

Efficient Population Diversity Handling Genetic Algorithm for QoS-Aware Web Services Selection*

Chengwen Zhang, Sen Su, and Junliang Chen

State Key Lab of Networking and Switching Technology,
Beijing University of Posts & Telecommunications(BUPT)
187# 10 Xi Tu Cheng Rd., Beijing 100876, China
zwjcbj2007@gmail.com, {susen, chjl}@bupt.edu.cn

Abstract. To maximize user satisfaction during composition of web services, a genetic algorithm with population diversity handling is presented for Quality of Service(QoS)-aware web services selection. In this algorithm, the fitness function, the selection mechanism of the population as well as the competition mechanism of the population are represented. The population diversity and population fitness are used as the primary criteria of the population evolution. By competing between the current population and the historical optimal population, the whole population evolution can be done on the basis of the whole population evolution principle of the biologic genetic theory. Prematurity is overcome effectively. Experiments on QoS-aware web services selection show that the genetic algorithm with population diversity handling can get more excellent composite service plan than the standard genetic algorithm.

1 Introduction

Web service is a software application identified by an URI [1]. How to create robust service compositions becomes the next step work in web services [2] and has attracted a lot of researches [3], [4], [16], [17]. Since the web services with the same functions and different QoS are increasing, and the web services requesters always express their functional requirements as well as their global QoS constraints, a determination needs to be made to select which services are used in a given composite service on the basis of multiple QoS attributes in order to maximize user satisfaction. Hence, QoS-aware web services selection plays an important role in web services composition [5], [6]. In the past years, the researches about web services selection have gained considerable momentums [9], [10], [11], [15], [16], [17], [18], [19].

To figure out QoS-aware web services selection, some approaches are made with the help of semantic web [9], [10], [11], and the others are based on the QoS attributes computation [15], [16], [17], [18], [19]. But it is obvious that the latter approaches are

* The work presented in this paper was supported by the National Basic Research and Development Program (973 program) of China under Grant No. 2003CB314806; the National Natural Science Foundation project of China under Grant No. 90204007; the National Natural Science Funds for Distinguished Young Scholar of China under Grant No. 60125101; the program for Changjiang Scholars and Innovative Research Team in University (PCSIRT); BUPT Education Foundation.

more suitable solutions to satisfy the global QoS requirement. It is a combinatorial optimization issue that the best combination of web services is selected in order to accord with the global constraints.

Genetic Algorithm is a powerful tool to solve combinatorial optimizing problems [12]. The design of a genetic algorithm has the greatest influence on its behavior and performance [20], especially the fitness function and the selection mechanism.

Since keeping the individual diversity is a perfect means to overcome the premature phenomenon, it may be a good way to add the diversity handling into the genetic algorithm. Some diversity studies were in [13], [14], [22], [23], which paid less attention to the population evolution making use of the population diversity.

Following the above analyses, we present a genetic algorithm with population diversity handling, which enables the population to evolve on the basis of the whole population evolution principle of the biologic genetic theory.

The remainder of this paper is organized as follows. After a review of the literature of the QoS-aware web services selection using QoS computation in section 2 and the diversity of genetic algorithm in section 3, Section 4 presents the genetic algorithm with population diversity handling we propose in detail. Section 5 describes the simulations of the proposed algorithm and discusses results aiming to support the work. Finally, section 6 is our conclusions.

2 Quality Computation-Based Selection of Web Services

According to Std. ISO 8402 [7] and ITU E.800 [8], QoS may include a number of nonfunctional properties such as price, response time, availability, reputation. Thus, the QoS value of a composition service can be achieved by the fair computation of QoS of every component web services. There are some techniques in the literatures, including some traditional optimization techniques [15], [16], [17] and strategies based on Genetic Algorithm (GA) [18], [19].

The QoS computation based on the QoS matrix is a representative solution. To normalize the QoS matrix to rank the web services was proposed in [15], however, it was only a local optimization algorithm but not a global one for services selection. Other works in the area of QoS computation include [16], [17], which proposed local optimization and global planning. But, both had limitation to some extent.

The genetic algorithm is well suitable for the QoS-aware services selection belonging to the class of NP-hard [19]. In [18] and [19], the authors proposed only a coding manner of chromosome or the fitness function for the service selection with little further information about the rest parts of the genetic algorithm, such as the selection mechanism.

3 Diversity of Genetic Algorithm

This section describes some diversity control means presented in [13], [14], [22], [23]. In [13], the affinity was defined by means of the information entropy of every gene of chromosome. Finally, the individual with higher fitness and lower affinity had higher priority to evolve. Also, the low diversity of the population was raised by a

procedure aiming at maintaining high diversity of the population in [14]. In [22], an individual diversity controlling approach was used to attain the global optimum without getting stuck at a local optimum. And, a fuzzy controller changed the crossover rate and the mutation rate in [23], in order to maintain the proper population diversity during the GA's operation.

