Towards the optimality of QoS-aware web service composition with uncertainty

Sen Niu

School of Computer and Information Engineering
Shanghai Polytechnic University,
Shanghai, China
and
School of Computer Engineering and Science,
Shanghai University,
Shanghai, 200444, China
Email: sens306314@gmail.com

Guobing Zou*

School of Computer Engineering and Science,
Shanghai University,
Shanghai, 200444, China
Email: gbzou@shu.edu.cn
*Corresponding author

Yanglan Gan

School of Computer Science and Technology,
Donghua University,
Shanghai, 201620, China
Email: ylgan@dhu.edu.cn

Yang Xiang and Bofeng Zhang

School of Computer Engineering and Science,
Shanghai University,
Shanghai, 200444, China
Email: xiang@shu.edu.cn
Email: bfzhang@shu.edu.cn

Abstract: Quality of service (QoS)-aware web service composition (QWSC) has recently become one of the most challenging research issues. Although much work has been investigated, they mainly focus on certain QoS of web services, while QoS with uncertainty exposes the most important characteristic in real and highly dynamic environment. In this paper, with the consideration of uncertain service QoS and user's preferences, we model the issue of uncertain QoS-aware WSC via interval number and translate it into a multi-objective optimisation problem with global QoS constraints of user's preferences. The encoded optimisation problem is solved by an non-deterministic multi-objective evolutionary algorithm, which exploits new genetic encoding schema, the strategy of crossover and uncertain interval Pareto comparison. To validate the feasibility, large-scale experiments have been conducted on simulated datasets. The results demonstrate that our proposed approach can effectively and efficiently find optimum composite service solutions set with satisfactory convergence.

Keywords: web services; uncertain QoS; WSC; web service composition; multi-objective optimisation.

Reference to this paper should be made as follows: Niu, S., Zou, G., Gan, Y., Xiang, Y. and Zhang B. (2019) 'Towards the optimality of QoS-aware web service composition with uncertainty', *Int. J. Web and Grid Services*, Vol. 15, No. 1, pp.1–28.

Biographical notes: Sen Niu is currently working as an Assistant Professor at the School of Computer and Information Engineering, Shanghai Polytechnic University, Shanghai, China. He received his PhD from the School of Computer Engineering and Science, Shanghai University. His research interests include web service composition, uncertain QoS, and automated planning. He has published around ten papers on international journals and conferences, including IEEE International Conference on Web Services, IEEE International Conference on Services Computing, *Journal of Internet Technology*, *International Journal of Web and Grid Services*, and *International Journal of Web Engineering and Technology*.

Guobing Zou is an Associate Professor in the School of Computer Engineering and Science, Shanghai University, China. He received his PhD in Computer Science from Tongji University, Shanghai, China, 2012. He has worked as a Visiting Scholar in the Department of Computer Science and Engineering at Washington University in St. Louis from 2009 to 2011, USA. His current research interests mainly focus on services computing, data mining, intelligent algorithms, and personalised recommendation. He has published around 50 papers on premier international journals and conferences, including *TSC*, *ICWS*, *SCC*, *IJWGS*, *AAAI*, *Knowledge-Based Systems*, *Applied Intelligence*, *Journal of Internet Technology*, *Journal of Web Semantics*, and *Soft Computing*. He served as a program committee member on several international conferences. He worked as a reviewer for *IEEE Transactions on Services Computing*, *Journal of Artificial Intelligence Research*, and *IEEE Transactions on Systems, Man and Cybernetics: Systems*.

Yanglan Gan is an Associate Professor in the school of computer science and technology, Donghua University, Shanghai, China. She received her PhD in Computer Science from Tongji University, Shanghai, China, 2012. Her research interests include bioinformatics, web services, and data mining. She has published more than 30 papers on premier international journals and conferences, including *Bioinformatics*, *BMC Bioinformatics*, *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, IEEE International Conference on Web Services, IEEE International Conference on Services Computing, Neurocomputing, and Knowledge-Based Systems. She served as a program committee member on varieties of international conferences. She worked as a reviewer for *BMC Bioinformatics*, *Knowledge-Based Systems*, APBC and ADMA.

Yang Xiang is currently a master student in the School of Computer Engineering and Science, Shanghai University, China. Before that, he received a Bachelor degree in Computer Science and Technology at Shanghai University, 2016. His research interests include Mashup service recommendation, and uncertain QoS. He has published a paper on the 24th IEEE International Conference on Web Services.

Bofeng Zhang is a Full Professor in the School of Computer Engineering and Science, Shanghai University. He received his PhD from the Northwestern Polytechnic University (NPU), China, 1997. He experienced a postdoctoral research at Zhejiang University from 1997 to 1999, China. He worked as a Visiting Professor at the University of Aizu from 2006 to 2007, Japan. He worked as a Visiting Professor at Purdue University from 2014 to 2015. His research interests include personalised service recommendation, intelligent human-computer interaction, and data mining. He has published more than 150 papers on international journals and conferences. He worked as the program chair for UUMA from 2011 to 2016. He also served as a program committee member for lots of international conferences.

This paper is a revised and expanded version of a paper entitled 'Towards uncertain QoS-aware service composition via multi-objective optimization' presented at IEEE 24th International Conference on Web Services (ICWS 2017), Honolulu, HI, USA, 25–30 June, 2017.

1 Introduction

Web services are platform-independent, modular, loosely coupled and self-describing distributed software components. They can be published, discovered and invoked over the internet using standard communication protocols. Thus, it has brought great potentials for service users to leverage online programming resources, which can accelerate software reuse and efficiently form loosely-coupled integrations among web services. As cloud computing and service-oriented architecture (SOA) paradigm play a key role in the development of enterprise application integration, web services are becoming the most important fundamental building blocks for fast establishing next generation applications. In many cases, however, no single web service in a service repository can satisfy a given service request with the demands of complex business processes. As a result, web service composition (WSC) technique is proposed and applied for the purpose of fulfilling a composite service request.

Web service composition is the task of combining a set of single web services together to create a complex, value-added and cross-organisational business process. Web services usually consist of two attributes, functional and non-functional properties. The non-functional property is also called Quality of Service (QoS). QoS is a broad concept that encompasses a group of non-functional properties, such as execute price, response time, latency time, availability, reliability, and successability (Hwang et al., 2015). For those web services providing the same functionality, QoS has recognised as one of the important criteria to differentiate them for efficient service discovery, optimal selection and dynamic composition. Especially, QoS-aware web service composition (QWSC) is becoming a hot research issue. To this end, the challenge is how to construct a composite service effectively and efficiently such that its overall QoS reaches global optimality, while all of the multiple QoS constraints can be satisfied for user's preference.

Many efforts on QWSC have been made in recent years. Several works calculate QoS value for composite web service by linear programming (LP) or mixed linear programming (MIP) (Zeng et al., 2003, 2004; Ardagna and Pernici, 2007; Huang et al., 2009; Yu et al., 2007; Jiang et al., 2010; Zheng et al., 2013) so as to get the global optimal solution. As the Artificial Intelligence (AI) algorithm applied on QWSC (Wu and Zhu, 2013), the authors in Cremene et al. (2015), Wagner et al. (2012) and Chen et al. (2014) model a QWSC problem into a multi-objective optimisation problem and solve it using a genetic algorithm. However, these approaches mainly focus on certain QoS of web services. The value of QoS is always uncertain in highly dynamic application environments on the internet.

Considering uncertain QoS, correlative approaches transform uncertain QoS into the certain QoS (Wang et al., 2011; Sun et al., 2014) or build the stochastic model (Chattopadhyay et al., 2016; Xia et al., 2015), and then find the optimal composite web service by MIP algorithm. Although some of the existing works have been carried on dynamic composition of web services with the consideration of QoS uncertainty, the disadvantage is three-fold for the problem solving of UQ-WSC. First, the most existing uncertain QWSC (UQ-WSC) problems have not yet considered multi-dimensional objective functions, while in most cases an UQ-WSC problem for a user's preferences should be modelled as the one with multi-dimensional QoS constraints and goal optimisation. In such case, solving an UQ-WSC problem is more adaptable for real-world applications. Therefore, it becomes a more challenging task to handle an UQ-WSC problem with multi-objective functions from user non-functional constraints. Second, it lacks of a formal mapping mechanism and modelling method for encoding an UQ-WSC problem with multiple global constraints to the one having multi-objective functions for optimality. Third, an effective and efficient algorithm is mandatory to solve the problem that integrates the features of multiple global constraints and objective functions. Consequently, how to design and propose a novel approach to deal with QWSC problem with multi-objective optimisation taking into account QoS uncertainty has become a challenging research issue.

To address above issue, this paper proposes a novel approach to find the solutions with optimum composite services for the first time. Initially, we formulate the WSC with QoS uncertainty as an UQ-WSC problem, where the QoS uncertainty of web services is modelled as different ways of expression, including interval number and real number. Then, an UQ-WSC problem is transformed as an interval number-based multi-objective optimisation problem with user's preference constraints (IMOPs). Finally, a new non-deterministic multi-objective evolutionary algorithm based on decomposition (NDmoea/d) is proposed for solving the IMOPs. The algorithm applies an interval Pareto theory for the comparison of interval number among candidate composite services that takes advantage of the Tchebycheff approach for decomposition based on the MOEA/D algorithm (Zhang and Li, 2007). Large-scale experimental evaluation has been carried to validate the feasibility of our proposed approach. The experimental results demonstrate that the proposed approach can be used to solve the UQ-WSC problem effectively and efficiently with satisfactory optimum composite services found for active users.

In a nutshell, the major contributions of our work are as the following:

- We propose a novel framework to solve an uncertain QWSC problem and can get all the optimal composite solutions effectively.

