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

Chapter 1 Introduction 1
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

Chapter 2 Literature Review and Foundation 15
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

Chapter 3 Coupled Neighbourhood-based Collaborative Filtering 50
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

Chapter 4 Coupled Matrix Factorization 71
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

Chapter 5 Two-level Matrix Factorization 105
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

Chapter 6 Conclusions and Future Work 126
6.1 Conclusions 126
6.2 Future Work 128

Bibliography 130

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 α and β . . . 65
3.7 Impact of Neighbour Size . . . 66
3.8 Impact of Parameter α’ . . . 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 α of CIMF on MovieLens1m . . . 97
4.7 Parameter α 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 α on MovieLens1m and BookCrossing for TLMF . . . 119

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

5.3 CF Comparisons on MovieLens1m and BookCrossing for TLMF 118
5.4 MF Comparisons on MovieLens1m and BookCrossing for TLMF 120
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, Zhendong 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.

Chapter 1

Introduction

1.1 Background

The rapid development of the information age has resulted in a massive amount of information being available for users, which can be of great assistance to them when engaging in various online activities, such as buying, selling, searching etc. However, it is still challenging for users to obtain valuable information, and it is common for users to suffer from the well-known information overload problem. Recommender Systems (RS) are proposed to help users tackle the information overload problem by suggesting potentially interesting items to users. RS are becoming increasingly important as they can impact our daily living, online, social, mobile and business activities. Typically, a set of users and items are involved, where each user rates various items according to his/her respective preferences. A new item is then recommended to a user, based on the rating behaviours of similar users on existing items. In performing this recommendation, an RS tries to predict the most suitable items to users, according to their preferences and dislikes. In order to provide more accurate recommendations, an RS usually first collects a user's explicit or implicit data, then tries to compute the user's preference for specific items, and then attempts to predict unknown ratings or to produce a ranked recommendation list of items.

RS are experiencing rapid development. The focus of interest on research into new RS has shifted from user preference understanding (Sarwar, Karypis, Konstan & Riedl 2001*a*) (Deshpande & Karypis 2004) to content analysis (Adomavicius & Tuzhilin 2005) (Pazzani & Billsus 2007), cross-domain recommendation (Hu, Cao, Xu, Wang, Gu & Cao 2013), and group-based recommendation (Li, Xu, Cao, Fan & Niu 2013). Increasing attention has been paid to typical challenges including cold-start (giving ratings to new users without knowledge of their previous rating behaviours), sparsity (sparse rate distribution for large-scale of users and items), cross-domain (gaps and linkages between domains), and tradeoff between group and individual preference difference.

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 (IID). 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.

Often recommendation algorithms produce outcomes based on the aggregated understanding of individual commonality. A rate is then predicted for a new item to a given user or a new user for a given item. The performance of applying such algorithms to real-time recommendations for specific users and items is often not very impressive. In addition to possible reasons for some of the above challenges, there are two other important aspects that have not been considered thoroughly in RS, (1) the heterogeneity between users and between items, namely users and items are personalized and thus a recommendation needs to be tailored according to individual characteristics; (2) the coupling relationships between users, between items, and between users and items, namely users and items are coupled and hence a rating needs to capture the underlying interactions. These two aspects together essentially bring the recommendation problem to a non-IID context (Cao & Yu 2015) (Cao 2015) (Cao 2014), namely users and items are not IID as usually assumed in the existing RS.

In reality, non-IID characteristics are inbuilt in business, such as online shopping, online broadcasting, IPTV and social media. On one hand, users and items, even those belonging to the same categories, share personal characteristics, just like two sons may have different flavor preferences. On the other hand, different types of intrinsic interactions may exist between users, items, and users-items. For example, users' behaviours and preferences may influence their friends, which further affect the behaviours of others. Item attributes such as item price and quantity are also often associated with each other. The price of one item may further affect the price of another. An item may influence the sale market of another.

The existing RS algorithms and systems such as collaborative filtering (CF) (Su & Khoshgoftaar 2009) and matrix factorization (MF) (Koren, Bell & Volinsky 2009) have been mainly built on the IID context, consequently they may overlook or may not fully capture the intrinsic heterogeneity and couplings. This is a very fundamental and critical issue for the RS community, as the large amount of recommendation data in online, social, mobile and business applications is essentially non-IID.

1.2 IID Recommender Systems

The most popular approach in RS, the traditional CF (Su & Khoshgoftaar 2009) (Sarwar et al. 2001*a*) (Deshpande & Karypis 2004) methods, simply focus on analysing the users' rating preferences on items, as shown in Table

Table 1.1: User-Item Ratings

	o_1	o_2	o_3
u_1	5	3	4
u_2	4	5	4
u_3	4	5	5

1.1, then try to predict an unkown rate for a given user on specific items. CF actually takes advantage of user rating history to predict users' interests and mainly involves neighbourhood-based and model-based CF methods. The neighbourhood-based CF often includes user-based CF and item-based CF methods, a very popular model-based CF method being MF. The basic idea of user-based CF is to recommend items of interest to active users according to the interests of the other users with whom they have close relationships. In the real world, people usually ask for suggestions from their friends, peers or family to decide which movie is worth watching, which book is interesting to read etc. RS are initiated from this very simple observation: individuals often rely on recommendations provided by others in making routine, daily decisions. User-based CF takes this observation as a basic recommendation principle by considering the relations between users based on their rating preferences. In contrast, item-based CF tries to recommend potentially interesting items which have close similarities to the historical items that the active user likes. Actually, when we decide which items to buy, the item itself is always a very important factor to consider. For example, we probably would buy a product which is very similar to the items we like, which is the basic idea of item-based CF. Specifically, item-based CF first calculates the relations between items based on the user-item ratings, then makes recommendations according to the item relations. In addition to the neighbourhood-based CF methods, model-based CF is also very successful in RS by modelling user behaviours and the relations between users and items. Actually, a very popular model-based CF is the MF approach which has been widely employed in many real industry applications. The goal of

MF is to learn the latent preferences of users and the latent characteristics of items from all known ratings, then predict unknown ratings through the inner product of the decomposed user and item latent factor matrices. MF first decomposes the user-item rating matrix into two latent matrices including the user latent factor matrix and item latent factor matrix through an optimization process, then computes a user's preferences on items from the optimized user and item latent matrices. Based on this idea, the prediction task of RS is transformed to decompose the user-item rating matrix by optimizing the objective function. Once this decomposition is completed, the inner product of the decomposed latent factor vectors can be easily utilized to predict the ratings or preferences given by the active user for the target item.

Table 1.2: User Attributes

User	Name	Sex	Age	City
u_1	John	M	45	Sydney
u_2	Cindy	F	42	Sydney
u_3	Julie	F	20	Sydney

Table 1.3: Item Attributes

Item	Price	Category	Subcategory
o_1	100	C1	C1.6
o_2	800	C2	C2.2
o_3	1200	C2	C2.3

Content-based Recommendation (CBR) (Adomavicius & Tuzhilin 2005) (Pazzani & Billsus 2007) is another successful method which recommends relevant items to users according to users' or items' attributes, as shown in Tables 1.2 and 1.3. Generally, attribute content is a kind of explicit context information in addition to the user-item ratings in RS. Different from the classic CF method which only utilizes user-item ratings, content-based methods (Mooney & Roy 2000) often first compute the similarities between items based on the items' various attributes, then recommend to the active user some interesting items that are similar to those that the user liked in the past. In particular, various candidate items are compared with items previously rated by the user and the best-matching items are recommended. The fundamental issue is to compute the similarities between items based on their content information. Different from CF methods, CBR mainly focuses on analysing the attribute context of users and items.

Although both CF and CBR are popular and have been successfully applied in real industry applications, there are still some problems for CF and CBF such as the well known cold-start and data sparsity problem. To enhance recommendation performance, many hybrid RS methods which combine multiple data sources, as shown in Tables 1.1 1.2 1.3 or recommendation techniques have also been proposed. For example, the hybrid filtering method integrating CF and CBR has been demonstrated to be more effective in Netflix [1], which makes recommendations by comparing the watching and searching habits of similar users using the CF approach as well as by offering movies that share characteristics with films that a user has rated highly using CBR methods. Hybrid approaches can be implemented in several ways: by making content-based and collaborative-based predictions separately and then combining them; by adding content-based capabilities to a collaborative-based approach (and vice versa); or by unifying the approaches into one model. Several studies empirically compare the performance of the hybrid with the pure collaborative and content-based methods and demonstrate that the hybrid methods can provide more accurate recommendations than single approaches. These methods can also be used to overcome some of the common problems in RS, such as the cold-start problem.

Typically, traditional CF algorithms only consider the user-item rating matrix, and the CBR methods mainly focus on analysing the user or item attributes, while the hybrid approaches combine different data sources or techniques together to improve recommendation quality. However, these

[1]https://www.netflix.com

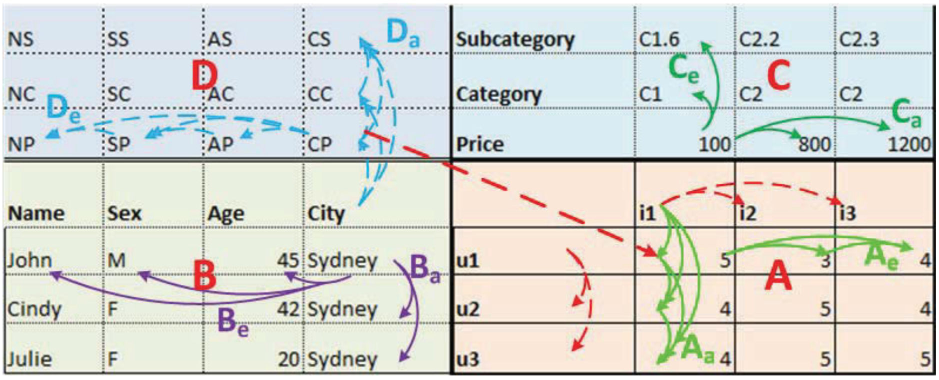

Figure 1.1: Non-IID Recommender System

recommendation methods usually overlook the heterogeneity and coupling relationships of users and items. Accordingly, we say these recommendation methods are handled in an IID context. Consequently, the existing methods cannot fundamentally solve the challenges of cold-start, cross-domain, and individual-group balancing, and the resultant prediction performance in the real world is usually marginal.

1.3 Non-IID Recommender Systems

To fundamentally solve the challenges in RS, the recommendation methods should be considered within the non-IID context, namely non-IID RS. A general framework of the non-IID RS from new perspectives different from IID situations is shown in Fig. 1.1, indicating a non-IID RS should essentially involve:

- *Four sources of data*: namely three explicit data sources including the user-item ratings, user attributes, item properties, and the implicit user-item interactions, which are respectively illustrated as A, B, C, D in Fig. 1.1. It should be noted that the implicit interactions between users and items indicated as D are invisible.

- *Both explicit and implicit interactions*: if we say a value in A, B and C represents the explicit interaction outcome of an object in one aspect, then the interactions between different rates in A for a user or item, and the influence of one value of an item property for an item on that of another item are hidden but exists. In addition, the complex interactions between user properties on an item property, item properties on a user property, and their combinations are fully implicit, as illustrated in D.

- *Both subjective and objective aspects*: the user-item ratings reflect the users' subjective preferences on items, while the user attributes and item properties and the implicit interactions between users and items indicate the objective factors driving the subjective preferences.

- *Both similar and dissimilar appearances*: each user and item essentially present respective characteristics in the community which make them heterogeneous, while the whole population or some subgroups of populations also share common interests or features which bring them together in a community.

The above perspectives enhance the interpretation of the non-IID nature of real-world recommendation tasks. More specifically, they allow the complex interactions and heterogeneity to be explored further, as they have not yet been considered thoroughly in existing RS. Different heterogeneity and coupling relationships exist in the four data sources in Fig. 1.1, the learning of comprehensive non-IIDness characteristics in a RS on such a multi-source coupled data build the foundation for the non-IID RS that is to extract and learn both visible and invisible heterogeneity and relationships and thus intrinsic driving forces for deep recommendation.

- *Rating Preference non-IIDness in A*: reflecting the explicit and subjective interactions and preferences of a user to an item, which is the most commonly explored area in current RS research.

- *User non-IIDness in B*: reflecting the heterogeneity between users, and the interactions and influence of one user on another through both internal factors of a user and external interactions, such as friendship between users.

- *Item non-IIDness in C*: indicating heterogeneity between items, and the relationships and connections between items through item properties and item-item connections.

- *Implicit User-item non-IIDness in D*: representing the implicit heterogeneity and interactions between user properties and item properties.

- *Aggregated non-IIDness*: involving all heterogeneity and couplings in A, B, C and D, and combining both explicit user-item couplings represented in preference couplings in A with implicit user-item couplings in D, as well as the impact of user couplings and item couplings in B and C on user-item couplings for the prediction.

1.4 Research Issues

A current challenge in RS is the cold-start problem. The essence of this problem is that classic recommendation techniques, such as neighbourhood-based CF and MF, mainly rely on the user-item rating matrix, which is sometimes not informative enough for predicting recommendations. For example, although CF has been widely applied in many real applications, e.g., Amazon, the effect of CF sharply decreases for new users and items. This is partly because, for new users and items, it is extremely difficult to infer the relationships between users and between items, only based on the user-item rating matrix. Similarly, traditional MF also does not take into account users' and items' properties which contribute to inferring the relationships between users and between items. This limitation partly motivates us to consider other relations between users and between items in RS. If we can obtain the users' or items' relations, no matter whether we have ample rating data, it may greatly enhance the effectiveness of recommendations.

Another critical issue is that most RS methods often assume that users, items, and their attributes are IID. Specifically, different kinds of data sources, such as the user-item ratings, users' and items' attributes are often considered in recommendation algorithms, however, most existing recommendation methods often ignore the underlying reasons driving specific user preferences (Breese, Heckerman & Kadie 1998) (Resnick, Iacovou, Suchak, Bergstrom & Riedl 1994) (Sarwar, Karypis, Konstan & Riedl 2001*b*) and simply assume that different attributes, users and items are IID, which is not always the case in reality, resulting in the challenging non-IID problem. Researchers in the data mining, machine learning and statistics communities have paid little attention to this challenging non-IID issue to date, and the emerging applications in big data are increasingly experiencing the emergence of gaps in the application of existing learning theories and techniques with IID assumptions. Two fundamental challenges in the non-IID learning problem are to analyse the heterogeneity and coupling relations. Several research outcomes such as (Wang, Cao, Wang, Li, Wei & Ou 2011) (Wang, She & Cao 2013*a*) (Wang, She & Cao 2013*b*) (Yu, Wang, Gao, Cao & Chen 2013) have been proposed to handle the challenging issues for clustering and classification problems. However, there is only limited research on non-IID learning for RS. Therefore, in this thesis we thoroughly analyse the couplings of users and items based on their attributes, and incorporate these couplings into an RS to improve recommendation quality.

