

FACULTY OF ENGINEERING AND INFORMATION  
TECHNOLOGY

# Tree Similarity Measure- based Recommender Systems

**Dianshuang Wu**

A thesis submitted for the Degree of  
Doctor of Philosophy



University of Technology, Sydney  
October, 2014

# CERTIFICATE OF AUTHORSHIP/ORIGINALITY

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Signature of Candidate

Production Note:  
Signature removed prior to publication.

---

# ACKNOWLEDGEMENTS

I would like to express my earnest thanks to my principal supervisor, Professor Guangquan Zhang, and my co-supervisor, Professor Jie Lu. Their comprehensive guidance has covered all aspects of my PhD study, including research methodology, research topic selection, experiments, academic writing skills and thesis writing, and even the sentence structure and formulas. Their critical comments and suggestions have strengthened my study significantly. Their strict academic attitude and respectful personality have benefited my PhD study and will be a great treasure throughout my life. Without their excellent supervision and continuous encouragement, this research could not have been finished on time. Thanks to you all for your kind help.

I am grateful to all members of the Decision Systems and e-Service Intelligent (DeSI) Lab in the centre for Quantum Computation and Intelligent Systems (QCIS) for their careful participation in my presentation and valuable comments for my research. I especially thank PhD students Mr Mingsong Mao, Mr Wei Wang, Master student Mr Yushi Zhou and other students for their contributions in the development of the recommender system softwares related to this study.

I would like to thank Ms. Barbara Munday and Ms. Sue Felix for helping me to correct English presentation problems in my publications and this thesis.

I am grateful to FEIT Travel fund, Vice-Chancellor's Postgraduate Conference Fund, and the International Postgraduate Research Scholarship (IPRS) and Australian Postgraduate Award (APA) scholarship.

Last but not least, I would like also to thank my family members. Thanks to my mother and father for their conscious encouragement and generous support.

# ABSTRACT

The rapid growth of web information provides excellent opportunities for developing e-services in many applications but also caused increasingly severe information overload problems whereby users are not able to locate relevant information to exactly meet their needs efficiently by using the current Internet search functions. A personalised recommender system aims to handle this issue.

A big challenge in current recommender system research is: the items and user profiles in many recommender system applications nowadays, such as the e-business and e-learning recommender systems, are so complex that they can only be described in complicated tree structures. Therefore, the item or user similarity measure, as the core technique of the recommendation approach, becomes a tree similarity measure, which existing recommender systems cannot provide. Another challenge is that in many real life situations, online recommendations to customers in selecting the most suitable products/services are often made under incomplete and uncertain information, which needs fuzzy set theory and techniques to deal with. Thus, how to use fuzzy set techniques to handle data uncertainty issues in tree-structured items or user profiles needs to be investigated.

This research aims to handle these two challenges in both theoretical and practical aspects. It first defines a tree-structured data model, which can be used to model tree-structured items, user profiles and user preferences. A comprehensive similarity measure on tree-structured data considering all the information on tree structures, nodes' concepts, weights and values is then developed, which can be used to compute the semantic similarity between tree-structured items or users, and the matching degree of items to tree-structured user requests. Based on the tree-structured data model, the tree-structured

items and user requirements are modelled as item trees and user request trees respectively. An item tree and user request tree-based hybrid recommendation approach is then developed. To model users' fuzzy tree-structured preferences, a fuzzy preference tree model is proposed. A fuzzy preference tree-based recommendation approach is then developed. Experimental results on an Australian business dataset and the Movielens dataset show that the proposed recommendation approaches have good performance and are well-suited in dealing with tree-structured data in recommender systems. By use of the proposed tree similarity measure and recommendation approaches based on that, two real world applications, a business partner recommender system, Smart BizSeeker, and an e-learning recommender system, ELRS, are designed and implemented, which demonstrate the applicability and effectiveness of the proposed approaches.