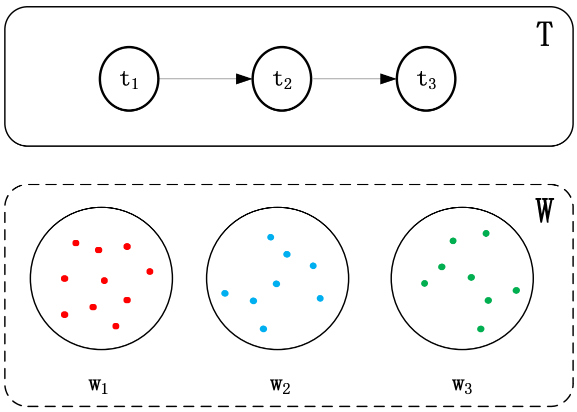
- A optimisation model strategy is presented, which translate an UQ-WSC problem into an interval number based multi-objective optimisation problem. Then, a improved evolutionary algorithm is proposed to solve the IMOPs problem.
- An extensive set of experiments are conducts to evaluate our approach in a benchmark QWS2 dataset, which includes measurements of 9 QoS attributes for 2507 real-world web services. The results show that our approach is very effective and has a good convergence.

The rest of this paper is organised as follows. A motivating example is illustrated in section 2. UQ-WSC problem is formulated in Section 3. The approach is presented in Section 4. The experiments are reported in Section 5. Finally, Section 6 reviews related work and Section 7 concludes the paper.

2 Motivating example

In this section, we introduce a motivating example to make the uncertain QWSC problem clearer and illustrate the necessity and significance of our work. Let's assume that there is a web service repository $W = \{w_1, w_2, w_3\}$ and a predefined composite service workflow $T = \{t_1, t_2, t_3\}$. When a user asks for a service request by providing a set of global constraints of QoS preferences $C = \{c_1, c_2, ..., c_n\}$. The goal of finding a feasible solution to an uncertain QWSC problem is to select a web service for each task from composite service workflow, when checking the compatibility of global QoS constraints to satisfy user preferences. A composite service workflow model and web service repository are illustrated in Figure 1.

Figure 1 A composite service workflow model and web service repository in an uncertain QWSC problem (see online version for colours)



As shown in Figure 1, the composite service workflow model consists of three business process tasks. Randomly two adjacent tasks t_1, t_2, and t_3 are dependently executed in sequence. Every task corresponds to a set of candidate services with the same functionality $w_1 = \{ws_{11}, ws_{12}\}$, $w_2 = \{ws_{21}, ws_{22}\}$ and $w_3 = \{ws_{31}, ws_{32}\}$ in web service repository W. We assume that every web service has three QoS criteria Q, including response time, availability, and latency time. They are invoked several times highly dynamic application environment on the internet. The invoked QoS transactional records of each web service on three non-functional criteria are shown in Table 1.

As shown in Table 1, each task has two candidate web services to select for a composite service. Every web service has several QoS transaction logs. The QoS values of a web service may vary in different times due to its execution in an uncertain and dynamic invocation environment. Now when a set of global QoS contraints $C = \{c_{rt}, c_a, c_{lt}\} = \{250, 0.6, 15\}$ are given by a user, an UQ-WSC problem is to achieve the object of finding a composite service where a web service is selected for a workflow task, making the comprehensive QoS of the generated composite service as lowest as possible, the RT and LT of the found composite service shortest and the availability highest, which can satisfy the global QoS constraints of user preferences C.

As we know, if a user does not provide the QoS constraints C, the example has eight feasible composite service solutions. They are the combinations of all the candidate services $< \{ws_{11}, ws_{12}\}, \{ws_{21}, ws_{22}\}, \{ws_{31}, ws_{32}\} >$. However, when the user gives the preferences $C = \{250, 0.6, 15\}$, it will get four feasible solutions satisfying users' constraints, including $< ws_{11}, ws_{21}, ws_{31} >$, $< ws_{11}, ws_{22}, ws_{31} >$, $< ws_{12}, ws_{21}, ws_{31} >$ and $< ws_{12}, ws_{22}, ws_{31} >$. Under the global QoS constraints, the user wants to get the optimal solutions in term of QoS dimension criteria, $< ws_{11}, ws_{21}, ws_{31} >$ and $< ws_{12}, ws_{21}, ws_{31} >$ are generated as the final solutions of the optimum composite services. They are not dominated with each other. For example, the comprehensive QoS of response time and availability for $< ws_{11}, ws_{21}, ws_{31} >$ is much more better than that of the others. But its' latency time is worst.

Table 1 Uncertain QoS transactional records on each candidate web service for a workflow task

T	*W*	*S*	*Resp. time (ms)*	*Availability*	*Late. time (ms)*
t_1	w_1	ws_{11}	40	0.85	3
			43	0.85	4
		ws_{12}	37	0.8	2
			38	0.8	3
			46	0.8	3
t_2	w_2	ws_{21}	82	0.9	1
			60	0.9	2
			70	0.9	3
			50	0.9	1
		ws_{22}	50	0.85	4
			70	0.85	3
			90	0.85	2
t_3	w_3	ws_{31}	108	0.85	4
			109	0.85	3
			108	0.85	5
		ws_{32}	122	0.9	8
			112	0.9	1
			134	0.9	1

That is very different from traditional QWSC problem, where each web service is only invoked once time from the perspective of QoS, i.e., its non-functional transactional record does not hold the feature of uncertainty. Since there are multiple QoS transaction logs and global constraints on user preferences in the motivation scenario, the approaches for traditional QWSC problem can not be directly applied to solve uncertain QWSC problem. In the following, to tackle the challenge, we focus on how to model, encode and solve the UQ-WSC problem.

3 Problem formulation

In this section, we focus on modelling the problem of uncertain QWSC which is formulated by a set of mutually correlative definitions.

Definition 1 (Web Service)**:** A web service ws is a 2-tuple $\langle F, NF\rangle$, where F and *NF* represent the functional and non-functional properties of web service.

QoS has been mostly applied to represent non-functional properties of web services. Here, ws can be denoted as $\langle F, Q\rangle$, where Q is the QoS and $ws.Q$ stands for the QoS of ws.

Definition 2 (Web Service Repository)**:** A web service repository consists of a set of available services. It can be denoted as $W = \{w_1, w_2, \cdots\} = \{ws_1, ws_2, \cdots\}$, where $\forall w_i \in W$ is a sub-service-repository, a set of web services with the same functionality and $\forall ws_i \in W$ is a web service with QoS characteristics.

In a web service repository, there are many single web services. When a group of web services with the same functionality, QoS can be applied to differentiate their quality on invocation and execution. For the better representation of QoS, we define QoS criteria as below.

Definition 3 (QoS Criteria)**:** Given a web service *ws*=$\langle F, Q\rangle$, the QoS value of ws originates from a set of QoS attributes, denoted as $ws.Q = \{q_1(ws), q_2(ws), \ldots, q_n(ws)\}$, where $q_i(ws) \in ws.Q$,$(i = 1, 2, \ldots, n)$ represents one facet on non-functional QoS values of ws.

QoS criteria for evaluating the quality of a web service can be modelled by a multi-dimensional vector, such as {*price, response time, latency time, reliability, availability, successability*}. QoS criteria can be divided into positive and negative categories. Positive QoS criteria denote better quality with higher values, while negative ones correspond to lower quality with higher values. For example, price, response time and latency time are negative criteria, whereas reliability, availability and successability are positive categories.

Definition 4 (Uncertain QoS of Web Service)**:** Given a web service $ws \in W$ and a set of QoS criteria with n attributes $Q = \{q_1, q_2, \cdots q_n\}$, the uncertain QoS of web service ws is formalised as a matrix $M_{m\times n}$ after invoking the service m times. The matrix $M_{m\times n}$ is denoted as

$$M_{m*n} = \begin{pmatrix} ws.Q_1 \\ ws.Q_2 \\ \vdots \\ ws.Q_m \end{pmatrix} = \begin{pmatrix} q_{11} & q_{12} & \cdots & q_{1n} \\ q_{21} & q_{22} & \cdots & q_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ q_{m1} & q_{m2} & \cdots & q_{mn} \end{pmatrix}$$

where $ws.Q_i$ $(i = 1, 2, \cdots m)$ represents the transactional QoS of web service ws on the i^{th} time, $q_{ij}(i = 1, 2, \ldots, m)(j = 1, 2, \ldots, n)$ represents value on the j^{th} QoS criterion during the invocation of the i^{th} time of ws.

Example 1: Let us assume a web service ws_1 called 'AmazonSearchService'. Its functionality is to search service on Amazon and its uncertain QoS during three times of invocation and execution are shown in Table 2.

Table 2 The QoS value of web service ws_1 during three times of invocation and execution

Times	*Price*	*Response time*	*Latency time*	*Reliability*	*Availability*	*Successability*
1	7	64.96	5.15	0.6	0.5	0.78
2	7	68.91	5.91	0.6	0.5	0.78
3	7	47.27	2	0.6	0.5	0.78

In Table 2, ws_1 has been invoked for 3 times and applied QoS criteria includes price, response time, latency time, reliability, availability and successability. In particular, price, response time and latency time are negative criteria, while others are positive criteria. The uncertain QoS matrix of ws_1 can be expressed by

$$M_{3*6}(ws_1) = \begin{pmatrix} 7 & 64.96 & 5.15 & 0.6 & 0.5 & 0.78 \\ 7 & 68.91 & 5.91 & 0.6 & 0.5 & 0.78 \\ 7 & 47.27 & 2 & 0.6 & 0.5 & 0.78 \end{pmatrix}$$

where each row and column represent an invoked transaction log across all of the QoS criteria and all of the values on a specified QoS criterion among the three invocation times, respectively.

Definition 5 (QoS Constraints of User Preferences)**:** Given a group of QoS criteria $Q = \{q_1, q_2, \cdots q_n\}$, QoS constraints of user preferences correspond to a set of values denoted as $C = \{c_1, c_2, \cdots c_n\}$. Each $c_i \in C$ ($c_j \in C$) is a lower bound on a positive QoS criterion q_i (an upper bound on a negative QoS criterion q_j).

Each QoS constraint of user preferences in C is used to restrict on its corresponding QoS criterion as a global satisfactory condition when finding optimum composite services for an active user. It is integrated into the definition of UQ-WSC problem as below.

