Web Service Composition Methods: A Survey
E. Pejman, Y. Rastegari, P. Majlesi Esfahani, A. Salajegheh

Abstract—Web service composition is one of the most challenging problems of recent years. The number of service providers is increasing, and along with that for a request they offer multiple services with the same functionality, so it makes the problem of composition quite complex. In this paper we present several methods on service composition. We group them into two categories: methods based on evolutionary algorithms (e.g. Genetic Algorithm) and methods based on non-evolutionary algorithms (e.g. Dynamic Programming, Heuristic Algorithms). In this paper we focus on syntactic service compositions based on Quality of Service (QoS).

Keywords-Web service; Quality of service; Service composition; Evolutionary algorithm.

I. INTRODUCTION

Web service is a promising technology that allows constructing and sharing independent and autonomous software [26]. Service interface publication, service discovery and service invocation are performed using XML-based standards, known as WSDL, UDDI and SOAP [15].

The Web service model consists of three entities, the service provider, the service registry and the service consumer. Fig. 1 shows a graphical representation of the traditional Web service model [24].

By adopting standard-based protocols SOAP, WSDL and UDDI, service components from different service providers can be conveniently integrated into a composite service regardless of their locations, platforms and/or execution speeds.

With growing demands, many service providers have started to offer different QoS service levels, to meet the need of different user groups [1]. Web service composition has received increasing attention by the research community in the past few years [25].

Fig. 2 gives a conceptual overview of the QoS-aware service composition problem. Given an abstract composition request, which can be stated in a workflow-like language (e.g. BPEL), the discovery engine uses existing infrastructure (e.g. UDDI) to locate available web services for each task in the workflow using syntactic (and probably semantic) functional matching between the tasks and service descriptions. As a result, a list of candidate web services is obtained for each task with different QoS attributes. The goal of QoS composition is to select one candidate service from each list such that the aggregated QoS values satisfy the user’s end-to-end QoS requirements (constraints). In service oriented environments, where deviations from the QoS estimates occur and decisions upon replacing some services has to be taken at run-time, the efficiency of the applied selection mechanism becomes crucial [22].

In this paper we discuss methods shown in Fig. 3. We group these methods in two categories: methods based on evolutionary algorithms (shown in dashed outline) such as Genetic Algorithm, Immune System Algorithm, PSO, etc. and methods based on non-evolutionary algorithms such as Dynamic Programming, Graph Theory, Integer Programming, etc. In this paper we discuss advantages and disadvantages of each group.
The reminder of this paper is organized as follows: In section II we discuss methods based on non-evolutionary algorithms. Section III we discuss methods based on evolutionary algorithms, and finally we conclude the paper with comparing these categories in section IV.

II. METHODS BASED ON NON EVOLUTIONARY ALGORITHMS

In this part we discuss several methods for service composition based on QoS attributes. These methods are based on non-evolutionary algorithms.

Yu, T. et al. propose several algorithms on service selection using QoS broker (A QoS broker collects QoS information about servers and makes decision (service candidates) for clients). In [1, 2] they propose two methods for service selection problem. The goal is to maximize the utility function while satisfying the constraints. Constraints are clients QoS requirements that should be satisfied in an end to end service composition (e.g. total Cost should not be more than 300$). There are several forms of to model a business process, like sequential, parallel, loop, etc. They propose their algorithm on sequential form and use only one constraint for composition problem. They use several QoS attributes like, Response time, Cost, Reliability and Availability.

They represent two different approaches for solving the problem. The first method for selection is called Combinatorial approach. They model the problem as Multiple-Choice Knapsack Problem (MCKP) that each item has a weight and a profit, and knapsack has a capacity. Each item in each service class is a candidate service that profit and weight presents service utility and QoS attribute respectively. Knapsack capacity is global constraint. The algorithm is to select an item from each service class such that the total utility under constraint is maximized. They formulate the problem as shown (1).

\[
\begin{align*}
\text{Max} & \quad \sum_{i=1}^{k} \sum_{j \in S_i} F_{ij} x_{ij} \\
\text{Subject to} & \quad \sum_{i=1}^{k} \sum_{j \in S_i} R_{ij} x_{ij} \leq R \\
& \quad \sum x_{ij} = 1
\end{align*}
\]

\[F_{ij}: \text{Utility value at step } i \text{ for candidate } j\]
\[r_{ij}: \text{Response time of candidate } j \text{ at step } i\]
\[R: \text{Total response time}\]

They use Dynamic programming to solve the problem and with using Pisinger’s algorithm they efficiently find a solution.