The third issue for RS methods which utilize textual information is that they do not fully consider the textual semantic couplings of users and items. In fact, the textual context of users and items, in combination with the users' and items' attributes, is additional valuable explicit information and can be applied to improve recommendation quality. To consider the textual context, it is very important to effectively analyse the textual semantic couplings. It is well known that latent factor models, such as Latent Semantic Analysis (LSA) (Landauer, Foltz & Laham 1998), Probabilistic Latent Semantic Analysis (PLSA) (Hofmann 1999), Latent Dirichlet Allocation (LDA) (Blei, Ng, Jordan & Lafferty 2003) are beneficial for identifying textual semantic couplings. However, textual semantic couplings have not yet been thoroughly studied. Recently, a novel Weighted Textual Matrix Factorization (WTMF) method (Guo & Diab 2012*b*) was proposed to compute the semantic similarities between textual sentences and achieved much better performance. However, this valuable semantic analysis method has not been applied in RS, which greatly motivates us to model the textual semantic couplings to improve recommendation quality and to solve the cold-start problem.

1.5 Research Contributions

The main contributions of the thesis are as follows.

- We have applied a novel coupled measure to capture the relationships between users and between items in RS, namely user coupling and item coupling, which consider the coupled interaction between attributes. (Chapter 3)

- We have proposed a hybrid Coupled User-based CF (CUCF) method incorporating user couplings into a user-based CF method with a hybrid user similarity measure. (Chapter 3)

- We have proposed a hybrid Coupled Item-based CF (CICF) method integrating the item couplings into an item-based CF method with a hybrid item similarity measure. (Chapter 3)

- We have proposed a novel hybrid Coupled CF (CCF) method integrating user couplings, item couplings and user-item ratings in the CF method with hybrid user similarity and item similarity measures. (Chapter 3)

- We have proposed a Coupled User-based Matrix Factorization (CUMF) model by incorporating user couplings and users' subjective rating preferences. (Chapter 4)

- We have proposed a Coupled Item-based Matrix Factorization (CIMF) model by incorporating item couplings and users' subjective rating preferences. (Chapter 4)

- We have proposed a Coupled Matrix Factorization (CMF) framework by incorporating user coupling, item coupling and users' subjective rating preferences. (Chapter 4)

- We have applied the novel WTMF model to infer the semantic couplings between items based on the textual context. (Chapter 5)

- We have proposed a Two-level Matrix Factorization (TLMF) recommendation model by incorporating textual semantic couplings between items and users' subjective rating preferences. (Chapter 5)

1.6 Thesis Structure

Chapter 2 introduces the related work and foundations regarding IID RS, such as CBR and CF, non-IIDness learning and non-IID RS.

Chapter 3 applies the significant coupling relations between users and between items to improve the recommendation quality of neighbourhood-based CF methods. We first infer the user couplings from users' attributes, then integrate them with the user's rating similarities to form a hybrid user similarity method. We then propose the CUCF method by incorporating the hybrid similarity method with the classic user-based CF method. We also analyse the item couplings from items' attributes and integrate them with the item's rating similarity to form a hybrid item similarity method which is then incorporated into the proposed CICF method. In addition, we also incorporate the three different relations, user couplings, item couplings, and user-item rating relations, into CF to form a combined CCF recommendation method.

Chapter 4 applies the coupling relations into the MF model. We first propose a CUMF model by incorporating the user couplings and the user-item ratings into MF. Specifically, we take advantage of the user couplings to update the decomposed latent factor matrix for users, then optimize the objective function to compute the best latent factor matrices for users and items. We also propose a CIMF model by integrating the item couplings and the user-item ratings into MF. Specifically, we first apply an attribute-based coupled similarity measure to capture the item couplings, then integrate the item couplings into MF to form the CIMF model. Lastly, we integrate the user couplings, item couplings and the user-item ratings to form the proposed CMF model.

Chapter 5 studies the textual semantic couplings between items to improve the recommendation quality of MF. In this chapter, we propose a novel TLMF model which not only considers semantic couplings between items in a lower level MF, but also utilizes the user-item ratings in an upper level MF. The lower level MF analyses the textual semantic relations between items using the textual MF method, and the upper level MF focuses on analysing the relations between users and items based on the user-item ratings using the MF method. The integration of the two-level MF methods enhances recommendation performance by considering the textual semantic couplings.

Chapter 6 concludes the thesis and outlines the scope for future work.

Figure 1.2 shows the research profile of this thesis.

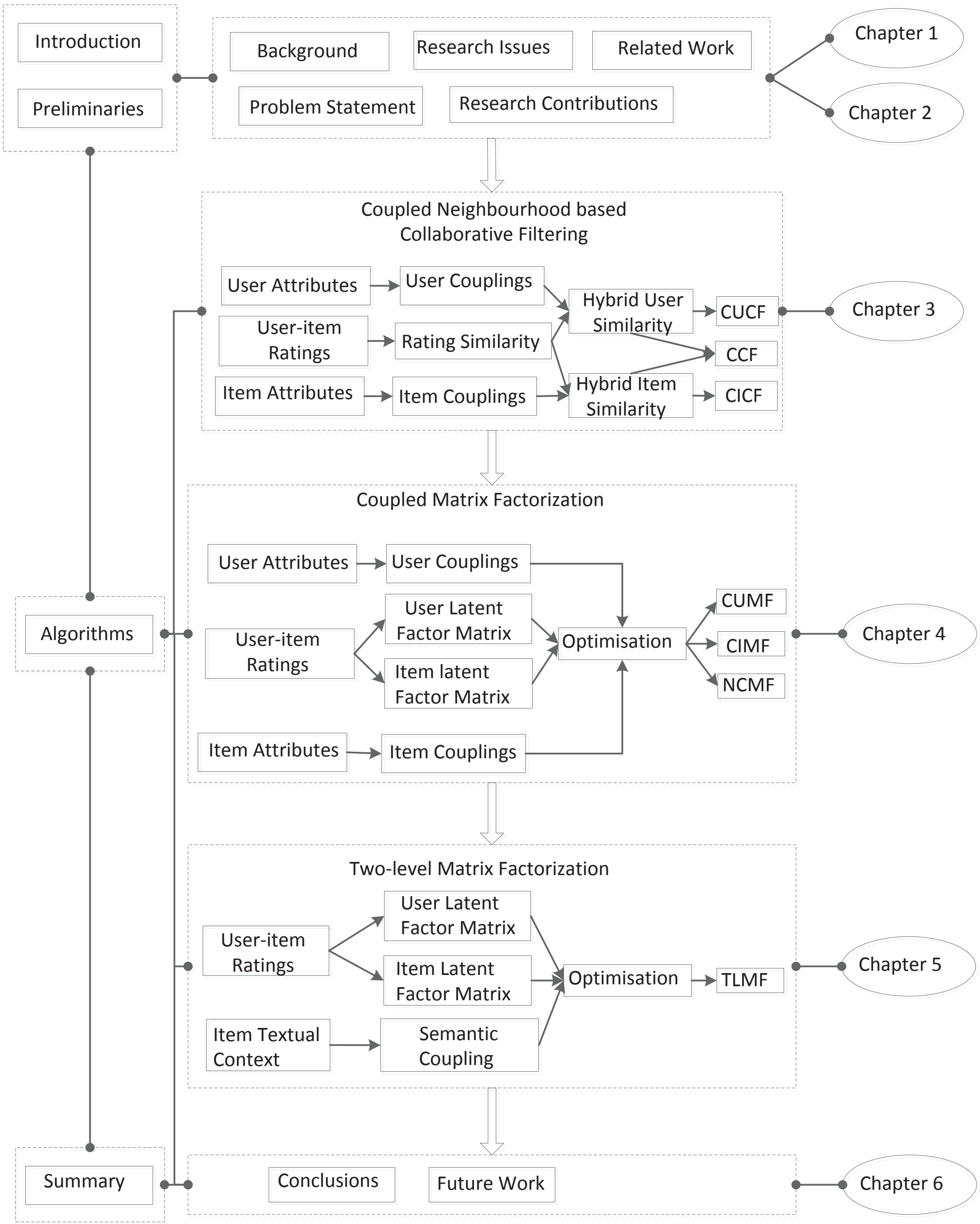

Figure 1.2: Thesis Structure

Chapter 2

Literature Review and Foundation

This chapter reviews the related work and foundations regarding IID RS, non-IIDness learning, and non-IID RS. In Section 1, we introduce traditional IIDness based recommendation methods including demographic filtering, collaborative filtering, content-based recommendation and hybrid filtering. Section 2 presents the concept and related work of non-IIDness learning. In Section 3, we review the related work regarding non-IID RS incorporating social and coupling relations.

2.1 IID Recommender Systems

2.1.1 Notations

In RS, a large number of user and item sets with attributes can be organized by a triple $C =< C_U, C_O, h >$, where $C_U =< U, A, V, f >$ describes the users' attribute space, $U = \{u_1, u_2, ..., u_m\}$ is a non-empty finite set of users, $A = \{A_1, ..., A_M\}$ is a finite set of attributes for users; $V = \cup_{j=1}^{J} V_j$ is a set of all attribute values for users, in which V_j is the set of attribute values of attribute $A_j (1 \leq j \leq J)$, V_{ij} is the attribute value of attribute A_j for user u_i, and $f = \wedge_{j=1}^{M} f_j (f_j : U \rightarrow V_j)$ is an information function which assigns a particular value of each feature to every user. Similar to C_U, $C_O =< O, A', V', f' >$ expresses the items' attribute space where $O = \{o_1, ..., o_n\}$, $A' = \{A'_1, ..., A'_{M'}\}$, $V' = \cup_{j=1}^{J'} V'_j$, $f' = \wedge_{j=1}^{M'} f'_j (f'_j : O \to V'_j)$ are all for items. In the triple $C =< C_U, C_O, h >$, $h(u_i, o_j) = R_{ij}$ expresses the subjective rating preference on item o_j for user u_i. User rating preferences on items are converted into a user-item matrix R, with m rows and n columns. Each element R_{ij} of R represents the rating given by user u_i on item o_j.

2.1.2 Demographic Filtering

Demographic filtering (Pazzani 1999) (Krulwich 1997) (Porcel, Tejeda-Lorente, Martnez & Herrera-Viedma 2012) is an early-age recommendation method based on the principle that individuals with certain common personal attributes (sex, age, country, etc.) will also probably have common preferences. In fact, demographic information, as shown in Table 1.2, can be used to divide the users into different clusters, where a cluster only involves the users with similar interests. For example, LifeStyle Finder (Krulwich 1997) first pre-defines 62 clusters, then tries to determine which cluster a user should belong to according to the demographic information. Then recommendation to this user can be made upon the interests of other users in the same cluster.

2.1.3 Collaborative Filtering

CF (Adomavicius & Tuzhilin 2005) (Herlocker, Konstan, Borchers & Riedl 1999) (Herlocker, Konstan & Riedl 2002) (Candillier, Meyer & Boullé 2007) (Su & Khoshgoftaar 2009) is another successful recommendation method including neighbourhood-based CF and model-based CF, which makes recommendations to the target user based on the user-item rating matrix, as shown in Table 1.1.

Neighbourhood-based Collaborative Filtering

One popular CF algorithm is the neighbourhood-based CF (Bobadilla, Hernando, Ortega & Bernal 2011) (Adomavicius & Tuzhilin 2005) (Schafer, Frankowski, Herlocker & Sen 2007) using k nearest neighbours (kNN) method. The neighbourhood-based CF algorithms usually use the user-item rating matrix to predict the preferences of a new user on items by identifying the neighbours of users or items. To identify the neighbourhood of users or items, the essence is to compute the similarity between users or items from the rating matrix. The neighbourhood-based CF algorithms include the following steps: (1) calculate the similarity, which reflects distance, correlation, or weight, between two users or two items; (2) produce a prediction for the active user by taking the weighted average of all the ratings of the user or item on a certain item or user.

Similarity computation between users or items is an essential step in neighbourhood-based CF algorithms. For item-based CF algorithms, the basic idea of the similarity computation between item o_i and item o_j is first to work on the users who have rated both of these items and then to apply a similarity computation to determine the similarity, S_{o_i,o_j}, between the two co-rated items of the users. For a user-based CF algorithm, we first calculate the similarity, S_{u_i,u_j}, between the users u_i and u_j who have both rated the same items. After computing the similarities between users or items, we can easily derive user's or item's neighbours. We then can predict the preference of user u_i on item o_j using the following prediction functions of user-based CF and item-based CF, respectively, as shown in Eqn. 2.1 2.2.

$$R_{ij} = \begin{cases} \frac{\sum_{u_k \in N(u_i)} S(u_i,u_k) R_{kj}}{\sum_{u_k \in N(u_i)} S(u_i,u_k)} & \sum |S(u_i, u_k)| > 0, \\ \overline{R}_{u_i} & \sum |S(u_i, u_k)| = 0. \end{cases} \tag{2.1}$$

$$R_{ij} = \begin{cases} \frac{\sum_{o_k \in N(o_j)} S(o_j,o_k) R_{ik}}{\sum_{o_k \in N(o_j)} S(o_j,o_k)} & \sum |S(o_j, o_k)| > 0, \\ \overline{R}_{o_j} & \sum |S(o_j, o_k)| = 0. \end{cases} \tag{2.2}$$

where $S(u_i, u_k)$ is the rating similarity from the user-item rating matrix between users u_i and u_k, $N(u_i)$ represents the neighbourhood of user u_i, and $\overline{R}_{u_i}$ is the average rating of the co-rated items for user u_i. $S(o_j, o_k)$ is the rating similarity from user-item rating matrix between items o_j and o_k, $N(o_j)$ is the neighbourhood of item o_j, and $\overline{R}_{o_j}$ is the average ratings by users on item o_j.

There are many different methods to compute the similarity between users or items, we here briefly introduce two frequently used similarity measures Pearson Correlation and Vector Cosine Similarity.

The similarity S_{u_i,u_j} between two users u_i and u_j, or S_{o_i,o_j} between two items o_i and o_j, can be measured by computing the Pearson correlation.

For the user-based CF algorithm, the Pearson correlation between users u_i and u_j is:

$$S_{u_i,u_j} = \frac{\sum_{o_k \in O}(R_{ik} - \overline{R}_{u_i})(R_{jk} - \overline{R}_{u_j})}{\sqrt{\sum_{o_k \in O}(R_{ik} - \overline{R}_{u_i})^2}\sqrt{\sum_{o_k \in O}(R_{jk} - \overline{R}_{u_j})^2} \tag{2.3}$$

where the $o_k \in O$ summations are over the items that both the users u_i and u_j have rated, $\overline{R}_{u_i}$ and $\overline{R}_{u_j}$ are respectively the average ratings of the co-rated items for user u_i and u_j.

For the item-based CF algorithm, denote the set of users $u_k \in U$ who rated both items o_i and o_j, then the Pearson Correlation is:

$$S_{o_i,o_j} = \frac{\sum_{u_k \in U}(R_{ki} - \overline{R}_{o_i})(R_{kj} - \overline{R}_{o_j})}{\sqrt{\sum_{u_k \in U}(R_{ki} - \overline{R}_{o_i})^2}\sqrt{\sum_{u_k \in U}(R_{kj} - \overline{R}_{o_j})^2} \tag{2.4}$$

where R_{ki}, R_{kj} are respectively the ratings of user u_k on item o_i and o_j, $\overline{R}_{o_i}$ and $\overline{R}_{o_j}$ are respectively the average ratings by those users on items o_i and o_j .

