2011 Third International Conference on Multimedia Information Networking and Security

A survey on QoS-aware web service composition

Shi Yulu
School of Management, Nanjing University, Nanjing, China
shiyulu9820@gmail.com

Chen Xi*
School of Management, Nanjing University, Nanjing, China
chenx@nju.edu.cn

Abstract - A comprehensive summary for QoS-aware web service composition, including QoS models and searching algorithms, has been presented in this paper. First, we have introduced the basic workflow models for service composition and their corresponding QoS aggregation functions. Then, traditional approaches for QoS-aware composition are divided into two groups: mathematical programming and heuristic algorithms. the dynamic service composition has also been discussed.

Key words - Service Composition, QoS, Workflow

I Introduction

According to Wikipedia, Web Service is defined as software system designed to support interoperable machine-to-machine interaction and it can be published, located and invoked across the Web. A lot of researches have focused on the process of Web Service Composition (WSC). Despite of the functional properties in the similar service groups, considerable attention has been attracted to the non-functional aspects, especially Quality of Service (QoS). In fact, the QoS perspective is considered as the most important crucial issue for service composition since it can sometimes determine the final performance of the composed service group and is directly related to the user satisfaction.

In this paper, we mainly discuss the QoS workflow models and current approaches for Web service composition. QoS models for basic composition workflows are summarized in Section II. In section III, relevant algorithms are broken down into mathematical programming and heuristic algorithms. Then dynamic service composition is discussed in section IV, followed by the final conclusion.

II QoS workflow models

Since traditional workflow technology performs well for modeling and coordination of business processes, it is believed that the workflow model can comprehensively depict the real situation of the information exchange in Service Oriented Architecture (SOA).

The concept of workflow and its application for Web service composition was first proposed by Cardoso(2004)[1]. As Cardoso explained, the workflow is usually composed of tasks and transitions which denote the dependencies between tasks and are associated with an enabling probability. According to recent studies, the elemental structures for workflow model include sequence, parallel, choice and loop.

Among the early studies on the end-to-end web service composition, various quality criteria have been proposed. Zeng et al. (2004)[2] set the QoS attributes as the execution price, execution duration, reputation, reliability and availability. Cardoso et al. (2004)[1] focused on three metrics of task time, task cost and task reliability. Response time, service cost, availability and reliability were considered in T. Yu et al.'s work (2005)[3]. In Table 1, we classify the QoS parameters for the single service in three categories according to their data sources.

As seen from Table 1, the task-based criteria are concerned with the specific request that the service gets and their values are usually based on the Web service description, while the others are more related with the

*Corresponding author: Chen Xi, E-commerce department, Nanjing University (E-mail: chenx@nju.edu.cn)

978-0-7695-4559-2/11 $26.00 © 2011 IEEE
DOI 10.1109/MINES.2011.118

service performance in the past. Specifically, the performance-based criteria can be directly calculated by using the objective data recorded by each execution, and the user-based criteria is more comprehensive since it takes the feedback of different users into account.

Compared to the other attributes, the task-based criteria are more popular for the current researches. According to Cardoso(2004)[1], the total cost when the service is invoked, is the sum of enactment cost(EC) and realization cost(RC), with the former associated with the management and monitoring for the service, and the latter representing the cost for the runtime execution. Similarly, the response time is also composed of process time (PT) and delay time (DT) which refers to the non-value-added time such as the queuing delay and the setup time. Here comes the computing expressions for these two criteria: $\begin{cases} C = EC + RC \\ T = DT + PT \end{cases}$

The reliability (R) of a service can be computed with the following formula[2] $R = \frac{N(s)}{K}$, where $N(s)$ is the number of times that the service has been successfully delivered within the maximum expected time, and K is the total number of invocations. And the availability (A) of a service can be similarly computed with another formula[2] $A = \frac{T(s)}{\theta}$, where $T(s)$ is the total amount of time (in seconds) in which service is available during the last θ seconds. θ, which is set by the service community, usually depends on the application environment and it varies according to the access frequency.

Recently, QoS aggregating functions for different models have reached the agreement. According to the current studies[4-5], the QoS aggregate functions for different workflow patterns is shown in Table 2.