# TABLE OF CONTENTS

|  |      |
|--|------|
| CERTIFICATE OF AUTHORSHIP/ORIGINALITY .....                        | i    |
| ACKNOWLEDGEMENTS .....   | ii   |
| ABSTRACT .....   | iii  |
| TABLE OF CONTENTS.....   | v    |
| LIST OF FIGURES.....   | x    |
| LIST OF TABLES.....  | xiii |
| CHAPTER 1 Introduction.....  | 1    |
| 1.1 Background.....  | 1    |
| 1.2 Research questions and objectives.....                         | 3    |
| 1.3 Research significance .....                                    | 6    |
| 1.4 Research methodology and process.....                          | 7    |
| 1.4.1 Research methodology.....                                    | 7    |
| 1.4.2 Research process.....  | 9    |
| 1.5 Thesis structure.....  | 10   |
| 1.6 Publications related to this thesis.....                       | 12   |
| CHAPTER 2 Literature Review.....                                   | 15   |
| 2.1 Recommendation techniques.....                                 | 15   |
| 2.1.1 Content-based recommendation techniques.....                 | 16   |
| 2.1.2 Collaborative filtering-based recommendation techniques..... | 18   |
| 2.1.3 Knowledge-based recommendation techniques.....               | 22   |

|   |    |
|---|----|
| 2.1.4 Hybrid recommendation techniques .....  | 24 |
| 2.1.5 Computational intelligence-based recommendation techniques .....                        | 25 |
| 2.1.6 Fuzzy set techniques in recommender systems.....  | 27 |
| 2.1.7 Social network-based recommendation techniques .....                                    | 31 |
| 2.1.8 Context awareness-based recommendation techniques .....                                 | 33 |
| 2.1.9 Group recommendation techniques .....   | 34 |
| 2.2 Recommender system applications.....  | 34 |
| 2.2.1 E-government recommender systems .....  | 34 |
| 2.2.2 E-business recommender systems.....   | 38 |
| 2.2.3 E-learning recommender systems .....  | 42 |
| 2.2.4 E-library recommender systems.....  | 44 |
| 2.2.5 Comprehensive analysis and findings.....  | 45 |
| 2.3 Tree similarity measure .....   | 48 |
| 2.3.1 Tree structured data.....   | 49 |
| 2.3.2 Tree similarity measure methods .....   | 50 |
| 2.3.3 Structural constraints and semantic similarity on tree-structured data measure<br>..... | 52 |
| 2.4 Summary.....  | 58 |
| CHAPTER 3 Similarity Measure on Tree-structured Data.....                                     | 60 |
| 3.1 Introduction.....   | 60 |
| 3.2 Tree-structured data model.....   | 62 |
| 3.3 Similarity measure method on tree-structured data.....                                    | 66 |
| 3.3.1 Conceptual similarity between two tree structured data .....                            | 67 |
| 3.3.2 Value similarity between two tree structured data.....                                  | 73 |
| 3.3.3 Similarity measurement algorithms .....   | 77 |

|   |     |
|---|-----|
| 3.4 Two illustrative examples and comparison with other approaches .....          | 86  |
| 3.4.1 Similarity measure computation between two tree-structured data .....       | 86  |
| 3.4.2 Similar tree-structured cases retrieval .....                               | 87  |
| 3.4.3 Comparison with other approaches .....                                      | 89  |
| 3.5 Summary .....   | 90  |
| CHAPTER 4 Item Tree and User Request Tree-based Hybrid Recommender System....     | 92  |
| 4.1 Introduction.....   | 92  |
| 4.2 Item tree and user request tree .....   | 94  |
| 4.2.1 Item tree and user request tree definitions .....                           | 94  |
| 4.2.2 An example of the item tree and user request tree .....                     | 95  |
| 4.3 An item tree and user request tree-based hybrid recommendation approach ..... | 100 |
| 4.4 A case study .....  | 105 |
| 4.5 Experimental evaluation .....   | 111 |
| 4.5.1 Evaluation data sets.....   | 111 |
| 4.5.2 Evaluation metrics.....   | 113 |
| 4.5.3 Benchmark recommendation approaches .....                                   | 114 |
| 4.5.4 Evaluation results on the Australian business dataset.....                  | 114 |
| 4.5.5 Evaluation results on the Movielens dataset .....                           | 116 |
| 4.6 Summary.....  | 119 |
| CHAPTER 5 A Fuzzy Preference Tree-based Recommender System .....                  | 120 |
| 5.1 Introduction.....   | 120 |
| 5.2 Fuzzy tree-structured preference model .....                                  | 121 |
| 5.2.1 Users' fuzzy preferences .....  | 121 |
| 5.2.2 Fuzzy tree-structured user preference .....                                 | 122 |
| 5.3 A fuzzy preference tree construction algorithm .....                          | 125 |



