

Faculty of Engineering and Information Technology  
University of Technology, Sydney

# **Incorporating Couplings into Collaborative Filtering**

A thesis submitted in partial fulfillment of  
the requirements for the degree of  
**Doctor of Philosophy**

by

Fangfang Li

April 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

Production Note:

Signature removed prior to publication.

---



# Acknowledgments

Foremost, I would like to express my sincere gratitude to my supervisors Prof. Longbing Cao and Dr. Guandong Xu for their continuous support of my Ph.D study and research, and for their patience, motivation, enthusiasm, and immense knowledge. Their guidance helped me at all times throughout the research and writing of this thesis.

I also would like to express my appreciation to Prof. Xiaozhong Fan and Prof. Zhendong Niu at Beijing Institute of Technology for their continuous support throughout my PhD study.

I am grateful for the expert and sincere help from Yanchang Zhao at the Department of Immigration and Border Protection in the cooperative project. His support secured the success of the cooperation. I am also grateful to my fellow friends, Wang Can, Zhong, Xueping Peng, Bin Fu, Chunming Liu, Mu Li, Xia Cui, Junfu, Yin Song and Wei Cao, for their discussions and help.

Last but not least, I would like to thank my family for their unconditional support, both financially and emotionally throughout the whole PhD study.

Fangfang Li

November 2015 @ UTS



# Contents

Certificate . . . . .	i
Acknowledgment . . . . .	iii
List of Figures . . . . .	ix
List of Tables . . . . .	xi
List of Publications . . . . .	xiii
Abstract . . . . .	xvii
<b>Chapter 1 Introduction . . . . .</b>	<b>1</b>
1.1 Background . . . . .	1
1.2 IID Recommender Systems . . . . .	3
1.3 Non-IID Recommender Systems . . . . .	7
1.4 Research Issues . . . . .	9
1.5 Research Contributions . . . . .	11
1.6 Thesis Structure . . . . .	12
<b>Chapter 2 Literature Review and Foundation . . . . .</b>	<b>15</b>
2.1 IID Recommender Systems . . . . .	15
2.1.1 Notations . . . . .	15
2.1.2 Demographic Filtering . . . . .	16
2.1.3 Collaborative Filtering . . . . .	16
2.1.4 Content-based Recommendation . . . . .	22
2.1.5 Hybrid Filtering . . . . .	24
2.2 Non-IIDness Learning . . . . .	26
2.2.1 Efforts of Non-IIDness Learning . . . . .	27

2.2.2	Similarity Computation . . . . .	29
2.2.3	Aspects of Non-IIDness Learning . . . . .	33
2.2.4	Challenges of Non-IIDness Learning . . . . .	35
2.3	Non-IID Recommender Systems . . . . .	38
2.3.1	Social RS . . . . .	38
2.3.2	Incorporating Coupling into RS . . . . .	45
 <b>Chapter 3 Coupled Neighbourhood-based Collaborative Fil-</b>		
	<b>tering . . . . .</b>	<b>50</b>
3.1	Introduction . . . . .	50
3.2	Coupled Neighbourhood-based CF . . . . .	54
3.2.1	Coupling Relations . . . . .	55
3.2.2	Coupled User-based CF . . . . .	56
3.2.3	Coupled Item-based CF . . . . .	58
3.2.4	Coupled CF . . . . .	60
3.2.5	Complexity Analysis . . . . .	61
3.3	Experiments and Results . . . . .	61
3.3.1	Data Set . . . . .	62
3.3.2	Experimental Settings . . . . .	62
3.3.3	Impact of Parameters . . . . .	64
3.3.4	Experimental Analysis . . . . .	67
3.3.5	Discussion . . . . .	70
3.4	Conclusion . . . . .	70
 <b>Chapter 4 Coupled Matrix Factorization . . . . .</b>		<b>71</b>
4.1	Introduction . . . . .	71
4.2	Coupled User-based MF Model . . . . .	74
4.2.1	Coupled User-based MF . . . . .	74
4.2.2	Complexity Analysis . . . . .	77
4.3	Coupled Item-based MF Model . . . . .	78
4.3.1	Coupled Item-based MF . . . . .	78
4.4	Coupled MF Model . . . . .	80

---

4.4.1	Coupled MF . . . . .	80
4.4.2	Complexity Analysis . . . . .	84
4.5	Experiments . . . . .	84
4.5.1	Data Sets . . . . .	85
4.5.2	Experimental Settings . . . . .	85
4.5.3	Experimental Evaluation for CUMF . . . . .	86
4.5.4	Experimental Evaluation for CIMF . . . . .	90
4.5.5	Experimental Evaluation for CMF . . . . .	98
4.6	Conclusion . . . . .	104
<b>Chapter 5 Two-level Matrix Factorization . . . . .</b>		<b>105</b>
5.1	Introduction . . . . .	105
5.2	Two-level Matrix Factorization . . . . .	107
5.2.1	Notations . . . . .	108
5.2.2	Weighted Textual Matrix Factorization . . . . .	109
5.2.3	Semantic Couplings for Items . . . . .	110
5.2.4	Two-level MF Model . . . . .	111
5.2.5	Complexity Analysis . . . . .	112
5.3	Experiments and Results . . . . .	115
5.3.1	Data Sets . . . . .	115
5.3.2	Experimental Settings . . . . .	116
5.3.3	Experiments and Discussions . . . . .	117
5.4	Conclusion . . . . .	125
<b>Chapter 6 Conclusions and Future Work . . . . .</b>		<b>126</b>
6.1	Conclusions . . . . .	126
6.2	Future Work . . . . .	128
<b>Bibliography . . . . .</b>		<b>130</b>