According to the biology genetic theory, the evolution of a population is a process during which all of individuals have higher and higher fitness to the environments. From the point of the view of the population, the population diversity evolves from high diversity at initial stage to low one at terminal stage during the population evolution. Hence, comparing population diversity between two populations produced at different generations is a feasible way to evaluate whether the current population has higher evolution extent than the former one, and the population evolution can be handled by the population diversity. Thus, we propose a genetic algorithm with population diversity handling to overcome the local optimal solution.

4 Genetic Algorithm with Population Diversity Handling

In this section, we present a genetic algorithm with diversity handling in order to resolve quality-driven selection. The main difference between our GA and the standard GA is that our GA includes the competition mechanism of the population that does not be expressed by the standard GA.

4.1 Fitness Function

Some QoS models and QoS computation formulas for composite service were available in [17], [19]. But, via comparison in experiments, the optimal solution based on the QoS computation formula in [19] is better than the one based on [17]. Consequently, the QoS computation formula in [19] is adopted, and the objective function is defined in (1):

$$f(g) = \frac{\sum_j (Q_j \times w_j)}{\sum_k (Q_k \times w_k)}. \quad (1)$$

Where $w_j, w_k \in [0,1]$, and w_j, w_k are real positive weight factors, represent the weight of criterion j and k for which end users show their favoritism concerning QoS by providing values respectively. The sum of all of them is 1. Q_j and Q_k denote the sum of QoS values of the individual j and k respectively.

End users usually assert their function requirements as well as the global constraints, e.g. $Cost < 60$, $Time < 150$. The individual whose QoS violates the constraints will be rejected in [19]. However, this approach belongs to the absolute rejection. It influences the diversity of the population seriously and always results in a local optimal solution. The individual disobeying the constraints should be selected proportionally into the next generation population on a basis of a certain technique. The most common method is penalty technique for constrained optimization problems. Hereby, the fitness function with penalty character is defined in (2):

$$Fit(g) = f(g) - \lambda \sum_{j=1}^n \left(\frac{\Delta P_j}{R_{jMax} - R_{jMin}} \right)^2 \tag{2}$$

Where R_{jMax} , R_{jMin} are the maximum value and minimal value of the No.j quality constraint respectively, n is the number of quality constraints, λ is a parameter used to adjust the scale of penalty value. P_j represents the calculation value of a Q_i or some Q_i s and these values are limited by a quality constraint. The following is the definition of ΔP_j :

$$\Delta P_j = \begin{cases} P_j - R_{jMax}; & P_j > R_{jMax} \\ 0 & R_{jMin} \leq P_j \leq R_{jMax} \\ R_{jMin} - P_j; & P_j < R_{jMin} \end{cases}$$

4.2 Design of Diversity

According to the definition of information entropy [14], $H_j(N)$ is the information entropy of the jth locus where N is the number of chromosome and $H(N)$ is the information entropy of all of N chromosomes. Clearly, the smaller the information entropy is, the worse the diversity is.

Firstly, the definition for individuals is (3):

$$d_i = \frac{1}{N} \sum_{j=1}^N H_j(2) \tag{3}$$

Where d_i is used for the diversity of the ith individual, N is the number of chromosomes, $H(2)$ is the information entropy of the ith individual between itself and all of the rest of the population.

Secondly, the population diversity can be acquired from $H(N)$.

4.3 Selection Mechanism of Individuals

The individual with high diversity and with high fitness will survive at higher rate. However, some sets of experiments show that introducing the information entropy to the diversity of the individuals always increases the running time greatly, and can not get obviously more optimal solution than without the diversity of the individuals. Thus, the individuals are selected only according to the fitness value.

4.4 Selection Mechanism of the Population

According to the evolution theory based on natural selection of Darwin, the species evolve in the form of the whole group but not individuals. During the evolution process of the population, in addition to the individuals evolution should be controlled, the whole population evolution should also be handled by means of some policies that prevent the population from degeneration and help to find the optimal solution. In this paper, a selection mechanism of the population is taken into consideration. The historical optimal population is always kept and participates in the competition process between itself and the current population. The population with lower expectation value is known as degeneration and is replaced by the one with higher expectation value.

The expectation value e of the population evolution is calculated in (4):

$$e = \frac{\overline{Fit(g)}}{1 + H(N)}. \tag{4}$$

From (4), we can see that the population with high average fitness and with low diversity will survive at higher rate. So, (4) represents the control mechanism of the diversity because the population with low diversity is promoted and the population with high diversity is suppressed.