The similarity between two users or items can be also measured by computing the cosine similarity of the rating vectors. Formally, if R is the $m \times n$ user-item matrix, then the similarity between two items, o_i and o_j, is defined as the cosine of the m dimensional vectors corresponding to the i_{th} and j_{th} column of matrix R. Vector cosine similarity between users u_i and u_j is described in Eqn. 2.5, and the similarity between items o_i and o_j is detailed in Eqn. 2.6.

$$S_{u_i,u_j} = cos(\vec{u_i}, \vec{u_j}) = \frac{\vec{u_i} \cdot \vec{u_j}}{\|\vec{u_i}\|\|\vec{u_j}\|} \tag{2.5}$$

$$S_{o_i,o_j} = cos(\vec{o_i}, \vec{o_j}) = \frac{\vec{o_i} \cdot \vec{o_j}}{\|\vec{o_i}\|\|\vec{o_j}\|} \tag{2.6}$$

where $\cdot$ denotes the dot-product of the two vectors.

Model-based Collaborative Filtering

In addition to the neighbourhood-based CF, the model-based CF method (Adomavicius & Tuzhilin 2005) (Su & Khoshgoftaar 2009) has also been widely used in RS by producing a model to generate the recommendations. The most common models used in CF include: Bayesian classifiers (Park, Hong & Cho 2007), neural networks (Roh, Oh & Han 2003), fuzzy systems (Yager 2003), genetic algorithms (Ho, Fong & Yan 2007), latent features (Zhong & Li 2010) and matrix factorization (Luo, Xia & Zhu 2012).

Generally, the model-based CF recommends items to users by estimating parameters of statistical models based on the explicit (Koren et al. 2009) or implicit (Pan & Scholz 2009) user ratings. For example, an approach is to map CF to the classification problem (Billsus & Pazzani 1998) by building a classifier for each active user with items represented as attributes and the ratings as labels. Dimensionality reduction techniques (Sarwar, Karypis, Konstan & Riedl 2000) are also applied to overcome data sparsity issues, and the reduction methods are based on MF (Koren et al. 2009) (Luo et al. 2012) (Luo, Xia & Zhu 2013) which is especially helpful for processing large RS databases and providing scalable approaches (Takács, Pilászy, Németh & Tikk 2009). For example, the model-based Latent Semantic Index (LSI) and the reduction method Singular Value Decomposition (SVD) are typically combined (Vozalis & Margaritis 2007)(Zhang, Wang, Ford, Makedon & Pearlman 2005)(Cacheda, Carneiro, Fernández & Formoso 2011) for recommendations. However, SVD method is computationally expensive for recommendation and usually can only be deployed in off-line settings where the known preference does not change with time, although it performs well to generate predictions.

Since MF techniques can be widely employed in RS, we here briefly introduce them first. The goal of MF is to learn the latent preferences of users and the latent characteristics of items from all known ratings, then predict the unknown ratings through the inner products of user latent factor vectors and item latent factor vectors. Formally, MF methods decompose the user-item rating matrix R into two low rank latent factor matrices P and Q. Then the matrix of predicted ratings $\hat{R} \in \mathbb{R}^{m \times n}$, where m, n respectively denote the number of users and the number of items, can be modelled as:

$$\hat{R} = PQ^T = \begin{bmatrix} P_1 \\ P_2 \\ ... \\ P_m \end{bmatrix} \begin{bmatrix} Q_1 & Q_2 & ... & Q_n \end{bmatrix} \tag{2.7}$$

where, matrices $P \in \mathbb{R}^{m \times d}$ and $Q \in \mathbb{R}^{n \times d}$, d is the rank (or dimension of the latent space) with $d \leq m, n$. The vectors P_i and Q_j represent the d-dimensional user-specific latent factor vector and item specific latent factor vector, respectively. Through this modelling, the prediction task of matrix $\hat{R}$ is transferred to compute the mapping of users and items to factor matrices P and Q. Once this mapping is completed, the inner product of P_i and Q_j can be easily utilized to predict the rating given by the active user u_i for target item o_j.

In order to learn the optimum latent factor vectors of users and items, one way is to optimize the objective function 2.8 as SVD (Wu, Chen, Liu, Xu, Bao & Zhang 2012),

$$\frac{1}{2} \left\| R - PQ^T \right\|_F^2 \tag{2.8}$$

where $\|.\|_F^2$ is the Frobenius norm. But this optimization method cannot effectively determine the latent semantic factors for a very sparse rating matrix. Hence, another direct optimization method is to factorize the observed ratings using the following objective function:

$$L = \min_{P,Q} \frac{1}{2} \sum_{(u_i,o_j)\in E} \left(R_{ij} - P_i Q_j^T\right)^2 + \frac{\lambda}{2}\left(\|P\|_F^2 + \|Q\|_F^2\right) \tag{2.9}$$

where E indicates the set of the (u_i, o_j) pairs for known ratings. To avoid over-fitting, two regularization terms on the sizes of P and Q are added in Eqn. 2.9 as constraints, and λ represents the regularization parameter impacting on latent semantic vectors.

To learn the latent factor matrices P and Q, an efficient stochastic gradient descent algorithm is often applied to optimize the objective function given in Eqn. 2.9 with an iteration process. The derivative of L with respect to P_i and Q_j are as follows:

$$\frac{\partial L}{\partial P_i} = \sum_{o_j} I_{i,j}(P_i {Q_j}^T - R_{ij})Q_j + \lambda P_i \tag{2.10}$$

$$\frac{\partial L}{\partial Q_j} = \sum_{u_i} I_{i,j}(P_i {Q_j}^T - R_{ij})P_i + \lambda Q_j \tag{2.11}$$

Furthermore, the updating rules for iteration are derived to learn the latent vectors P_i and Q_j:

$$P_i \leftarrow P_i + \eta((R_{ij} - P_i {Q_j}^T)Q_j - \lambda P_i) \tag{2.12}$$

$$Q_j \leftarrow Q_j + \eta((R_{ij} - P_i {Q_j}^T)P_i - \lambda Q_j) \tag{2.13}$$

This optimization process enables us to get the optimum latent factor matrices P and Q from the user-item rating history, then the unknown ratings can be predicted by the multiplication of the two decomposed latent matrices.

In the MF process, different loss functions regularizes and additional model constraints have been used in RS. For example, the maximum margin MF (Rennie & Srebro 2005) approach uses margin-based loss functions, such as the hinge loss used in SVM classification, and its ordinal extensions for handling multiple ordered rating categories. For ratings that span over K values, this reduces to finding $K-1$ thresholds that divide the real line into consecutive intervals specifying rating bins to which the output is mapped, with a penalty for insufficient margin of separation. Rennie and Srebro (Rennie & Srebro 2005) suggest a nonlinear conjugate gradient algorithm to minimize a smoothed version of this objective function.

Another class of techniques is the non-negative MF popularized by the work of Lee and Seung (Lee & Seung 1999) where non-negativity constraints are imposed on P, Q. NMF is in fact essentially equivalent to probabilistic latent semantic analysis (pLSA) which has also previously been used for CF tasks (Hofmann 2004). Many different weighted extensions of NMF have also been applied in RS. The rating behaviour of each user may be viewed as being a manifestation of different roles, for example, a composition of prototypical behaviour in clusters of users bound by interests or community. Thus, the ratings of each user are an additive sum of basis vectors of ratings in the item space. By disallowing subtractive basis, non-negativity constraints lend a "part-based" interpretation to the model. NMF can be solved with a variety of loss functions, but with the generalized KL-divergence loss defined as follows:

$$L = \min_{P,Q} \frac{1}{2} \sum_{(u_i,o_j)\in E} R_{ij} \log \frac{R_{ij}}{P_i Q_j^T} - R_{ij} + P_i Q_j^T \tag{2.14}$$

Generally, CF has its own advantages and disadvantages. For example, CF methods are built on the user-item rating matrix, hence, they are domain-independent and can recommend any items. However, CF methods also have their own disadvantages such as the cold-start problems for new users and new items and the data sparsity problem (Adomavicius & Tuzhilin 2005) (Su & Khoshgoftaar 2009).

2.1.4 Content-based Recommendation

Pure CF algorithms only utilize the user-item rating matrix to generate a collaborative model by assuming all the users and items are atomic units.

These methods do not consider the specific characteristics of the users and items. However, RS can be much better personalized by knowing more about a user, such as the discussed demographic information (Pazzani 1999) of users, or about an item, such as the director, actors and genre of a movie (Melville, Mooney & Nagarajan 2002). The CBR method (Lang 1995) (Salter & Antonopoulos 2006) utilizes the items' attributes, as shown in Table 1.3, to make recommendations based on an assumption that a user probably likes the items that are similar to those that the user liked in the past. For example, in a web-based e-commerce RS, if the user purchased some fiction films in the past, the RS will probably recommend a recent fiction film that he has not yet purchased on this website to this user. The commonly used content information of items in CBR usually contains demographic attributes, text, images and sound. The essence of the CBR is to compute the similarities between items based on their content information, which are then used for recommending potentially interesting items similar to the items that the user has bought, visited, watched and ranked accordingly.

Some researchers only focus on the textual content of items, such as web pages, user reviews, item's descriptions, and consider the CBR as an information retrieval task, where the content associated with the user's preferences is treated as a query, and the unrated documents are scored with relevance/similarity to this query (Balabanović & Shoham 1997). For example, NewsWeeder (Lang 1995) first converts all the documents into the $TF * IDF$ word vectors using the vector space model, and then similarity of the documents are computed from the cosine similarity of their represented vectors.

Alternatively, CBR can be also considered as a classification task, where each example represents the content of an item, and a user's past ratings are used as labels for these examples. For instance, the content of books, such as title, author, synopses, reviews, and subject terms, are utilized to train a multinomial naive bayes classifier (Mooney & Roy 2000). The ratings (Melville et al. 2002) (Mooney & Roy 2000) can also be trained to build classifiers. In addition, other classification algorithms including k-nearest neighbours, decision trees, and neural networks (Pazzani & Billsus 1997) have also been applied in CBR methods.

The main limitations of CBR as discussed in (Balabanović & Shoham 1997) (Adomavicius & Tuzhilin 2005) include:

- Limited content analysis - these methods are difficult to be applied in some domains which have an inherent problem with automatic feature extraction such as multimedia data;
- Over-specialization - items recommended to a user are limited to those similar to items the user already rated;
- Cold-start users - CBR is difficult to recommend for cold-start users who do not rate sufficient items.

2.1.5 Hybrid Filtering

From the above discussions, we can see that both CBR and CF methods have their disadvantages. To improve the recommendation quality and overcome the limitations of CBR and CF methods, hybrid filtering approaches integrating CBR and CF are also proposed (Cotter & Smyth 2000) (Melville et al. 2002). There are three categories of strategies for combining CBR and CF, which are reviewed as follows.

- Combining different predictions: This strategy first implements the CBR and CF separately, then their predictions are combined to generate the final recommendation list (Cotter & Smyth 2000). Various ways are proposed, such as a voting scheme (Pazzani 1999) and a linear combination of ratings (Claypool, Gokhale, Miranda, Murnikov, Netes & Sartin 1999), to combine predictions from content and CF based methods. For example, Claypool et al. proposed a hybrid method to combine the prediction results with an adaptive weighted average

(Claypool et al. 1999), where the weight of the collaborative component increases as the number of users accessing an item increases. A content-boosted CF framework was also proposed (Melville et al. 2002), where the content-based predictions were applied to change the sparse user-item rating matrix, and the CF component was utilized to produce recommendations. Specifically, they first trained a Naive Bayes classifier on the documents of the rated items of each user, then replaced the unrated items by predictions from this classifier. After that, the resulting pseudo rating matrix was used to compute the neighbours of the active user and to generate recommendations. Experiments demonstrated that this approach performed much better than pure CF, pure CBR, and a linear combination of CF and CBR. In addition, Su, Greiner, and Zhu (Su, Greiner, Khoshgoftaar & Zhu 2007) showed that a content-boosted CF method using a stronger content-predictor, TAN-ELR, and an unweighted Pearson CF can improve the recommendation results. In addition, clustering algorithms (George & Merugu 2005) are also beneficial for improving the recommendation quality of hybrid filtering methods and solving the cold-start challenges (Shinde & Kulkarni 2012) (Yao & Zhang 2009).

- Adding CBR characteristics to CF models: this strategy calculates the similarities between users or items based on their content information, especially for the users and items without ratings, which can partly overcome the sparsity-related problems of CF methods and recommend items directly when item scores highly against the users profiles (Pazzani 1999) (Good, Schafer, Konstan, Borchers, Sarwar, Herlocker & Riedl 1999). For example, Pazzani (Pazzani 1999) tried to represent each user's profile by a vector of weighted words derived from positive training examples using the Winnow algorithm, then applied CF directly on the user-profile matrix to predict the recommendations. In addition, user's relevant feedback, topic information (Balabanović & Shoham 1997) (Good et al. 1999), and co-occurrence (Popescul, Ungar, Pennock & Lawrence 2001) (Hofmann 1999) can also be applied in hybrid filtering for improving recommendation quality and solving the cold-start problem (Schein, Popescul, Ungar & Pennock 2002).

- Adding CF characteristics to CBR models: this strategy often uses CF method to generate different attributes of users and items, which are then exploited by the CBR models. The most popular approach of this strategy is to use a dimensionality reduction technique on the content profile matrix. For example, LSI (Nicholas & Nicholas 1999) is used to create a collaborative view of a set of user profiles, which improves recommendation performance compared to pure CBR approaches. A content-profile matrix is produced from a term-document matrix representing all item content with the user-item rating matrix, using LSI. Then term vectors of the user's relevant documents are averaged to produce a user's profile, and new documents are ranked against each user's profile in the LSI space. In addition, a rule induction system (Basu, Hirsh & Cohen 1998) is applied to learn a function and generate recommendations, considering the collaborative and content information with created new features, such as comedies liked by users and users who liked movies of a specific genre.

Generally, the hybrid filtering methods indeed perform better than the pure CF, pure CBR methods, however, most existing hybrid methods are still based on the assumption of IID, which is not always held in reality.

2.2 Non-IIDness Learning

Generally, the recommendation methods discussed above can be classified to the IID RS. However, most data in the communities of data mining, machine learning and statistics are essentially non-IID, attracting more and more researchers concentrating on non-IID data to effectively handle complicated challenges with strong heterogeneity and correlation. Relevant research regarding non-IIDness learning (Cao 2014) involves statistical relation learning and similarity enhancement under the existing learning frameworks.

