OPTIMAL QOS-AWARE MULTIPLE PATHS WEB SERVICE COMPOSITION USING HEURISTIC ALGORITHMS AND DATA MINING TECHNIQUES

OSAMA KAYED TAHER QTAISH

DOCTOR OF PHILOSOPHY
UNIVERSITI UTARA MALAYSIA
2014
Permission to Use

In presenting this thesis in fulfillment of the requirements for a postgraduate degree from Universiti Utara Malaysia, I agree that the Universiti Library may make it freely available for inspection. I further agree that permission for the copying of this thesis in any manner, in whole or in part, for a scholarly purpose may be granted by my supervisor(s) or, in their absence, by the Dean of Awang Had Salleh Graduate School of Arts and Sciences. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain is not allowed without my written permission. It is also understood that due to recognition shall be given to me and to Universiti Utara Malaysia for any scholarly use, which may be made of any material from my thesis.

Requests for permission to copy or to make other uses of materials in this thesis, in whole or in part, should be addressed to:

Dean of Awang Had Salleh Graduate School of Arts and Sciences
UUM College of Arts and Sciences
Universiti Utara Malaysia
06010 UUM Sintok
Abstrak


Kata kunci: Penggubahan perkhidmatan web, QoS, Pemilihan perkhidmatan, Algorithma heuristik, Perlombongan data.
Abstract

The goal of QoS-aware service composition is to generate optimal composite services that satisfy the QoS requirements defined by clients. However, when compositions contain more than one execution path (i.e., multiple path's compositions), it is difficult to generate a composite service that simultaneously optimizes all the execution paths involved in the composite service at the same time while meeting the QoS requirements. This issue brings us to the challenge of solving the QoS-aware service composition problem, so called an optimization problem. A further research challenge is the determination of the QoS characteristics that can be considered as selection criteria. In this thesis, a smart QoS-aware service composition approach is proposed. The aim is to solve the above-mentioned problems via an optimization mechanism based upon the combination between runtime path prediction method and heuristic algorithms. This mechanism is performed in two steps. First, the runtime path prediction method predicts, at runtime, and just before the actual composition, execution, the execution path that will potentially be executed. Second, both the constructive procedure (CP) and the complementary procedure (CCP) heuristic algorithms computed the optimization considering only the execution path that has been predicted by the runtime path prediction method for criteria selection, eight QoS characteristics are suggested after investigating related works on the area of web service and web service composition. Furthermore, prioritizing the selected QoS criteria is suggested in order to assist clients when choosing the right criteria. Experiments via WEKA tool and simulation prototype were conducted to evaluate the methods used. For the runtime path prediction method, the results showed that the path prediction method achieved promising prediction accuracy, and the number of paths involved in the prediction did not affect the accuracy. For the optimization mechanism, the evaluation was conducted by comparing the mechanism with relevant optimization techniques. The simulation results showed that the proposed optimization mechanism outperforms the relevant optimization techniques by (1) generating the highest overall QoS ratio solutions, (2) consuming the smallest computation time, and (3) producing the lowest percentage of constraints violated number.

Keywords: Web service composition, QoS, Service selection, Heuristic algorithm, Data mining.
Acknowledgement

All thanks to Allah who gave me the ability to finish this work. Without the mercy of Allah, I could not achieve anything. I would like to gratefully acknowledge the enthusiastic supervision of my thesis supervisor, Professor Dr. Zulikha Bt Jamaludin and my Co-supervisor Dr. Massudi bin Mahmuddin. I could not have imagined having a better adviser and mentor for my Ph.D. Without their inspiration, stimulating suggestions, sound advice, guidance, and active participation throughout the process of work, I would never have finished.

I am truly and deeply indebted to so many people that there is no way to acknowledge them all or even any of them properly. I am grateful to my lovely wife, Dua’a. I’ll forever be indebted to this great woman. Very special thanks to your practical and emotional support. Without your endless patience, understanding, support, and help, I would never have been able to finish this thesis. Your supreme trust is always the most efficient motivation to accomplish my ultimate goal. No word can describe what you have done for me. I love you. I would like to give special thanks to my cute son Yammn, who came to this life while I was studying for my PhD. His gorgeous smile blessed and encouraged me to continue my studies.