The second approach is graph theory. They model service composition to a constrained shortest path problem. In this approach each service candidate in each service class represents a node. They move QoS parameter from nodes to their corresponding edges and construct a DAG. Then by using several algorithms like Constrained Bellman-Ford (CBF), Constrained Shortest Path (CSP) they solve the problem.


Zeng et al. [6] present a platform named AgFlow that enables the quality-driven composition of Web services. In this platform QoS of web services is evaluated and then it selects web services that can optimize total QoS, it also takes into account users requirements. They consider several QoS attributes such as execution price, execution duration, reputation, successful execution rate and availability. They also consider several forms of workflow such as AND/OR and loop. They propose two approaches for service selection in the composition, local optimization and global planning. In local optimization approach they use SAW [7] technique to compute the overall QoS of each web service in order to
choose the optimal one. In global planning they take into account the global constraints. They use IP (integer programming) [8] to solve the problem. Although it is better than exhaustive searching but running time in IP is considerable. By unlooping loops in workflow, they also solve this form of workflows in the business process.

In local optimization approach, service selection is performed individually for each task and a service with best score is chosen. In global planning they take into account the global constraints. They use IP (Integer Programming) [8] to solve the problem. Although it is better than exhaustive searching but running time in IP is considerable. By unlooping loops in workflow, they also solve this form of workflows in the business process.

In paper [15] they use Genetic algorithm to tackle the composition problem. They use fitness function to compare solutions. Constraints are considered in fitness function and all forms of workflow in business process are taken into account. The approach for computing the overall QoS in composition is similar to Cardoso [16] except for loop that they use the method in [6].

First they encode the problem in to a genome and then Genetic operation performed on that. Fitness function considered multiple constraints in their method. In [20] they also try to tackle the problem with the idea of convex hull.

Alrifai et al. [21] propose an algorithm based on heuristics that decompose the main problem into sub problems and with solving these, according to their experiments, they found near to optimal solution. They use aggregation functions for computing overall QoS attributes. They take into account global constraints and because of the decomposition of QoS problem, they also decompose global constraints into n local constraints and these local constraints are checked in each of local selections. Evaluation shows that the algorithm performance is better than IP programming [6] and heuristic algorithm WSHEU [4]. In [22] they use the same approach except beside using heuristics, they use mixed integer programming (MIP) to find the optimal composition.

So far we discuss several methods and algorithms in service composition. Approaches that use Integer Programming, Dynamic Programming or Graph Theory, definitely find best solution, But complexity in these algorithms is high. Although, as stated earlier, some heuristic algorithms helps these algorithms to converge sooner.

III. METHODS BASED ON EVOLUTIONARY ALGORITHMS

In this part we discuss methods based on evolutionary algorithms.

In paper [15] they use Genetic algorithm to tackle the composition problem. They use fitness function to compare solutions. Constraints are considered in fitness function and all forms of workflow in business process are taken into account. The approach for computing the overall QoS in composition is similar to Cardoso [16] except for loop that they use the method in [6].