# List of Figures

1.1	Non-IID Recommender System . . . . .	7
1.2	Thesis Structure . . . . .	14
3.1	Item-based CF . . . . .	51
3.2	Couplings in Recommender Systems . . . . .	55
3.3	Coupled User-based CF . . . . .	57
3.4	Coupled Item-based CF . . . . .	59
3.5	Coupled CF . . . . .	60
3.6	Impact of Parameter $\alpha$ and $\beta$ . . . . .	65
3.7	Impact of Neighbour Size . . . . .	66
3.8	Impact of Parameter $\alpha'$ . . . . .	67
4.1	Coupled User-based MF Model . . . . .	74
4.2	Coupled Item-based MF Model . . . . .	78
4.3	Coupled MF Model . . . . .	81
4.4	Performance of CIMF over Hybrid Methods on MovieLens1m .	95
4.5	Performance of CIMF over Hybrid Methods on BookCrossing .	96
4.6	Parameter $\alpha$ of CIMF on MovieLens1m . . . . .	97
4.7	Parameter $\alpha$ of CIMF on BookCrossing . . . . .	98
5.1	Two-level Matrix Factorization Model . . . . .	108
5.2	Weighted Textual Matrix Factorization . . . . .	109
5.3	Impact of Parameter $\alpha$ on MovieLens1m and BookCrossing for TLMF . . . . .	119

*LIST OF FIGURES*

---

5.4	Impact of Parameter $d$ on MovieLens1m and BookCrossing for TLMF . . . . .	121
5.5	Impact of Parameter $K$ on MovieLens1m and BookCrossing for TLMF . . . . .	122
5.6	Impact of Neighbourhood Size of Items on MovieLens1m and BookCrossing for TLMF . . . . .	123

# List of Tables

1.1	User-Item Ratings . . . . .	4
1.2	User Attributes . . . . .	5
1.3	Item Attributes . . . . .	5
2.1	Object Attribute Space . . . . .	47
3.1	User Attributes . . . . .	52
3.2	Item Attributes . . . . .	52
3.3	User-Item Ratings . . . . .	53
3.4	Performance on MovieLens100K . . . . .	69
4.1	Performance of CUMF on MovieLens100K and BookCrossing Data Sets . . . . .	89
4.2	Performance of CUMF for Cold-start Users . . . . .	89
4.3	MF Comparisons on MovieLens1m and BookCrossing Data Sets for CIMF . . . . .	92
4.4	CF Comparisons on MovieLens1m and BookCrossing Data Sets for CIMF . . . . .	93
4.5	Performance of CMF on MovieLens100K and BookCrossing Data Sets . . . . .	101
4.6	Performance of CMF for Cold-start Users . . . . .	103
4.7	Performance of CMF for Cold-start Items . . . . .	103
5.1	A Toy Example . . . . .	106
5.2	Basic Statistics for Data Sets . . . . .	115

*LIST OF TABLES*

---

5.3	CF Comparisons on MovieLens1m and BookCrossing for TLMF118	
5.4	MF Comparisons on MovieLens1m and BookCrossing for TLMF120	
5.5	MF Comparisons on Cold-start Items for TLMF . . . . .	120

# List of Publications

## Papers Published

- **Fangfang Li**, Guandong Xu, Longbing Cao (2015), Coupled Matrix Factorization within Non-IID Context. *in* 'Proceedings of the 19th Pacific-Asia Conference on Knowledge Discovery and Data Mining(PAKDD2015)', pp. 707-719. (CORE2014 A).
- **Fangfang Li**, Guandong Xu and Longbing Cao (2015), Two-level Matrix Factorization for Recommender Systems, *Neural Computing and Applications*, doi: 10.1007/s00521-015-2060-3, pp. 1-12. (SCI indexed).
- Chuyuan Wei, **Fangfang Li**, Yu Mao, Dakui Zhang, Xiaozhong Fan (2015), Coupled Matrix Factorization for Sentence Similarity in Community Question Answering, *Chinese Journal of Electronics*. (Accepted, SCI indexed).
- **Fangfang Li**, Guandong Xu and Longbing Cao (2015) CSAL: Self-adaptive Labeling based Clustering Integrating Supervised Learning on Unlabeled Data, <http://arxiv.org/abs/1502.05111>.
- **Fangfang Li**, Guandong Xu, Longbing Cao (2014), Coupled Item-based Matrix Factorization. *in* 'Proceedings of the 15th International Conference on Web Information System Engineering (WISE14)', pp. 1-14. (CORE2014 A)
- **Fangfang Li**, Guandong Xu, Longbing Cao, Xiaozhong Fan, Zhen-dong Niu (2013), CGMF: Coupled Group-based Matrix Factorization