I would like to thank my Father, who passed away during my PhD study, for his support and encouragement to continue my study. I know his soul is very happy about my successful achievement. Then I would like to thank my beloved mother, who is always supporting me and praying for me to obtain this degree. I respect her deep faith, unconditional love, and support at each time of my life made me this man whom I am today. Furthermore, I would like to thank my stepfather Dr. Abdul Raheem Issa and stepmother for their support. Warm thanks go to my dearest brothers Mohammed and Emad, sisters Suhad, Sahar, Raghda, Wafa’ and Abeer, and friends, especially Dr. Yousif Faza, Dr. Ammar Yasser, Dr. Mossab Al Hunaity, Wael Abo Rahma, Mohammed Eriqat, and Bashar Barakat for giving me their unequivocal support throughout, as always, for which my mere expression of thanks, likewise, does not suffice.
Table of Contents

Permission to Use........................................................................................................... i
Abstrak.......................................................................................................................... ii
Abstract .......................................................................................................................... iii
Acknowledgement ......................................................................................................... iv
Table of Contents ......................................................................................................... v
List of Tables .................................................................................................................. ix
List of Figures ................................................................................................................ xi
List of Appendices ......................................................................................................... xiii
List of Abbreviations ..................................................................................................... xiv

CHAPTER ONE: INTRODUCTION .............................................................................. 1
1.1 Introduction ............................................................................................................... 1
   1.1.1 Web Service Composition Technology for Building Business Processes .......... 4
   1.1.2 QoS-Aware Service Composition ...................................................................... 6
1.2 Problem Background ............................................................................................... 9
1.3 The Problem Statement ........................................................................................... 12
1.4 The Motivation ......................................................................................................... 15
1.5 Research Questions ................................................................................................. 16
1.6 Research Objectives ............................................................................................... 18
1.7 Research Scope ...................................................................................................... 19
1.8 Research Design ...................................................................................................... 20
1.9 Thesis Layout .......................................................................................................... 22

CHAPTER TWO: RELATED WORKS ...................................................................... 24
2.1 QoS for Web Service ............................................................................................... 24
2.2 QoS for Web Service Composition .......................................................................... 25
2.3 Multiple Paths Composition .................................................................................... 27
2.4 Optimization Approaches for QoS-Aware Service Composition ......................... 28
   2.4.1 Single Path Composition Approaches ............................................................... 29
   2.4.2 Multiple Paths Composition Approaches ......................................................... 33
      2.4.2.1 A Separate Path Optimization Technique .................................................... 33
2.4.2.2 All Paths Optimization Technique.................................38

