

FACULTY OF ENGINEERING AND INFORMATION
TECHNOLOGY

Enhanced Group Recommender System and Visualization

Wei Wang

A thesis submitted for the Degree of
Doctor of Philosophy



University of Technology Sydney
March, 2016

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

ACKNOWLEDGEMENTS

This thesis represents not only my work at UTS, it is a milestone for almost four years of work at UTS and specifically within the Decision Systems and e-Service Intelligence (DeSI) Lab. This thesis is also the result of many experiences I have encountered at UTS from dozens of remarkable individuals who I also wish to acknowledge.

First and foremost I wish to thank my principal supervisor, Professor Guangquan Zhang, and my co-supervisor, Professor Jie Lu. They have been supportive since the days I began working on DeSI as an undergraduate and have 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 personalities 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 would like to express my sincerest gratitude to Prof. Dacheng Tao, Dr. Wei Bian and Dr. Tianyi Zhou for their helping me come up with and sharing code of the matrix decomposition topic; to Prof. Maolin Huang and Miss Wenbo Wang for data visualization advice; to Dr. Dianshuang Wu, Dr. Mingsong Mao, for efficient, and most important, friendly assistance in recommender systems and to Ms. Barbara Munday and Ms. Sue Felix for helping me to correct English presentation problems in my publications and this thesis.

The most difficult ones to thank are my family; my mother and father, my wife and son. I appreciate your understanding, your encouragement and enthusiasm, and

your help. I couldn't have done it without your support. Without you nothing in my life would be possible.

Last but not least, I am grateful to the FEIT Travel fund, the Vice-Chancellor's Postgraduate Conference Fund, and the ARC scholarship.

ABSTRACT

Requirement of group recommender systems (GRSs) is experiencing a dramatic growth due to intelligent services being applied more broadly and involved in more and more domains. However, effectivity and interpretability are still two challenges in GRSs. A typical scenario is: a group is formed randomly without active organizing in advance and sufficient negotiation between members before recommending, such as e-shopping and e-tourism. Therefore, deeply modeling the group profile is the first key part to generate recommendations. Moreover, accurately predicting should be a problem under biased and limited information provided by users. The interpretability challenge is that most of GRSs are black boxes for providing no necessary explanation of recommendations but only a list. It is quite important to convince members to make them understand why the specific recommendations are reasonable. Thus, explaining the reason generated recommendations and relationships between members needs to be investigated.

This research aims to handle these two challenges in both theoretical and practical aspects. A novel group recommendation approach is developed and aims to maximize satisfaction within random groups by modeling the group profiles through the analysis of contributed member ratings alone. First, the Contribution Score is defined to numerically measure each member's importance in terms of the sub-rating matrix which makes it practical even when the matrix is highly incomplete and sparse. Second, a local collaborative filtering method is developed to address the biased rating problem caused by severe preference conflicting in random groups. An adaptive average rating calculating model is proposed taking into consideration of the target item by reducing the set to those which are highly relevant to it. By integrating these two models, a Contribution Score-based Group Recommendation (CS-GR) approach is developed to efficiently depict groups. Also, a novel hierarchy graph-based visualization method, based on data visualization techniques, which are

powerful tools to offer intuitive abstractions of concepts, is suggested to offer explanations for users. First a higher level of abstraction of the overall recommender modules, such as group profile modeling and prediction calculating, is presented using a hierarchy graph. To do this, all the entities involved in a group recommender process are summarized and visualized as nodes in the graph and the edges in the graph represent information inherited. Second, the layout provides detailed information for individual members to track their influences in the system by adding pie charts at each single node to show individual influences for all involved members. This enables members to track and compare their influences with others in every single procedure.

This research provides the GRSs effectivity for the biased and sparse information which can be handled to model the group and generate the predictions. The scalability and efficiency are also guaranteed because only rating information is needed and matrix decomposition technique is employed. The visualization is used to provide both overall and detailed explanation for users.

