EDA-Based Approach to Comprehensive Quality-Aware Automated Semantic Web Service Composition

Chen Wang
Victoria University of Wellington
Wellington, New Zealand
chen.wang@ecs.vuw.ac.nz

Hui Ma
Victoria University of Wellington
Wellington, New Zealand
hui.ma@ecs.vuw.ac.nz

Gang Chen
Victoria University of Wellington
Wellington, New Zealand
aaron.chen@ecs.vuw.ac.nz

ABSTRACT

In the domain of Service-Oriented Architecture, web services are selected and composed to meet users' functional and non-functional requirements. A few researchers have proposed Evolutionary Computation (EC) techniques for service composition problems, where semantic matchmaking quality and Quality of Service (QoS) are individually or jointly optimised. However, these EC-based approaches allow best individuals (i.e., composition solutions) to be carried across generations, but does not consider the knowledge of the promising individuals for breeding better solutions. Due to its capability of learning the knowledge of good solutions, Estimation of Distribution Algorithm (EDA) has been successfully applied to tackle semi-automated QoS-aware service composition, where an abstract composition workflow is assumed known. However, this assumption is not always satisfied, so learning the knowledge of solutions without pre-given workflows is very challenge. We propose an EDA-based service composition approach to jointly optimize QoS and semantic matchmaking quality in a fully automated way, and our experiment evaluation demonstrates the efficiency and the effectiveness of our proposed approach.

CCS CONCEPTS

• **Mathematics of computing** → **Combinatorial optimization**; • **Information systems** → *Web services*; • **Theory of computation** → Evolutionary algorithms;

KEYWORDS

Web service composition, QoS optimisation, Combinatorial optimisation, EDA

ACM Reference Format:
Chen Wang, Hui Ma, and Gang Chen. 2018. EDA-Based Approach to Comprehensive Quality-Aware Automated Semantic Web Service Composition. In *GECCO '18 Companion: Genetic and Evolutionary Computation Conference Companion, July 15–19, 2018, Kyoto, Japan.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3205651.3205734

1 INTRODUCTION

*C*omprehensive quality-aware semantic web service composition is aimed at loosely coupling semantic web services while optimizing *semantic matchmaking quality* and *quality of services*, see the formal definition of the problem in [3, 4]. EDA is good at extracting useful knowledge from past experiences in the form of distribution models for more effective searches of promising solutions. Existing research work [1] only considers a possible use of EDA for semi-automated service composition. They assume one composition workflow is given. However, this assumption is not always satisfied, so learning the knowledge of solutions without pre-given workflows is very challenge in automated service composition. In this paper, we will propose a EDA-based approach to fully automated service composition, and compare it with one recent work [3] on the same problem.

2 PROPOSED APPROACH

To enable EDA to learn the knowledge of service composition, we proposed a decoding and encoding algorithm for our vector-based individuals, which contribute to a more reliable and accurate learning of probabilistic distribution model for our permutation-based populations. Apart from that, to enhance the searching ability of EDA, two updating methods are proposed to update the distribution model adaptively.

Representation and Population. Let $\mathcal{G} = (V, E)$ be a DAG-based service composition solution. Let $[S_0, \ldots, S_t]$ be a sequence of services discovered by Breadth-First search (BFS) performed on $\mathcal{G}$, $[S_{t+1}, \ldots, S_{n-1}]$ be a sequence of remaining services in service repository $\mathcal{SR}$ but not utlized by $\mathcal{G}$. We use $sol_k^g = [I_k^g(S_0), \ldots, I_k^g(S_t), \ldots, I_k^g(S_{n-1})]$ to represent the k^{th} (out of m) service composition solution, and $P(g) = [sol_0^g, \ldots, sol_k^g, \ldots, sol_{m-1}^g]$ to represent a population of solutions of generation g. $I_k^g(S_x)$, $x \in \{1, \ldots, n-1\}$, represents the index of service S_x in $\mathcal{SR}$.

We use Node Histogram Matrix (NHM) [2] to represent the knowledge of a $P(g)$ as a NHM^g, and Node Histogram Based Sampling Algorithm (NHBSA) [2] is employed to the NHM^g to breed solutions represented in vectors.

A Decoding and Encoding Algorithm. The decoding algorithm is extended from [3], which decodes vector-based individuals (i.e., bred vectors from NHBSA) into DAGs for fitness evaluations. To make it possible to learn the distribution over service composition solutions in each population, BFS is applied to encode DAGs into a sequence of vectors. These resulted vectors enable a reliable and accurate learning of NHM.