Definition 6 (UQ-WSC Problem)**:** An uncertain QWSC problem is defined as a 5-tuple, denoted as *UQ-WSC*=$\langle W, T, f, C, Q\rangle$, where W is a service repository, $T = \{t_1, t_2, \cdots\}$ consists of a set of abstract workflow tasks, which each task t_i corresponds to a candidate web service repository w_i, $f : t_i \to t_j$ represents a mapping for relationship between randomly two tasks, $C = \{c_1, c_2, \ldots, c_n\}$ is designated as multi-dimensional global constraints of user preferences, which is based on a set of given QoS criteria $Q = \{q_1, q_2, \ldots, q_n\}$.

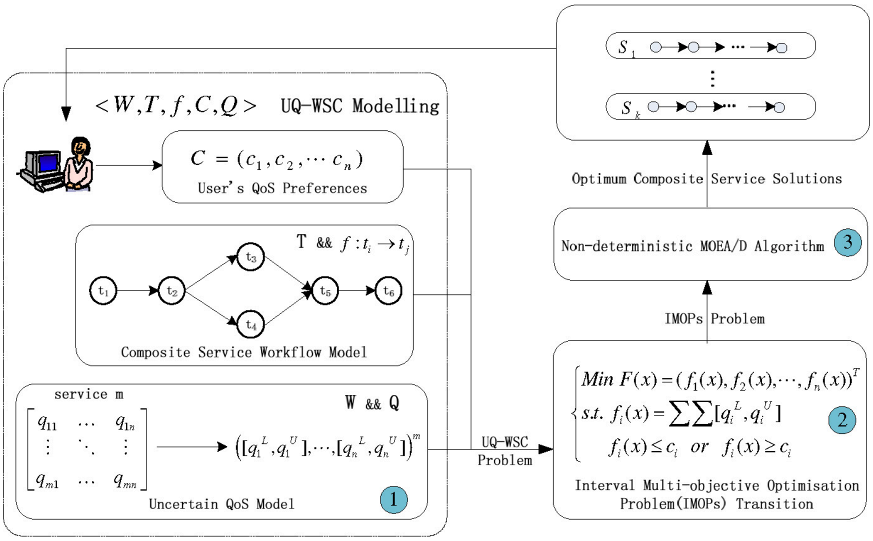
An UQ-WSC problem defines a composition problem where a user can specify multiple QoS preference constraints based on a service repository and workflow model. Note that every service in an UQ-WSC problem has uncertain QoS formulated as a matrix to model multiple service transactional records, which are totally different from traditional certain QWSC.

Example 2: In the motivating example, the UQ-WSC problem above *UQ-WSC*=$\langle W, T, f, C, Q\rangle$ can be represented by a available web service repository $W = \{w_1, w_2, w_3\} = \{ws_{11}, ws_{12}, ws_{21}, ws_{22}, ws_{31}, ws_{32}\}$, a set of abstract service tasks $T = \{t_1, t_2, t_3\}$, the relationship function of invocation dependency between two tasks $f(t_1 \to t_2)$ and $f(t_2 \to t_3)$ (sequence structure), global QoS constraints of user preferences $C = \{250, 0.6, 15\}$ and QoS criteria $Q = \{q_{rt}, q_a, q_{lt}\}$.

4 Approach

To solve an UQ-WSC problem, we propose an approach including three components. In this section, a framework is initially illustrated for the whole architecture of our approach. Then, uncertain QoS model of composite service is presented. Next, we translate an UQ-WSC problem into an interval number based multi-objective optimisation problem with user's preference constraints. Finally, an interval non-deterministic MOEA/D algorithm is presented to solve the problem and find the optimum composite services.

Figure 2 Overview of our approach for solving an UQ-WSC problem using multi-objective optimisation transition and evolutionary algorithm (see online version for colours)



4.1 Framework of solving an UQ-WSC problem

We develop an approach for uncertain QWSC using multi-objective optimisation transition and evolutionary algorithm. Figure 2 illustrates an overview of how to solve an UQ-WSC problem using the approach. It goes through three crucial steps:

- modelling uncertain QoS of composite service via interval number

- encoding an UQ-WSC problem as an interval number based multi-objective optimisation problem with global constraints (IMOPs)
- solving the transformed IMOPs problem using an improved non-deterministic MOEA/D evolutionary algorithm.

More specifically, we first take advantage of a matrix to formulate multiple execution process of a web service in real-world applications. It is shrunk to a vector where each dimension calculates an interval number to represent the uncertainty of web service on a QoS criterion. Based on the uncertain model of singleton web service, we model uncertain QoS of composite service, where varieties of invocation relationships among web services have been taken into consideration. Then, given user's QoS preferences and the workflow of composite service, we transform an uncertain QWSC problem into an interval number based multi-objective optimisation problem with global constraints. To solve the problem, we finally propose an improved evolutionary algorithm, named Non-deterministic MOEA/D algorithm, to generate the optimum composite service solutions.

4.2 *Uncertain QoS model of composite service*

To transform an UQ-WSC problem into an interval number based multi-objective optimisation problem with global constraints, we need to model uncertain QoS of composite service. First, the uncertain QoS of singleton web service is represented via matrix and interval number. Then, uncertain QoS model of composite service is built on correlative singleton services with global constraints.

4.2.1 *Uncertain QoS model of singleton service*

During the invocation and execution of a web service, it produces multiple uncertain transactional QoS records that has been formulated as a matrix. In this section, we model the uncertain QoS of singleton service as a QoS vector via interval number as follows.

Definition 7 (Interval Number)**:** Let $\Re$ is a set of real numbers. A closed interval X is an interval number, if $X = [x^L, x^U] = \{x | x \in \Re, x^L \leq x \leq x^U\}$, where x^L is the lower bound and x^U is the upper bound of interval number.

When an interval number $X = [x^L, x^U]$ satisfies $x^L = x^U$, it is compressed as a real number. The general rules of interval number on mathematical calculation are similar to the rules of the set. Given a matrix M_{m*n}, we apply interval number to model uncertain QoS of singleton service.

Definition 8 (Interval Number of Uncertain QoS)**:** Given a web service $ws \in W$ and its uncertain QoS matrix $M_{m\times n}$, the matrix $M_{m\times n}$ can be modelled as a vector, where each dimension is calculated as an interval number for uncertain QoS. It is formulated as below.

$$M_{m*n} = \begin{pmatrix} ws.Q_1 \\ ws.Q_2 \\ \vdots \\ ws.Q_m \end{pmatrix} = \begin{pmatrix} q_{11} & q_{12} & \cdots & q_{1n} \\ q_{21} & q_{22} & \cdots & q_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ q_{m1} & q_{m2} & \cdots & q_{mn} \end{pmatrix}$$

$$= \left([{q_1}^L, {q_1}^U] \, [{q_2}^L, {q_2}^U] \cdots [{q_n}^L, {q_n}^U] \right)$$

where ${q_i}^L$ and ${q_i}^U$ are the lowest and highest QoS value for its corresponding QoS criterion $q_i(i = 1, 2, ..., n)$ of ws, respectively.

In Example 1, the uncertain QoS attributes are represented as interval number and others are expressed by real number. So, uncertain QoS matrix of web service ws_1 can be modelled by a vector via interval number as $M_{ws_1} = \{7, [47.27, 68.91], [2, 5.91], 0.6, 0.5, 0.78\}$.

Definition 9 (Arithmetic on Interval Number)**:** Given two interval numbers $A = [a^L, a^U]$ and $B = [b^L, b^U]$, the arithmetic operations of addition and multiplication between A and B are denoted as $A + B = [a^L + b^L, a^U + b^U]$ and $A * B = [min(a^L b^L, a^L b^U, a^U b^L, a^U b^U), max(a^L b^L, a^L b^U, a^U b^L, a^U b^U)]$, respectively.

Assume that there are two interval numbers $A = [1, 2]$ and $B = [3, 5]$. Following the definition, the addition is $A + B = [1 + 3, 2 + 5] = [4, 7]$ and the multiplication is $A * B = [\min(3, 5, 6, 10), \max(3, 5, 6, 10)] = [3, 10]$.

4.2.2 Uncertain QoS composite model with global constraints

Given an UQ-WSC problem, the QoS value of composite service relies on the invocation patterns of predefined workflow model. There are usually four patterns, including sequence, parallel, switch and loop. In this paper, we mainly investigate composite service workflow structure in sequence. Since response time and latency time may be changed at every invocation process, they are expressed by interval number to represent the uncertainty of QoS. However, for other QoS criteria (e.g., availability, reliability and successability), they should be calculated as a real number by statistics from uncertain invocation QoS values within a period of time. So, given a composite service, the aggregative value on each QoS criterion within a composite service is calculated by arithmetic as below.

- *Price*(*p*). Given a workflow model for uncertain WSC, its price Q_p is the sum of execution price from all the services selected for tasks in workflow model. When a global constraint on execution price of composite service c_p is given by a user, the inequality of uncertain QoS composite model on execution price is formulated as

$$Q_p = \sum_{i=1}^{k} q_p(ws_i) = \sum_{i=1}^{k} q_{p_i} \leq c_p \tag{1}$$

  where k is the number of web services selected for workflow model, $q_p(ws_i)$ is the execution price of ws_i and c_p is an upper bound global QoS constraint on execution price that a user can afford to. q_p is the QoS on execution price of ws_i.

- *Response time*(*rt*). Given a workflow model for uncertain WSC, the response time Q_{rt} of composite service is accumulated by all the services selected. Under the global constraint c_{rt} from a user preference, the inequality of uncertain QoS composite model on response time is formulated as

$$Q_{rt} = \sum_{i=1}^{k} q_{rt}(ws_i) = \sum_{i=1}^{k} [{q_{rt}}^L, {q_{rt}}^U]_i \leq c_{rt} \tag{2}$$

  where $q_{rt}(ws_i)$ is the value of response time of ws_i and c_{rt} is an upper bound global QoS constraint on response time desired by a user.