First they encode the problem in to a genome and then Genetic operation performed on that. Fitness function maximizes some QoS attributes while minimizing some of them and also a penalty for meeting constraints is also considered. The fitness function they use is shown in formula 3.

\[
F(g) = \frac{w_1 \text{Cost}(g) + w_2 \text{Response Time}(g)}{w_3 \text{Availability}(g) + w_4 \text{Reliability}(g)} + w_5 D(g)
\]

Where:

- \(w_1\): User define weight for each QoS attribute
- \(D(g)\): a penalty for meeting constraints

Jiuyun et al. [17] use immune algorithm to tackle the composition problem. First they encode the problem in to an antibody. This method has two steps: the Immune selection operation and Clonal selection operation. In immune selection, antibodies are proliferated and suppressed in order to control their density in the mating pool and also make sure antibodies that are helpful and potential (vaccine) will not be destroyed. In Clonal operation they use antibodies with high fitness as heuristic information for speeding up convergence. They consider several control flow operators of business process and take into account several QoS

\[
\text{Score}(S_i) = \sum_{j=1}^{N_S}(V_{ij} \ast W_j)
\]

Where:

- \(V_{ij}\): Matrix of QoS Attributes for services
- \(W_j\): User define weight for each QoS attribute
- \(i\): Represent number of tasks
- \(j\): Represent number of QoS attributes

In local optimization approach, service selection is performed individually for each task and a service with best score is chosen. Global planning they take into account the global constraints. They use IP (Integer Programming) [8] to solve the problem. Although it is better than exhaustive searching but running time in IP is considerable. By unlooping loops in workflow, they also solve this form of workflows in the business process.

In local optimization approach, service selection is performed individually for each task and a service with best score is chosen. Global planning they take into account the global constraints. They use IP (Integer Programming) [8] to solve the problem. Although it is better than exhaustive searching but running time in IP is considerable. By unlooping loops in workflow, they also solve this form of workflows in the business process.

In local optimization approach, service selection is performed individually for each task and a service with best score is chosen. Global planning they take into account the global constraints. They use IP (Integer Programming) [8] to solve the problem. Although it is better than exhaustive searching but running time in IP is considerable. By unlooping loops in workflow, they also solve this form of workflows in the business process.
attributes in fitness function like service cost, service response time, service availability and service reliability. Antibodies with best fitness are considered as vaccine in the algorithm.

[19] proposes an algorithm named ACAGA_WSC. They use the combination of ant colony and Genetic algorithm for solving service composition problem. First they model service composition in to Ant Colony algorithm but, as they mentioned, there are some important parameters in Ant Colony algorithm that have great effect on the algorithm and ones when we select them trough experiments, those are not suitable for other experiments, so they use Genetic algorithm to set the key parameters of the Ant Colony algorithm in order to obtain the great efficiency.

In [23] authors solve the problem with combination of Genetic and Tabu search algorithm. They compare their method only with genetic algorithm.

Chen, M. et al. [27] solve the problem of composition by using DPSO (Discrete Particle Swarm Optimization). In this method they consider each particle as a solution. Each particle has a position and a velocity. Particles try to change their positions according to two elements: first their last best particle has a position and a velocity. Particles try to change using DPSO (Discrete Particle Swarm Optimization). In this method only with genetic algorithm.

They consider different forms of workflow. They use formula (2) to compute the score.

\[ V = C_1 (X_{\text{best}} - X) + C_2 (X_{\text{best}} - X) \]  
\[ X = X + V \]  

(4)

IV. CONCLUSION

Interaction between services transform the world of information oriented internet to a cloud of services. In this wide and distributed environment we need service composition to answer different requests. Service composition is a NP-hard problem. The problem is by getting a composite plan and service candidates for each task in workflow languages (e.g. BPEL) as inputs, we need to find the best solution for the composition.

In this paper we studied several methods in service composition. We grouped them in two categories: methods based on evolutionary algorithms and methods based on non-evolutionary algorithms. Methods in first category definitely find an optimal solution to the problem but, complexity in these methods is high. Although, some heuristics help them to speed up and improve the complexity. Methods in second category converge faster but the main problem in this category is that the final solution is not necessarily optimal.

So we suggest for small scale environment or locations that business processes are simple or candidates are limited, using non-evolutionary algorithms is more efficient and in environments that the business processes are complex and service candidates are distributed, using methods based on evolutionary algorithms can achieve best solution in a reasonable time.

REFERENCES


