A Multi-Criteria QoS-aware Trust Service Composition Algorithm in Cloud Computing Environments

Weina Lu¹², Xiaohui Hu¹, Shangguang Wang³ and Xiaotao Li¹

¹School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China
²Department of Mechanics and Electronics, Hebei Normal University of Science and Technology, Changli, Hebei, 066600, China
³State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China
E-mail: luweina1981@126.com

Abstract

In an open and dynamic cloud computing environment, the randomness of cloud services with unreliable quality leads to low accuracy in service composition. To address the above problem, this paper proposes a global trust service composition approach based on random QoS and trust evaluation, considering the multi-criteria assessment of service quality. Firstly, statistical test is employed to remove the uncertain outliers and to estimate the ideal value of the collected objective QoS data. Secondly, subjective QoS evaluations of providers and users are aggregated according to direct trust and recommended trust. Finally, services are composed through global QoS optimization. Experiments for each stage show that the approach improves the accuracy and precision of service composition.

Keywords: cloud computing, service composition, quality of service, multi-criteria, hypothesis test, global optimization

1. Introduction

Cloud computing is an internet-based service mode to achieve various IT capabilities and flexible invocations. Configurable resources are provided to computers and other devices as services (e.g., IaaS, PaaS, and SaaS). The traditional service computing is evolved towards a more flexible, dynamic, scalable and business-oriented cloud computing environment. There are more and more popular cloud applications which are composed by software components and cloud services (such as Web service). As single service cannot satisfy ever-changing functional requirements of user-sides, many enterprises aim at combining multiple services into a more powerful new service in the cloud. Moreover, high quality of service (QoS) must be critically considered in service selection and composition, since there are a number of functionally equivalent services. However, in the open and variable cloud environment, the unstable situation such as dynamic network topology, transmission channel interference and node performance changing makes QoS to present dynamic and random values. Hence, the authenticity of service quality is badly influenced. Some abnormal emergency of system, or false malicious evaluation beyond the normal levels of QoS also results in the false attributes of service quality. Therefore, the dynamic quality of cloud services supplied by more and more public, private and hybrid cloud, along with ways to build credible and accurate combination of services become an important research problem.
In this study, we present a dynamic trust service composition algorithm with multi-criteria assessments of QoS. Our aim is to ensure the stability and reliability of the objective and subjective QoS attributes, so that the needs of customers, providers and third-party net environment can be better satisfied. We considered two types of attributes of QoS. One type is objective QoS (such as execution time, reliability, availability, throughout) and the second type is subjective QoS (such as QoS evaluation from service provider and client). In first phase, hypothesis testing method is applied to remove the uncertain outliers of the collected QoS objective data, and the ideal QoS value is estimated. In second phase, the credibility of the subjective evaluation from the providers and customers is analyzed to obtain the comprehensive evaluation of service. Finally, with the objective and subjective QoS data, the global dynamic service composition model is built and the optimal results are calculated. Simulation results demonstrate the proposed approach can greatly improve the accuracy of service composition.

The remaining paper is organized as follows: Section 2 reviews related work. Section 3 describes a t-test method for objective and random QoS. Section 4 presents QoS evaluation based on direct and recommended trust for subjective QoS. Section 5 proposes dynamic optimal service composition with global QoS constraints. Section 6 presents experiments and Section 7 concludes the paper.

2. Related Work

Some unsolved problems in QoS-aware services composition in cloud computing environment cause widespread concerns of scholars. So far, there have been a lot of significant research results. In [1], the global QoS constraints are adaptively decomposed to local constraints with preferences of users, and the optimal combination of services is obtained by mixed integer programming method with the real-time requirement. In [2], the problem of QoS-aware service composition is taken in another way by modeling as an extended flexible constraint satisfaction framework and is solved by the branch and bound algorithm. In [3], the author classified the QoS evaluation criteria into two categories: objective and subjective factors. Combining user preferences and QoS global constraint with the Brown-Gibson heuristic algorithm is found useful to calculate the optimal aggregation value of QoS. All of the above methods have some progress in effectiveness and real-time of services optimal selection, but they generally lack considerations of dynamic QoS variability [4]. Consequently, the accuracy and precision of services optimal composition cannot be guaranteed. In [5], cloud model is employed to compute the uncertainty of QoS and Skyline algorithm is proposed to prune redundant services. As the range of searching is reduced, the efficiency of service composition is greatly increased. In [6], the uncertain attributes and non-parametric test was used to detect historical QoS information on the clients-side. Then the detected result is taken as the criteria of service selection. Although above mentioned two methods considered the uncertainty of objective QoS, they just use the uncertainty degree for screening services. The stability of QoS is not substantially improved and the trust of Service composition cannot be guaranteed. The collaborative filtering approach is designed in [7, 8]. It takes advantage of the past QoS evaluations from users to predict the current QoS values, and then recommends the best service. In [9], the authors discussed the trusted QoS evaluation and service selection with environment aware. Although the QoS evaluation structure of trust reasoning evolution is given by the characteristic vector, the credibility is only expressed through gradient modulus of the difference between two evaluations, lacking of consideration on their own characteristics of evaluations. This results in the instability of success rate in services composition.
To summarize, the research on QoS-aware service composition made progresses in service composition optimization, QoS uncertainty criterion and service reputation. However, few studies consider both QoS objective randomness and subjective trust evaluation together, and there are often large deviations between the actual demands and the results. The success rate and accuracy of service composition is not high.