2.5 Solutions for Multi-dimensional Multi-choice Knapsack Problem (MMKP) .....49

2.6 The Path Prediction ..................................................................51

2.7 Chapter Summary ......................................................................53

CHAPTER THREE: RESEARCH METHODOLOGY ...........................................56

3.1 Introduction .............................................................................56

3.2 Analyzing the Research Problem .............................................59

3.3 Multiple Paths QoS-Aware Service Composition Approach ............60

3.3.1 The Problem and Composition Structure Model .........................60

3.3.2 The Selection Criteria ..........................................................61

3.3.3 QoS Computation for Web Service Composition .......................62

3.3.4 The Utility Function ..............................................................62

3.3.5 New Optimization Mechanism .................................................63

3.3.5.1 Prediction of Execution Path ...............................................63

3.3.5.2 The Computation of Optimization ......................................65

3.4 Evaluation of the Proposed Approach ........................................71

3.4.1 Evaluation of Runtime Path Prediction Method ..........................71

3.4.1.1 An Experiment Design .......................................................71

3.4.1.2 Data Preparation ..............................................................72

3.4.1.3 The Data Mining Tool .......................................................77

3.4.1.4 Datasets Preparation for WEKA ........................................77

3.4.1.5 Machine Learning Algorithms .........................................79

3.4.1.6 The Performance Evaluation .............................................80

3.4.1.7 Experiments .....................................................................83

3.4.2 Evaluation of the Optimization Mechanism ................................83

3.4.2.1 An Experimental Design ..................................................83

3.4.2.2 Simulation Prototype .......................................................84

3.4.2.3 The Evaluation Measure ..................................................84

3.4.2.4 Evaluation Methods .......................................................85

3.4.2.5 Experiments ..................................................................90

3.5 Chapter Summary ....................................................................92
CHAPTER FOUR: QUALITY OF SERVICE FOR WEB SERVICE COMPOSITION ................................................................................................................................. 94
4.1 QoS Models and QoS Characteristics ................................................................. 94
4.2 Analysis on QoS Selection Criteria ................................................................... 98
4.3 Priority for QoS Criteria .................................................................................. 103
4.4 Chapter Summary ............................................................................................ 104

CHAPTER FIVE: A SMART QOS-AWARE SERVICE COMPOSITION APPROACH ................................................................................................................................. 106
5.1 The Proposed Approach for QoS-Aware Service Composition ...................... 106
5.2 The Problem Model ......................................................................................... 109
5.3 Modeling the Composition Structures ............................................................... 110
5.4 Selection Criteria ............................................................................................ 111
5.5 QoS Computation for Web Service Composition ............................................ 112
5.6 A Utility Function ........................................................................................... 113
5.7 Multiple Paths Composite Service Scenarios .................................................. 115
  5.7.1 An Auto Insurance Composite Service ....................................................... 116
  5.7.2 A Bank Loan Composite Service ............................................................... 117
5.8 New Optimization Mechanism ........................................................................ 119
  5.8.1 Predicting the Execution Path .................................................................... 119
  5.8.1.1 A Runtime Path Prediction Method ....................................................... 120
  5.8.2 Computing the Optimization .................................................................... 128
  5.8.2.1 Mapping the Selection Problem to Multi-dimensional Multi-choice
  Knapsack Problem (MMKP) ............................................................................. 128
  5.8.2.2 Initial Feasible Solution of Constructive Procedure ............................ 132
  5.8.2.3 Using CCP to Improve the Initial Feasible Solution ............................ 137
5.9 Chapter Summary ........................................................................................... 139

CHAPTER SIX: PERFORMANCE EVALUATION OF THE PROPOSED APPROACH ................................................................................................................................. 141
6.1 Evaluation of Runtime Path Prediction Method ............................................... 141
  6.1.1 Datasets Description ................................................................................ 141
  6.1.2 Experiments and Results .......................................................................... 145
  6.1.2.1 The Accuracy of Path Prediction ......................................................... 146
  6.1.2.2 Scalability of the Prediction Method .................................................... 149
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.2</td>
<td>Evaluation of Optimization Mechanism</td>
<td>153</td>
</tr>
<tr>
<td>6.2.1</td>
<td>A Simulation Prototype</td>
<td>153</td>
</tr>
<tr>
<td>6.2.1.1</td>
<td>Parameters</td>
<td>156</td>
</tr>
<tr>
<td>6.2.1.2</td>
<td>Implementations</td>
<td>157</td>
</tr>
<tr>
<td>6.2.1.3</td>
<td>Software and Hardware Simulation</td>
<td>161</td>
</tr>
<tr>
<td>6.2.2</td>
<td>Experiments and Results</td>
<td>161</td>
</tr>
<tr>
<td>6.2.2.1</td>
<td>QoS Ratio</td>
<td>163</td>
</tr>
<tr>
<td>6.2.2.2</td>
<td>Constraints Violated Number</td>
<td>166</td>
</tr>
<tr>
<td>6.2.2.3</td>
<td>Computation Time</td>
<td>171</td>
</tr>
<tr>
<td>6.2.2.4</td>
<td>Practical Composite Service Scenario</td>
<td>175</td>
</tr>
<tr>
<td>6.3</td>
<td>Conclusions</td>
<td>179</td>
</tr>
<tr>
<td>6.4</td>
<td>Chapter Summary</td>
<td>182</td>
</tr>
</tbody>
</table>

CHAPTER SEVEN: CONCLUSION AND FUTURE RESEARCH WORK | 183 |
| 7.1 | Conclusion of the Research | 183 |
| 7.2 | Contributions of the Research | 187 |
| 7.3 | Research Limitation | 190 |
| 7.4 | Future Works | 191 |