Table 1 QoS criteria for single web service

Categories	QoS Criteria	Definition
Task-based	Cost (C)	The price that a service requester has to pay for invoking the service.
	Response time (T)	The time interval between the moment when a service is invoked and the moment when it is finished.
Performance-based	Reliability (R)	The probability that a request is correctly responded within the maximum expected time.
	Availability (A)	The probability that a service is available during the request.
User-based	Reputation (Rep)	The average ranking given to the service by end users according to their own experiences.

Table 2 Workflow QoS aggregation

QoS criteria	Sequence	Parallel	Choice	Loop
Cost (C)	$\sum_{i=1}^{n} C(s_i)$	$\sum_{i=1}^{n} C(s_i)$	$\sum_{i=1}^{n} p_i C(s_i)$	$kC(s)$
Response time (T)	$\sum_{i=1}^{n} T(s_i)$	$Max\{T(s_i)_{i\in\{1...n\}}\}$	$\sum_{i=1}^{n} p_i T(s_i)$	$kT(s)$
Reliability (R)	$\prod_{i=1}^{n} R(s_i)$	$\prod_{i=1}^{n} R(s_i)$	$\sum_{i=1}^{n} p_i R(s_i)$	$R(s)^k$
Availability (A)	$\prod_{i=1}^{n} A(s_i)$	$\prod_{i=1}^{n} A(s_i)$	$\sum_{i=1}^{n} p_i A(s_i)$	$A(s)^k$
Reputation (Rep)	$\frac{\sum_{i=1}^{n} Rep(s_i)}{n}$	$\frac{\sum_{i=1}^{n} Rep(s_i)}{n}$	$\sum_{i=1}^{n} p_i Rep(s_i)$	$Rep(s)$

III Current approaches for Web service composition

A. Mathematical programming

The QoS-aware service composition problem has been defined as mathematical problems such as Integer Linear Planning (LIP), Single Objective Problem with QoS constraints and Multiple Objective Problem with QoS constraints are the most common ones.

For the service composition, LIP[6-7] can help to get the good solution without constructing all the possible composite services[8]. However, it can only be available for the composition problems with small volumes since the traditional Branch and Bound technique has its computation limitations. On the other hand, it also asks for the linearization of the objective function and corresponding constraints.

The single objective problems[9-12] assume that the weights for QoS attributes are given, but transformation from multi-objective to single-objective usually can not be achieved by simply weighted summation. The multiple objective problems have no requirements for the weights of QoS parameters, and the objective function for each QoS parameter is combined to generate a group of Pareto optimal solutions which can be selected by users in terms of their own preferences. Compared to the former, the latter seems to be more flexible and can be better applied to service composition.

The typical mathematical solution for the composition problem was proposed by Zeng et al.(2004)[2], which has been adopted by relevant researchers. Approaches for both local selection and global planning were discussed in their paper. The local selection is based on SAW dealing with the evaluation of single WS while the global planning mainly includes three steps, the determination of single execution path, the comparison between multiple execution paths and the employment of integer programming.

B. Heuristic algorithms

QoS-aware service composition is NP-hard, and it is often defined as a Multi-dimension Multi-choice Knapsack Problem (MMKP) which searches for the composition that has the highest total fitness when satisfying QoS constraints. It usually include numerous constraints and can not perform well when large amounts of services are involved. The greedy approach is used to help to select the best candidate service suitable for the execution[13]. However, it belongs to local strategy which can only get access to the approximate solution rather than the optimal one. Thus, the heuristic algorithms are becoming more and more popular for QoS-aware service composition, among which the genetic algorithm (GA) is claimed to be the most pervasive one.

GA[14] is based on the idea of Darwinian evolution and it usually includes operations of crossover, mutation and selection operations. By gradual evolution, the most adaptable chromosome can be achieved. In this way, GA can well reflect the optimizing process and the final chromosome is equal to the best solution. For QoS-aware service composition, the chromosome represents a specific composition plan and the genes on it depict the corresponding abstract services. According to the QoS values for different services, the fitness function can be determined.