2.2.1 Efforts of Non-IIDness Learning

Although it is still in the very early stage of non-IIDness learning, a lot of effort has already put on this topic. It can be easily observed that non-IID data are widely existed in complex business, behaviour, social and big data applications. To analyse and solve the non-IID problems, relevant effort regarding non-IIDness learning includes: rule relation analysis (Cao 2013*a*) (Cao, Li, Wang & Yu 2013), active learning (Olsson 2009) (Rubens, Kaplan & Sugiyama 2011) (Settles 2009), group/community analysis (Fortunato 2010) (Girvan & Newman 2002) (Porter, Onnela & Mucha 2009), mix-structured data analysis (Losee 2006), multisource analysis (Peddle & Ferguson 2002) (Huopaniemi, Suvitaival, Nikkila, Oresic & Kaski 2010), mixture network analysis (Newman 2003) (Amaral, Scala, Barthelemy & Stanley 2000), hypothesis testing (Lehmann & Romano 2005), learning with graphical models (Jordan 1999) (Lauritzen 1996), mining graph data (Cook & Holder 2006), social network analysis (Kumar 2013), social media analysis (Boden, Karnstedt, Fernández & Markl 2013)(Fan & Gordon 2014), multimedia analysis (Troncy, Huet & Schenk 2011), cross-market analysis (Aggarwal 2011)(Cao, Cao & Song 2013), high frequency and evolving data analysis (Gama 2010), multi-task (Caruana 1993) (Kumar & III 2012) and cross-domain (Hu et al. 2013) analysis, and language processing (Manning & Schütze 1999). These learning problems go far beyond the dominative theories and tools learnt from IID data.

Although it is at its very early stage, a lot of researchers have made contributions to explore non-IID data in statistics, information theory, and machine learning communities. These include concentration inequalities for dependent random variables (Kontorovich, Ramanan et al. 2008) or via coupling (Kontorovich & Ramanan 2008), least-square regularized regression (Pan & Xiao 2009), divergence estimation (Prez-cruz 2008) and generalized online learning (Agarwal & Duchi 2013) on dependent data, rademacher complexity bounds (Mohri & Rostamizadeh 2008), stability bounds for mixing processes (Mohri & Rostamizadeh 2010)(Roussas 1990), chromatic PAC-Bayes bounds (Ralaivola, Szafranski & Stempfel 2010), group Lasso (Jacob, Obozinski & Vert 2009), coupled hidden Markov model for coupled behaviour analysis (Cao, Ou & Yu 2012*a*), structure learning (Tillman 2009), kernel measure (Zhang, Song, Gretton & Smola 2009), active learning and generalization bounds (Cohn, Atlas & Ladner 1994), combined mining for structural pattern pairs or clusters (Cao, Zhang, Zhao, Luo & Zhang 2011), pattern relation analysis (Cao 2013*b*), and dimensionality reduction and feature selection (Ping, Xu, Ren, Chi & Furao 2010) (Song, Smola, Gretton, Bedo & Borgwardt 2012), as well as applications for classification (Dundar, Krishnapuram, Bi & Rao 2007) (Guo & Shi 2011), multi-instance learning (Zhou, Sun & Li 2009), and image categorization (Cinbis, Verbeek & Schmid 2012).

In addition, statistical relation learning (SRL) (Getoor & Taskar 2007) is concerned with models of domains that exhibit both uncertainty and complex relational structures between objects. The main approaches and tools explored for SRL include: Bayesian logic program (Kersting & Raedt 2008), Markov logic networks (Richardson & Domingos 2006), multi-entity Bayesian networks (Laskey 2008), probabilistic relational models (Friedman, Getoor, Koller & Pfeffer 1999), conditional random fields (Lafferty, McCallum & Pereira 2001), relational Bayesian networks (Jaeger 1997), relational dependency networks (Neville & Jensen 2007), first-order probabilistic languages (Milch & Russell 2007), learning infinite relational models (Xu, Tresp, Yu & Kriegel 2006), statistical predicate invention (Kok & Domingos 2007), and relational Markov networks (Taskar, Abbeel, Wong & Koller 2007).

In particular, a series of research outcomes have been reported on coupled object learning (Wang et al. 2011) (Wang et al. 2013*b*) (Wang et al. 2013*a*) by incorporating intra-coupling within an object, inter-coupling between objects, and their combination (Cao, Dai & Zhou 2009). Such concepts have been expanded to coupled similarity learning (Wang et al. 2011), coupled clustering for categorical data (Wang et al. 2011), coupled ensemble clustering (Wang et al. 2013*b*), coupled classification for numeric data (Wang et al. 2013*a*), coupled behaviour analysis (Cao, Ou & Yu 2012*b*), term coupling in text mining (Cheng, Miao, Wang & Cao 2013), item coupling in CF (Yu et al. 2013), and couplings in MF (Li et al. 2013) (Li, Xu & Cao 2014)(Li, Xu & Cao 2015*a*).

Other promising techniques and aspects include active learning (Olsson 2009) (Rubens et al. 2011) (Settles 2009), transfer learning (Pan & Yang 2010), combined mining (Cao 2013*a*) (Cao et al. 2011), probabilistic graphical modeling (Jordan 1999) (Lauritzen 1996), tensor theory (Itskov 2009), Copula (Nelsen 2006) (Wei, Fan, Li & Cao 2012), group Lasso (Meier, van de Geer & Bühlmann 2008), multivariate generalized autoregressive conditional heteroscedasticity (Tse & Tsui 2002), regret analysis (Loomes & Sugden 1982), coupled Hidden Markov Model (Cao et al. 2012*b*), convex analysis (Bertsekas, Nedić & Ozdaglar 2003), and online learning (Littlestone 1988).

2.2.2 Similarity Computation

Similarity computation (Gibson, Kleinberg & Raghavan 2000) is often an essential task in data mining, machine learning communities. To enhance the similarity computation methods, different approaches considering matching, frequency and neighbourhood have been explored in learning the relations between objects. Due to the difference between categorical data and numerical data, and between varied learning purposes, various efforts are being made to improve the existing similarity or dissimilarity metrics.

For categorical data, frequency-based approaches, such as SMS (Kaufman & Rousseeuw 1990) and Jaccard coefficients (Ribeiro & Harder 2011), have been frequently applied in unsupervised learning tasks. In many learning problems, the frequency distribution of attribute values has been considered for similarity measures, such as frequent pattern mining and association rule mining (Zhang, Zhao, Cao & Zhang 2008) (Zhao, Zhang, Figueiredo, Cao & Zhang 2007), frequent sequence analysis (Zhao, Zhang, Wu, Pei, Cao, Zhang

& Bohlscheid 2009) (Cao, Zhao & Zhang 2008), and similarity enhancement (Boriah, Chandola & Kumar 2008).

Neighbourhood-based similarity (Guha, Rastogi & Shim 1999) (Houle, Oria & Qasim 2010) was also explored to measure the proximity of objects by using functions that operate on the intersection of two neighbourhoods. They present the similarity between a pair of objects by considering only the relationships among data objects, which are built on the similarity between attribute values simply quantified by the variants of SMS. All the above methods are attribute-independent since similarity is calculated separately for two categorical values of individual attributes.

An increasing number of researchers argue that the attribute value similarity is also dependent on the couplings of other attributes (Boriah et al. 2008) (Cao et al. 2012*b*). The Pearson correlation coefficient (Houle et al. 2010) measures only the strength of linear dependence between two variables (e.g., nominal attributes). The Iterated Contextual Distances algorithm proves that attribute, object and sub-relation similarities are inter-dependent (Das & Mannila 2000). They convert each object with binary attribute values to a continuous vector by a kernel smoothing function, and define the similarity between objects as the Manhattan distance between continuous vectors. Andritsos et al. (Andritsos, Tsaparas, Miller & Sevcik 2004) introduced a context-sensitive dissimilarity measure between attribute values based on the Jensen-Shannon divergence.

Co-occurrence is another important framework for considering object couplings. For example, word co-occurrence features (Figueiredo, Rocha, Couto, Salles, Gonçalves & Meira Jr. 2011) are introduced for text classification. Similarly, an algorithm ADD (Ahmad & Dey 2007) was proposed to compute the dissimilarity between attribute values by considering the co-occurrence probability between each attribute value and the values of another attribute. The concept of coupled object analysis is proposed (Cao 2013*a*), with comprehensive discussions on the problem of coupled behaviour analysis (Cao et al. 2012*a*). They consider intra-couplings between values, inter-couplings between attributes, and the integration of intra-couplings and inter-couplings in analysing coupled object relations. Coupled nominal similarity (Wang et al. 2011) measures are then proposed to address the intra-couplings, inter-couplings, and coupled object similarity in categorical data for clustering. In addition, similarity for coupled numerical data (Wang et al. 2013*a*) is also introduced.

When textual information is considered in RS, text similarity is always a fundamental and important research. To date, many text similarity approaches are based on word similarity, such as term frequency-inverse document frequency (TF*IDF) (Wu, Luk, Wong & Kwok 2008). The TF*IDF is a very popular method for computing the text similarity which is commonly used in information retrieval and text mining. The TF*IDF weight is a statistical measure used to evaluate how important a word is to a document in a collection or corpus. The importance increases proportionally to the number of times a word appears in the document but is offset by the frequency of the word in the corpus. Typically, the TF*IDF weight is composed by two terms: the first computes the normalized Term Frequency (TF), the number of times a word appears in a document, divided by the total number of words in that document; the second term is the Inverse Document Frequency (IDF), computed as the logarithm of the number of the documents in the corpus divided by the number of documents where the specific term appears.

The TF value measures how frequently a term occurs in a document. Since every document is different in length, it is possible that a term would appear much more times in long documents than shorter ones. Thus, the term frequency is often divided by the document length as a way of normalization, as shown in Eqn. 2.15.

$$TF(t,d) = \frac{f(t,d)}{|d|} \tag{2.15}$$

where the f(t,d) denotes the frequency of term t in the document d, $|d|$ represents the length of the document.

The IDF value measures how important a term is in the document. While computing TF, all terms are considered equally important. However it is known that certain terms, such as “is”, “of”, and “that”, may appear a lot of times but have little importance. Thus we need to weigh down the frequent terms while scale up the rare ones, by Eqn. 2.16.

$$IDF(t, D) = ln(\frac{N}{|d \in D : t \in d|}) \tag{2.16}$$

where N denotes the total number of documents in the corpus, and $|d \in D : t \in d|$ denotes the number of documents where the term t appears.

Then we can organize the textual documents into a term-document matrix with TF*IDF values using Eqn 2.17, saying, rows and columns represent terms and documents, respectively. Then all the documents can be represented by vectors with each column as a document vector. After this step, the document similarity can be easily computed by the vector cosine similarity.

$$TFIDF(t, d, D) = TF(t, d) * IDF(t, D) \tag{2.17}$$

Although TF*IDF method is successful in many real applications of text mining, it still does not consider the semantic interaction between the words and the sentential context, which negatively impact the performance. It is well known that semantic analysis can help solve the issue of traditional word similarity. For example, latent factor models, such as Latent Semantic Analysis (LSA) (Landauer et al. 1998), Probabilistic Latent Semantic Analysis (PLSA) (Hofmann 1999), Latent Dirichlet Allocation (LDA) (Blei et al. 2003) can model the semantics of words and sentences simultaneously in the low-dimensional latent space. However, these methods did not model the missing words of a sentence which are helpful for restricting the relations between words and sentence. Recently, a Weighted Textual Matrix Factorization (WTMF) approach (Guo & Diab 2012*b*) was proposed to address this issue by modelling missing words and achieved better performance of semantic analysis. The details of WTMF and the textual semantic relations are described in Chapter 5.

2.2.3 Aspects of Non-IIDness Learning

As discussed above, most machine learning and data mining theories and tools are built on the IID principle, which is really not always held in many emerging applications of big data. In fact, it has already been shown that simply upgrading the extant algorithms and frameworks (Cao 2013*b*)(Cao et al. 2012*a*) probably cannot solve the real challenges of non-IIDness learning. In fact, gaps in the application of existing IIDness-based learning theories and techniques for the aforementioned problem-solving are becoming bigger and bigger without considering the non-IID characteristics. Two fundamental challenges in non-IIDness learning are heterogeneity and coupling hidden (Cao 2014) in complex behavioural/social applications and big data.

As concluded by Prof. Longbing Cao (Cao 2014), the main aspects in non-IIDness learning include: entity, property, interaction, learning and context.

- Entity: refers to what the learning is conducted on; a learning entity consists of multiple levels from information source, data object, object instance, object attribute, to attribute values. For instance, an RS probably includes three data sources: (a) user-item rating preferences, as shown in Table 1.1; (b) user attributes, as shown in Table 1.2; and (c) item attributes, as shown in Table 1.3. In this example, the objects of users and items with attributes are the entities considered in non-IID RS.

- Property: refers to the entity property; which is further described by data distribution, structure (type), semantics, logic, linkage and dynamics etc. For instance, in RS, users and items often follow certain distributions, each user or item attribute has specific data structure and meaning, users and items are linked through user preferences which are rated in terms of certain business logic, and the values of each attribute follow respective dynamical change.

- Interaction: refers to the associations between data entities; entity interaction may occur on the entity or property levels. Such interactions may take different forms, including dependency, correlation, association, matching, semantics, co-occurrence, neighbourhood, distance, density, uncertainty (such as entropy), logic (such as inductive logic programming), overlapping (Palla, Dernyi, Farkas & Vicsek 2005) (or coverage), networking (including reference and citation) or connection (such as linkage analysis). Some of these perspectives, such as dependency, distance, and uncertainty, have been widely used for building learning models to partition objects. One example is the textual semantic relations between items in Chapter 5.

- Context: refers to additional entities, entity properties and interactions surrounding the target entities. Context may refer to a certain domain, and constraints on entities, properties and interactions. Domain-specific context creates special behavioural, social, organizational, cultural, economic or emotional linkage between entities.

- Learning: refers to the learning process explicitly or implicitly; non-IIDness may appear at learning aspects, such as method, objective, task, level, dimension, process, measure and outcome. For instance, the heterogeneity and couplings may exist in ensemble clustering(Wang et al. 2013*b*), multiple objectives, multiple tasks, multiple levels, multiple dimensions, different processes, multiple measures, and between outcomes (such as pattern combinations (Cao 2013*a*) (Cao et al. 2011) (Cao et al. 2008)).

Generally, the non-IIDness may be embodied in the above aspects regarding heterogeneity or couplings. For example, in this dissertation, the users and items in RS can be referred as entities, the user or item attributes are properties, the implicit or explicit relations between users and items are with regard to interaction aspect, the textual relations between items are about context aspect, the CF or MF models incorporating the couplings are about the learning aspect.

2.2.4 Challenges of Non-IIDness Learning

After introducing the main aspects of non-IIDness learning, the main challenges of heterogeneity and coupling are also explored (Cao 2014).The heterogeneity challenge could be studied from different perspectives such as dimensions, aspects, interactions and hierarchy. Some heterogeneity analysis ideas, such as combined mining (Cao et al. 2011) (Cao 2013*b*), subspace learning (Cao 2014)(Cai, He & Han 2007) and discriminative learning (Lacoste-Julien, Sha & Jordan 2008), are helpful for non-IIDness learning. In addition to the heterogeneity challenge, the coupling problems are studied in this dissertation, and we briefly introduce the concept of coupling as follows.