REFERENCES | 193 |
List of Tables

Table 2.1 A Summary of the State of the Art Approaches Proposed to Handle the Multiple Paths Composition Problem ................................................................. 48

Table 3.1 Research Methodology ........................................................................ 57

Table 3.2 An Example of Extended Composition Log ............................................. 64

Table 3.3 Value Ranges of Loan Type, Loan Amount, and Loan Year Attributes ... 75

Table 3.4 A Confusion Matrix of Two Classes ..................................................... 81

Table 3.5 A Comparison Between the Algorithms Used for the Performance Evaluation ........................................................................................................ 90

Table 4.1 A Summary of the Considered QoS Characteristics in the Domain of Web Services ................................................................................................. 99

Table 4.2 The Priority of QoS Criteria and Justifications ....................................... 104

Table 5.1 System Notations .................................................................................. 110

Table 5.2 QoS Aggregation Functions ................................................................... 112

Table 5.3 Training Dataset Structure for Auto Insurance Problem ....................... 123

Table 5.4 Example of Auto Insurance Training Dataset ........................................ 123

Table 6.1 Datasets Description for Bank Loan Composite Service ....................... 144

Table 6.2 Path Description for Auto Insurance Composite Service ..................... 144

Table 6.3 Path Description for Bank Loan Composite Service ............................. 145

Table 6.4 Evaluation Criteria Results Achieved by Using the J48, NB, and SMO Classifiers When Applied to Auto Insurance Dataset ........................................ 147

Table 6.5 Average Accuracy and Number of Involved Paths for Dataset1, Dataset5, Dataset4, and Dataset9 .................................................................................. 151

Table 6.6 Precision and Recall Results for Dataset2 .............................................. 152

Table 6.7 Precision and Recall Results for Dataset4 .............................................. 152

Table 6.8 Parameter Value Ranges of the Simulation ........................................... 157
Table 6.9 Cost and Response Time Intervals.......................................................... 158
Table 6.10 Setup for the Resulting QoS Ratio Test........................................... 163
Table 6.11 The Results of QoS Ratio Test............................................................. 164
Table 6.12 Setup for the Constraints Violated Number Test.............................. 166
Table 6.13 The Average of the Aggregated QoS Characteristics in Each Run and the
Average of the Constraint Imposed for Each Considered QoS Characteristic ...... 167
Table 6.14 The Results of the Constraints Violated Number Test ..................... 169
Table 6.15 A Setup for the Computation Time Test............................................ 172
Table 6.16 The Results of Computation Time Test............................................. 173
Table 6.17 Setup for Test Experiment ................................................................. 177
Table 6.18 Results for Test Experiment.............................................................. 177
Table 6.19 Average QoS ratio, Total Number of Considered Constraint, and
Constraints Violated Number............................................................................. 179
List of Figures

Figure 1.1. SOA model (Source: Newcomer & Lomow, 2004) ...........................................2
Figure 1.2. Web service composition phases ................................................................8
Figure 1.3. An example of multiple paths composition ..................................................10
Figure 2.1. Illustration of competing web services .........................................................26
Figure 3.1. Flowchart for CP ................................................................. ............................... 68
Figure 3.2. Flowchart for CCP ......................................................................................70
Figure 3.3. The main steps for the experimental procedure used for evaluating the runtime path prediction method ....................................................................................72
Figure 3.4. A fragment of the auto insurance data collected from First Insurance Company ..........................................................................................................................74
Figure 3.5. An illustration of dataset5 created by using a spreadsheet application ... 78
Figure 3.6. A fragment of dataset5 in ARFF format .......................................................79
Figure 3.7. A graphical illustration of 10-fold cross-validation method using a dataset which consists of 1000 instances, test subsets (gray), train subsets (white) .. 82
Figure 3.8. The main steps for the experimental procedure used for evaluating the optimization mechanism .................................................................................................84
Figure 4.1. Software product quality model ....................................................................95
Figure 4.2. Suggested QoS criteria for web service selection .........................................101
Figure 5.1. The proposed approach ............................................................................107
Figure 5.2. A multiple execution paths composition .....................................................111
Figure 5.3. A typical auto insurance composite service scenario ..................................117
Figure 5.4. A typical bank loan composite service scenario .........................................118
Figure 5.5. A typical online application form for requesting auto insurance ............126
Figure 5.6. The runtime path prediction method .........................................................127
Figure 5.7. Representation for a solution ....................................................................132
Figure 5.8. The constructive procedure (CP) .................................................. 136