Both the maximum of the population fitness and the result of subtracting the minimal fitness of the individuals from the maximal fitness of the individuals are not adopted at the e calculation formula, because the average fitness can cover all of individuals and the others can not reflect the real status of the population, especially in the worst concentricity.

4.5 Structure of Genetic Algorithm with Population Diversity Handling

Hitherto, some important elements in the GA for QoS-driven web services selection have been proposed. Here, the structure of the GA is available in sequence as figure 1.

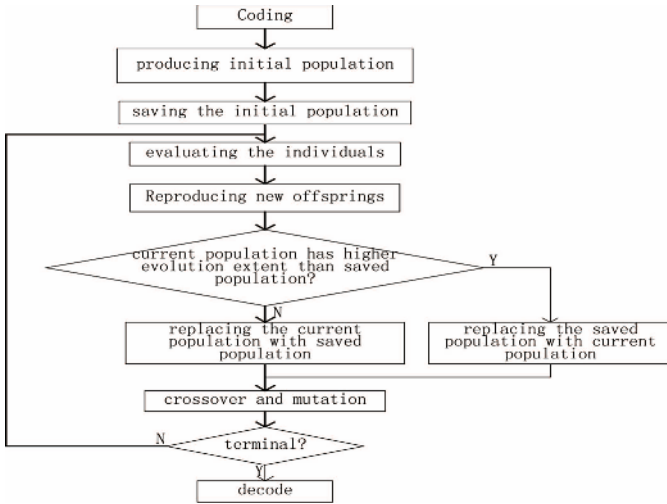


Fig. 1. Structure of the genetic algorithm with population diversity handling

The outstanding point of this paper includes the holding mechanism and competition mechanism of the historical optimal population that ensure the whole population evolution.

5 Experiments

To verify the excellence of the GA of the population diversity handling we have proposed, numerous simulation comparisons between itself and the standard GA had been performed on QoS-aware web services selection. All the experiments were taken

on the same software and hardware, which were Pentium 1.6GHz processor, 512MB of RAM, Windows XP Pro, development language JAVA, IDE Eclipse 3.1. The same data were adopted, including workflows of different sizes, 15-50 candidate web services for each task and 5 QoS data for each web service. A simplified representation of web service was used, including an ID number, some QoS data that were retrieved according to some principles in the range of defined values.

The two GAs were set up with the same population size, crossover operation and probability, mutation operation and probability, the fitness function, the selection algorithm of the individuals and the selection mechanism of the individuals. QoS model in [17] was used for both of them. The population diversity handling is the only difference between the two GAs.

The data of the experiments were collected with two methods: statistic data and process data. Figure 2 plotted the fitness function evolution across GA generation. Table 1 presented statistic experiment results of the average fitness and time for the maximal fitness.

Figure 2 was the population size 200, crossover probability 0.7, mutation probability 0.1. The upper curve stood for the GA of population diversity handling and the lower one stood for the standard GA in figure 2. This means the GA of population diversity handling could have larger fitness value than the standard GA.

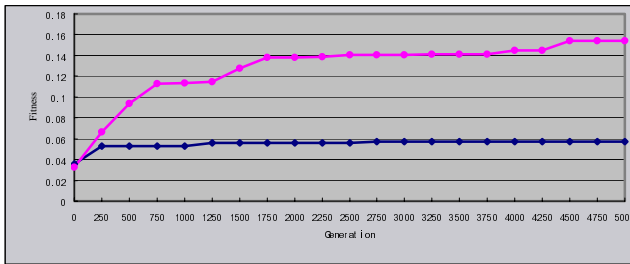


Fig. 2. Fitness comparison when tasks number 30

As shown in table 1, the statistic data were collected after the two GAs running for 50 times with the population size 200, iterations 500, crossover probability 0.7 and mutation probability 0.1. In table 1, the N(ew) represented the GA of population diversity handling and the O(ld) for the standard GA, the time comparison of N and O was the time when the maximal fitness value appears with unit as ms.

Table 1. Statistic Data

Tasks Num	Average fitness (N:O)	Time (N:O)
20	0.125:0.076	7882:2569
30	0.093:0.053	11307:2976

As described above, it is the introduction of the holding mechanism and competition mechanism of the historical optimal population on the basis of the

population diversity that ensures that the GA of population diversity handling can get more optimal solution than the standard GA, and to some extent overcome the premature phenomenon of the standard GA. The service composed by the GA of population diversity handling is better than the one composed by the standard GA.

However, it can be seen that the execution time of the GA of population diversity handling is longer than the standard GA at the cost of the holding mechanism and competition mechanism of the historical optimal population. The general indication given by the simulations is that the weaknesses are including the long running time and slow convergence. These would require refining the convergence and shrinking the running time by means of the original population policy and mutation policy, etc.