3. Outlier Test of Objective QoS

For the randomness of objective QoS data, we employ hypothesis test to remove outliers and correct the QoS values. All attributes of QoS are calculated through statistical analysis, so that the accuracy of QoS can be improved.

After executing service several times, a large number of QoS data are collected. We assume all the data are derived from normal population and they follow a normal distribution. However, a few larger or smaller QoS data values also exist in the collected QoS. They are suspicious and they need to be t-tested. Considering the tested value $Y$ and other value $X_1, X_2, \cdots X_{n-1}$ as samples from two independent normal population $N(\mu, \sigma^2)$ and $N(\mu, \sigma^2)$, to determine whether the value $Y$ is credible, we judge whether $\mu_1$ is equal to $\mu$ under some significance level $\alpha (\alpha \in [0,1])$. Therefore, we need to test the hypotheses as follows

$$H_0: \mu_1 = \mu \leftrightarrow H_1: \mu_1 \neq \mu$$

The test statistics is

$$T = \frac{Y - \bar{X}}{S}$$

where, $\bar{X} = \frac{1}{n-1} \sum_{i=1}^{n-1} X_i$, $S = \frac{1}{\sqrt{n-2}} \sum_{i=1}^{n-1} (X_i - \bar{X})^2$.

Under the condition of hypothesis $H_0$, $T \sim t(n-2)$. Given significance level $\alpha = 0.05$, let

$$P\left( |T| \geq t_\alpha(n-2) \right) = \alpha$$

So the rejection region of test is

$$W = \left\{ \frac{|Y - \bar{X}|}{S} \geq t_\alpha(n-2) \right\}$$

When $|Y - \bar{X}| \geq t_\alpha(n-2)$, $Y$ is considered as an outlier and should be removed, otherwise kept.

After removing the outliers, the objective QoS data follow normal distribution $N(\mu, \sigma^2)$. For some attribute i (e.g., reliability), we can calculate its mathematic expectation $E(x) = \mu$ with mean method and take this result as the QoS attribute value $\bar{q}_i(s)$.

4. Evaluation Models of subjective QoS

There are two types of QoS evaluations: Service Provider (SP) evaluation and Service Client (SC) evaluation. Both the evaluations belong to subjective QoS attributes with human experiences of using service [10]. As influenced by various factors, the service evaluation may have some unreliable information. For example, some SPs may publish the QoS evaluation beyond actual level to attract more clients. Under the influences of network environment, user context and preference of users, different users have different evaluations on the same cloud service. Furthermore, there may be false and malicious clients to raise or lower the evaluation scores of some services. Consequently, during service selection,
requestors should firstly identify the credibility of SP and SC, and then adopts or refuses their suggestions. In our study, we propose respectively QoS evaluation models based on direct and recommended trust, so that false evaluations of providers and users can be corrected.

4.1. QoS Evaluation based on Direct Trust

The direct trust is about service provider. It is obtained through the invoking records of service. That is to say, the direct trust is generated from multiple history interactions between users and the services supplied by SPs. Here, the maximum likelihood method is used to estimate the direct trust $C_{sp}$.

Given the direct trust $\theta_{ij}$, it represents user's (SC$_i$) trust degree to service provider (SP$_j$). Each time the user SC$_i$ executes the service provided by SP$_j$, he records an invoking result. The corresponding record is $x_{ij} = \{s_{ij}, f_{ij}\}$. Where, $s_{ij}$ is the times of successful execution, and $f_{ij}$ is the times of fail execution. After collecting all users' records $x_{ij}(i = 1, 2, \cdots m)$, the likelihood function for the global credibility $\theta_j$ of provider is

$$L(x_1j, x_2j, \cdots x_mj; \theta_{ij}) = \prod_{i=1}^{m} \theta_j^{s_{ij}}(1 - \theta_j)^{f_{ij}}$$

(4)

The log-likelihood function is

$$\ln L(\theta_j) = \sum_{i=1}^{m} s_{ij} \ln \theta_j + \sum_{i=1}^{m} f_{ij} \ln (1 - \theta_j)$$

(5)

The maximum likelihood estimate $\hat{\theta}_j$ of $\theta_j$ is the direct trust $C_{sp}$ of the service provider. Hence, the final QoS evaluation of service provider $R_{spj} = C_{sp} \ast R_{sj}$, where, $R_{sj}$ is the initial QoS evaluation published by SP.