Adaptive Updating of NHM. To tract the promising searching area, we attempt to select a proper learning discount rate α in Eq.(1) for updating NHM. This formula defines a mechanism to update the newly generate NHM^{g+1} by considering NHM^g generated

GECCO '18 Companion, July 15–19, 2018, Kyoto, Japan

ACM ISBN 978-1-4503-5764-7/18/07.
https://doi.org/10.1145/3205651.3205734

from a prior population to prevent from premature and low rate convergence.

$$NHM^{g+1} \leftarrow (1-\alpha) \times e^{g}_{i,j} \in NHM^{g} + \alpha \times e^{g+1}_{i,j} \in NHM^{g+1} \quad (1)$$

To balance exploration and exploitation for the searching process, we propose two methods to predict proper α using Eq. (2) and (4). These two methods named as Entropy-based EDA (E-EDA) and Linear decrement EDA (L-EDA). Eq. (2) decides α based on its linear relationship to the changes in knowledge of NHM. This relationship is established for a given effective moving range [A, B].

$$\alpha = \frac{\overline{\mathcal{H}(P(g+1))} - min\{\mathcal{H}(P(g)), \mathcal{H}(P(g+1))\}}{max\{\mathcal{H}(P(g)), \mathcal{H}(P(g+1))\} - min\{\mathcal{H}(P(g)), \mathcal{H}(P(g+1))\}} \times (B-A) + A \quad (2)$$

$$\overline{\mathcal{H}(P(g))} = \frac{1}{n} \sum_{i \in C} \sum_{j \in C} - \frac{e^{g}_{i,j}}{\sum_{j \in C} e^{g}_{i,j}} \log_2 \frac{e^{g}_{i,j}}{\sum_{j \in C} e^{g}_{i,j}} \quad (3)$$

$\overline{\mathcal{H}(P(g))}$ indicates the entropy of of NHM extended from the Shannon entropy using Eq. 3. n is the dimension size of composition solutions , and C is a list, $[0, 1, \ldots, n-1]$. [A, B] is a moving range of α subsumed by [0, 1]. $min\{\mathcal{H}(P(g)), \mathcal{H}(P(g+1))\}$ and $max\{\mathcal{H}(P(g)), \mathcal{H}(P(g+1))\}$ are theoretically minimum and maximum entropy values of NHM that are calculated based on the historically found values during the run.

Eq. (4) decides α based on a linear decrement strategy. A linear decrement strategy is applied to adjust α along with generations.

$$\alpha = \frac{m-g}{m} \times (B-A) + A \quad (4)$$

m is the maximum generation, and g is a counter for the current generation in EDA. [A, B] is the same interval discussed in Eq. 2

3 FITNESS FUNCTION

To evaluate each individual we employ the fitness function in [3, 4]. The fitness of an individual can be calculated based on a weighted sum of all the quality criteria using Eq. (5).

$$Fitness = w_1 \hat{MT} + w_2 \hat{SIM} + w_3 \hat{A} + w_4 \hat{R} + w_5(1-\hat{T}) + w_6(1-\hat{C}) \quad (5)$$

with $\sum_{k=1}^{6} w_k = 1$. This objective function is defined as a *comprehensive quality model*, which considers quality aspects in semantic matching type $\hat{MT}$, semantic similarity $\hat{SIM}$, availability $\hat{A}$, reliability $\hat{R}$, time $\hat{T}$, and cost $\hat{C}$. $\hat{T}$ and $\hat{C}$ are offset by 1 that results in higher scores for better quality. The goal of comprehensive quality-aware semantic service composition is to maximize the objective function Eq. (5) to find the best solution.

4 EXPERIMENTS

We conducted an experiment to evaluate the performance of three methods (i.e., EDA, L-EDA and E-EDA) with the PSO-based method [3]. The weights settings of fitness function followed strictly these in [3]. The datasets used in this comparison were based on the one proposed in [3]. The population size is set to 200, number of generations equals to 100, and b_{ratio} is 0.0002. The interval [A, B] is set to [0.2, 0.9]. We run the experiment with 30 independent repetitions with the results shown in Table 1 and Table 2. Table 1 shows that the fitness of EDA-based approaches is generally higher than those produced using the PSO-based approach [3]. L-EDA and E-EDA slightly outperform EDA, and are comparable to each other. Table 2 shows that three EDA-based approaches consistently require less execution time (in seconds) compared to [3]. L-EDA and E-EDA require less time for execution for most tasks.