- *Latency time*(lt). Latency Time is the time taken by the server to process a given request. Given a workflow model for uncertain WSC, the latency time of a composite service Q_{lt} is evaluated by the sum of the period for processing a request from all the web services selected. Based on the given user's preference c_{lt} as an upper bound, the inequality of uncertain QoS composite model on latency time is formulated as

$$Q_{lt} = \sum_{i=1}^{k} q_{lt}(ws_i) = \sum_{i=1}^{k} [{q_{lt}}^L, {q_{lt}}^U]_i \leq c_{lt} \tag{3}$$

- *Reliability*(r). Reliability is the ratio of the number of error messages to total messages, which is a positive QoS criterion. Given a workflow model for uncertain WSC, its reliability Q_r for a composite service is the product of probability form all the services selected. Using global QoS constraint c_r as a lower bound, we have the inequality of uncertain QoS composite model on reliability as

$$Q_r = \prod_{i=1}^{k} q_r(ws_i) = \prod_{i=1}^{k} q_{ri} \geq c_r \tag{4}$$

- *Availability*(a). Given a workflow model for uncertain WSC, its availability Q_a is evaluated by the product of probability from all the services included. we have the inequality of uncertain QoS composite model on availability as

$$Q_a = \prod_{i=1}^{k} q_a(ws_i) = \prod_{i=1}^{k} q_{ai} \geq c_a \tag{5}$$

  As a lower bound, c_a is represented as a global QoS constraint from a user preference.

- *Successability*(s). Given a workflow model for uncertain WSC, the successability of a composite service Q_s is calculated by the product of probability from all the services selected. Given a global QoS constraint c_s as a lower bound, we have the inequality of uncertain QoS composite model on successability as

$$Q_s = \prod_{i=1}^{k} q_s(ws_i) = \prod_{i=1}^{k} q_{si} \geq c_s \tag{6}$$

Following the above uncertain QoS model for both singleton web service and a composite service, we can identify the satisfiability of each possibly composite service solution under multiple global QoS constraints of user preferences.

4.3 UQ-WSC problem transition

Given an UQ-WSC problem $< W, T, f, C, Q >$, we assume that a user provides QoS preferences C and asks for a composite service, which originates from T workflow tasks and the relationship function f between two randomly adjacent tasks produces sequence. Now we have a service repository W, where m disjoint candidate services maps to an abstract workflow task in terms of equivalent functionality. Every web service in W exploits six QoS attributes in Q to make transactional records, including price, response time, latency time, reliability, availability and successability. We aim to find a composite service solution making its comprehensive QoS value as lowest as possible. The price, response time and latency time of composite service are lowest and shortest, the reliability, availability and successability highest. To solve the UQ-WSC problem, we encode it into an interval number based multi-objective optimisation problem with global QoS constraints.

A traditional multi-objective optimisation problem includes a minimisation function and constraint function, which is described as

$$\begin{cases} \dot{M}in\ F(x) = (f_1(x), f_2(x), \ldots, f_n(x))^T \\ s.t.\ x \in \Omega \end{cases} \tag{7}$$

where Ω refers to the decision variable space, $F : \Omega \to R^m$ consists of n real-valued objective functions $f_i(x)(i = 1, 2, \ldots, n)$ and the objective space R^m.

Following the formulation of traditional multi-objective optimisation model, an UQ-WSC problem can be encoded as an interval number based multi-objective optimisation problem based on uncertain QoS model of composite service as below.

To determine the best combination, we firstly define a variable s_{ij}, which represents the jth candidate service ws_j from web service repository w_i in W for task i in a given composite workflow model $< T, f >$.

$$s_{ij} = \begin{cases} 1, \text{ if service } j \text{ is selected for task } i; \\ 0, \text{ otherwise} \end{cases} \tag{8}$$

In a composite workflow model, only one service is selected for a task, when finding a feasible solution to an UQ-WSC problem. Therefore, the variables must obey the following conditional constraint.

$$\sum_{j=1}^{m} s_{ij} = 1,\ 1 \le i \le |T| \tag{9}$$

where m is the number of candidate services with same functionality for task i. The constraint equation above ensures that only one candidate service is selected for each task in workflow T.

A multi-objective optimisation problem has several sub-objective functions. As for our UQ-WSC problem, including six QoS criteria, the interval number based encoding problem corresponds to six sub-objective functions, making the price, response time and latency time of composite service lowest and reliability, availability and successability highest. They are combined together as the optimisation function. These objective functions are described as below.

$$
\begin{cases}
Min\ F(x) = (f_p(x), f_{rt}(x), f_{lt}(x), -f_r(x), -f_a(x), -f_s(x))^T \\
f_p(x) = \sum\limits_{i=1}^{|T|} \sum\limits_{j=1}^{m} (s_{ij} * q_{p_{ij}}) \\
f_{rt}(x) = \sum\limits_{i=1}^{|T|} \sum\limits_{j=1}^{m} (s_{ij} * [{q_{rt}}^L, {q_{rt}}^U]_{ij}) \\
f_{lt}(x) = \sum\limits_{i=1}^{|T|} \sum\limits_{j=1}^{m} (s_{ij} * [{q_{lt}}^L, {q_{lt}}^U]_{ij}) \\
f_r(x) = \prod\limits_{i=1}^{|T|} \prod\limits_{j=1}^{m} (q_{r_{ij}})^{s_{ij}} \\
f_a(x) = \prod\limits_{i=1}^{|T|} \prod\limits_{j=1}^{m} (q_{a_{ij}})^{s_{ij}} \\
f_s(x) = \prod\limits_{i=1}^{|T|} \prod\limits_{j=1}^{m} (q_{s_{ij}})^{s_{ij}}
\end{cases}
\tag{10}
$$

where $f_p(x)$ is the sum of price with all candidate services in the composite workflow; $f_{rt}(x)$ and $f_{lt}(x)$ are the sum of response time and latency time; $f_r(x)$, $f_a(x)$ and $f_s(x)$ are the product of reliability, availability and successability, respectively; T and m are the number of the tasks in composite workflow model and the candidate services for each task; s_{ij} is the variable to demonstrate the status of service selection on ws_{ij} and $[{q_{rt}}^L, {q_{rt}}^U]_{ij}$ represents an interval number of service j for task i in response time. When $s_{ij} = 0$, $s_{ij} * [{q_{rt}}^L, {q_{rt}}^U]_{ij} = 0$ and ${q_{r_{ij}}}^{s_{ij}} = 1$; When $s_{ij} = 1$, $s_{ij} * [{q_{rt}}^L, {q_{rt}}^U]_{ij} = [{q_{rt}}^L, {q_{rt}}^U]_{ij}$ and ${q_{r_{ij}}}^{s_{ij}} = q_{r_{ij}}$.

In our defined UQ-WSC problem, a user specifies a set of global QoS constraints $C = \{c_p, c_{rt}, c_{lt}, c_r, c_a, c_s\}$, including global user preferences on execution price c_p, response time c_{rt}, latency time c_{lt}, reliability c_r, availability c_a, and successability c_s. As a result, they are applied as the global QoS upper or lower bounds, which lead to the following six inequalities as conditional constraints in our transformed optimisation problem.

$$
\begin{cases}
f_p(x) \leq c_p \\
f_{rt}(x) \leq c_{rt} \\
f_{lt}(x) \leq c_{lt} \\
f_r(x) \geq c_r \\
f_a(x) \geq c_a \\
f_s(x) \geq c_s
\end{cases}
\tag{11}
$$

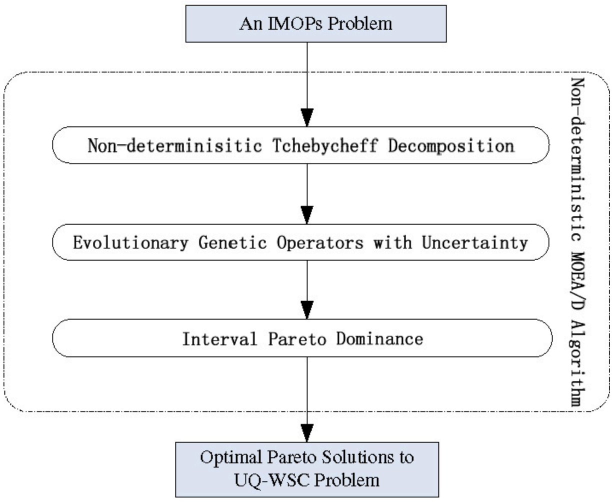
By the above encoding, a UQ-WSC problem is translated into an interval number based multi-objective optimisation problem, which consists of a main objective function as well as six derivable sub-objective functions (10) and multiple global constraints (8)–(11). To solve the encoded multi-objective optimisation problem with uncertainty, we present an improved evolutionary algorithm in the next section.

4.4 Solving IMOPs with improved evolutionary algorithm

Given a multi-objective optimisation problem, many methods have been proposed in recent years, such as mathematical model (Ruzika and Wiecek, 2005; Wiecek et al., 2001), boundary intersection method (Messac et al., 2003; Mattson et al., 2004) and evolutionary algorithm (Deb et al., 2002; Coello et al., 2004). However, these highly effective and efficient approaches cannot be directly applied to solve an interval number based multi-objective optimisation problem modelled in this paper.

Based on classic MOEA/D algorithms (Zhang and Li, 2007)), we present an improved evolutionary algorithm called NDmoea/d (Non-deterministic multi-objective optimisation based decomposition) algorithm to solve an IMOPs encoded from an UQ-WSC problem. The theory architecture of NDmoea/d algorithm is shown in Figure 3.

Figure 3 The theory architecture of NDmoea/d algorithm (see online version for colours)



As shown in Figure 3, the input of algorithm is an IMOPs problem, which encoded from an Uncertain QWSC problem and the output is optimal composite service solutions for users. More specifically, the encoding problem is first decomposed into several single objective optimisation problems by Tchebycheff decomposition strategy. Then, genetic operators (crossover and mutation) in classic evolutionary algorithm are applied to generate new solutions. Finally, an interval Pareto comparison approach is presented to reduce the solutions space from feasible composite service solutions and find all the optimal composite solutions satisfied user's global QoS preference constraints.