4.2. QoS Evaluation based on Recommended Trust

QoS evaluation based on recommended trust is about feedback information of clients after executing service. For the same service, the more similar evaluations of the requestor and user are, the higher recommending credibility of the users is. In our study, we regard the similarity between the requestor and user as the recommended trust of QoS evaluation, and then aggregate evaluations of different users though the recommended trust. The accurate QoS evaluation of current service is obtained finally.

Given the service requestor $i$ and the evaluating user $j$, $I_i$ is the service set evaluated by the requestor in the past, and $I_j$ is the service set evaluated by the user. $I_{ij}$ is the service set evaluated both by the requestor $i$ and user $j$. So the similarity between requestor $i$ and user $j$ is:

$$sim(i,j) = \frac{\sum_{s \in I_{ij}}(R_{i,s} - \bar{R}_i)(R_{j,s} - \bar{R}_j)}{\sum_{s \in I_i}(R_{i,s} - \bar{R}_i)^2 \sum_{s \in I_j}(R_{j,s} - \bar{R}_j)^2}$$

(7)

where, $R_{i,s}$ is the evaluation of service $s$ for the user $i$, $R_{j,s}$ is the evaluation of service $s$ for the user $j$, $\bar{R}_i$ is the mean evaluation of services for the requestor $i$, and $\bar{R}_j$ is the mean
evaluation of services for the user \( j \). \( sim(i, j) \) is the evaluation similarity between the requestor \( i \) and user \( j \), and the range of its value is \([-1, 1]\). If \( sim(i, j) > 0 \), the recommended trust \( C_{i,j} \) of the user \( j \) is equal to \( sim(i, j) \), while if \( sim(i, j) \leq 0 \), the requestor \( i \) hasn’t trusted the user \( j \) and the recommended trust \( C_{i,j} \) of the user \( j \) is equal to zero.

Having calculated the recommended trust of clients, the formula of QoS evaluation \( R_{sc} \) for the requestor is:

\[
R_{sc} = \frac{\sum_{j=1}^{n} C_{i,j} R_{j,s_i}}{\sum_{j=1}^{n} C_{i,j}}
\]  

(8)

Where, \( C_{i,j} \) is the recommended trust of the user \( j \) for the requestor \( i \) and \( R_{j,s_i} \) is the using evaluation of service \( s_i \) for the user \( j \).

After weighted aggregation of the credible evaluation from providers and clients, the comprehensive evaluation of service quality is obtained. It is calculated by the following formula:

\[
R = \mu R_{sp} + (1 - \mu) R_{sc}
\]  

(9)

Where, \( \mu = \frac{1}{1+b/N} \), here \( b \) is the number of evaluating users. \((1 - \mu)\) makes the weight of users’ evaluation increase along with the growing number of users. Here, constant \( N \) controls the increasing speed of the weight. The value of \( N \) can be designed by different needs. When \( b=4N \) and \( \mu=0.2 \), the comprehensive evaluation has been determined basically by evaluations of users. If the user evaluation cannot be collected, let \( b=0 \). Then the service selection is only based on the provider’s evaluation.

5. Global Optimal Service Composition

In this part of study, we propose a dynamic service composition with global QoS constraints. According to the requirements of service function and QoS attributes, the most appropriate services are selected from many candidate services to form a greater combined service.