- for Recommender System. *in* 'Proceedings of the 14th International Conference on Web Information System Engineering (**WISE13**)', pp. 189-198. (CORE2014 A)
- **Fangfang Li**, Xiaoming Liu (2012), A Specific Relation Extraction Approach Combining with Pointwise Mutual Information and Linguistics Information. 'Journal of Information and Computational Science', 8(16): 4115-4122.
  - **Fangfang Li**, Quan Qi, Yue Chen (2012), User Recommendation Based on Semantic Pattern. *in* 'Proceedings of the International Conference on Communication Systems and Network Technologies', pp. 992-995.
  - Li Liu, **Fangfang Li** (2011), 3-layer Ontology Based Query Expansion for Searching. *in* 'Proceedings of the 8th International Symposium on Neural Networks (**ISNN11**)', pp. 621-628.
  - Xiaoming Liu, Jinzhong Xu, **Fangfang Li** (2011), Domain-Specific Ontology Construction from Hierarchy Web Documents. *in* 'Proceedings of the 7th International Conference on Semantics Knowledge and Grid (**SKG11**)', pp. 160-163.
  - **Fangfang Li**, Li Liu (2010), The Construction and Maintenance of the Frequently Asked Question. *in* 'Proceedings of the 2nd IEEE Symposium on Web Society (**SWS10**)', pp. 296-300.
  - Li Liu, Quan Qi, **Fangfang Li** (2010), Ontology-based Interactive Question and Answering System. *in* 'Proceedings of the International Conference on Internet Technology and Applications (**iTAP10**)', pp. 1-4.
  - Li Liu, Xiaoming Liu, **Fangfang Li**, Quan Qi (2010), A Semi-automatic Method of Deriving "is-a" Relations from Text. *in* 'Proceedings of the 3rd International Conference on Computer and Automation Engineering (**ICCAE11**)'.

**Papers to be Submitted/Under Review**

- **Fangfang Li**, Longbing Cao (2015), Incorporating Couplings into Collaborative Filtering.
- **Fangfang Li**, Longbing Cao (2015), Incorporating Couplings into Matrix Factorization for Recommendation.
- **Fangfang Li**, Longbing Cao (2015), Coupled User-based Matrix Factorization for Recommendation.

**Research Reports of Industry Projects**

- **Fangfang Li**, Longbing Cao, Yanchang Zhao. Early Detection and Intervention of Adverse Behaviours in Immigration Transaction Data using Behaviour Analytics, Adverse Behaviour Analytics Project, Australian Department of Immigration and Border Protection, July 2015.
- **Fangfang Li**, Longbing Cao, Yanchang Zhao. Social Media and Call Centre Data Analysis, Correlation Analysis Project, Australian Department of Immigration and Border Protection, July 2014.



# Abstract

Recommender Systems (RS) have been proposed to help users tackle information overload by suggesting potentially interesting items to users. A typical RS usually has a set of users and items with various rating preferences. The key task of RS is to predict an unknown rating or to recommend relevant items to a given user. Many existing recommendation methods such as Collaborative Filtering (CF), Content-based Recommendation, and Hybrid Filtering often assume that users, items and their attributes are identically and independently distributed. In the real world, however, these objects and their attributes are often coupled with each other through explicit or implicit relations. On one hand, users are often connected through social or trust relations, and items are interacted with linkage or citation relations. On the other hand, the attributes of users or items are also more or less coupled with each other. These dependent relations clearly demonstrate that the users, items, and their attributes in RS are not identically and independently distributed (non-IID), which is rarely considered in most existing recommendation methods. The non-IID RS have emerged with the consideration of non-IID characteristics into RS. A main challenge in non-IID RS is to analyse and model the coupling relations between users and between items.

In this dissertation, we aim to improve recommendation effectiveness by incorporating the coupling relations into RS. The main contributions of the dissertation are summarized as follows:

- (1) We propose three novel neighbourhood-based CF methods including cou-

pled user-based CF, coupled item-based CF, and coupled CF. Specifically, we first apply a novel coupled object similarity to compute the coupling relations between users and between items based on their attributes. We then integrate the user and item couplings into the neighbourhood-based CF to produce the proposed methods by inventing new similarity measures.

- (2) We propose three novel model-based CF methods including coupled user-based matrix factorization (CUMF), coupled item-based matrix factorization (CIMF), and coupled matrix factorization (CMF). CUMF and CIMF respectively integrate the attribute-based user couplings and item couplings into MF, and CMF incorporates the user couplings, item couplings, and the user-item rating matrix together into MF.
- (3) We propose a two-level matrix factorization recommendation model which integrates the textual semantic couplings between items and the user-item rating matrix together.
- (4) We conduct experiments to evaluate the effectiveness of incorporating the couplings into non-IID RS.