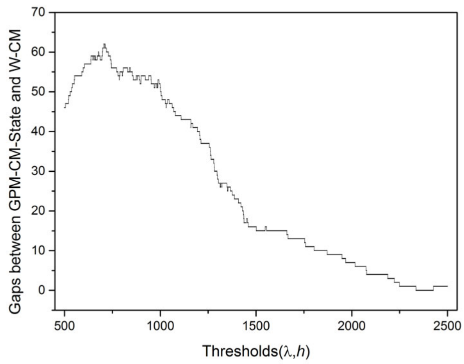
As we can see in Figure 6, three approaches of Web service reliability via coefficient of variance spend shorter time than those via cloud model. The reason is that it takes more complex computing steps in cloud model to evaluate the reliability of a Web service. Furthermore, the smaller the granularity we select, the longer computation time it costs. However, the absolute increment of execution time is still very small. For example, it costs 328.6ms for W-CM/CV and 370.2ms for granularity Country Partitioning via cloud model. In addition, The reason of computing of State Partitioning is shorter than the other two approaches is that we ignore some transaction logs, which cannot find out the state information in datasets.

4.5 The Influences of λ, h and α

To test the influences of thresholds on the evaluation of Web service reliability among three approaches, we compute the gaps between Country Partitioning or State Partitioning with W-CM/CV for both cloud model and coefficient of variance. The influence results are shown in Figure 7 and Figure 8, respectively.

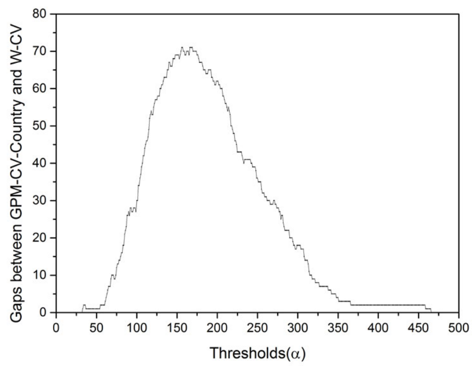


(a) Gaps between GPM-CM-Country and W-CM

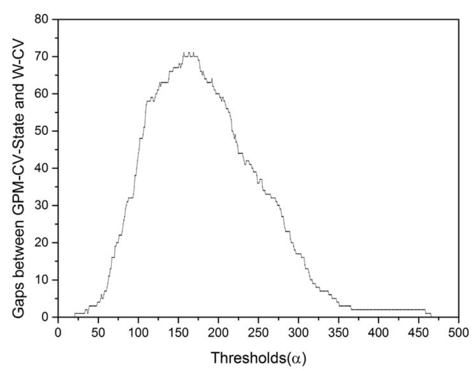


(b) Gaps between GPM-CM-State and W-CM

Figure 7. Influences of λ and h on the evaluation of Web service reliability with QoS uncertainty



(a) Gaps between GPM-CV-Country and W-CV



(b) Gaps between GPM-CV-State and W-CV

Figure 8. Inflences of α on the evaluation of Web service reliability with QoS uncertainty

From the above results of influences from three different parameters, we can see that they share the same significant changes. More specifically, the gaps of between Country Partitioning or State Partitioning with W-CM/CV increase with the growth of three threshold parameters. However, they begin to decline after reaching their own peaks. For example, the peaks are the points (980, 44) and (700, 61) in Figure 7, while the peaks are the points (160, 70) and (170, 70) in Figure 8, respectively. These values of peak points present the gaps of the number of reliable Web services between our methods and traditional one. We find that the larger the values are, the more effective our methods are. According to the above analysis, in specific applications the value of λ and h should be set a value between 700 and 980. Moreover, the value of α should be set a value between 160 and 170.

5 Related Work

In recent years, a variety of QoS-aware hot research issues have been comprehensively studied around Web services in the field of service computing. With the consideration of QoS performance of Web services, various approaches have been proposed for Web service selection, service composition, and its applications. In the aspect of those certain QoS related works, we review QoS-aware Web service selection and composition. From the perspective of uncertain QoS, we mainly focus on the evalution of Web services and its applications.