4.4.1 Tchebycheff decomposition strategy for IMOPs

Before applying an algorithm to find composite service solutions by encoded IMOPs, we decompose it into a set of single objective optimisation problems by effective existing approaches. In this paper, we take advantage of Tchebycheff decomposition approach (Jaszkiewicz, 2002) to decompose the main objective optimisation function of IMOPs into multiple single optimisation problems. The decomposed single objective optimisation problem can be described as

$$\begin{cases} \dot{M}in\ g^{te}(x|\lambda, z^*) = \max\limits_{1\le i\le n} \{\lambda_i |f_i(x) - z_i{}^*\} \\ s.t.\ x \in \Omega \end{cases} \tag{12}$$

Where $g^{te}(x|\lambda, z^*)$ is a objective function, x is the variables to be optimised, λ is a coefficient vector, and $z^* = (z_1{}^*, z_2{}^*, \ldots, z_n{}^*)^T$ is the reference point. For example, $z_i{}^* = \min\{f_i(x)|x \in \Omega\}, i = 1, 2, \ldots, n$. In an IMOPs, $f_i(x)$ is corresponding to six sub-objective functions, $f_p(x), f_{rt}(x), f_{lt}(x), f_r(x), f_a(x), f_s(x)$ respectively. For each interval Pareto optimal solution $x^* \in \Omega$ of an IMOPs, there exists a weight vector λ that makes x^* the solution optimal to a single objective optimisation problem. Similarly, each interval optimal solution of a single objective optimisation problem is also an interval Pareto solution to an IMOPs problem. By the way of finding different weight vectors, a set of interval Pareto solutions to IMOPs problem are obtained as the optimum composite services.

4.4.2 *Interval number based uncertain Pareto comparison*

Pareto theory is the most classic model for the evaluation of feasible solutions when solving a multi-objective optimisation problem. Meanwhile, possibility method for interval number comparison has been widely applied. By the integration, we present an interval number based uncertain Pareto comparison strategy to differentiate the quality of candidate composite services.

Definition 10 (Interval Possibility)**:** Given two interval numbers $a = [a^L, a^U]$ and $b = [b^L, b^U]$, let $L(a) = a^U - a^L$ and $L(b) = b^U - b^L$. The possibility of two interval numbers $p(a \geq b)$ is measured as (Gang and Feng, 2008):

$$p(a \geq b) = \begin{cases} 1 & b^L \leq b^U \leq a^L \leq a^U \\ 1 - \frac{1}{2}\frac{(b^U - a^L)^2}{L(a)L(b)} & b^L \leq a^L \leq b^U \leq a^U \\ \frac{1}{2}\frac{(a^L + a^U - 2b^L)}{L(b)} & b^L \leq a^L \leq a^U \leq b^U \\ \frac{1}{2}\frac{(2a^U - b^U - b^L)}{L(a)} & a^L \leq b^L \leq b^U \leq a^U \\ \frac{1}{2}\frac{(a^U - b^L)^2}{L(a)L(b)} & a^L \leq b^L \leq a^U \leq b^U \\ 0 & a^L \leq a^U \leq b^L \leq b^U \end{cases} \tag{13}$$

It has been proved that the interval possibility holds five great properties as follows.

- $0 \leq p(a \geq b) \leq 1$
- Iff $b^U \leq a^L$, $p(a \geq b) = 1$
- Iff $a^U \leq b^L$, $p(a \geq b) = 0$
- $p(a \geq b) + p(b \geq a) = 1$
- If $p(a \geq b) \geq \frac{1}{2}$, $p(b \geq c) = \frac{1}{2}$, then $p(a \geq c) \geq \frac{1}{2}$.

Given two interval numbers A and B, we can compare it for dominance recognition by using classic Pareto strategy via interval possibility, which is defined as below.

Definition 11 (Interval Pareto Dominance)**:** Given a feasible solution set Ω, let $x_A, x_B \in \Omega$, where x_A and x_B are two random feasible solutions to a multi-objective optimisation problem. We say that x_A is interval Pareto dominance to x_B, if and only if $\forall i \in \{1, 2, \ldots, n\}, f_i(x_A) \leq f_i(x_B)$ and $\exists j \in \{1, 2, \ldots, n\}, f_j(x_A) < f_j(x_B)$, denoted as $x_A \prec x_B$.

There are generally a set of feasible solutions for an interval number based multi-objective optimisation problem. When a feasible solution is better than another one in all of the sub-objective functions, that is the former feasible solution dominates the latter by applying interval Pareto dominance.

Definition 12 (Interval Pareto Solution)**:** Given a decision variable space Ω, a feasible solution $x^* \in \Omega$ is called as an interval Pareto solution for an IMOPs problem, if and only if it does not exist a feasible solution x, which satisfies $x \in \Omega$ and $x \prec x^*$.

Based on the interval Pareto dominance strategy, an interval number based multi-objective optimisation problem may exist multiple interval Pareto solutions, where they cannot dominate with each other.

4.4.3 Improved non-deterministic MOEA/D algorithm

To solve the problem of our encoded IMPOs, we present an improved non-deterministic MOEA/D algorithm, which is based on multi-objective evolutionary algorithm in combination with decomposition (MOEA/D). The algorithm integrates Tchebycheff decomposition approach to partition objective function and interval Pareto comparison strategy for optimum solutions choice.

As the same as MOEA/D algorithm, NDmoea/d algorithm decomposes an encoded IMOPs problem into n number of sub-objective optimisation problems, which can be solved simultaneously in a single run. At each generation, the algorithm needs to maintain the information as below.

- A population of N points: $x^1, x^2, \ldots, x^N \in \Omega$, where x^i is the current solution to the ith subproblem;
- $FV^1, FV^2, \ldots, FV^N$, where $FV^i = F(x^i)$ is the multi-objective optimisation function F-value of x^i, $i = 1, 2, \ldots, N$;
- $z = (z_1, z_2, \ldots, z_n)^T$, where z_i is the best value found so far for objective f_i;
- EP, an external population, which stores the Pareto solutions during the search for finding the optimum composite services.

More specifically, NDmoea/d algorithm as shown in Algorithm 1 as follow:

The NDmoea/d algorithm consists of three steps: initialisation, update and stopping criterion.

- In initialisation phrase, the algorithm initialises the population and the reference points, and obtains the neighbourhoods by calculating Euclidean distances between any two weight vectors.
- In update phrase, the algorithm reproduces a new feasible composite service solution by genetic operators, including crossover and mutation. On this basis, it updates the reference points, neighbouring composite service solutions and *EP* using the new feasible composite service solution.
- In stopping criterion phase, the algorithm can be stopped and outputs the optimum interval Pareto solutions, when the stopping criterion is satisfied.

Algorithm 1: NDmoea/d algorithm

Input:

- An IMOPs problem
- C: user QoS preferences
- N: the number of the subproblems considered in NDmoea/d (population size)
- $\lambda^1, \lambda^2, \cdots, \lambda^N$: a uniform spread of N weight vectors
- T: the number of the neighborhood of the each weight vector
- S: a stopping criterion

Output:

- *EP*: an external population, which is used to store non-dominated solutions found during the search

Step 1 Initialization:

Step 1.1 Set *EP* = $\varnothing$.

Step 1.2 Compute the Euclidean distances between any two weight vectors, and find out the T closest weight vectors to each weight vector as the neighborhoods. Set $B(i) = \{i_1, i_2, \cdots, i_T\}$, where $\lambda^{i_1}, \lambda^{i_2}, \cdots, \lambda^{i_T}$ are the T closest weight vectors to λ^i.

Step 1.3 Generate an initial population $x^1, x^2, \cdots, x^N \in \{0,1\}^n$ randomly, where $x^1, x^2, \cdots, x^N$ meets the constraint equation (8) and (9). Set $FV^i = F(x^i)$, $i = 1, 2, \cdots, N$.

Step 1.4 Initialize $z = (z_1, z_2, \cdots, z_n)^T$ randomly.

Step 2 Update:
For $i = 1, 2, \cdots N$, do

Step 2.1 **Reproduction:** Select two neighborhoods $\{k, l\}$ from $B(i)$ randomly, and then generate a new solution y' from x^k and x^l by using genetic operators.

Step 2.2 **Update of z:** For each $j = 1, 2, \cdots, m$, if $z_j < f_j(y')$, then set $z_j = f_j(y')$.

Step 2.3 **Update of Neighboring Solutions:** Following the ranking methods about interval numbers, if $g^{ws}(y|\lambda^j) \geq g^{ws}(x^j|\lambda^j)$ for each index $j \in B(i)$, then set $x^j = y$ and $FV^j = F(y)$.

Step 2.4 **Update of *EP*:** Remove all the vectors unsatisfied user QoS preferences C; Remove all the vectors dominated by $F(y)$ from *EP*; Add $F(y)$ to *EP*, if no vectors in *EP* dominate $F(y)$.

Step 3 **Stopping Criterion:** If stopping criterion S is satisfied, then stop and output *EP*. Otherwise, go to **Step 2**.

Different from the traditional MOEA/D algorithm, our proposed NDmoea/d algorithm has its own advantages for solving an IMOPs problem, where it integrates the feature of multi-objective optimisation functions and interval number formulation on each web service with its QoS uncertainty.

- In GA encoding schema, an individual in a population is encoded by the format as below,

$$(s_{11}\ s_{12} \cdots s_{1m}\ \ \&\ \ s_{21}\ s_{22} \cdots s_{2m}\ \ \&\ \ \cdots\ \ \&\ \ s_{T1}\ s_{T2} \cdots s_{Tm})$$

  where T is the number of tasks in a composite service workflow; m is the number of candidate services for every task; $s_{ij} \in \{0, 1\}$ represents that if the jth service in the ith abstract task is selected, $s_{ij} = 1$; otherwise $s_{ij} = 0$.

- When generating an initial population, the problem is an interval number based multi-objective optimal problem with global constraints. Therefore, each initial population for the algorithm must meet the constraint equations for an IMOPs problem encoded from UQ-WSC problem, $s_{ij} \in \{0, 1\}$ and $\sum_{j=1}^{m} s_{ij} = 1$.

- As for genetic operations, the algorithm partitions the encoding individuals into T tasks. In the process of crossover and mutation, all candidate services $(s_{i1}\ s_{i2} \cdots s_{im})$ for each task are as a whole genetic block. The new population is generated by crossover and mutation among blocks, which must satisfy the constraint equations.

- On comparison with dominated solutions, the algorithm apples the interval Pareto comparison strategy for uncertain QoS and generates the optimum composite services between two feasible solutions during the procedure of update.