Informally, coupling refers to any relationship between two or more aspects, such as entity, property, interaction, context and learning. For the above aspects, we use the word "coupling" to refer to any forms of non-IIDness aspects between the underlying entities, properties, interactions, contexts and learning modules (e.g., object, method or pattern). In practice, couplings may be presented in different forms, types and formats, such as numeric correlation, syntactic relation, logical relation, probabilistic relation, and semantic relation for categorical objects. The couplings probably can be studied from the following perspectives.

- From the data format side, coupling may be presented in different forms corresponding to data formats, including numerical, categorical, textual, graphic, and mixed-structure. For example, the couplings within the attributes of users and items in Chapter 3 are studied on categorical data, and the semantic couplings in Chapter 5 are about textual data.

- From the knowledge representation aspect, coupling may be presented in terms of syntactic, semantic, graphic, or inferential specifications. For example, different temporal logic-based, ontological and inferential couplings are discussed for coupled group behaviours (Wang & Cao 2012) (Cao 2013*a*).

- From the domain perspective, coupling may be syntactic, semantic, organizational, social, cultural, and economic.
- From the characteristic perspective, coupling may be explicit vs. implicit, qualitative vs. quantitative, descriptive vs. deep, specific vs. comprehensive, local vs. global, certain vs. uncertain, and known vs. unknown/latent etc.

According to aspects of non-IIDness learning, there are also many different varieties of couplings which are discussed as follows.

- Entity coupling: Couplings exist in many entities, such as users and items, in RS. For example, intra-coupling occurs within an entity (Wang et al. 2011), and inter-coupling appears between entities (Wang et al. 2011). Coupling also has different forms for different types of entities, for example, couplings in numeric data (Wang et al. 2013*a*) are different from those in categorical data (Wang et al. 2011).
- Property coupling: Property coupling may take place within entity property distribution, structure, semantics, logic, linkage or dynamics, for example, intra-term and inter-term coupling for disclosing the semantic linkage between terms in document analysis (Cheng et al. 2013). In addition, one property of an entity may more or less depend on another property, which is called the inter-coupling of properties.
- Context coupling: The context surrounding an underlying learning problem may include aspects related to user sentiment and behaviour, as well as society, organization, culture, economy, domain or particular region or partition, and specific constraint. Corresponding context couplings exist in these aspects.
- Interaction coupling: From the interaction perspective, couplings may take the form of dependence, correlation, association, matching, semantic similarity, co-occurrence, neighbourhood, distance, density, uncertainty, logic relation, overlapping, networking or connection.

- Learning coupling. Learning coupling refers to learn the interactions between different components, including method, objective, task, level, dimension, process, measure and outcome, especially when the learning involves multiples aspects, for instance, multi-methods or multi-tasks. Typically, we consider to learn couplings on different levels, such as values, attributes, objects, methods and measures. Such couplings, which are more comprehensive and complex than correlation and association, refer to the relations that exist explicitly or implicitly between different entities.

- Value coupling: Couplings between the values within an attribute are considered as intra-couplings of an attribute. Values are usually dependent on each other.

- Object coupling: Data objects share similarity and dissimilarity, which forms object couplings. The measure of object couplings is embodied in attribute and value couplings. The heterogeneity between objects needs to be considered in object relation analysis. For example, the similarity metrics (Wang et al. 2011) are designed to handle coupled categorical objects.

- Method coupling: When multiple methods are involved, different methods may only capture a partial picture of the underlying problem, and there may be an overlapping area of different methods. Therefore, method relation analysis needs to consider method coverage, difference and complementarity. For instance, the coupling between different clustering methods are studied to form the ensemble clustering framework (Wang et al. 2013*b*).

- Pattern coupling: From the perspective of pattern mining, the discovered patterns also have coupling relations (Cao et al. 2011)(Cao 2013*b*), which could be explored from semantic, structural and knowledge representation perspectives (Cao 2013*b*).

- Coupling measurement: Different measures probably also have couplings. For example, correlation coefficient and distance-based similarity metrics are widely used for numeric couplings. However, for categorical data, matching, instance-based overlapping, co-occurrence frequency etc., are typical paradigms for considering coupling measurement. Big data is often multi-structured, leading to the need of considering the measure couplings for different methods, sources, and features.

It has already been demonstrated that coupling learning is beneficial for data mining, machine learning communities, although it is also a challenging learning process. Some issues need to be considered for learning couplings. First, we need to first determine what sorts of couplings should be considered for the specific data. Second, incorporating more coupling into the learning process may incur less gain but at a substantially higher cost. Hence, the learning model of coupling relations for specific tasks, such as fraud detection (Cao, Li, Wang & Yu 2013), need to be carefully designed.

2.3 Non-IID Recommender Systems

We have introduced the main related work of the traditional recommendation methods and the non-IIDness learning, we then review the relevant work of non-IID RS incorporating the social relations and coupling relations. As mentioned above, both of the social relations from online social networks and the coupling relations are the aspects of non-IIDness learning (Cao 2014).

2.3.1 Social RS

The increasing popularity of social media greatly enriches people's social activities with their families, friends, and colleagues, which produces rich social relations, such as friendships in Facebook[1], following relations in Twitter[2]

[1] https://www.facebook.com/
[2] https://twitter.com/

and trust relations in Epinions[3]. Online social relations provide a different way for individuals to communicate digitally and allow online users to share ideas and opinions with their connected users. With the advent of online social networks, more and more social information is incorporated to RS, and social RS are becoming an active area in RS (Ma, King & Lyu 2009) (Jamali & Ester 2010*a*).

The main motivation behind social RS is to leverage the auxiliary friend relations of users to tackle the common challenges in RS, e.g., cold-start and sparsity. For a new user to RS, it is usually difficult to find the like-minded users due to the limitations of traditional RS which assume that users are IID. However, online users are inherently connected via various types of relations such as friendships and trust relations. In addition to the user-item rating matrix in traditional RS, users in social RS are connected, providing social information. Since they are connected, users are correlated rather than IID, essentially non-IID. For example, users with following relations (Weng, Lim, Jiang & He 2010) are more likely to share similar interests in topics than two randomly chosen users, and users with trust relations (Tang, Gao, Hu & Liu 2013) are more likely to have similar preferences in item ratings. Through the social information known from social networking, this non-IID difficulty could be partly overcome. The underlying assumption of the social recommendation approach is that users' taste is influenced by their friends in social networking (McPherson, Smith-Lovin & Cook 2001)(Marsden & Friedkin 1993). Specifically, users with similar preferences are more likely to be connected, and users who are connected are more likely to have similar preferences. Analogous to the fact that users in the physical world are likely to seek suggestions from their friends before making a purchase decision and users' friends consistently provide good recommendations (Sinha & Swearingen 2001), social relations can be potentially exploited to improve the performance of online RS (Golbeck 2006) (Jamali & Ester 2009) (Ma, Yang, Lyu & King 2008)(Ma, Zhou, Liu, Lyu & King 2011).

[3]http://www.epinions.com/

Accordingly, assigning more weights to items that the friends are interested in will potentially improve the satisfaction of recommendations. For example, models incorporating social relations into MF (Koren 2008) (Koren et al. 2009), such as Social Recommendation (SoRec) (Ma et al. 2008), Social Trust Ensemble (STE)(Ma et al. 2009), and Recommender Systems with Social Regularization (Ma, Zhou, Liu, Lyu & King 2011), have attracted a lot of attention from academia and industry, since MF ignores the social activities between users, which is not consistent with the reality that we normally ask friends for recommendations. One of the earliest social RS appeared in 1997 (Kautz, Selman & Shah 1997). Myriads of social media services, such as Facebook and Twitter, have emerged in recent years to allow people to easily communicate and express themselves conveniently. The pervasive use of social media generates social information at an unprecedented rate. For example, Facebook, the largest social networking site produces 35,000,000,000 online friendships (Guy & Carmel 2011) and the most popular user on Twitter, the largest microblogging site, has 37,974,138 followers. The rapid development of social media has greatly accelerated the development of social RS (Guy & Carmel 2011)(King, Lyu & Ma 2010).

In addition, RS comprising social information are helpful for improving the following areas: tagging (Shepitsen, Gemmell, Mobasher & Burke 2008) (Feng & Wang 2012), explicit social links (Pham, Cao, Klamma & Jarke 2011) and explicit trust information (Pitsilis, Zhang & Wang 2011). In addition, social recommendation also involves the independent research field of social network analysis (SNA) (Wasserman & Faust 1994)(Scott 2011)(Scott 2000)(Davis, Lichtenwalter & Chawla 2013). SNA is the methodical analysis of social networks and has emerged as a key technique in modern sociology, contributing the research of social RS. It has gained a significant following in various disciplines, such as communication studies, economics, geography, and computer science; consumer tools for it are now commonly available (Scott 2000). Social RS can take advantage of research results from social network analysis, such as social correlation theories (McPherson et al. 2001)(Marsden & Friedkin 1993), status analysis (Page, Brin, Motwani & Winograd 1998)(Kleinberg 1999), community detection (Leskovec, Huttenlocher & Kleinberg 2010)(Tang & Liu 2010), online trust (Tang, Gao & Liu 2012)(Massa 2006)(Jøsang, Ismail & Boyd 2007)(Sherchan, Nepal & Paris 2013) and heterogeneous networks (Sun & Han 2012) for recommendation.

Social recommendation can potentially solve some challenging problems of traditional RS, such as the data sparsity problem and the cold-start problem, and has attracted broad attention from both academia and industry. On the one hand, social recommendation has been studied for many years in literature. Many successful systems have been proposed and recommendation performance improvement is reported. On the other hand, successful systems in industry are rare and there is a lot of work reporting unsuccessful experiences in applying social RS in academia and industry. Therefore, the advantages and disadvantages of social RS are also discussed (Tang, Hu & Liu 2013).

The success of social RS is contributed from the following aspects:

(1) Social RS can be improved with the accommodation of SNA techniques based on the complementary social relations. Online users rarely make decisions independently and usually seek advice from their friends before making purchase decision, because they are usually inherently correlated (Ma, Zhou, Lyu & King 2011) As mentioned, a users preference is more likely to be similar to that of her social network than to those of randomly chosen users. This phenomenon is widely observed in many online social networks, such as following relations in Twitter (Weng et al. 2010) and trust relations in Epinions (Tang, Gao, Hu & Liu 2013). Connected users provide different information from similar users for recommendation, which can be exploited to improve the quality of recommendations (Jamali & Ester 2009)(Crandall, Cosley, Huttenlocher, Kleinberg & Suri 2008).

(2) Users opinions and tastes can be propagated via social networks, which

can reduce the size of cold-start users. To create good quality recommendations, traditional RS need enough historical ratings from each user. Since the user-item rating matrix is usually very sparse due to most users rating few of the millions of items, two users don't have enough of the number of items rated in common required by user similarity metrics to compute similarity. Therefore, the system is forced to choose neighbours in the small portion of comparable users and is probably going to miss other non-comparable but relevant users (Massa & Avesani 2007). Most of these systems are not able to generate accurate recommendations for users with few or no ratings. Social recommendation can make recommendations as long as the user is connected to a large enough component of the social network (Jamali & Ester 2010*b*), hence social recommendation can significantly reduce cold-start users. For example, traditional RS totally fail for new users, however, by considering ratings of trusted users (Massa & Avesani 2007), social recommendation achieves a very small error and is able to produce a recommendation for almost 17% of the users.

(3) Social recommendation can significantly improve the coverage of recommendation. On a very sparse dataset that contains a large portion of cold-start users and of items rated by just one user, coverage becomes an important issue since many of the ratings become hardly predictable (L, Medo, Yeung, Zhang, Zhang & Zhou 2012). Coverage refers to the fraction of ratings for which, after being hidden, the RS are able to produce a predicted rating. By propagating trust, it is possible to reach more users; hence, to compute a predicted trust score among them and to count them as neighbours, social recommendation can improve the coverage of recommendation especially for new items. For example, traditional RS cannot be applied to new locations; however, social recommendation can achieve more than 20% recommendation accuracy for new locations (Gao, Tang & Liu 2012*b*) when exploiting social relations.

Due to the potential value of social information in RS, social recommendation is aggressively pursued in industry and academia. However, it is still difficult to really apply the social RS successfully in applications (Cho, Myers & Leskovec 2011) (Gao, Tang & Liu 2012*a*)(Leskovec, Adamic & Huberman 2007). The limitations of social RS (Tang, Hu & Liu 2013) include.

(1) Social relations are too noisy and may have a negative impact on RS. The low cost of formation of connections allows one to have an inordinate number of friends in the online world. For example, a Facebook user has 130 friends on average, and an average Twitter user has 126 followers. Research by Robin Dunbar indicates that 100 to 150 is the approximate natural group size in which everyone can really know each other because our minds are not designed to allow us to have more than a very limited number of people in our social world. The emotional and psychological investments that a close relationship requires are considerable, and the emotional capital we have available is limited (Dunbar 2010). Since the social network comprises of valuable friends, casual friends and event friends, users are not necessarily all that similar and social relations mixed with useful and noise connections may introduce negative information into RS. For example, social RS simply using all available relations perform worse than traditional RS (man Au Yeung & Iwata 2011)(Tang et al. 2012).

(2) Cold-start users are likely to also have few or no social relations and it is difficult for social RS to improve recommendation performance for these users. The available social relations are extremely sparse, and the distribution of the number of social relations follows a power-law-like distribution (Newman 2005), suggesting that a small number of users specify many social relations while a large proportion of users specify a few relations. Users with many social relations are likely to be active users (Agarwal, Liu, Tang & Yu 2008) and they are likely to have many ratings, while users with fewer ratings are likely to also have fewer connections (Tang, Gao, Hu & Liu 2013). For users with enough ratings,

traditional RS already perform well, while for users with fewer ratings, they are likely to also have fewer social relations and social RS cannot help much. For example, some social RS (Cho et al. 2011) (Gao et al. 2012*a*)(Leskovec et al. 2007) gain a little or even no improvement compared to traditional RS.

(3) Different types of social relations have different effects on social RS, and successful experiences in exploiting trust relations may not be applicable to other types of relations. Most existing successful social RS use trust relations and recommends items to a user from her trusted users. Trust plays a central role in exchanging relationships involving unknown risk (Gefen, Karahanna & Straub 2003), which provides information about with whom we should share information and from whom we should accept information (Golbeck 2009). The role of trust is especially critical in some online communities, such as e-commerce sites and product review sites, which has been described as the "wild wild west" of the 21st century (McKnight, Choudhury & Kacmar 2002). These successful social RS assume that trust relations are formed when users have similar opinions to similar products. With this assumption, users are likely to seek recommendations from their trusted friends and social RS are likely to be successful. However, trust is a complex concept and has different interpretations in different contexts (Dellarocas, Zhang & Awad 2007)(Falcone & Castelfranchi 2010)(Adali 2013). A user in the context of product review sites, such as Epinions, will trust the users if she agrees with their opinions about products, while a user in the context of P2P networks will trust others because of their reliability. Different interpretations of trust may result in different solutions for social recommendation, which suggests that more nuanced approaches are needed to exploit trust relations for recommendation. Trust relations are not necessarily equivalent to other types of social relations. For example, following a user in Twitter does not indicate one's trust in the user. The success of exploiting trust relations for recommendation may not be ap-

plied to other relations, and different types of social relations may have very different impacts on social RS (Yuan, Chen & Zhao 2011).