In certain QoS-aware research work, Mohammad et al. [8] present an efficient heuristic algorithm for the QoS-based service composition that combines global optimization with local selection techniques to benefit from the advantages of both worlds. They first use mixed integer programming (MIP) to find the optimal decomposition of global QoS constraints into local constraints. Especially, distributed local selection is used to find the best Web service that can satisfy these local constraints. In our previous work [16] we proposed a novel planning-based approach that can automatically convert a QoS-aware composition task to a planning problem with temporal and numerical features. Yu et al. [17] propose two novel models for solving the problem of Web service selection with multiple QoS constraint, including a combinatorial model and a graph model. The combinatorial model defines the problem as a multidimensional multichoice 0-1 knapsack problem (MMKP), while the graph model defines the problem as a multiconstraint optimal path (MCOP) problem. Zeng et al. [18] present a middleware platform which addresses the issue of selecting Web services for the purpose of dynamic composition in a way that maximizes users' satisfaction expressed as utility functions over QoS attributes. Within the platform, they propose two selection approaches, including a local selection strategy of Web services and a global allocation of tasks to Web services using the techniques of integer programming. Although much work has been done on certain QoS-aware applications, the network environment is dynamically changing for the invocation of Web services which may lead to uncertain QoS values. The uncertainty of QoS of a Web service plays an important role in dynamic service composition, evaluaiton and its applications.

In most cases, non-functional QoS of Web services is dynamically changing and has become its natural characteristics. Some of recent efforts have been made on uncertain QoS evaluation and dynamic selection of Web services. With the consideration of QoS uncertainty, Wang et al. [12] proposed an effective and efficient approach for QoS-aware service selection via uncertain QoS evaluation. They first employed cloud model to compute the reliability of QoS uncertainty for pruning redundant services, then a mixed integer programming technique was utilized to select optimal services. Although three standard metrics in cloud model have been applied for the filtering of reliable Web services, they only directly made the computation of uncertainty of a Web service without the consideration of partitioning to QoS transaction logs, which leads to the decrease of evalution quality of Web service reiliability. In addition, Sun et al. [19] recently proposed a fast and reliable Web service selection approach that attempts to select the best reliable composited service. They employs information theory and variance theory to abandon those high QoS uncertainty services and downsize the solution spaces. To filter out those low reliable Web services, a reliability fitness function is designed to evaluate the QoS uncertainty and select reliable Web services.

Taking QoS uncertainty as the application scene on service skyline filtering, a novel approach was investigated for computing the service skyline from uncertain QoWS [13]. They first modeled the uncertainty of a Web service via the probability for each QoS property and then presented a p-R-tree indexing structure and a dual-pruning scheme to efficiently compute the p-dominant skyline. Furthermore, to tackle the problem of skyline on uncertain QoS [5], K. Benouaret et al represented each QoS property of a Web service using a possibility distribution. Under this QoS modeling, two skyline extensions have been proposed on uncertain QoS, called pos-dominate skyline and nec-dominate skyline, respectively. They stands for two different kinds of skyline evaluation based on the uncertainty of service QoS.

In addition to uncertainty of non-functional QoS of Web services, some of research work has been done in terms of functional uncertainty for service selection and composition [20-21]. In these works, advanced automated planning techniques and heuristic search algorithms in artificial intelligence have been applied for solving the issues on Web service selection and composition with the uncertainty of Web service functionality.

6 Conclusion and Future Work

QoS has played an important role in determining the invocation success or failure in real composition applications. To extend existing work on computing the uncertainty of QoS, we propose a novel approach via granularity-based partitioning model for the evaluation of Web service reliability. First, we build a granularity-based partitioning model (GPM) that divides the whole uncertain QoS matrix of a Web service into multiple independent regions of service transaction logs via different desired granularities. Second, two effective algorithms based on cloud model and coefficient of variance are integrated into our uncertain QoS partitioning model GPM to evaluate the reliability of a Web service. Extensive experiments have been conducted on a real-world large-scale dataset with 1,558,214 Web service invocation transaction logs. The experimental results demonstrate our proposed approach can significantly outperform the existing state-of-the-art approach for the evaluation of Web service reliability with the consideration of QoS uncertainty.