5 Experimental evaluation

5.1 Experimental setup and datasets

In order to evaluate the effectiveness and efficiency of our proposed approach to generate the optimum composite services, we have designed and developed a prototype system where uncertain QoS model, UQ-WSC transition and NDmoea/d algorithm are implemented in Python. The experimental experiments have carried on a PC with Intel Dual Core 2.8 GHZ processor and 3G RAM in Windows 7.

We took advantage of response time, latency time, availability, reliability and successability as uncertain QoS criteria and generated a synthetic uncertain QoS dataset based on QWS2 dataset (Al-Masri, 2007), which includes the measurements of 9 QoS criteria for 2507 real-world web services. In our experiments, we firstly generated a normal distribution on interval number for uncertain QoS values of web services, which have 50 QoS transaction logs for every service in QWS2. It totals 2507*50 QoS invoked record in web service repository. Then, for applying the NDmoea/d algorithm, the crossover probability is set as 0.8 and the mutation probability is set as 0.1. Moreover, we dynamically set the number of abstract services tasks T in a composite workflow model, the number of candidate services N for each task in $< T, f >$, the iteration stopping criterion M, and the population P to evaluate the NDmoea/d algorithm. Finally, we evaluate and analyse the algorithm from two aspects, including the quality of the generated optimum solutions and the efficiency of finding these composite services.

5.2 The experimental results and analysis

From previous investigations, we observe that there are only few research efforts on uncertain QWSC and they mainly take advantage of comprehensive QoS value to find solutions, which means that they do not consider multi-objective optimisation problem for satisfying multi-dimensional user's preferences. Therefore, we have not made comparisons in our experiments. We assume there exists an abstract service workflow model where the number of tasks $T = 5$, the number of candidate services for each task $N = 3$. There are three QoS criteria for the description of nun-functional values of each web service, including response time, availability, and latency time. Let the interval possibility be $\lambda = 1/2$, the population be $P = 100$ and the iteration numbers of running the NDmoea/d algorithm be $M = 500$. Under above parameters setting, when given different QoS preferences, denoted as A, B, C, D, E, F, by a user, our proposed NDmoea/d algorithm can effectively find the optimum solutions of composite services. They are shown in the Table 3.

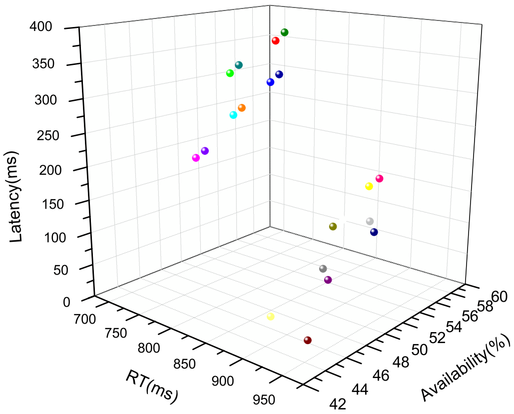
Table 3 The optimum interval Pareto solutions of composite services found by applying NDmoea/d algorithm in terms of different user's QoS preferences

ID	*C*	*Numbers*	*Solutions*	*Uncertain QoS of composite services*
A	(700, 0.7, 500)	0	–	–
B	(800, 0.5, 400)	4	$< ws_1, ws_5, ws_9, ws_{10}, ws_{13} >$	$([713.24, 727.69], 0.58, [348.84, 364.33])$
			$< ws_1, ws_6, ws_9, ws_{10}, ws_{13} >$	$([714.16, 728.10], 0.53, [309.98, 324.74])$
			$< ws_1, ws_5, ws_9, ws_{10}, ws_{15} >$	$([733.31, 747.40], 0.56, [289.31, 304.18])$
			$< ws_1, ws_6, ws_9, ws_{10}, ws_{15} >$	$([734.23, 747.82], 0.52, [250.45, 264.59])$
C	(900, 0.5, 300)	1	$< ws_1, ws_6, ws_9, ws_{10}, ws_{15} >$	$([734.23, 747.82], 0.52, [250.45, 264.59])$
D	(1000, 0.5, 200)	3	$< ws_2, ws_5, ws_9, ws_{10}, ws_{13} >$	$([892.85, 906.42], 0.55, [162.24, 177.58])$
			$< ws_2, ws_6, ws_9, ws_{10}, ws_{13} >$	$([893.77, 906.83], 0.51, [123.38, 137.99])$
			$< ws_2, ws_5, ws_9, ws_{10}, ws_{15} >$	$([912.93, 926.13], 0.54, [102.71, 117.43])$
E	(1000, 0.4, 100)	2	$< ws_2, ws_6, ws_9, ws_{10}, ws_{15} >$	$([913.84, 926.55], 0.49, [63.856, 77.847])$
			$< ws_2, ws_6, ws_9, ws_{11}, ws_{15} >$	$([953.35, 966.06], 0.44, [32.469, 45.774])$
F	(—, —, —)		$< ws_1, ws_5, ws_9, ws_{10}, ws_{13} >$	$([713.24, 727.69], 0.58, [348.84, 364.33])$
			$< ws_1, ws_6, ws_9, ws_{10}, ws_{13} >$	$([714.16, 728.10], 0.53, [309.98, 324.74])$
			$< ws_1, ws_5, ws_9, ws_{10}, ws_{15} >$	$([733.31, 747.40], 0.56, [289.31, 304.18])$
			$< ws_1, ws_6, ws_9, ws_{10}, ws_{15} >$	$([734.23, 747.82], 0.52, [250.45, 264.59])$
			$< ws_1, ws_6, ws_9, ws_{11}, ws_{15} >$	$([773.74, 787.33], 0.46, [219.07, 232.52])$
			$< ws_2, ws_5, ws_9, ws_{10}, ws_{13} >$	$([892.85, 906.42], 0.55, [162.24, 177.58])$
			$< ws_2, ws_6, ws_9, ws_{10}, ws_{13} >$	$([893.77, 906.83], 0.51, [123.38, 137.99])$
			$< ws_2, ws_5, ws_9, ws_{10}, ws_{15} >$	$([912.93, 926.13], 0.54, [102.71, 117.43])$
			$< ws_2, ws_6, ws_9, ws_{10}, ws_{15} >$	$([913.84, 926.55], 0.49, [63.856, 77.847])$
			$< ws_2, ws_6, ws_9, ws_{11}, ws_{15} >$	$([953.35, 966.06], 0.44, [32.469, 45.774])$

From Table 3, the approach based on NDmoea/d algorithm generates different numbers of interval Pareto solutions in terms of different user's QoS preferences. They make the response time and latency time lowest and availability highest. Especially, when without given any preferences by the user, 10 number of interval Pareto solutions can be found for the UQ-WSC problem, which include all the results by using above other user's QoS preferences.

We observe that in Table 3, all the solutions of composite services for a user's QoS preferences consist of an optimal solution set and cannot be dominated by each other with respect to the comprehensive QoS of composite services. To show the QoS value of composite service solutions in 3D space, we divide an interval number of QoS of composite services into a minimum number and a maximum number, which represent the lower bound and the upper bound of interval number, respectively. When a user does not give any QoS preferences, the QoS distribution space of composite services is shown in Figure 4.

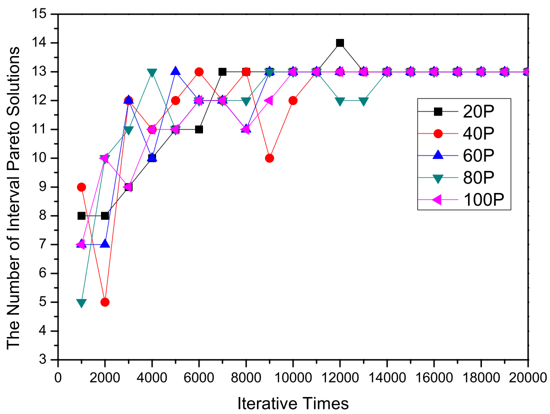
Figure 4 The QoS distribution space of composite service solutions (see online version for colours)



5.3 The algorithm analysis on convergence

The convergence speed is one of the most important index to evaluate the efficiency of our NDmoea/d algorithm. In the experiments, we evaluate the metric of convergence by calculating and observing the number of interval Pareto solutions to an original UQ-WSC problem. Let's assume that the number of abstract service tasks and candidate services for each task are $T = 5$ and $N = 5$, respectively. The population size ranges from 20 to 100 with an interval of 20. The iterative times of running our NDmoea/d algorithm vary from 1000 to 20000 with an interval of 1000. The experimental result, the number of generated interval Pareto solutions of composite services, is shown in Figure 5.

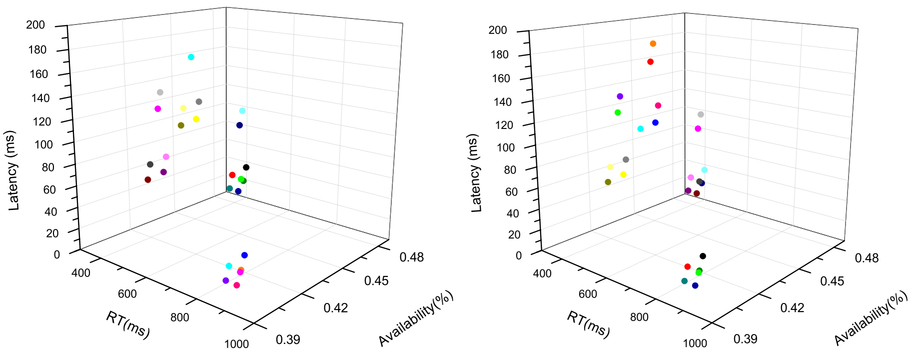
Figure 5 The experimental results on the number of interval Pareto solutions with the changes of different populations (see online version for colours)



Observed from Figure 5, we find that along with the changes of population, the number of interval Pareto solutions of composite services maintains 13 without any variation during iterative times from 14,000 to 20,000. As a result, we conclude that the number of interval Pareto solutions to an given IMOPs problem keeps steady at last, as the population and iterative times both grow up. Therefore, from the above experiments results, we conclude that the number of the optimum interval Pareto solutions of composite services keeps steady after a set of iterative times for running our NDmoea/d algorithm and the algorithm is convergent.