Figure 5.9. The Complementary Constructive Procedure (CCP) ..................... 139

Figure 6.1. The prediction accuracy per each fold achieved using J48, NB, and SMO classifiers when applied to auto insurance dataset .............................................. 147

Figure 6.2. A number of correctly/incorrectly classified instances achieved by using J48, NB, and SMO classifiers when applied to auto insurance dataset .......... 149

Figure 6.3. An average prediction accuracy acheived by using NB, J48, and SMO classifiers when applied to 9 different datasets .............................................. 150

Figure 6.4. A screen capture of generation of 5 candidate services and their QoS characteristics values and utility ............................................................... 160

Figure 6.5. Average QoS ratios of the different algorithms in each run .......... 164

Figure 6.6. Constraints violated numbers of the different algorithms in each run... 169

Figure 6.7. Computation times of the different algorithms each run .............. 173

Figure 6.8. A typical travel agency scenario ..................................................... 176
List of Appendices

Appendix A: Test Experiments for Evaluating the Optimization Mechanism……208

Appendix B: Data Collection……………………………………………………………212

Appendix C: Validation and Verification………………………………………………215

Appendix D: Performance Evaluation Results………………………………………..234
**List of Abbreviations**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACO</td>
<td>Ant Colony Optimization</td>
</tr>
<tr>
<td>ARFF</td>
<td>Attribute Relation File Format</td>
</tr>
<tr>
<td>BPI</td>
<td>Business Process Intelligence</td>
</tr>
<tr>
<td>BPMS</td>
<td>Business Process Management System</td>
</tr>
<tr>
<td>CACO</td>
<td>Continuous Ant Colony Optimization</td>
</tr>
<tr>
<td>CIAC</td>
<td>Continuous Interacting Ant Colony</td>
</tr>
<tr>
<td>CCP</td>
<td>Complementary Constructive Procedure</td>
</tr>
<tr>
<td>CP</td>
<td>Constructive Procedure</td>
</tr>
<tr>
<td>CSV</td>
<td>Comma Separated Value</td>
</tr>
<tr>
<td>DAG</td>
<td>Directed Acyclic Graph</td>
</tr>
<tr>
<td>ESGA</td>
<td>Elitist Selection Genetic Algorithm</td>
</tr>
<tr>
<td>FN</td>
<td>False Negative</td>
</tr>
<tr>
<td>FP</td>
<td>False Positive</td>
</tr>
<tr>
<td>FS</td>
<td>Feasible State</td>
</tr>
<tr>
<td>GAELS</td>
<td>Genetic Algorithm Embedded Local Searching</td>
</tr>
<tr>
<td>GA</td>
<td>Genetic Algorithm</td>
</tr>
<tr>
<td>GSA</td>
<td>Gravitational Search Algorithm</td>
</tr>
<tr>
<td>HGA</td>
<td>Hybrid Genetic Algorithm</td>
</tr>
<tr>
<td>HR</td>
<td>Harmony Research</td>
</tr>
<tr>
<td>IDE</td>
<td>Integrated Development Environment</td>
</tr>
<tr>
<td>ILP</td>
<td>Integer Linear Programming</td>
</tr>
<tr>
<td>IP</td>
<td>Integer Programming</td>
</tr>
<tr>
<td>ISO</td>
<td>International Organization for Standardization</td>
</tr>
<tr>
<td>IT</td>
<td>Information Technology</td>
</tr>
<tr>
<td>MCDM</td>
<td>Multiple Criteria Decision Making</td>
</tr>
<tr>
<td>MCOP</td>
<td>Multi-Constraint Optimal Path</td>
</tr>
<tr>
<td>MILP</td>
<td>Mixed Integer Linear Programming</td>
</tr>
<tr>
<td>MMKP</td>
<td>Multi-dimensional Multi-choice Knapsack Problem</td>
</tr>
<tr>
<td>NB</td>
<td>Naïve Base</td>
</tr>
<tr>
<td>NP</td>
<td>Non-deterministic Polynomial-time</td>
</tr>
<tr>
<td>OASIS</td>