Table 1: Comparison results of Mean fitness

Task	EDA	$L-EDA$	$E-EDA$	PSO [3]
WSC 08-1	0.50621 ± 0.0096	0.50692 ± 0.0112	0.50639 ± 0.0100	**0.52216 ± 0.0044**
WSC 08-2	**0.61433 ± 0.0000**	**0.61433 ± 0.0000**	**0.61433 ± 0.0000**	0.61355 ± 0.0030
WSC 08-3	0.45509 ± 0.0001	**0.45513 ± 0.0001**	**0.45513 ± 0.0001**	0.45415 ± 0.0005
WSC 08-4	0.46447 ± 0.0001	0.46447 ± 0.0001	0.46450 ± 0.0001	0.46451 ± 0.0001
WSC 08-5	0.46908 ± 0.0003	**0.46910 ± 0.0002**	**0.46908 ± 0.0003**	0.46863 ± 0.0011
WSC 08-6	**0.47422 ± 0.0001**	**0.47424 ± 0.0001**	**0.47423 ± 0.0001**	0.47326 ± 0.0006
WSC 08-7	0.48075 ± 0.0001	**0.48077 ± 0.0000**	**0.48077 ± 0.0000**	0.47900 ± 0.0005
WSC 08-8	**0.46182 ± 0.0000**	**0.46182 ± 0.0000**	**0.46182 ± 0.0000**	0.46156 ± 0.0003
WSC 09-1	0.56690 ± 0.0085	0.56835 ± 0.0076	0.56857 ± 0.0086	0.56944 ± 0.0089
WSC 09-2	0.47114 ± 0.0000	0.47115 ± 0.0000	**0.47116 ± 0.0000**	0.47110 ± 0.0003
WSC 09-3	**0.55116 ± 0.0000**	**0.55116 ± 0.0000**	**0.55116 ± 0.0000**	0.55109 ± 0.0003
WSC 09-4	**0.47255 ± 0.0003**	**0.47242 ± 0.0004**	**0.47246 ± 0.0003**	0.47129 ± 0.0008
WSC 09-5	**0.47041 ± 0.0000**	**0.47041 ± 0.0000**	**0.47041 ± 0.0000**	0.47008 ± 0.0003

Table 2: Comparison results of Mean execution time

Task	EDA	$L-EDA$	$E-EDA$	PSO [3]
WSC 08-1	56.32 ± 3.59	**53.81 ± 3.88**	56.62 ± 2.18	62.76 ± 33.27
WSC 08-2	36.36 ± 1.16	**35.23 ± 2.06**	35.91 ± 2.15	41.72 ± 22.91
WSC 08-3	7815.85 ± 2040.97	8449.16 ± 2355.53	**7106.93 ± 1156.56**	12152.72 ± 1971.99
WSC 08-4	36.25 ± 0.91	**35.73 ± 1.53**	35.96 ± 1.07	118.53 ± 29.69
WSC 08-5	**410.84 ± 66.05**	**431.95 ± 52.27**	**421.86 ± 53.11**	1174.28 ± 380.70
WSC 08-6	6419.64 ± 257.47	**6185.60 ± 369.85**	6338.24 ± 355.60	11321.95 ± 2269.06
WSC 08-7	**954.36 ± 167.92**	**1022.54 ± 175.85**	**1026.04 ± 158.96**	2133.10 ± 753.53
WSC 08-8	1729.35 ± 157.70	1657.01 ± 156.29	**1632.35 ± 193.30**	4864.01 ± 1141.94
WSC 09-1	**56.69 ± 3.32**	**57.41 ± 2.68**	**57.63 ± 3.59**	91.61 ± 47.36
WSC 09-2	1180.21 ± 144.15	1116.08 ± 135.96	**1105.67 ± 92.66**	2201.56 ± 522.42
WSC 09-3	**805.82 ± 28.33**	**790.33 ± 23.15**	**803.40 ± 26.19**	1298.32 ± 445.03
WSC 09-4	26741.33 ± 1464.03	26386.08 ± 2818.80	**25733.08 ± 1333.06**	36804.51 ± 7670.98
WSC 09-5	**5861.03 ± 366.95**	**6006.96 ± 472.10**	**5892.08 ± 419.45**	9556.08 ± 2194.68

5 CONCLUSION

In this paper, we proposed an EDA-based approach to comprehensive quality-aware semantic web service composition. In particular, we proposed a decoding and encoding algorithm and two adaptive updating methods. Our experimental evaluation shows that EDA-based approaches are more effective and efficient compared to the PSO-based approach [3]. This demonstrates that learning the knowledge of promising composition solutions can help find near-optimal solutions. In addition, two updating methods proposed in E-EDA and L-EDA achieve a reasonable good results compared to EDA. As the interval [A, B] for updating α plays an important role for these two updating methods, we can investigate the influence of different intervals for our adaptive updating methods in the future. Besides that, we can investigate other methods to decide α based on its non-linear relationship to the entropy of NHM for EDA.

REFERENCES

[1] Koramit Pichanaharee and Twittie Senivongse. 2008. QoS-based service provision schemes and plan durability in service composition. In *IFIP International Conference on Distributed Applications and Interoperable Systems*. Springer, 58–71.

[2] Shigeyoshi Tsutsui. 2006. A comparative study of sampling methods in node histogram models with probabilistic model-building genetic algorithms. In *Systems, Man and Cybernetics, 2006. SMC'06. IEEE International Conference on*, Vol. 4. IEEE, 3132–3137.

[3] Chen Wang, Hui Ma, Aaron Chen, and Sven Hartmann. 2017. Comprehensive quality-aware automated semantic web service composition. In *Australasian Joint Conference on Artificial Intelligence*. Springer, 195–207.

[4] Chen Wang, Hui Ma, Aaron Chen, and Sven Hartmann. 2017. GP-Based Approach to Comprehensive Quality-Aware Automated Semantic Web Service Composition. In *Asia-Pacific Conference on Simulated Evolution and Learning*. Springer, 170–183.