For a single service, after the above process, we have obtained the real and credible attributes of objective and subjective QoS. The aggregated function of QoS weighted by different attributes requirements is described as

\[
M_{s_i} = \alpha \hat{Q}(s_i) + \beta R(s_i)
\]  

(10)

Where, \( \alpha + \beta = 1 \), \( 0 \leq \alpha \leq 1 \), \( 0 \leq \beta \leq 1 \), \( R(s_i) \) is the comprehensive evaluation of the service \( s_i \) and \( \hat{Q}(s_i) \) is the objective QoS utility value of service \( s_i \). Let \( \hat{Q}(s_i) = \sum_{j=1}^{z} \omega_j \hat{q}_j \), where \( \hat{q}_j \) is the estimated value of objective attribute \( j \), \( \omega_j \) is the weight and \( z \) is the number of objective attributes. Therefore, formula (10) turns into the following function

\[
M_{s_i} = \alpha \sum_{j=1}^{z} \omega_j \hat{q}_j + \beta R(s_i)
\]  

(11)

The service composition problem can be described as follows: Given \( m \) tasks (that is, \( m \) kinds of services needed to compose) in a workflow and \( n \) candidate services in each task. The variable \( x_{ij} \) (\( 1 \leq i \leq m, 1 \leq j \leq n \)) is introduced. If the i-th task is performed by the j-th service (namely service \( s_j \)), then \( x_{ij} = 1 \), else \( x_{ij} = 0 \). In this paper, the objective attributes of QoS include execution time, reliability, availability and throughput. The subjective attributes
are evaluations of SP and SC. Therefore, the total set of QoS attributes is \( Q = \{ T(\text{execution time}), D(\text{reliability}), U(\text{availability}), H(\text{throughput}), R(\text{comprehensive evaluation}) \} \). Take the sequence execution for example, the decision objective functions of QoS in service composition is as follows:

1) execution time \( T \):
\[
\min \sum_{i=1}^{m} \sum_{j=1}^{n} x_{ij} t_{ij}
\]

2) reliability \( D \):
\[
\max \prod_{i=1}^{m} \sum_{j=1}^{n} x_{ij} d_{ij}
\]
After taking logarithm, that is
\[
\max \sum_{i=1}^{m} \sum_{j=1}^{n} \ln(d_{ij}) \cdot x_{ij}
\]

3) availability \( U \):
\[
\max \prod_{i=1}^{m} \sum_{j=1}^{n} x_{ij} u_{ij}
\]
After taking logarithm, that is
\[
\max \sum_{i=1}^{m} \sum_{j=1}^{n} \ln(u_{ij}) \cdot x_{ij}
\]

4) throughput \( H \):
\[
\max \sum_{i=1}^{m} \sum_{j=1}^{n} x_{ij} h_{ij}
\]

5) comprehensive evaluation \( R \):
\[
\max \sum_{i=1}^{m} \sum_{j=1}^{n} x_{ij} R_{ij}
\]

After service selection, the QoS attributes of the services involved in combination should be all optimal ideally. But actually, they are restricted and conflicting with each other. So, not all decision objective functions can be optimized at the same time. This is the NP-hard problem with multiple dimension and choice [13]. Accordingly, a compromise is needed between the objective function of attributes and the global optimum. And then a relatively optimal or approximate optimal solution of QoS is obtained for the service composition. After normalization, the objective function of optimization is
\[
Qos_{\text{total}} = \max \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{k=1}^{z} x_{ij} \cdot \omega_{ij}^k q_{ij}^k
\]

Where, \( q_{ij}^k \) is the k-th quality attribute of service \( s_{ij} \), \( \omega_{ij}^k \) is the weight coefficient of the k-th quality attribute.
In each instance of the service composition, user preferences about the quality attributes of the service are decided in advance. Accordingly, for each service $s_{ij}$ in sequence composition, its $\sum_{k=1}^{r} a_{ij} q_{ij}^{k}$ is an established utility value. Therefore, given

$$a_{ij} = \sum_{k=1}^{r} a_{ij} q_{ij}^{k}$$

(18)

In conclusion, after normalization, the global constraint optimization model of service composition is as follows:

$$\begin{aligned}
&\text{Objective function: } \max \sum_{i=1}^{m} \sum_{j=1}^{n} x_{ij} a_{ij} \\
&\text{Conditions of Constraints: } \\
&\sum_{i=1}^{m} \sum_{j=1}^{n} x_{ij} t_{ij} \geq T, \quad \sum_{i=1}^{m} \sum_{j=1}^{n} x_{ij} \ln d_{ij} \geq \ln D, \\
&\sum_{i=1}^{m} \sum_{j=1}^{n} x_{ij} \ln u_{ij} \geq \ln U, \quad \sum_{i=1}^{m} \sum_{j=1}^{n} x_{ij} R_{ij} \geq R, \\
&\sum_{j=1}^{n} x_{ij} = 1, \quad x_{ij} = 0 \text{ or } 1
\end{aligned}$$

(19)

Mixed integer programming method is used to solve the above optimization model. For each task, the best service suitable for global constraints can be found quickly and reliably from each service class. The optimal service composition is obtained finally.