<td>Organization for the Advancement of Structured Information Standards</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>PACA</td>
<td>Particle-Ant Colony Algorithm</td>
</tr>
<tr>
<td>PAIS</td>
<td>Process-Aware Information System</td>
</tr>
<tr>
<td>PSO</td>
<td>Particle Swarm Optimization</td>
</tr>
<tr>
<td>QoE</td>
<td>Quality of Experience</td>
</tr>
<tr>
<td>QoS</td>
<td>Quality of Service</td>
</tr>
<tr>
<td>QP</td>
<td>Quadratic Programming</td>
</tr>
<tr>
<td>QQDSGA</td>
<td>Quality of Experience (QoE)/Quality of Service (QoS) Driven Simulated Annealing-based Genetic Algorithm</td>
</tr>
<tr>
<td>SA</td>
<td>Simulated Annealing</td>
</tr>
<tr>
<td>SAW</td>
<td>Simple Additive Weight</td>
</tr>
<tr>
<td>SLA</td>
<td>Service Level Agreement</td>
</tr>
<tr>
<td>SMO</td>
<td>Sequential Minimal Optimization</td>
</tr>
<tr>
<td>SN</td>
<td>Solution</td>
</tr>
<tr>
<td>SOA</td>
<td>Service Oriented Architecture</td>
</tr>
<tr>
<td>SOAP</td>
<td>Simple Object Access Protocol</td>
</tr>
<tr>
<td>SOC</td>
<td>Service Oriented Computing</td>
</tr>
<tr>
<td>SVM</td>
<td>Support Vector Machines</td>
</tr>
<tr>
<td>TN</td>
<td>True Negative</td>
</tr>
<tr>
<td>TP</td>
<td>True Positive</td>
</tr>
<tr>
<td>TS</td>
<td>Tabu Search</td>
</tr>
<tr>
<td>UDDI</td>
<td>Universal Description Discovery and Integration</td>
</tr>
<tr>
<td>URL</td>
<td>Uniform Resource Locator</td>
</tr>
<tr>
<td>US</td>
<td>Unfeasible State</td>
</tr>
<tr>
<td>W3C</td>
<td>World Wide Web Consortium</td>
</tr>
<tr>
<td>WEKA</td>
<td>Waikato Environment For Knowledge Analysis</td>
</tr>
<tr>
<td>WFMS</td>
<td>Workflow Management System</td>
</tr>
<tr>
<td>WS-BPEL</td>
<td>Web Services Business Process Execution Language</td>
</tr>
<tr>
<td>WSDL</td>
<td>Web Service Description Language</td>
</tr>
<tr>
<td>WSQM</td>
<td>Web Service Quality Model</td>
</tr>
<tr>
<td>XML</td>
<td>Extensible Markup Language</td>
</tr>
</tbody>
</table>
CHAPTER ONE
INTRODUCTION

1.1 Introduction

Service Oriented Computing (SOC) recently has gained a considerable momentum from both industry and academia as a new emerging paradigm to develop rapid, low cost, and loosely coupled software systems. This vision is captured by Service Oriented Architecture (SOA) through the provision of an architectural style (Michlmayr, Rosenberg, Platzer, Treiber & Dustdar, 2006). SOA is “a way of designing a system so that it can provide services to end users and/or other applications in the network” (Baryannis et al., 2008).

The SOA model illustrated in Figure 1.1 consists of three core entities: service provider, service consumer (also called requester), and service registry. The service provider implements the web service and describes it using a standard format. And then it publishes the description in the service registry. The service consumer queries the registry about a specific web service. The service registry checks, whether the requested web service is available or not. If it is available, the registry returns descriptions of the matched web services back to the service consumer. The service consumer obtains the location of the selected web service from the returned descriptions. Finally, the service consumer binds and invokes the web service.
The contents of the thesis is for internal user only
REFERENCES


199


Miremadi, & S. Hessabi (Eds.), *Advances in Computer Science and Engineering* (pp. 265-282). Springer Berlin Heidelberg.