Although a Web service is reliable, it may not hold high quality QoS values. In our future work, on one hand, we will further validate the effectiveness of our proposed approach by using larger uncertain QoS datasets under real-life application environment. On the other hand, we also plan to select those skyline of Web services through the values of their uncertain QoS by designing novel query optimization algorithms. By doing so, we extend our work to the applicaton of QoS-aware Web service composition with multiple global constraints.

Acknowledgements

This work was partially supported by National Natural Science Foundation of China (No. 61772128, 61303096), Shanghai Natural Science Foundation (No. 18ZR1414400, 17ZR1400200), Fundamental Research Funds for the Central Universities (No. 16D111208), platform fund of PAPD and CICAEET, and Chen Guang project supported by Shanghai Municipal Education Commission.

We would like to appreciate all of the anonymous reviewers for their constructive suggestions and insightful comments that will significantly improve the quality of our manuscript. Meanwhile, we also thank Dr.Zibin Zheng for providing us large-scale uncertain QoS dataset of Web services.

References

[1] F. Curbera, M. Duftler, R. Khalaf, W. Nagy, N. Mukhi, S. Weerawarana, Unraveling the Web Services Web: An Introduction to SOAP, WSDL, and UDDI, *IEEE Internet Computing*, Vol. 6, No. 2, pp. 86-93, March-April, 2002.

[2] Y. Ren, J. Shen, J. Wang, J. Han, S. Lee, Mutual Verifiable Provable Data Auditing in Public Cloud Storage, *Journal of Internet Technology*, Vol. 16, No. 2, pp. 317-323, March, 2015.

[3] Z. Xia, X. Wang, X. Sun, Q. Wang, A Secure and Dynamic Multi-keyword Ranked Search Scheme over Encrypted Cloud Data, *IEEE Transactions on Parallel and Distributed Systems*, Vol. 27, No. 2, pp. 340-352, February, 2016.

[4] Z. Fu, X. Sun, Q. Liu, L. Zhou, J. Shu, Achieving Efficient Cloud Search Services: Multi-keyword Ranked Search over Encrypted Cloud Data Supporting Parallel Computing, *IEICE Transactions on Communications*, Vol. 98. B, No. 1, pp. 190-200, January, 2015.

[5] K. Benouaret, D. Benslimane, A. Hadjali, Selecting Skyline Web Services from Uncertain QoS, *Proceedings of the Ninth International Conference on Services Computing (SCC)*, Honolulu, HI, 2012, pp. 523-530.

[6] L. Zeng, B. Benatallah, M. Dumas, J. Kalagnanam, Q. Z. Sheng, Quality Driven Web Services Composition, *Proceedings of the 12th International Conference on World Wide Web (WWW)*, Budapest, Hungary, 2003, pp. 411-421.

[7] M. Alrifai, D. Skoutas, T. Risse, Selecting Skyline Services for QoS-based Web Service Composition, *Proceedings of the 19th International Conference on World Wide Web (WWW)*, Raleigh, NC, 2010, pp. 11-20.

[8] M. Alrifai, T. Risse, Combining Global Optimization with Local Selection for Efficient QoS-aware Service Composition, *Proceedings of the 18th International Conference on World Wide Web (WWW)*, Madrid, Spain, 2009, pp. 881-890.

[9] T. Yu, K.-J. Lin, Service Selection Algorithms for Web Services with End-to-end QoS Constraints, *Information Systems and E-Business Management*, Vol. 3, No. 2, pp. 103-126, July, 2005.

[10] Z. Zheng, Y. Zhang, M. R. Lyu, CloudRank: A QoS-driven Component Ranking Framework for Cloud Computing, *Proceedings of the 29th IEEE Symposium on Reliable Distributed Systems*, New Delhi, India, 2010, pp. 184-193.

[11] Z. Zheng, T. C. Zhou, M. R. Lyu, I. King, FTCloud: A Component Ranking Framework for Fault-tolerant Cloud Applications, *Proceedings of the 21th IEEE International Symposium on Software Reliability Engineering (ISSRE 2010)*, San Jose, CA, 2010, pp. 398-407.