6. Experiment Results and Discussions

In order to evaluate the proposed approach, we did related experiments by stages: (I) the comparative experiment about expectation and standard deviation of objective QoS data; (II) the comparative experiment about QoS subjective evaluation; (III) the comparative experiment about optimal degree and deviation of service composition. One part of the used data is from the real service set, while the other part of data is from the simulated data set. All experiments were performed in the same hardware and software environment (Pentium Dual 2.4GHz, 2.0GB RAM, Windows 7, MATLAB7.0, Java 1.7).

6.1. Comparing Subjective Evaluation of QoS Data

We employed real service set for the experiment of QoS objective randomness. This data set is collected from 142 distributed computers with 4,532 Web services. It contains two attributes of execution time and throughout [14].

In order to illustrate the algorithm can overcome influences of the random environment, we compare the proposed approach to the statistical method based on mean. As shown in Figure 1, our t-test method can remove outliers of the QoS data. Accordingly, the relatively stable expectations (TME) and standard deviation (TSTD) is obtained. However, in Figure 2, the result of mean method has obvious and large fluctuations on either the expectation (TME) or the standard deviation (MSTD). Moreover, its deviation value is already hundred times more than ours.
6.2. Comparing Subjective Evaluation of QoS Data

In this part of experiment, 2000 evaluations of users are simulated. Where, 500 users is credible and other users are incredible. The evaluations of providers is from 100 Web service.

We compare the experiments of the proposed algorithm (PRC) with other two algorithms (the distance method (DIS) [15]) and the difference method (DIF) [16]). The result of each experiment is expressed by Mean Absolute Error (MAE) of user evaluations. The formula of MAE is defined as follows:

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |R_{s_{i}} - R_{s_{i}}|$$

Where, $R_{s_{i}}$ is ideal evaluation of requested service $s_{i}$, $R_{s_{i}}$ is the evaluation of service $s_{i}$ for requestor $s_{i}$.
In Figure 3, we can see that the MAE of proposed algorithm (PRC) is lower and more stable than other two methods (DIS and DIF). Fig.4 shows the comprehensive evaluation (CER) of providers and clients based trust. Given the ideal evaluation of each service is 0.5, N=100. The initial evaluations of clients are all supposed to be higher than real values. From the graph, along with the lower rate (σ) of false, the comprehensive evaluations are closer to the ideal 0.5.

6.3. Comparing Optimal Degree and Deviation of Service Composition

Two data sets are employed in this part of experiment. One set is QWS real data [17,18,19], the other set is simulated data set. The optimal degree of service composition is defined as follows:

\[ \psi = \frac{O_{global}}{O_{pure}} \]

Where, \( O_{global} \) is object value of global optimization, \( O_{pure} \) is object value of local pure optimization. In Figure 5, we compare the result of proposed global dynamic optimal composition (GCP) with the local pure optimal method (LPP). Where, GPC(3) represents optimal degree of composition with 3 constraints (execution time, reliability and availability), and GCD(5) represents optimal degree of composition with 5 constraints (execution time, reliability, availability, throughout and comprehensive evaluation). Obviously, GCP(3) is higher than GCP(5), because the more constraints are, the stricter service selection is. This is also a good illustration that our dynamic optimization algorithm is quite effective.

The deviation rate of optimal degree is defined as follows:

\[ \chi = \left| \frac{O_{value} - O_{true}}{O_{true}} \right| \times 100\% \]

Where, \( O_{value} \) is the optimal value of global dynamic composition, and \( O_{true} \) is the real QoS aggregation value of service composition. Fig.6 shows the changes of the deviation rate of optimal degree in two composition patterns. GCPD represents the result of our method and
UTED represents the result without t-test and evaluation model. It is obviously that GCP is lower than UTED, and our method is proved to have a higher accuracy. As the QoS false rate increasing, both the deviation rates of GCPD and UTED have increasing trends. However, the rise in UTED is significantly larger and quicker than GCPD. This proves GCPD algorithm has better stability. Our approach of GCPD is more close to the optimal goal. This because GCPD QoS data are more centralized and precise after filtered by t-test and revised by the trust evaluation model, while the UTED QoS data contains obvious outliers and false values, which result in the inaccuracy of data.

7. Conclusion

This paper proposed a global dynamic service composition based on objective QoS randomness and subjective trust evaluation in cloud computing environment. In contrast with previous studies, we designed a more comprehensive multi-criteria assessment for service quality, considering dynamic influences of unstable network factors, providers and clients. After rejecting QoS outliers values and revising false evaluations, the services are globally optimized and composed with objective and subjective QoS constraints. The experimental results proved that the algorithm is suitable for dynamically changing cloud environment and is able to guarantee the true quality of service composition.

References