TABLE OF CONTENTS

CERTIFICATE OF AUTHORSHIP/ORIGINALITY	i
ACKNOWLEDGEMENTS	ii
ABSTRACT	iv
TABLE OF CONTENTS.....	vi
LIST OF FIGURES.....	x
LIST OF TABLES.....	xiv
CHAPTER 1 Introduction.....	1
1.1 Background.....	1
1.2 Research questions.....	4
1.3 Research objectives.....	4
1.4 Research significance	7
1.4.1 Theoretical significance	7
1.4.2 Practical significance	8
1.5 Research methodology and process.....	8
1.5.1 Research methodology.....	8
1.5.2 Research process	9
1.6 Thesis structure.....	10
1.7 Publications related to this thesis.....	10
CHAPTER 2 Literature Review.....	13
2.1 Basic recommender systems.....	13
2.1.1 Concepts.....	13

2.1.2 Techniques	14
2.1.3 Recommender applications	24
2.2 Group recommender systems.....	35
2.2.1 Concepts	35
2.2.2 Techniques	36
2.2.3 Recommender applications	41
2.3 Matrix factorization	46
2.3.1 Matrix factorization in recommender systems.....	46
2.3.2 Non-negative matrix factorization in recommender systems.....	48
2.3.3 Separable non-negative matrix factorization in recommender systems	50
2.4 Data visualization	51
2.4.1 Concepts.....	51
2.4.2 Data visualization in recommender systems	52
CHAPTER 3 Local Collaborative Filtering Approach	59
3.1 Introduction.....	59
3.2 Local collaborative filtering approach.....	60
3.2.1 Local average rating estimation approach.....	61
3.2.2 Group local average rating estimation approach.....	67
3.3 A case study	72
3.3.1 Leave-one-out cross validation	72
3.3.2 Scenario 1	74
3.3.3 Scenario 2.....	75
3.3.4 Discussion	76
3.4 Summary	77
CHAPTER 4 Entropy-driven User Similarity for Collaborative Filtering.....	78
4.1 Introduction.....	78

4.2 Information entropy-based collaborative filtering.....	80
4.2.1 Framework of method.....	80
4.2.2 Entropy-driven user similarity measure.....	81
4.2.3 Hybrid user similarity.....	84
4.2.4 Group local collaborative filtering.....	85
4.3 Experiments and evaluation.....	85
4.3.1 Data set.....	86
4.3.2 Experiment design.....	87
4.3.3 Experiment result.....	88
4.4 Summary.....	94
CHAPTER 5 Contribution Score-Based Group Recommender System.....	95
5.1 Introduction.....	95
5.2 Contribution score measure for members.....	97
5.3 Contribution score-based group recommendation method.....	98
5.3.1 Contribution score calculation in single sample.....	99
5.3.2 Global member contribution in all samples.....	102
5.3.3 Group modeling using contributions scores.....	103
5.3.4 Unknown group ratings calculation using local collaborative filtering method.....	107
5.4 Experiments and evaluation.....	108
5.4.1 Datasets and pre-processing.....	109
5.4.2 Group generation protocol.....	110
5.4.3 Metrics.....	110
5.4.4 Experiment design.....	112
5.4.5 Results and discussion.....	113
5.5 GroTo: a contribution score-based group recommender system for tourism	121

5.6 Summary	123
CHAPTER 6 Hierarchy Visualization for Group Recommender Systems	126
6.1 Introduction.....	126
6.2 Hierarchy visualization method for group recommender systems	128
6.2.1 Layout	128
6.2.2 Components	129
6.3 Hierarchy visualization implementation.....	135
6.3.1 Usability	136
6.3.2 Interactivity	138
6.3.3 Adaptability.....	139
6.3.4 Expansibility	139
6.4 Hierarchy visualization on SmartBizSeeker	140
6.4.1 Architecture.....	144
6.4.2 Hierarchy visualization module	145
6.4.3 Similarity visualization module	149
6.5 Summary.....	151
CHAPTER 7 Conclusions and Further Study	152
7.1 Conclusions.....	152
7.2 Further study	154
References	156
Abbreviations	175

LIST OF FIGURES

Figure 1-1. Research Methodology	9
Figure 1-2. The structure and contents of this thesis.....	12
Figure 2-1. The two basic approaches to making group recommendations. The top approach aggregates individual preferences and the bottom approach aggregates individual recommendations.	37
Figure 2-2. CATS tourism system critiquing interface (McCarthy, Salamó, et al. 2006a).....	45
Figure 2-3. INTRIGUE system preference specification interface (Ardissono et al. 2003).....	46
Figure 2-4. Foxtrot system. The profile is illustrated at the top of the page.	52
Figure 2-5. TasteWeights system for music recommendation.	53
Figure 2-6. An example of product tree.	54
Figure 2-7. A dependency network in (Heckerman et al. 2001) to show the relationships between demographic information and internet-use data.	55
Figure 2-8. An example of chord and Sankey diagram.....	56
Figure 2-9. Fan lens diagram is presented in the main window when line chart, parallel coordinates and bar chart are shown at the side bar as alternatives.	57
Figure 2-10. SOM visualization.	57
Figure 2-11. (a) is an example of hierarchy relationship on a movie. Producer, director and actor are there different types of nodes that a movie inherits from. (b) represents general hierarchy graph and use different level to represent different type of nodes when different line type to represent different edge type.	58