More specifically, we analyse and compare the QoS value of the optimum composite services with the setting of population size $P = 20$ and $P = 80$. The QoS distribution space of composite services is shown in Figure 6.

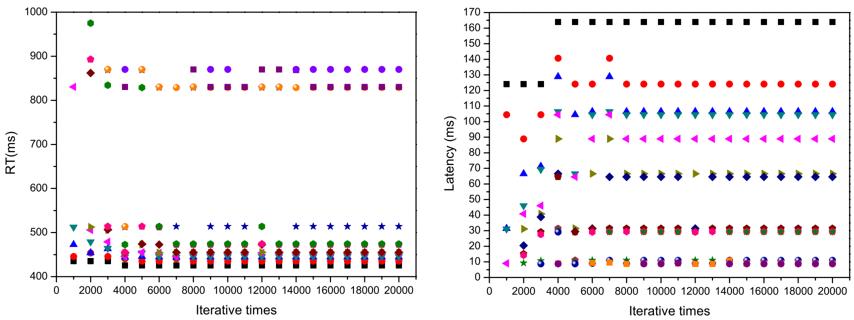
Figure 6 The QoS distribute value of composition services with $P = 20$ and $P = 80$ (see online version for colours)



From the experimental results in Figure 6, the QoS distribution space of composite service solutions is the same between the population $P = 20$ and $P = 80$ and the value of optimal composite service is the same.

To further verify the convergence of the NDmoea/d algorithm, we analyse the changes of the comprehensive QoS value of each composite service included in interval Pareto solutions to an UQ-WSC problem. As the iterative times vary from 1000 to 20,000 with an interval of 1000, the experimental results are shown in Figure 7.

Figure 7 The comprehensive QoS values of composite services on RT and Latency (see online version for colours)

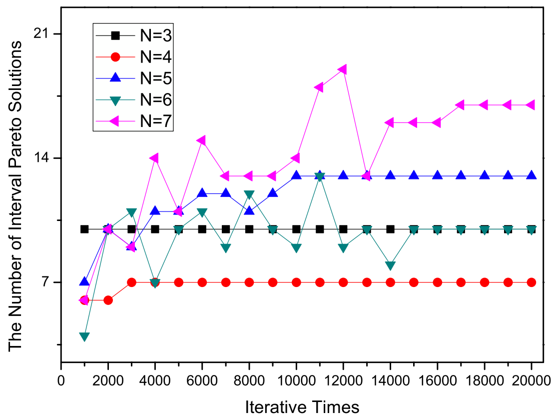


From the results as shown in Figure 7, we can find that the comprehensive QoS values of composite services on response time and latency time change when the iterative times are less than 14,000. However, the comprehensive QoS values of composite services on response time and latency time keep stable after iterative times are equal to and more than 14,000. In other words, the algorithm makes interval Pareto solutions to the problem convergent in a constant set.

5.4 Impact of candidate services on the number of composite service solutions

To demonstrate the impact of candidate services on the variations of the number of interval Pareto solutions to an UQ-WSC problem, we set the abstract service tasks $T = 5$ and the number of candidate services N changes from 3 to 7. We observe that the changes of the number of interval Pareto solutions along with different iterative times, which are shown in Figure 8.

Figure 8 The experimental results on the number of interval Pareto solutions with different number of candidate services (see online version for colours)



As shown in Figure 8, we conclude that with the increase of iterative times on running NDmoea/d algorithm, the number of interval Pareto solutions to a given UQ-WSC problem comes to a constant in the end from $N = 3$ to $N = 7$. In addition, the more the number of candidate services becomes, the more complex an UQ-WSC problem is. Therefore, as the number of candidate services N changes from 3 to 7, the convergence speed on finding the optimum composite services becomes lower and lower. Specifically, when the number of candidate services N is set as 3, the convergence speed is the fastest and the iterative times of finding the stable number of composite services starts from 700. When N becomes 7, the convergence speed becomes the slowest and the iterative times of finding the stable number of composite services starts from 17000.

From the experimental results above, we conclude that an uncertain QWSC problem is encoded as an interval number based multi-objective optimisation problem, which can be effectively and efficiently solved by the application of our improved Non-deterministic MOEA/D algorithm. The interval Pareto solutions of composite services can be generated with great convergence speed.

5.5 Discussion

With the consideration of both uncertain QoS and user's multiple preferences, this work presented a framework for solving the issue on uncertain QWSC, which is modelled into an interval number based multi-objective optimisation problem with global constraints. To solve encoded problem, an improved non-deterministic multi-objective evolutionary algorithm based on decomposition is proposed and get all the optimum composite services for users. These solutions can satisfy user's preference from multiple QoS criteria through interval Pareto comparison. Experimental results demonstrate the effectiveness of our approach, when generating the optimal composite services holding an efficient algorithm convergence.

Compared with prior works, our approach has its advantages. We mainly focus on solving uncertain QWSC problem in terms of multi-objective optimisation tasks. By doing so, we can find multiple composite service solutions, instead of generating only one composite service by a comprehensive QoS utility function in the current work. Although most of the existing approaches have a better solution-solving efficiency, they cannot satisfy user's multi-dimensional QoS preferences, which have been taken into account in our research work. Furthermore, how to select the weights among multiple QoS criteria in single utility-score metric is still a challenging problem, which can be solved in our improved NDmoea/d algorithm via multi-objective optimisation task decomposition.

However, our approach still has its own disadvantage. Since the Ndmoea/d algorithm needs to compare all of the feasible solutions via interval Pareto strategy so as to find the optimum composite services, it consumes additional time. Consequently, as the increasing number of web services, the efficiency of our approach is slower than that of the existing ones. In the future work, we need to improve the algorithm and make it more efficient.

6 Related work

QWSC problem has become one of the most popular research issues during the last few years. To effectively and efficiently find the optimal composite services, related methods have been proposed. We review highly correlative works with us from two aspects, QWSC and uncertain QWSC, respectively.

For certain QoS-aware WSC, Zeng et al. (2003) and Zeng et al. (2004) proposed a method based on linear programming. Under a predefined workflow model, it transforms a QoS-aware WSC problem to an integer programming problem from which it generates the global optimal composite service solution. With the consideration of loops peeling and stateful web services, Ardagna and Pernici (2007) modelled a Q-WSC problem into a mixed integer linear programming problem and solved it in the QoS constrained web service selection problem. Based on QoS measurement metrics, Huang et al. (2009) proposed a multiple criteria decision making and integer programming approach to select the optimal candidate service. To select service components with various QoS levels, Yu et al. (2007) modelled a WSC problem into a combinatorial model and graph model. It took advantage of some heuristic algorithms in finding the solution. To improve the scalability and accuracy, Jiang et al. (2010) implemented a tool, called QSynth by pruning service search spaces. Taking into consideration of the probability and conditions of each execution path, Zheng et al. (2013) proposed an approach to calculate QoS for composite services with complex structure and deal with different workflow composite patterns effectively.

As the application of Artificial Intelligence algorithms, Canfora et al. (2005) proposed the use of Genetic Algorithms (GA) for Q-WSC problem. Furthermore, Wagner et al. (2012) adopted the multi-objective optimisation approach to consider multiple workflows for WSC at the same time. In addition, Chen et al. (2014) proposed a distributed online service selection algorithm by Lyapunov optimisation techniques for large-scale web service systems. Cremene et al. (2015) performed an analysis on several state-of-the-art multi-objective evolutionary algorithms to solve QWSC problem. Laleh et al. (2017) defined a constraint-based composite service model that not only considers service constraints, but also adapts composite plans according to new constraints that might add new restriction to composite service at run time. Peng et al. (2017) proposed an adaptive approach based on Restricted Boltzmann Machine, that maintains the diversity of alternative solutions and captures the potential solutions. However, these approaches are designed to focus on certain QoS, instead of the QoS with uncertainty of web services.

As for uncertain QWSC, Wang et al. (2011) employed cloud model to compute the QoS uncertainty for pruning redundant services, while extracting reliable services. Then, it found the optimal composite service by mixed integer programming. After that, Jian et al. (2016) proposed a novel QoS interval model, which evaluate profit and stability in a fuzzy way. Sun et al. (2014) proposed a method using information theory and variance theory to abandon high QoS uncertainty services and downsize the solution spaces, and then selected the best reliable composite service by 0-1 integer programming. Considering the QoS activities to be fluctuating, Li et al. (2015) introduce a dynamic framework to predict the runtime QoS by employing an autoregressive moving average model and QoS reduction rules. Taking the consideration of stochastic dynamic optimisation strategy, Xia et al. (2011), Xia et al. (2012) and Xia et al. (2013) proposed a model-driven approach for service reliability evaluation and prediction. The approach employed a stochastic Perti net as fundamental model. Chattopadhyay et al. (2016) presented a stochastic integer linear programming formulation and a scalable heuristic algorithm for Q-WSC problem. These approaches mainly analysed the uncertainty of service QoS and select reliable web services for composite service. However, they are still lack of the integral, global, and multi-objective optimisation with the consideration of multiple QoS user preferences.

Based on above investigations on WSC, we proposed a non-deterministic multi-objective evolutionary algorithm (NDmoea/d) for solving an interval number based multi-objective optimisation problem, which is translated from an original uncertain QWSC problem. The method can find interval Pareto solutions of composite services with global QoS optimality and the satisfiability of multiple user preferences. Consequently, we can provide different composite services for users, when they perform decision-making for the selection of their satisfactory composite service.

7 Conclusion and future work

This paper presents a novel approach for solving uncertain QWSC problem. The method firstly formulates UQ-WSC problem where uncertain QoS of composite service is modelled with matrix and interval number. Then, an UQ-WSC problem is encoded as an interval number based multi-objective optimisation problem (IMOPs) by using user's QoS preferences. Finally, an improved non-deterministic MOEA/D algorithm is applied to solve the encoded IMOPs problem and produce interval Pareto solutions with QoS optimality.

Experimental results indicate that our proposed approach can effectively and efficiently solve UQ-WSC problem with great convergence.

For our future work, we will improve the approach to handle more multi-dimensional uncertain QoS criteria and validate its effectiveness and efficiency on new large-scale uncertain web service repository. We also consider how to integrate the functionality matchmaking of composite service into the current UQ-WSC problem for more compatibility in real-world service-oriented applications.