2.3.2 Incorporating Coupling into RS

In addition to the social relations, there are also many other couplings have been incorporated into the non-IID RS. In this subsection, we introduce the related work of incorporating couplings into non-IID RS.

The social RS approaches consider the social activities of users, but social relations are mixed and treated equally, which violate with a fact that user's social interests are intrinsically multifaceted. As a result, it is impossible to differentiate social recommendations from different friends in terms of their preferred areas. In fact, everyone has specific preference in particular groups. This indicates that a user may trust different subsets of friends in different groups. More specially, a user may have friends working in different domains, and join in activities across different domains. This is evidenced by that the extant social networks, such as Google+, Facebook, and Twitter, already have such mechanisms to divide users into groups for sharing different information with different groups. Undoubtedly, in social RS, utilizing such social group information will be able to provide better personalized services for users. But most of extant social Web applications, such as Tweeter, Sina Weibo, and Delicious, do not provide reliable mechanisms to allow users to differentiate social connections from individual groups.

Some recent researches integrate the distinguished group information into recommendation algorithms, e.g. (Yang, Steck & Liu 2012) leverages the social trust circles from item-category information for social recommendation. Despite of the superior results demonstrated from the given multi-category rating data sets, this approach has a major limitation that it relies on the explicit item category information to form user circles, upon which social recommendation is made. However, such information is not always available in existing social networks e.g., Facebook or Twitter might not have such explicit category information, resulting in difficulties in applying the proposed algorithm. To address this problem, a coupled group-based MF has been proposed (Li et al. 2013) by leveraging the user and item groups learned by topic modeling and incorporating couplings between users and items and within users and items.

In addition, there are also other non-IID RS methods. For example, item couplings computed from the coupled object similarity (Yu et al. 2013) are integrated into a hybrid recommendation algorithm combing the CBR and CF techniques. Specifically, this method firstly partitions items into several item groups by using a coupled version of k-modes clustering algorithm (Wang et al. 2011) (Chaturvedi, Green & Caroll 2001), where the similarity between items is measured by the coupled object similarity (Wang et al. 2011) considering couplings between items. The CF technique is then used to produce the recommendations for active users. This work shows that item couplings are beneficial for improving recommendation performance by incorporating them into the hybrid recommendation method. However, the couplings between users are still ignored. To deeply analyse the non-IID problem, we incorporate the coupling relations between users and between items into CF recommendation methods, and propose three hybrid neighbourhood-based CF methods CUCF, CICF and CCF. Specifically, we first apply a novel coupled object similarity approach to compute the similarities between users and between items, then integrate the inferred user and item couplings into CF to form the proposed CUCF, CICF and CCF methods, which will be detailed in Chapter 3. In addition, we also propose novel MF models integrating item couplings (Li et al. 2014), user couplings (Li et al. 2015*a*) derived from the attributes, and textual semantic couplings (Li, Xu & Cao 2015*b*), which will be further discussed in Chapter 4 and 5.

In order to capture the couplings among objects, a novel coupled object similarity (Wang et al. 2011) measure based on the attribute space of objects showed as Table 2.1 was introduced. Here are some relevant definitions of the coupled object similarity, which will be further applied to compute the couplings between users and between items in the latter chapters.

Table 2.1: Object Attribute Space

Objects	A_1	A_2	...	A_n
Ob_1	V_{11}	V_{12}	...	V_{1n}
Ob_2	V_{21}	V_{22}	...	V_{2n}
...	...	...	...	...
Ob_m	V_{m1}	V_{m2}	...	V_{mn}

Definition 2.1 (Information Functions) *Given an attribute space of objects $C_{Ob} =< Ob, A, V, f >$ with m rows and n columns, A and V are the corresponding attribute and values sets for object set Ob. The Information Functions are defined as:*

$$f_j^*(\{Ob_{k_1}, ..., Ob_{k_t}\}) = \{f_j(Ob_{k_1}), ..., f_j(Ob_{k_t})\} \tag{2.18}$$

$$g_j(x) = \{Ob_i | f_j(Ob_i) = x, 1 \leq j \leq n, 1 \leq i \leq m\} \tag{2.19}$$

$$g_j^*(W) = \{Ob_i | f_j(Ob_i) \in W, 1 \leq j \leq n, 1 \leq i \leq m\} \tag{2.20}$$

where $Ob_i, Ob_{k_1}, ..., Ob_{k_t} \in Ob$, and $W \subseteq V_j$.

Definition 2.2 (Intra-coupled Attribute Value Similarity) *Given an attribute space of objects $C_{Ob} =< Ob, A, V, f >$, the Intra-coupled Attribute Value Similarity (IaAVS) between values x and y of attribute A_j for objects is defined as:*

$$\delta_j^{Ia}(x, y) = \frac{|g_j(x)|.|g_j(y)|}{|g_j(x)| + |g_j(y)| + |g_j(x)|.|g_j(y)|} \tag{2.21}$$

where $g_j(x)$ is the subset of objects Ob with corresponding attribute A_j having attribute value x, and $|g_j(x)|$ is the size of the subset.

Definition 2.3 (Inter-information Function) *Given an attribute space of objects $C_{Ob} =< Ob, A, V, f >$, its Inter-information Function (IIF) $\varphi_{j \to k}$:*

$V_j \rightarrow 2^{V_k}$ is defined:

$$\varphi_{j\rightarrow k}(x) = f_k^*(g_j(x)) \tag{2.22}$$

This IIF $\varphi_{j\rightarrow k}$ is the composition of f_k^* and g_j.It contains the kth attribute value subset for the corresponding objects, which are derived from the jth attribute value x.

Definition 2.4 (Information Conditional Probability) *Given an attribute space of objects $C_{Ob} =< Ob, A, V, f >$, the kth attribute value subset $W \subseteq V_k$, and the jth attribute value $x \in V_j$, the Information Conditional Probability (ICP) of W with respect to x is $P_{k|j}(W|x)$:*

$$P_{k|j}(W|x) = \frac{|g_k^*(W) \cap g_j(x)|}{|g_j(x)|} \tag{2.23}$$

Definition 2.5 (Inter-coupled Relative Similarity) *Given an attribute space of objects $C_{Ob} =< Ob, A, V, f >$, the Inter-coupled Relative Similarity (IRS) between attribute values x and y of attribute A_j based on another attribute A_k is:*

$$\delta_{j|k}(x, y) = \sum_{w\in\cap} min\{P_{k|j}(w|x), P_{k|j}(w|y)\} \tag{2.24}$$

where $w \in \varphi_{j\rightarrow k}(x) \cap \varphi_{j\rightarrow k}(y)$.

Definition 2.6 (Inter-coupled Attribute Value Similarity) *Given an attribute space of objects $C_{Ob} =< Ob, A, V, f >$, the Inter-coupled Attribute Value Similarity (IeAVS) between attribute values x and y of attribute A_j for objects is:*

$$\delta_j^{Ie}(x, y) = \sum_{k=1,k\neq j}^{n} \gamma_k \delta_{j|k}(x, y) \tag{2.25}$$

where γ_k is the weight parameter for attribute A_k, $\sum_{k=1,k\neq j}^{n} \gamma_k = 1$, $\gamma_k \in [0, 1]$, and $\delta_{j|k}(x, y)$ is inter-coupled relative similarity.

Definition 2.7 (Coupled Attribute Value Similarity) *Based on IaAVS and IeAVS, the Coupled Attribute Value Similarity (CAVS) between attribute* *values x and y of attribute A_j is defined as follows.*

$$\delta_j^A(x,y) = \delta_j^{Ia}(x,y) * \delta_j^{Ie}(x,y) \quad (2.26)$$

Definition 2.8 (Coupled Object Similarity) *For two objects described by the attribute space $C_{Ob} =< Ob, A, V, f >$, the Coupled Object Similarity (COS) is defined to measure the similarity between different objects.*

$$COS(ob_i, ob_j) = \sum_{k=1}^{J} \delta_k^A(V_{ik}, V_{jk}) \quad (2.27)$$

where V_{ik} and V_{jk} are the values of attribute j for objects ob_i and ob_j, respectively; and δ_k^A is Coupled Attribute Value Similarity.

From the definition of coupled object similarity (Wang et al. 2011), we clearly see that the intra-couplings between values within an attribute and inter-couplings between attributes are incorporated for measuring object coupling. This coupled object similarity can partly help to uncover the intrinsic relations within objects rather than considering them independently. And the coupling relations between objects can be used to derive couplings between users and between items for RS.

Chapter 3

Coupled Neighbourhood-based Collaborative Filtering

3.1 Introduction

RS are extremely useful in overcoming the information overload problem by predicting users preferences and recommending potentially interesting items to users (Melville & Sindhwani 2010). A popular and interesting research topic, RS have been successfully applied in many industry applications, such as LinkedIn, Amazon, Taobao and Last.Fm. In the extant recommendation algorithms, CF is a popular method of recommending interesting items for users. The essence of a CF (Su & Khoshgoftaar 2009) algorithm is to compute the similarities between users or items based on historical ratings. However, it is sometimes difficult to accurately infer the relations among users or items only based on the user-item ratings, as it often suffers from the cold-start problem for new users or new items in RS.

To better illustrate this issue, we first briefly discuss the traditional item-based CF method based on a user-item rating matrix including m users and n items represented by rows and columns, respectively, as shown in Fig. 3.1. To compute the similarity between items o_i and o_j, we first obtain the co-rated users u_2, u_l and u_m, and then organize their ratings for items o_i and o_j

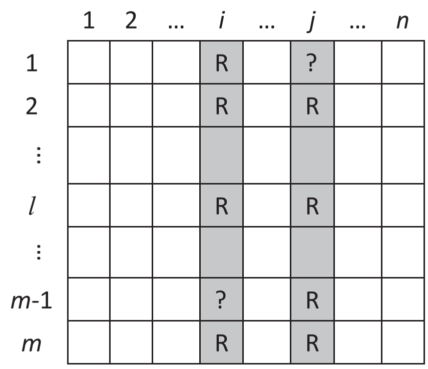

Figure 3.1: Item-based CF

into two vectors $\vec{i} =< R_{2i}, R_{li}, R_{mi} >$ and $\vec{j} =< R_{2j}, R_{lj}, R_{mj} >$, respectively. After this, the similarity between items o_i and o_j can easily be computed by Vector Cosine Similarity or Pearson Correlation Similarity measures, after which the active user can receive recommendations for interesting items which are similar to the items they currently like. Now imagine item o_n is a new product which has not been rated by any users. This is known as a cold-start item, and it is impossible to infer similar items only from the user-item rating matrix.

To overcome this cold-start challenge, many hybrid methods have been proposed which consider different types of data to improve recommendation quality. For example, content-based recommendation methods utilizing the users' or items' valuable attributes are often integrated with CF algorithms (Gantner, Drumond, Freudenthaler, Rendle & Schmidt-Thieme 2010). Generally, leveraging other valuable information into RS is beneficial for enhancing recommendation performance. Nevertheless, most of the existing methods assume that the attributes of users/items are IID. This is a very fundamental and critical issue for the RS community, as the huge amount of recommendation data in online, social, mobile and business applications is essentially not identically and independently distributed. For example, a user's preference may influence his/her friends, users' attributes are often associated with each other via explicit or implicit relationships (Wang et al. 2013*a*) (Wang et al. 2013*b*) (Cao et al. 2012*b*). There are two important aspects that have not been considered thoroughly regarding non-IID issues, (1) the heterogeneity between users and between items, namely users and items are personalized and thus ratings need to be tailored according to individual characteristics; (2) the coupling relationships between users, between items, and between users and items, namely users and items are coupled and hence ratings need to capture the underlying interactions. These two aspects bring the recommendation problem to the non-IID RS (Cao 2015) (Cao 2014). Therefore in this chapter, we thoroughly analyse the coupling relationships between users and between items based on their attributes, and incorporate the coupling relations into CF methods to improve recommendation quality.

Table 3.1: User Attributes

User	Occupation	Age	ZipCode	Country	Sex
u_1	Analyst	20	10081	China	M
u_2	Engineer	40	2007	Australia	F
u_3	Developer	20	2008	Australia	M

Table 3.2: Item Attributes

Item	Director	Actor	Genre
$GodFather$	Scorsese	De Niro	Crime
$GoodFellas$	Coppola	De Niro	Crime
$Vertigo$	Hitchcock	Stewart	Thriller
$NbyNW$	Hitchcock	Grant	Thriller

To illustrate the coupling relationships between users and between items in RS, we give a toy example in Tables 3.1 3.2 3.3. Assume there is a rating matrix consisting of three users and four items with their attributes. Most traditional CF methods only utilize the rating information for recommendation and ignore the attributes of users and items. However, when the rating matrix is very sparse, the attribute information of users and items may also

Table 3.3: User-Item Ratings

	God Father	Good Fellas	Vertigo	N by NW
u_1	1	3	5	4
u_2	4	2	1	5
u_3	-	2	-	4

contribute to solving the challenges. Specifically, we can infer the relationship between u_1 and u_2 from the "Age", "ZipCode", "Country", "Gender" and "Occupation" attribute space. Similarly, we can obtain the movies' relationship from the "Director", "Actor" and "Genre" attributes. Intuitively, the existing similarity methods such as Pearson Correlation can be applied to compute the similarities between users and between items, based on the IID assumption. In reality, however this assumption is not always held and there are more or less coupling relations between instances and attributes. We know that the coupling relationships can be aggregated from the similarity of attribute values for all the attributes. From the perspective of IID assumption, different attribute values are independent, and one attribute value will not be influenced by others. However, in reality we observe that one attribute value will often be dependent on other values of the same attribute. Specifically, two attribute values are similar if they present an analogous frequency distribution on one attribute, which leads to another so-called intra-coupled similarity within an attribute. For example, two job occupations "Analyst" and "Engineer" or two directors "Scorsese" and "Coppola" are considered similar because they appear with the same frequency in the data. On the other hand, the similarity of two attributes values is dependent on other attribute values from different attributes, for example, two occupations' relationship is dependent on "Country" and "Gender" attributes over all the users, and two directors' relationship is also dependent on "Actor" and "Genre" attributes over all the items. This dependent relation is called the inter-coupled similarity between attributes. We believe that the intra and inter-coupled similarities with a non-IID perspective will contribute to analysing the relationships between users and between items, namely user coupling and item coupling, to further improve recommendation qualities by incorporating user coupling and item coupling into RS.

The contributions of the chapter are summarized as follows:

- We apply a novel coupled object similarity to compute the user couplings and item couplings which consider the coupled interaction between attributes.

- We propose a novel hybrid CF recommendation method, CUCF, incorporating the user couplings into a user-based CF method with a hybrid user similarity measure.

- We propose a novel hybrid CF recommendation method, CICF, integrating the item couplings into item-based CF method with a hybrid item similarity measure.