6 Conclusions

The QoS-aware web services selection is an active research area. In this paper, we present a structure of genetic algorithm characterized by the population competition mechanism. Prematurity is overcome effectively through the conservation of the historical optimal population and the competition between the historical optimal population and the current population.

We also verify the formulas that we use in the genetic algorithm through experiments and the results show that the genetic algorithm with population diversity handling can get more excellent composite service plan than the standard GA.

Providing adaptive capability of genetic algorithms is an active research area [21]. Therefore, how to design a self-adaptive genetic algorithm for QoS-aware selection is one of our future works. Work-in-process is devoted to better extend the approach as follows: accelerating the convergence of the GA of population diversity handling, supplying the self-adaption.

References

1. W3C. Web Services Architecture. <http://www.w3.org/TR/2004/NOTE-ws-arch-20040211/> (2004)
2. F. Curbera, R. Khalaf, N. Mukhi, etc.: The Next Step in Web Services. *Communication of the ACM*, 46(10) (2003) 29-34
3. Nikola Milanovic, Miroslaw Malek: Current Solutions for Web Service Composition. *IEEE Internet Computing* (2004) 51-59
4. B. Orriens, J. Yang, M P Papazoglou: Model Driven Service Composition. In the First International Conference on Service Oriented Computing (ICSOC'03) (2003)
5. D. A. Menascé: QoS Issues in Web Services. *IEEE Internet Computing*, 6(6) (2002) 72-75
6. D. A. Menascé: Composing Web Services: A QoS View. *IEEE Internet Computing* (2004) 88-90
7. ISO 8402, Quality Vocabulary
8. ITU-T Recommendation E.800 (1994), Terms and Definitions Related to Quality of Service and Network Performance Including Dependability
9. M. Tian, A. Gramm, H. Ritter, J. Schiller: Efficient Selection and Monitoring of QoS-Aware Web Services with the WS-QoS Framework. *IEEE/WIC/ACM International Conference on Web Intelligence (WI04)* (2004)

10. A. Soydan Bilgin, Munindar P. Singh: A DAML-Based Repository for QoS-Aware Semantic Web Service Selection. Proceedings of the IEEE International Conference on Web Services (ICWS'04) (2004)
11. Chen Zhou, Liang-Tien Chia, Bu-Sung Lee: DAML-QoS Ontology for Web Services. IEEE International Conference on Web Services (ICWS'04) (2004)
12. M. Srinivas, L. M. Patnaik: Genetic Algorithm: a Survey. IEEE (1994) 17-26
13. Chun J S, Kim M K, Jung H K. Shape Optimization of Electromagnetic Devices Using Immune Algorithm. IEEE Trans on Magnetics, 33(2) (1997) 1876-1879
14. Yasuhiro TSUJIMURA, Mitsuo GEN: Entropy-Based Genetic Algorithm for Solving TSP. The Second International Conference on Knowledge-Based Intelligent Electronic Systems (1998) 285-290
15. Y. Liu, A. H. Ngu, L. Zeng: QoS Computation and Policing in Dynamic Web Service Selection. In Proceedings of the 13th International Conference on World Wide Web (WWW), ACM Press (2004) 66-73
16. L. Zeng, B. Benatallah, M. Dumas, J. Kalagnanam, Q. Z. Sheng: Quality Driven Web Services Composition. Proc. 12th Int'l Conf. World Wide Web (WWW) (2003)
17. Liangzhao Zeng, Boualem Benatallah, Anne H. H. Ngu, etc.: QoS-Aware Middleware for Web Services Composition. IEEE Transactions on Software Engineering, 30(5) (2004) 311-327
18. LiangJie Zhang, Bing Li, Tian Chao, etc.: On Demand Web Services-Based Business Process Composition. IEEE (2003) 4057-4064
19. G. Canfora, M. Di Penta, R. Esposito, M. L. Villani: A Lightweight Approach for QoS-Aware Service Composition. ICSSOC (2004)
20. R. Ignacio, G. Jesús, P. Héctor, etc.: Statistical Analysis of the Main Parameters Involved in the Design of a Genetic Algorithm. IEEE Transactions on Systems, Man, and Cybernetics—Part C: Applications and Reviews, 32(1) (2002) 31-37
21. R. Hinterding, Z. Michalewicz, A. E. Eiben: Adaptation in Evolutionary Computation: a Survey. IEEE EC (1997) 65-69
22. Hisashi Shimodaira: DCGA: A Diversity Control Oriented Genetic Algorithm. The Ninth IEEE International Conference on Tools with Artificial Intelligence (1997) 367-374
23. Kejun Wang: A New Fuzzy Genetic Algorithm Based on Population Diversity. IEEE International Symposium on Computational Intelligence in Robotics and Automation (2001) 108-112