The main problem for GA is that the result based on this algorithm may be local optimal that will leads to its practical obstacle in the future. And Compared to IP, GA is more efficient only when the volume of relevant services is large[15]. Thus, several methods have occurred. Liu et al. (2005) [16] converted the NP-hard problem into a multi-objective optimization problem with constraints and proposed GODSS to realize the service selection with QoS global optimization. The elitist selection genetic algorithm (ESGA) is presented by Dong et al. (2009) [17], where the integer encoding is used as the encoding rule and the initial population selection strategy ensures that the fitness of selected service could not be worse than the average fitness.

On the other hand, the design of coding scheme of chromosome, fitness function, evolution operations and selection mechanism have direct effect on the efficiency and global astringency[18]. Zhang et al. (2003)[19] has proposed the binary strings of chromosome, where every gene in chromosome represented a service candidate with values 0 and 1. The one dimension coding scheme of chromosome was presented to describe the service composition by Canfora et al. (2004)[20]. Since the multiple paths and service replanning problem were usually not taken into account in one dimension coding scheme, the relation matrix coding scheme was proposed[19][21][22]. Zhang et al.(2007)[19] has also presented a new evolution function of population as well as the population policy based on the combination of population diversity and simulated annealing which can converge asymptotically to the optimal solution in a local area and is proved to be a powerful stochastic optimization technique.

IV Service composition in dynamic environment

Due to the dynamic execution environment, more and more studies have been into the changing contexts for the service composition. The QoS-aware environment is classified into four categories by Shen et al. (2011)[23]: service context, device context, policy context and user satisfaction enhancement. According to Yu et al. (2009) [24], the context constraints for service composition can be divided into three groups including execution context, coordination context and composition policy context, among which the last two groups are considered to be static while the first one is more dynamic.

Some studies connect these dynamic factors with service replanning. Jaeger et al. (2005) [25] discussed the improvement of service composition by redundant services which can be used when the system weak point is found. Both Zhang et al. (2006) [22] and Gong et al. (2008) [26] have referred to the service replanning by GA. Dai et al. (2009)[27] pointed out that the self-healing strategies mainly included re-selecting the service and preparing the backup.

V Conclusion

In this paper, we focus on QoS models and searching algorithms for service composition.

Based on workflow theory, we have introduced four elemental composition structure including sequence, parallel, choice and loop. And QoS criteria are summarized in three categories: task-based, performance-based and user-based.

Then, we have discussed mathematical programming and heuristic algorithm. Meanwhile, service composition in dynamic environment has also been summarized.

As a result, we have provided a comprehensive summary for current studies and our discussion for the latest researches will contribute to the future studies.

Acknowledgements

The work was supported in part by the National Natural Science Foundation of China under Grant No. 70901039 and 70732002, National Postdoctoral Science Foundation of China under Grant No. 20090450144 and National Postdoctoral Science Special Foundation of China under Grant No.201003565, Jiangsu Postdoctoral Science Foundation under Grant No. 0901104C.

References

[1] Cardoso, J., A. Sheth, et al. Quality of service for workflows and web service processes. Web Semantics: Science, Services and Agents on the World Wide Web, 2004, 1(3): 281-308.

[2] Zeng, L., B. Benatallah, et al. QoS-aware middleware for web services composition. IEEE Transactions on Software Engineering, 2004, 30(5) : 311-327.

[3] Yu, T. and K. J. Lin. Service selection algorithms for Web services with end-to-end QoS constraints. Information Systems and e-Business Management, 2005, 3(2): 103-126.

[4] Kobti, Z., Zhiyang, W. 2007. An adaptive approach for QoS-aware web service composition using cultural algorithms. AI'07 Proceedings of the 20th Australian joint conference on Advances in artificial intelligence, 2007: 140-149.

[5] Canfora, G., Di Penta, M., Esposito, R., Villani, M. L. A framework for QoS-aware binding and re-binding of composite web services. Journal of Systems and Software, 2008, 81(10): 1754-1769.

[6] Zeng Liang zhao，Benatallah B, Dumasm, et al. Quality driven Web services composition. Proc of the 12th International Conference on World Wide Web. New York: ACM Press, 2003: 411-421．