- We propose a novel hybrid CF recommendation method, CCF, integrating the user couplings, item couplings and the user-item ratings into the CF method with the hybrid user similarity and item similarity measures.

- We conduct experiments to evaluate the superiority of couplings and the effectiveness of proposed non-IID CUCF, CICF and CCF methods.

The rest of the chapter is organized as follows. Section 2 first analyses the couplings in RS, then details the proposed CUCF, CICF and CCF methods. The experimental results and the analysis are presented in Section 3, followed by the conclusion.

3.2 Coupled Neighbourhood-based CF

In this section, we propose the coupled neighbourhood-based CF methods, CUCF, CICF and CCF, which incorporate the coupling relations between

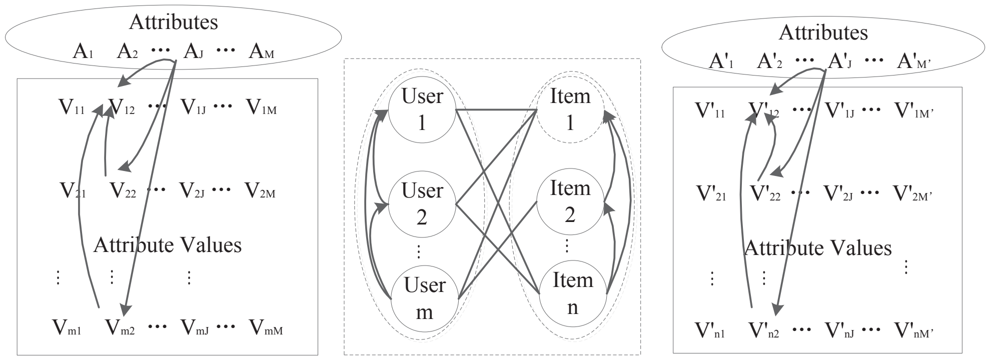

Figure 3.2: Couplings in Recommender Systems

users and between items into CF. In addition, we analyse the complexity of computing user couplings and item couplings in the final part of this section.

3.2.1 Coupling Relations

The extant similarity methods for computing the relationships between users and between items often assume that the attributes are independently distributed. However, all the attributes should be coupled together and further impact each other. The user couplings and item couplings in RS are illustrated in Fig.3.2, which clearly shows that within an attribute A_j, there is a dependent relation between values V_{lj} and V_{mj} $(l \neq m)$, while a value V_{li} of an attribute A_i is further influenced by the values of other attributes A_j $(j \neq i)$. For example, attributes A_1, A_3, ... to A_J all more or less influence the values of V_{12} to V_{m2} of attribute A_2.

As mentioned, user coupling should reflect the non-IID relationships between different users. For two users described by the attribute space $C_U =< U, A, V, f >$, the Coupled User Similarity (CUS) is derived from the coupled object similarity to measure the similarity between users.

Definition 3.1 (Coupled User Similarity) *Formally, given user attribute space $C_U =< U, A, V, f >$, the Coupled User Similarity (CUS) between two* *users u_i and u_j is defined as follows.*

$$CUS(u_i, u_j) = COS(u_i, u_j) = \sum_{k=1}^{J} \delta_k^A(V_{ik}, V_{jk}) \tag{3.1}$$

where V_{ik} and V_{jk} are the values of attribute k for user u_i and u_j, respectively; and δ_k^A is Coupled Attribute Value Similarity.

Similarly, item coupling should also reflect the non-IID relationships between different items. For two items described by the attribute space $C_O < O, A', V', f' >$, the Coupled Item Similarity (*CIS*) can also be derived from the coupled object similarity to measure the similarity between items.

Definition 3.2 (Coupled Item Similarity) *Formally, given item attribute space $C_O =< O, A', V', f' >$, the Coupled Item Similarity (CIS) between two items o_i and o_j is defined as follows.*

$$CIS(o_i, o_j) = COS(o_i, o_j) = \sum_{k=1}^{J} \delta_k^{A'}({V'}_{ik}, {V'}_{jk}) \tag{3.2}$$

where ${V'}_{ik}$ and ${V'}_{jk}$ are the values of attribute k for items o_i and o_j, respectively; and $\delta_k^{A'}$ is Coupled Attribute Value Similarity.

From the above two definitions, we clearly see that the computations of user couplings and item couplings involve the intra-couplings between values within an attribute and the inter-couplings between attributes which partly uncover the intrinsic relations between users and between items, rather than considering them independently. These user couplings and item couplings are employed and incorporated into the neighbourhood-based CF in this chapter, and will be integrated into MF methods in Chapter 4.

3.2.2 Coupled User-based CF

On top of the traditional user-based CF method, we propose a novel CUCF method as shown in Fig. 3.3 which takes not only the user-item rating matrix, but also the user couplings into account. In Fig. 3.3, S represents the

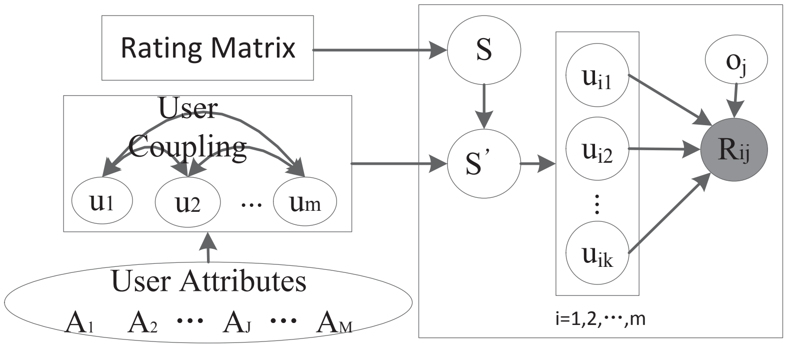

Figure 3.3: Coupled User-based CF

similarity matrix computed from the user-item rating matrix, S' is the hybrid similarity matrix which integrates user couplings and the basic similarity matrix S together, and $u_{i1}, ..., u_{ik}$ represent the neighbours of user u_i. We first compute the user couplings based on their attributes. We then incorporate user couplings and users' rating preferences to compute the hybrid similarity between users which is further used to derive users' neighbourhoods. After the active user's neighbours are determined, the rating prediction R_{ij} for user u_i on specific item o_j can be aggregated from the ratings of the neighbours of this user. The CUCF method has the following advantages: (1) the user-item rating matrix and user's attributes are combined together to compute the similarity between users; (2) the user couplings based on the user's attributes disclose the common IID assumption and consider real world non-IID situations; (3) the hybrid similarity measure considering different data sources is beneficial for overcoming the challenge of cold-start users.

Basically, the predicted rating values should be as close as possible to the observed rating values for RS, which have been studied well in most existing recommendation algorithms with the consideration of the similarity between users from the user-item rating matrix. However, the prediction should also be related to the user couplings computed from the user attributes, which are ignored by most user-based CF algorithms or without considering the non-IID characteristics. Therefore, we propose this hybrid similarity approach to integrate the user couplings and the user-item rating similarity, and to derive users' neighbourhoods and aggregate their ratings to predict the preference of the target user, as shown in Eqn. 3.3 and 3.4.

$$S'(u_i, u_j) = \alpha * CUS(u_i, u_j) + (1 - \alpha) * S(u_i, u_j) \tag{3.3}$$

$$R_{ij} = \begin{cases} \frac{\sum_{u_k \in N(u_i)} S'(u_i, u_k) R_{kj}}{\sum_{u_k \in N(u_i)} S'(u_i, u_k)} & \sum |S'(u_i, u_k)| > 0, \\ \overline{R}_{u_i} & \sum |S'(u_i, u_k)| = 0. \end{cases} \tag{3.4}$$

where $CUS(u_i, u_j)$, $S(u_i, u_j)$ and $S'(u_i, u_j)$ are respectively the coupled similarity, the rating similarity from user-item rating matrix, and the hybrid similarity between users u_i and u_j, $N(u_i)$ represents the neighbourhood of user u_i.

3.2.3 Coupled Item-based CF

From the above discussion, we can clearly see that the CUCF method considers both user coupling and user-item ratings to infer user's relations. Items' relations are also very important for improving recommendation qualities, for example, users will probably like a new item o_i which is similar to other items that the users like. Therefore, similar to the CUCF method, we also propose a novel CICF method which integrates the item couplings and the user-item rating matrix, as shown in Fig. 3.4. In this figure, S represents the similarity matrix between the items computed from the user-item rating matrix, S' is the hybrid similarity matrix which integrates the item couplings and the basic similarity matrix S for items, and $o_{j1}, ..., o_{jk}$ represent the neighbours of item o_j. We first compute the item couplings based on their attributes, then incorporate the item couplings and users' rating preferences to compute the hybrid similarity between items which are further used to derive the items' neighbourhoods. After the target item's neighbours are determined, the rating prediction R_{ij} for user u_i on the specific item o_j can be aggregated from the ratings of the neighbours of this item. The CICF method also has

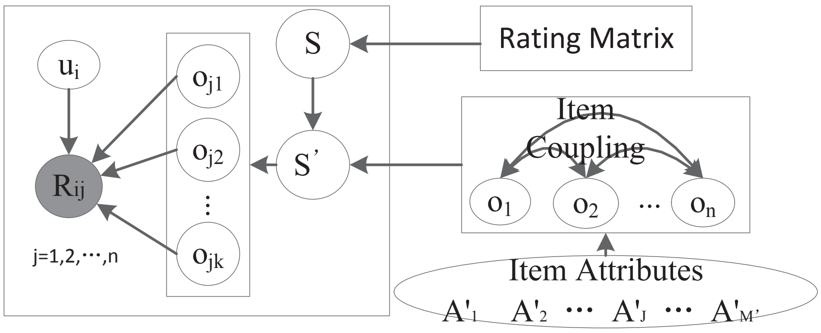

Figure 3.4: Coupled Item-based CF

the following advantages: (1) the user-item rating matrix and the item's attributes are combined to compute the similarity between items; (2) the item couplings based on the item's attributes disclose the common IID assumption and consider real world non-IID situations; (3) the hybrid similarity measure considering different data sources is beneficial for overcoming the challenge of cold-start items.

Different from CUCF, the prediction of CICF considers the item couplings computed from the item attributes, which are ignored by most item-based CF algorithms without considering the non-IID characteristics. Therefore, we propose a hybrid similarity approach to integrate the item couplings and the user-item rating similarity, and to derive items' neighbourhoods and aggregate the relevant ratings to predict a preference for the target user, as shown in Eqn. 3.5 and 3.6.

$$S'(o_i, o_j) = \beta * CIS(o_i, o_j) + (1 - \beta) * S(o_i, o_j) \tag{3.5}$$

$$R_{ij} = \begin{cases} \frac{\sum_{o_k \in N(o_j)} S'(o_j, o_k) R_{ik}}{\sum_{o_k \in N(o_j)} S'(o_j, o_k)} & \sum |S'(o_j, o_k)| > 0, \\ \overline{R}_{o_j} & \sum |S'(o_j, o_k)| = 0. \end{cases} \tag{3.6}$$

where $CIS(o_i, o_j)$, $S(o_i, o_j)$ and $S'(o_i, o_j)$ are respectively the coupled similarity, the rating similarity from the user-item rating matrix, and the hybrid similarity between items o_i and o_j, $N(o_i)$ represents the neighbourhood of

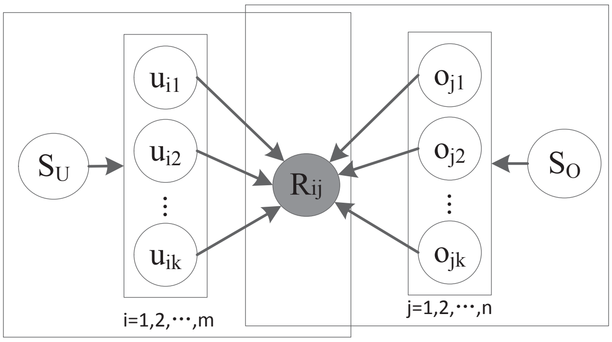

Figure 3.5: Coupled CF

item o_i.

3.2.4 Coupled CF

To this point, we have discussed the CUCF and CICF methods which respectively integrate user couplings and item couplings into CF methods. CUCF considers user couplings and user-item ratings to infer user's relations, and CICF incorporates item couplings and user-item ratings to infer item's relations to produce recommendations using the inferred user or item relations. However, we can clearly observe that the user's relations and item's relations are still separately incorporated into the CF methods. In fact, the user's relations and item's relations should be considered and balanced together to impact the recommendation strategies. Therefore, to completely consider the non-IID relation in RS, we propose another novel CCF method by incorporating the user couplings, item couplings, and the user-item ratings, as shown in Fig. 3.5. This CCF method balances the impact of the non-IID relations to improve the recommendation quality. In Fig. 3.5, S_U and S_O respectively represent the hybrid similarity matrix for users and items computed from Eqn. 3.3 and 3.5, and $u_{i1}, ..., u_{ik}$ represent the neighbours of user u_i, $o_{j1}, ..., o_{jk}$ represent the neighbours of item o_j. Different from CUCF and CICF, CCF integrates user couplings, item couplings and user-item ratings into the CF method.

After incorporating the different relations into CF, the prediction function of the proposed method can be revised as follows:

$$R_{ij} = \alpha' \frac{\sum_{o_k \in N(o_j)} S'(o_j, o_k) R_{ik}}{\sum_{o_k \in N(o_j)} S'(o_j, o_k)} + (1 - \alpha') \frac{\sum_{u_k \in N(u_i)} S'(u_i, u_k) R_{kj}}{\sum_{u_k \in N(u_i)} S'(u_i, u_k)} \tag{3.7}$$

where $S'(o_j, o_k)$ is the hybrid similarity between items o_j and o_k, $S'(u_i, u_k)$ is the hybrid similarity between users u_i and u_k, $N(o_j)$ and $N(u_i)$ respectively represent the neighbourhood of item o_j and user u_i. The prediction function clearly shows that the user's hybrid relations, the item's hybrid relations, and the user-item ratings are incorporated to balance their impact and further improve the recommendation quality.

3.2.5 Complexity Analysis

For the proposed CUCF, CICF and CCF methods, the most time-consuming step is the calculation of coupling relations for users and items from their respective attributes. The time cost for computing the coupled similarities between users/items is $O(D^2 w^3)$, where D is the number of user or item attributes, and w is the maximal number of attribute values for all the attributes in the attribute information matrix for users and items. In practice, to speed up the online learning process, the computation of the coupled user or item similarity can be offline, which does not decrease the efficiency of online CF recommendation, as the user or item information space is usually fixed or stable for a given recommendation problem.

3.3 Experiments and Results

In this section, we evaluate our proposed methods and compare them to the existing approaches on MovieLens100K data set.

3.3.1 Data Set

The MovieLens100K data set was collected by the GroupLens Research Project at the University of Minnesota and has been widely explored in RS research in the last decade. The MovieLens100K data set consists of 100,000 ratings from 943 users on 1682 movies. The ratings are made on a 5-star scale and each user has rated at least 20 movies. In addition, simple demographic information on the users such as age, gender, occupation and zip code is also included in the data, which can be used to compute the user couplings. Similarly, the genre attributes of items such as Crime, Documentary, Drama are used to infer the item couplings.