|   |     |
|---|-----|
| 5.3.1 Algorithm description .....                                       | 125 |
| 5.3.2 An example .....  | 129 |
| 5.4 A fuzzy preference tree-based recommendation approach.....          | 131 |
| 5.4.1 Approach description .....  | 131 |
| 5.4.2 An example .....  | 135 |
| 5.5 Experimental evaluation .....                                       | 136 |
| 5.5.1 Evaluation data sets.....   | 136 |
| 5.5.2 Evaluation metrics.....   | 137 |
| 5.5.3 Benchmark recommendation approaches .....                         | 138 |
| 5.5.4 Evaluation results .....  | 138 |
| 5.6 Summary.....  | 142 |
| CHAPTER 6 Smart BizSeeker – A Business Partner Recommender System ..... | 144 |
| 6.1 Introduction.....   | 144 |
| 6.2 Requirements of e-business recommender systems .....                | 145 |
| 6.3 System architecture.....  | 146 |
| 6.3.1 Databases .....   | 147 |
| 6.3.2 Web application components .....                                  | 149 |
| 6.4 System implementation .....   | 154 |
| 6.5 A case study.....   | 156 |
| 6.6 Summary.....  | 160 |
| CHAPTER 7 ELRS – An E-Learning Recommender System.....                  | 161 |
| 7.1 Introduction.....   | 161 |
| 7.2 Requirements of e-learning recommender systems.....                 | 162 |
| 7.3 Learning activity tree.....   | 164 |
| 7.3.1 Node concept of the learning activity tree .....                  | 164 |

|   |     |
|---|-----|
| 7.3.2 Fuzzy category tree and the fuzzy category similarity .....               | 165 |
| 7.3.3 The pedagogical relations between learning activities .....               | 169 |
| 7.4 Learner profile tree .....  | 171 |
| 7.4.1 Node concept of the learner profile tree.....                             | 171 |
| 7.4.2 The similarity measures related to the fuzzy required category tree ..... | 173 |
| 7.5 A modified IUTH recommendation approach for learning activities.....        | 175 |
| 7.6 System architecture.....  | 178 |
| 7.6.1 Databases .....   | 180 |
| 7.6.2 Web application components .....  | 180 |
| 7.7 System implementation .....   | 182 |
| 7.8 A case study.....   | 183 |
| 7.9 Summary.....  | 190 |
| CHAPTER 8 Conclusions and Further Study .....                                   | 191 |
| 8.1 Conclusions.....  | 191 |
| 8.2 Further study.....  | 194 |
| References .....  | 196 |
| Abbreviations .....   | 221 |