[12] S. Wang, Z. Zheng, Q. Sun, H. Zou, F. Yang, Cloud Model for Service Selection, *Proceedings of the 2011 IEEE INFOCOM Workshops on Computer Communications*, Shanghai, China, 2011, pp. 666-671.

[13] Q. Yu, A. Bouguettaya, Computing Service Skyline from Uncertain QoWS, *IEEE Transactions on Services Computing*, Vol. 3, No. 1, pp. 16-29, January-March, 2010.

[14] D. Li, D. Cheung, X. Shi, V. Ng, Uncertainty Reasoning

based on Cloud Models in Controllers, *Computers and Mathematics with Applications*, Vol. 35, No. 3, pp. 99-123, February, 1998.

[15] Z. Zheng, M. R. Lyu, Collaborative Reliability Prediction of Service-oriented Systems, *Proceedinds of the 32nd ACM/IEEE International Conference on Software Engineering (ICSE)*, Cape Town, South Africa, 2010, pp. 35-44.

[16] G. Zou, Q. Lu, Y. Chen, R. Huang, Y. Xu, Y. Xiang, QoS-aware Dynamic Composition of Web Services using Numerical Temporal Planning, *IEEE Transactions on Services Computing*, Vol. 7, No. 1, pp. 18-31, January- March, 2014.

[17] T. Yu, Y. Zhang, K.-J. Lin, Efficient Algorithms for Web Services Selection with End-to-end QoS Constraints, *ACM Transactions on the Web (TWEB)*, Vol. 1, No. 1, pp. 1-26, May, 2007.

[18] L. Zeng, B. Benatallah, A. H. H. Ngu, M. Dumas, J. Kalagnanam, H. Chang, QoS-aware Middleware for Web Services Composition, *IEEE Transactions on Software Engineering*, Vol. 30, No. 5, pp. 311-327, May, 2004.

[19] L. Sun, S. Wang, J. Li, Q. Sun, F. Yang, QoS Uncertainty Filtering for Fast and Reliable Web Service Selection, *Proceedings of the IEEE International Conference on Web Services (ICWS)*, Anchorage, AK, 2014, pp. 550-557.

[20] G. Zou, Y. Chen, Y. Xu, R. Huang, Y. Xiang, Towards Automated Choreographing of Web Services Using Planning, *Proceedings of the 26th AAAI Conference on Artificial Intelligence (AAAI)*, Toronto, Canada, 2012, pp. 178-184.

[21] S. Niu, G. Zou, Y. Gan, Z. Zhou, B. Zhang, UCLAO* and BHUC: Two Novel Planning Algorithms for Uncertain Web Service Composition, *Proceedings of the International Conference on Services Computing (SCC)*, San Francisco, CA, 2016, pp. 531-538.

Biographies

Guobing Zou is an associate professor in School of Computer Engineering and Science at Shanghai University, China. His research interests include service computing and data mining. He has published around 45 papers on international journals and conferences, including IEEE TSC, Knowledge-Based Systems, Applied Intelligence, Soft Computing, AAAI, and SCC.

Zhimin Zhou received his undergraduate degree from School of Internet of Things Engineering, Jiangnan University, China. He is currently a master student in School of Computer Engineering and Science, Shanghai University, China. His research interests include service computing, uncertainty of service QoS recommendation.

Mei Zhao received her master degree from the School of Computer Engineering and Science, Shanghai University, China. She is currently working for a IT company called Qiniu, Shanghai. Her research interests include service computing, cloud storage. She published papers on SNPD-16 and International Journal of Web Engineering and Technology.

Sen Niu is currently a Ph.D. candidate in the School of Computer Engineering and Science, Shanghai University, China. His research interests include Web service composition, service computing, artificial intelligence and automated planning. He published papers on SCC-16 and International Journal of Web Engineering and Technology.

Yanglan Gan is an associate professor in School of Computer Science and Technology at Donghua University, China. Her research interests include data mining and bioinformatics. She has published around 20 papers on international journals and conferences, including Bioinformatics, BMC Bioinformatics, IEEE/ACM Transactions on Computational Biology and Bioinformatics, and Neurocomputing.