Figure 3-1. Explanation of calculating the local average rating. The relevance between the target item and a non-target item is measured according to distance between two corresponding item rating vectors.	62
Figure 3-2. A 2D example for item relevance measuring.	65
Figure 4-1. Flowchart of the approach.	80
Figure 4-2 A page of SBS system show a list of recommended suppliers.	86
Figure 4-3. MAE results on MovieLens when using different α under given threshold T	90
Figure 4-4. MAE results on MovieLens when using different T under given threshold α	90
Figure 4-5. MAE results on SBS when using different α under given threshold T	92
Figure 4-6. MAE results on SBS when using different T under given threshold α	92
Figure 4-7. Neighbours vs MAE on MovieLens.....	93
Figure 4-8. Neighbours vs MAE on SBS.....	93
Figure 5-1. Contribution Score-based Recommender System Architecture.....	98
Figure 5-2. Explanation of sampling and aggregating of CS model to compute the contributions of the group members.....	99
Figure 5-3. nDCG result of MovieLens100K.	114
Figure 5-4. nDCG result of MovieLens1M.....	114
Figure 5-5. nDCG result of Jester.	115
Figure 5-6. F result of MovieLens100K.....	116
Figure 5-7. F result of MovieLens1M.....	116
Figure 5-8. F scores F result of Jester.	117
Figure 5-9. nDCG results when using the CS model alone and when combining the LCF model using 0.2, 0.3 and 0.4 thresholds to produce local average ratings on 100K.....	118

Figure 5-10. nDCG results when using the CS model alone and when combining the LCF model using 0.2, 0.3 and 0.4 thresholds to produce local average ratings on 1M.	118
Figure 5-11 nDCG results when using the CS model alone and when combining the LCF model using 0.2, 0.3 and 0.4 thresholds to produce local average ratings on Jester.	119
Figure 5-12. F results when using the CS model alone and when combining the LCF model using 0.2, 0.3 and 0.4 thresholds to produce local average ratings on 100K.	119
Figure 5-13. F results when using the CS model alone and when combining the LCF model using 0.2, 0.3 and 0.4 thresholds to produce local average ratings on 1M.	120
Figure 5-14. F results when using the CS model alone and when combining the LCF model using 0.2, 0.3 and 0.4 thresholds to produce local average ratings on Jester.	120
Figure 5-15. Architecture of the tourism recommender system GroTo.	122
Figure 6-1. A visualization example on real data set MovieLens 100K.	138
Figure 6-2. Visualization supports zoom and pan to enable interactivity.	139
Figure 6-3. The login page of SBS.	142
Figure 6-4. Buying request management page.	142
Figure 6-5. The supplier recommendation results.	143
Figure 6-6. The buyer recommendation results.	143
Figure 6-7. SBS system architecture with visualization module.	145
Figure 6-8. Trackable hierarchy visualization result example.	146
Figure 6-9. Floating business information over node.	147
Figure 6-10. Floating business information over node.	148
Figure 6-11. Highlighting for a recommendation node.	148
Figure 6-12. Highlighting for a profile node.	148

Figure 6-13. Higher tree similarity example for sharing common Party product.	149
Figure 6-14. Higher tree similarity example for sharing common Accommodation product.....	150
Figure 6-15. Lower tree similarity example for sharing Pizza product.....	150

LIST OF TABLES

Table 3-1. Rating matrix of the case.	72
Table 3-2. Prediction results of the case.....	73
Table 3-3. MAE results of the case.	73
Table 4-1. Two vectors for similarity calculating.	79
Table 4-2. MAE with different approaches.....	88
Table 5-1. Features of three test data sets.....	109
Table 5-2. Ratings of group members on the activities Nature and Sport: each row represents a member.	124
Table 5-3. Observed ratings of non-member users for the activities Nature and Sport	125
Table 6-1. Visualization Components Summarization.....	137