# LIST OF FIGURES

|  |     |
|--|-----|
| Figure 1-1. Reasoning in the general design cycle (Vaishnavi & Kuechler Jr 2007) .....   | 7   |
| Figure 1-2. Thesis structure .....   | 11  |
| Figure 2-1. The recommendation list of potential business partners generated by BizSeeker (Lu et al. 2010).....  | 37  |
| Figure 2-2. Plan and package recommendation for a customer in the Fuzzy-based Telecom Product Recommender System (Zhang et al. 2013).....  | 41  |
| Figure 2-3. (a) an isomorphic mapping and (b) an edit distance mapping.....  | 54  |
| Figure 2-4. Three examples of different mappings: (a) is a constrained edit distance mapping; (b) is a less constrained edit distance mapping; (c) is an edit distance mapping which is neither constrained nor less-constrained. (Bille 2005) ..... | 56  |
| Figure 3-1. (a) (b) Two examples of tree-structured business data.....   | 61  |
| Figure 3-2. (a) (b) Two examples of tree-structured business data.....   | 69  |
| Figure 3-3. (a) a bipartite graph $G_{ij}$ and (b) its maximum weighted bipartite matching ...   | 72  |
| Figure 3-4. The maximum conceptual similarity tree mapping between tree-structured data $T_1$ and $T_2$ in Figure 3-1.....   | 73  |
| Figure 3-5. The matching tree between trees $T_1$ and $T_2$ in Figure 3-1 .....  | 75  |
| Figure 3-6. Flowchart to compute the similarity between two tree-structured data.....  | 78  |
| Figure 3-7. A new case $T_a$ and five existing cases in a case base .....  | 88  |
| Figure 4-1. A product tree structure (a) and a buying request tree structure (b).....  | 96  |
| Figure 4-2. A product category tree .....  | 97  |
| Figure 4-3. Product tree and buying request tree examples .....  | 99  |
| Figure 4-4. The matching tree between $T_r$ and $T_p$ in Figure 4-3 .....  | 100 |
| Figure 4-5. An item tree and user request tree-based hybrid recommendation approach .....  | 102 |

|   |     |
|---|-----|
| Figure 4-6. The product tree and buying request trees of business users.....  | 106 |
| Figure 4-7. The product trees of the businesses .....   | 107 |
| Figure 4-8. The buying request trees of the businesses.....   | 107 |
| Figure 4-9. The movie tree structure .....  | 112 |
| Figure 4-10. Recommendation accuracy comparison between the IUTH recommendation approach and other benchmark approaches on different numbers of neighbours with the Australian business dataset.....                          | 115 |
| Figure 4-11. Recommendation accuracy comparison between the IUTH recommendation approach and other benchmark approaches on different numbers of neighbours with all items in Movielens dataset .....                          | 117 |
| Figure 4-12. Recommendation accuracy comparison between the IUTH recommendation approach and HSR approach on different numbers of neighbours with new items in Movielens dataset .....  | 117 |
| Figure 4-13. Recommendation accuracy comparison between the IUTH recommendation approach and other benchmark approaches on different numbers of neighbours with items rated less than twenty times in Movielens dataset ..... | 118 |
| Figure 5-1. Intentionally expressed user preference.....  | 124 |
| Figure 5-2. Two tree-structured data examples.....  | 130 |
| Figure 5-3. The maximum conceptual similarity tree mapping between $T_1$ and $T_2$ .....  | 130 |
| Figure 5-4. The constructed fuzzy preference tree .....   | 131 |
| Figure 5-5. An item tree .....  | 135 |
| Figure 5-6. The maximum conceptual similarity tree mapping between $T_u$ and $T_3$ .....  | 135 |
| Figure 5-7. The MAE comparison with the Australian business data set .....  | 139 |
| Figure 5-8. The precision comparison with the Australian business data set .....  | 140 |
| Figure 5-9. The recall comparison with the Australian business data set .....   | 140 |
| Figure 5-10. The F1 comparison with the Australian business data set.....   | 140 |
| Figure 5-11. The MAE comparison with the MovieLens data set .....   | 141 |
| Figure 5-12. The precision comparison with the MovieLens data set .....   | 141 |
| Figure 5-13. The recall comparison with the MovieLens data set .....  | 142 |
| Figure 5-14. The F1 comparison with the MovieLens data set.....   | 142 |
| Figure 6-1. The architecture of Smart BizSeeker .....   | 148 |