Acknowledgements

This work was partially supported by National Natural Science Foundation of China (Grant No. 61772128), Shanghai Natural Science Foundation (Grant No. 18ZR1414400 and 17ZR1400200), National Key Research and Development Program of China (2017YFC0907505), and Fundamental Research Funds for the Central Universities (No. 16D111208).

We would like to appreciate the anonymous reviewers for their insightful suggestions and constructive comments. We thank Professor Qingfu Zhang et al. for their open sources of MOEA/D algorithm and E.AI-Masri et al. for their QoS dataset of web services.

References

Al-Masri, E. and Mahmoud, Q.H. (2007) 'Discovering the best web service', *Proceedings of the 16th International Conference on World Wide Web*, ACM, 8–12 May, Banff, Alberta, Canada, pp.1257–1258.

Ardagna, D. and Pernici, B. (2007) 'Adaptive service composition in flexible processes', *IEEE Transactions on Software Engineering*, Vol. 33, No. 6, pp.369–384.

Canfora, G., Penta, M.D., Esposito, R. and Villani, M. L. (2005) 'An approach for QoS-aware service composition based on genetic algorithms', *Proceedings of the 7th Annual Conference on Genetic and Evolutionary Computation*, ACM, 25–29 June, Washington DC, USA, pp.1069–1075.

Chattopadhyay, S. and Banerjee, A. (2016) 'QSCAS: QoS aware web service composition algorithms with stochastic parameters', *Proceedings of the IEEE International Conference on Web Services (ICWS)*, IEEE, 27 June–2 July, San Francisco, CA, USA, pp.388–395.

Chen, Y., Huang, J., Xiang, X. and Lin, C. (2014) 'Energy efficient dynamic service selection for large-scale Web service systems', *Proceedings of the IEEE International Conference on Web Services (ICWS)*, IEEE, 27 June–2 July, Anchorage, AK, USA, pp.558–565.

Coello, C.A.C., Pulido, G.T. and Lechuga, M.S. (2004) 'Handling multiple objectives with particle swarm optimization', *IEEE Transactions on Evolutionary Computation*, Vol. 8, No. 3, pp.256–279.

Cremene, M., Suciu, M., Pallez, D. and Dumitrescu, D. (2015) 'Comparative analysis of multi-objective evolutionary algorithms for QoS-aware Web service composition', *Applied Soft Computing*, Vol. 39, pp.124–139.

Deb, K., Pratap, A., Agarwal, S. and Meyarivan, T. (2002) 'A fast and elitist multiobjective genetic algorithm: Nsga-ii', *IEEE Transactions on Evolutionary Computation*, Vol. 6, No. 2, pp.182–197.

Gang, Q. and Feng, X. (2008) 'Ranking approaches for interval numbers in uncertain multiple attribute decision making problems', *Proceedings of the 27th Chinese Control Conference*, IEEE, 16–18 July, Kunming, China, pp.280–284.

Huang, A.F.M., Lan, C.W. and Yang, S.J.H. (2009) 'An optimal QoS-based Web service selection scheme', *Information Sciences*, Vol. 179, No. 19, pp.3309–3322.

Hwang, S.Y., Hsu, C.C. and Lee, C.H. (2015) 'Service selection for web services with probabilistic QoS', *IEEE Transactions on Services Computing*, Vol. 8, No. 3, pp.467–480.

Jaszkiewicz, A. (2002) 'On the performance of multiple-objective genetic local search on the 0/1 knapsack problem -a comparative experiment', *IEEE Transactions on Evolutionary Computation*, Vol. 6, No. 4, pp.402–412.

Jian, X., Zhu, Q.S. and Xia, Y.N. (2016) 'An interval-based fuzzy ranking approach for QoS uncertainty-aware service composition', *Optik - International Journal for Light and Electron Optics*, Vol. 127, No. 4, pp.2102–2110.

Jiang, W., Zhang, C., Huang, Z., Chen, M., Hu, S. and Liu, Z. (2010) 'QSynth: A Tool for QoS-aware Automatic Service Composition', *Proceedings of the IEEE International Conference on Web Services (ICWS)*, IEEE, 5–10 July, Miami, Florida, USA, pp.42–49.

Kang, G., Liu, J., Tang, M., Cao, B. and Xu, Y. (2015) 'An effective web service ranking method via exploring user behavior', *IEEE Transactions on Network and Service Management*, Vol. 12, No. 4, pp.554–564.

Laleh, T., Paquet, J., Mokhov, S. and Yan, Y.H. (2017) 'Constraint adaptation in web service composition', *Proceedings of the IEEE International Conference on Services Computing*, IEEE, 25–30 June, Honolulu, HI, USA, pp.156–163.

Li, J., Luo, X., Xia, Y.N., Han, Y.K. and Zhu, Q.S. (2015) 'A time series and reduction-based model for modeling and QoS prediction of service compositions', *Concurrency and Computation Practice and Experience*, Vol. 27, No. 1, pp.146–163.

Li, S., Wen, J., Luo, F., Gao, M., Zeng, J. and Dong, Z. (2017) 'A new QoS-aware web service recommendation system based on contextual feature recognition at server-side', *IEEE Transactions on Network and Service Management*, Vol. 14, No. 2, pp.332–342.

Mattson, C., Mullur, A. and Messac, A. (2004) 'Smart pareto filter: obtaining a minimal representation of multiobjective design space', *Engineering Optimization*, Vol. 36, No. 6, pp.721–740.

Messac, A., Ismail-Yahaya, A. and Mattson, C.A. (2003) 'The normalized normal constraint method for generating the pareto frontier', *Structural and Multidisciplinary Optimization*, Vol. 25, No. 2, pp.86–98.

Peng, S.S., Wang, H.B. and Yu, Q. (2017) 'Estimation of distribution with restricted Boltzmann machine for adaptive service composition', *Proceedings of the IEEE International Conference on Web Services (ICWS)*, IEEE, 25–30 June, Honolulu, HI, USA, pp.114–121.

Ruzika, S. and Wiecek, M.M. (2005) 'Approximation methods in multiobjective programming', *Journal of Optimization Theory and Applications*, Vol. 126, No. 3, pp.473–501.

Sun, L., Wang, S., Li, J., Sun, Q. and Yang, F. (2014) 'QoS uncertainty filtering for fast and reliable Web service selection', *Proceedings of the IEEE International Conference on Web Services (ICWS)*, IEEE, 27 June–2 July, Anchorage, AK, USA, pp.550–557.

Tang, M., Zheng, Z., Kang, G., Liu, J., Yang, Y. and Zhang, T. (2016) 'Collaborative web service quality prediction via exploiting matrix factorization and network map', *IEEE Transactions on Network and Service Management*, Vol. 13, No. 1, pp.126-137.

Wagner, F., Klein, A., Klopper, B., Ishikawa, F. and Honiden, S. (2012) 'Multi-objective service composition with time- and input-dependent QoS', *Proceedings of the IEEE International Conference on Web Services (ICWS)*, IEEE, 24–29 June, Honolulu, HI, USA, pp.234–241.

Wang, S., Zheng, Z., Sun, Q. and Zou, H. (2011) 'Cloud model for service selection', *Proceedings of the IEEE Conference on Computer Communications Workshops*, IEEE, 10–15 April, Shanghai, China, pp.666–671.

Wiecek, M.M., Chen, W. and Zhang, J. (2001) 'Piecewise quadratic approximation of the non-dominated set for bi-criteria programs', *Journal of Multi-criteria Decision Analysis*, Vol. 10, No. 1, pp.35–47.

Wu, Q. and Zhu, Q. (2013) 'Transactional and QoS-aware dynamic service composition based on ant colony optimization', *Future Generation Computer Systems*, Vol. 29, No. 5, pp.1112–1119.

Xia, Y.N., Dai, G., Li, J., Sun, T.H. and Zhu, Q.S. (2011) 'A model-driven approach to predicting dependability of WS-CDL based service composition', *Concurrency and Computation Practice and Experience*, Vol. 23, No. 10, pp.1127–1145.

Xia, Y. N., Liu, Y., Liu, J. and Zhu, Q.S. (2012) 'Modeling and performance evaluation of BPEL processes: a stochastic-petri-net-based approach', *IEEE Transactions on Systems Man and Cybernetics Part A Systems and Humans*, Vol. 42, No. 2, pp.503–510.

Xia, Y. N., Luo, X., Li, J. and Zhu, Q.S. (2013) 'A Petri-net-based approach to reliability determination of ontology-based service compositions', *IEEE Transactions on Systems Man and Cybernetics Part B*, Vol. 43, No. 5, pp.1240–1247.

Xia, Y.N., Zhou, M.C., Luo, X., Pang, S.C. and Zhu, Q.S. (2015) 'Stochastic modeling and performance analysis of migration-enabled and error-prone clouds', *IEEE Transactions on Industrial Informatics*, Vol. 11, No. 2, pp.495–504.

Yu, T., Zhang, Y. and Lin, K.J. (2007) 'Efficient algorithms for web services selection with end-to-end QoS constraints', *Acm Transactions on the Web*, Vol. 1, No. 1, pp.1–26.

Zeng, L., Benatallah, B., Dumas, M., Kalagnanam, J. and Sheng, Q.Z. (2003) 'Quality driven web services composition', *Proceedings of the 12th International Conference on World Wide Web*, ACM, 20–24 May, Budapest, Hungary, pp.411–421.

Zeng, L., Benatallah, B., Ngu, A.H.H., Dumas, M., Kalagnanam, J. and Chang, H. (2004) 'Qos-aware middleware for web services composition', *IEEE Transactions on Software Engineering*, Vol. 30, No. 5, pp.311–327.

Zhang, Q. and Li, H. (2007) 'Moea/d: a multiobjective evolutionary algorithm based on decomposition', *IEEE Transactions on Evolutionary Computation*, Vol. 11, No. 6, pp.712–731.

Zheng, H., Zhao, W., Yang, J. and Bouguettaya, A. (2013) 'QoS analysis for Web service compositions with complex structures', *IEEE Transactions on Services Computing (TSC)*, Vol. 6, No. 3, pp.373–386.