|   |     |
|---|-----|
| Figure 6-2. The product tree structure (a) and the buying request tree structure (b) in Smart BizSeeker .....   | 151 |
| Figure 6-3. The Smart BizSeeker site map.....   | 153 |
| Figure 6-4. The login page of Smart BizSeeker .....   | 156 |
| Figure 6-5. The product tree and buying request tree .....  | 156 |
| Figure 6-6. Profile management page .....   | 157 |
| Figure 6-7. Product management page.....  | 158 |
| Figure 6-8. Buying request management page.....   | 158 |
| Figure 6-9. The supplier recommendation results.....  | 159 |
| Figure 6-10. The buyer recommendation results.....  | 159 |
| Figure 6-11. Comment and rating page.....   | 160 |
| Figure 7-1. The structure of a learning activity tree.....  | 164 |
| Figure 7-2. The learning activity category tree .....   | 165 |
| Figure 7-3. The fuzzy category trees of two learning activities: (a) is the fuzzy category tree of the subject <i>Business Intelligence</i> . (b) is the fuzzy category tree of the subject <i>Marketing Management</i> . ..... | 166 |
| Figure 7-4. The combination of two fuzzy category trees in Figure 7-3.....  | 169 |
| Figure 7-5. The structure of a learner profile tree .....   | 171 |
| Figure 7-6. Two fuzzy required category trees .....   | 172 |
| Figure 7-7. The architecture of the e-learning recommender system.....  | 179 |
| Figure 7-8. The homepage of the e-learning recommender system .....   | 182 |
| Figure 7-9. Five learner profiles.....  | 184 |
| Figure 7-10. The fuzzy category trees of the subjects.....  | 185 |
| Figure 7-11. The student profile page .....   | 185 |
| Figure 7-12. The student's study room .....   | 186 |
| Figure 7-13. Student rating and comment input page .....  | 187 |
| Figure 7-14. Learning activity recommendation results .....   | 187 |

# LIST OF TABLES

|  |     |
|--|-----|
| Table 2-1. Summary of recommendation techniques in main application domains .....                      | 46  |
| Table 2-2. Summary of the recommender systems developed, the techniques applied and users suited ..... | 46  |
| Table 3-1. The conceptual similarity between node labels .....   | 65  |
| Table 3-2. The attribute definition of a matching tree node.....                                       | 82  |
| Table 3-3. The value ranges of four value types .....  | 87  |
| Table 3-4. Similarity between $T_a$ and cases in the case base.....                                    | 89  |
| Table 3-5. Comparison between our proposed method and other methods .....                              | 90  |
| Table 4-1. Products of the businesses .....  | 106 |
| Table 4-2. Buying requests of the businesses.....  | 107 |
| Table 4-3. User supplier rating matrix .....   | 108 |
| Table 4-4. Request matching degrees.....   | 108 |
| Table 4-5. Semantic similarities between businesses.....   | 109 |
| Table 4-6. Item-based CF similarities between businesses.....  | 109 |
| Table 4-7. Total similarities between businesses .....   | 109 |
| Table 4-8. Predicted ratings to suppliers.....   | 110 |
| Table 4-9. Predicted ratings to buyers.....  | 110 |
| Table 4-10. Coverage rate of the recommendation algorithms for the Australian business dataset.....    | 115 |
| Table 4-11. Coverage rate of the recommendation algorithms .....                                       | 119 |
| Table 6-1. Business categories in Smart BizSeeker .....  | 154 |
| Table 6-2. Product categories in Smart BizSeeker.....  | 155 |
| Table 7-1. Linguistic terms and related fuzzy numbers for learner requirement.....                     | 172 |

|   |     |
|---|-----|
| Table 7-2. Learner-subject rating matrix.....                                   | 188 |
| Table 7-3. The matching degrees of the learning activities to the learners..... | 189 |
| Table 7-4. The semantic similarity between learners.....                        | 189 |
| Table 7-5. The CF similarity between learners.....                              | 189 |
| Table 7-6. The total similarity between learners.....                           | 189 |
| Table 7-7. The predicted ratings.....   | 189 |