3.3.2 Experimental Settings

We perform five-fold cross validation in our experiments for the MovieLens100K data set. We first split the original data into five equal sized samples, then we keep a single sample as the test set, and the remaining four samples are used as training sets. In this way, the original data is converted to a five-fold data set with each fold being 80% as the training set and 20% as the test set. Then, the cross-validation process is repeated five times on the sampled data for each fold. Finally, the estimation on the whole data can be averaged from the five results for each fold. Similarly, we also test other different percentages such as 60%, 40% and 20% as the training set in each fold for the cross validation process. We use Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) as the evaluation metrics:

$$RMSE = \sqrt{\frac{\sum_{(u_i,o_j)|R_{test}} \left(R_{ij} - \hat{R}_{ij}\right)^2}{|R_{test}|}} \tag{3.8}$$

$$MAE = \frac{\sum_{(u_i,o_j)|R_{test}} \left|R_{ij} - \hat{R}_{ij}\right|}{|R_{test}|} \tag{3.9}$$

where R_{test} is the set of all pairs (u_i, o_j) in the test set.

To evaluate the performance of our proposed CUCF, CICF and CCF, we consider eight different benchmark methods. These models include:

- User-based CF ($UBCF^p$) (Su & Khoshgoftaar 2009). This method first computes users' similarity by Pearson Correlation on the rating matrix, then recommends relevant items to the given user according to the users who have strong relationships;

- User-based CF ($UBCF^c$) (Su & Khoshgoftaar 2009). This method is also a type of user-based CF like $UBCF^p$, but different from $UBCF^p$, this method computes users' similarity by Cosine Similarity on the rating matrix;

- Item-based CF ($IBCF^p$) (Deshpande & Karypis 2004). This method first considers items' similarity by Pearson Correlation on the rating matrix, then recommends relevant items which have strong relationships with the given user's interesting items.

- Item-based CF ($IBCF^c$) (Deshpande & Karypis 2004). This method is also a type of item-based CF like $IBCF^p$, but the difference is that $IBCF^c$ considers items' similarity by Cosine Similarity on the rating matrix;

- Hybrid User-based CF method ($HUCF^p$). This method integrates the relations between users based on their attributes and the rating preferences together into the user-based CF algorithm. The relations between users based on their attributes are defined as the similarities between users computed by Pearson Correlation Coefficient with the same incorporation method as our proposed $CUCF$ method.

- Hybrid User-based CF method ($HUCF^c$). This method is similar to the hybrid $HUCF^p$ method, the only difference being that the relations between users based on their attributes are computed by Cosine Similarity.

- Hybrid Item-based CF method ($HICF^p$). This method integrates the relations between items based on their attributes and the rating preferences together into the hybrid item-based CF algorithm. The relations between items based on their attributes are defined as the similarities between items computed by Pearson Correlation Coefficient with the same incorporation method as our proposed $CICF$ method.

- Hybrid Item-based CF method ($HICF^c$). This method is similar to the hybrid $HICF^p$ method, the difference being that the relations between items based on their attributes are computed by Cosine Similarity.

3.3.3 Impact of Parameters

Parameters α and β respectively control the influence of user couplings and item couplings in the proposed CUCF and CICF methods. A bigger value of α indicates a higher impact of the user couplings in CUCF, and a bigger value of β implies a higher impact of the item couplings in CICF. To select the optimum parameters α and β, we depict the MAE and RMSE changing trends of the proposed CUCF and CICF methods when α and β ranges in [0,1]. Fig. 3.6 shows the impacts of parameter α and β on the MovieLens100K data set. Experimental results show that the proper values of α and β for the MovieLens100K data are 0.4 and 0.6, respectively.

In addition to the parameters, the neighbourhood size of users and items also impact the recommendation qualities of the proposed CUCF and CICF methods. Fig. 3.7 shows the effect of the neighbourhood size of users and items for the MovieLens100K data set. The experimental results indicate that the MAE and RMSE decrease sharply with the increase of the neighbourhood size of users or items until reaching a steady point. Continually increasing the neighbourhood size does not improve the performance after the steady point. From the experiments, we can see that the best neighbourhood size of users and items for the MovieLens100K data are respectively 30 and 25.

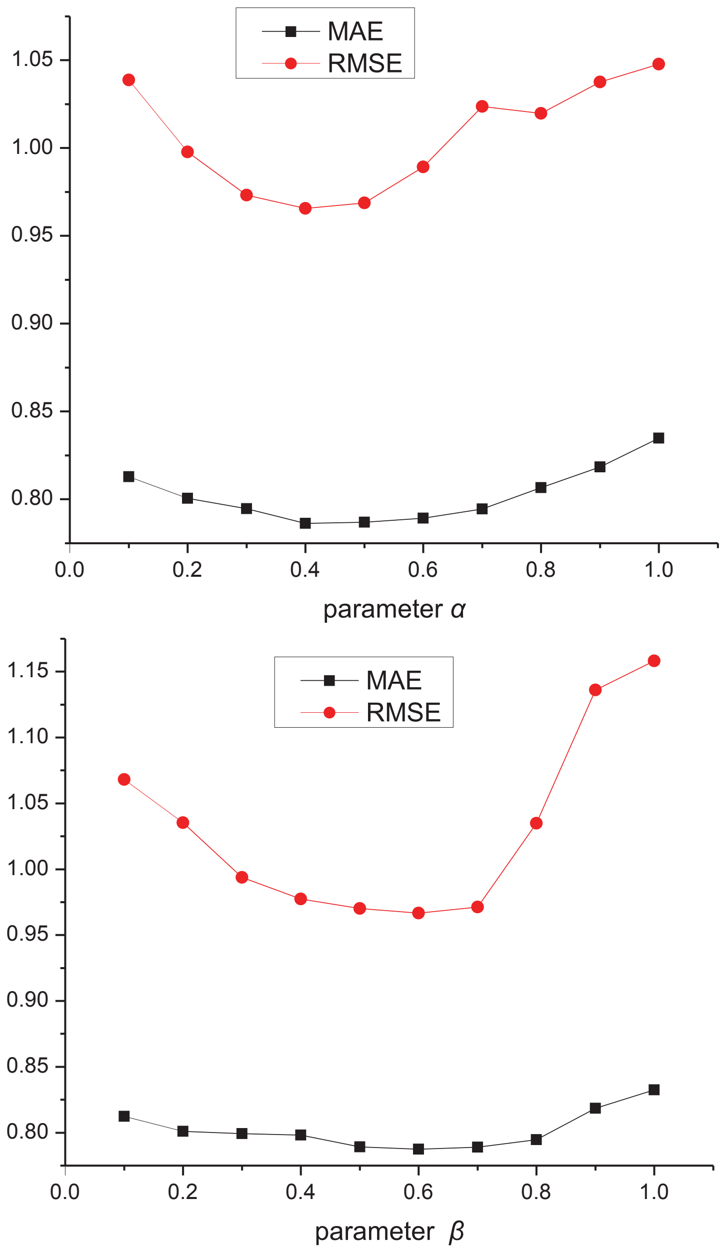

Figure 3.6: Impact of Parameter α and β

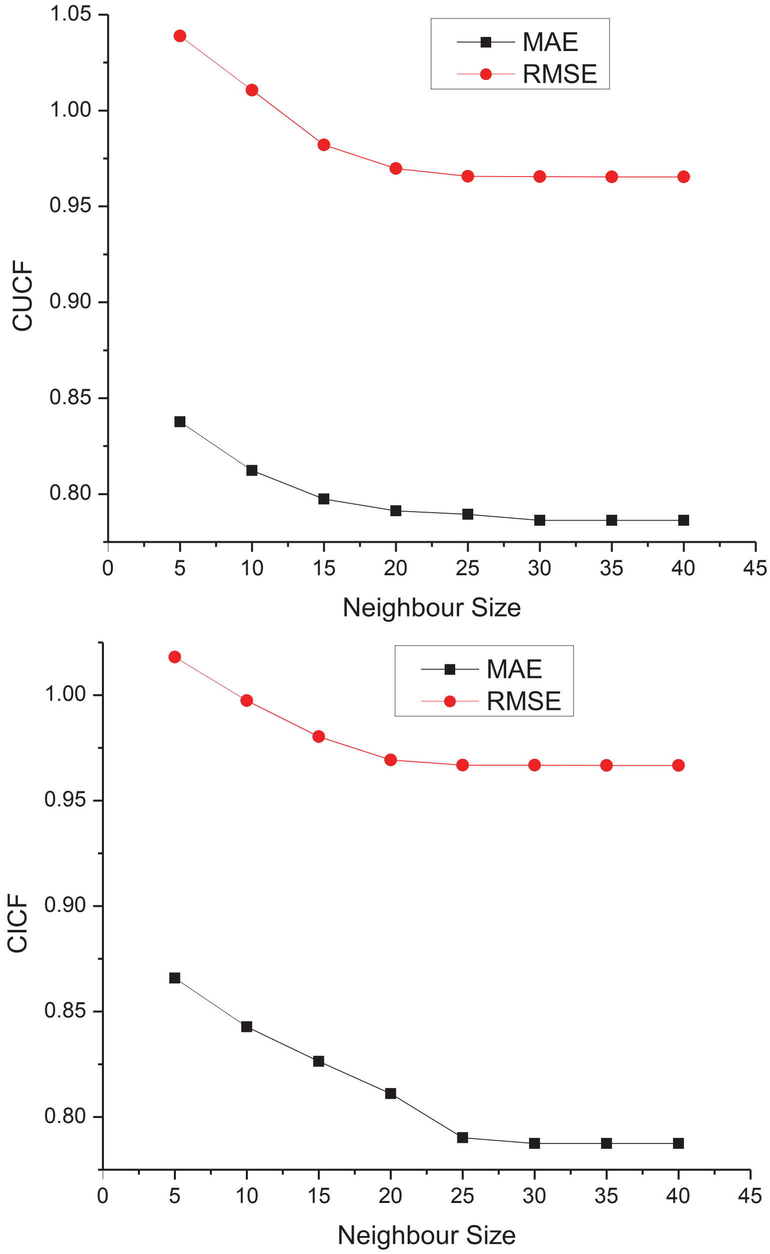

Figure 3.7: Impact of Neighbour Size

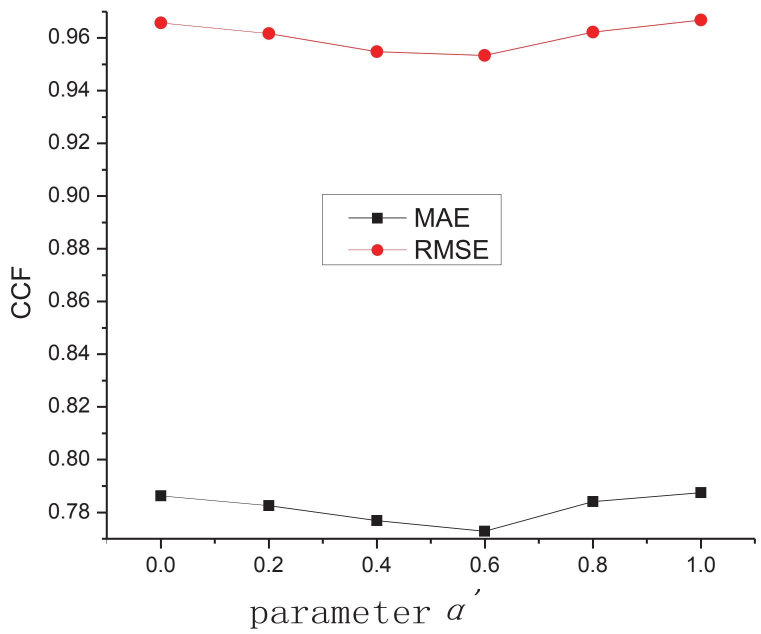

Figure 3.8: Impact of Parameter α'

After the above experiments, we first fix different parameters by the following settings. $\alpha = 0.4$, $\beta = 0.6$, the neighbourhood size of users and items are respectively 30 and 25. We then test the impact of the balanced parameter α' as shown in Fig. 3.8. The experimental results show that the optimum value of the parameter α' is 0.6.

3.3.4 Experimental Analysis

We evaluate the effectiveness of the proposed CUCF, CICF and CCF methods by comparing them with the above benchmark methods on the MovieLens100K data set. Two evaluation metrics MAE and RMSE are considered to evaluate the proposed CUCF, CICF and CCF methods with the results shown in Table 3.4. This comparison result is divided into two parts, the top part contains the results in terms of MAE, and the bottom part contains the results in terms of RMSE. The experiments demonstrate that the proposed coupled neighbourhood-based CF methods CUCF, CICF and CCF outperform the methods which only utilize the user-item rating matrix, such as user-based CF and item-based CF with different similarity methods such as Pearson Correlation and Vector Cosine Similarity for computing the relations between users or items. In addition, we also evaluate our proposed CUCF, CICF and CCF methods by comparing them with another four hybrid CF models $HUCF^p$, $HUCF^c$, $HICF^p$ and $HICF^c$ with Pearson Correlation and Vector Cosine Similarity to infer user or item relations. These four hybrid CF methods not only utilize the rating preferences, but also use different attributes of users or items. The experimental results clearly show that the proposed CUCF, CICF and CCF can perform better than these four hybrid methods. In addition, it is also observed that it is uncertain which method performs better for CUCF and CICF with different settings, while CCF always outperforms CUCF and CICF, which is due to the combined consideration of the user couplings, item couplings and the user-item ratings. In addition, we also notice that the proposed methods outperform the benchmark methods for cold-start situations with only 20% or 40% of the whole data as the training set, which demonstrates that the discussed coupling relations are beneficial for solving cold-start problems and improving the recommendation quality of RS.

Based on the experimental results on the MovieLens100K data set, we can conclude that our proposed methods CUCF, CICF and CCF not only outperform the traditional user-based CF and item-based CF methods which only utilize user-item ratings, but also perform better than the hybrid methods utilizing user-item ratings, users' attributes and items' attributes in terms of MAE and RMSE metrics. In addition, the experiments also demonstrate that the coupling relations between users or items are indeed helpful for overcoming the cold-start challenge. For the proposed three methods, CCF usually performs better than CUCF and CICF because of the complete consideration of the coupling relations in RS.

Table 3.4: Performance on MovieLens100K

Data	$UBCF^p$	$UBCF^c$	$IBCF^p$	$IBCF^c$	$HUCF^p$	$HUCF^c$	$HICF^p$	$HICF^c$	CUCF	CICF	CCF
80%	0.8371	0.8351	1.1954	0.8324	0.8348	0.8337	1.0427	0.8315	0.7863	0.7875	**0.7728**
60%	0.8344	0.8366	1.1474	0.8510	0.8327	0.8335	1.0589	0.8410	0.7885	0.7894	**0.7785**
40%	0.8364	0.8390	1.1140	0.8674	0.8356	0.8374	1.0608	0.8439	0.7964	0.7938	**0.7821**
20%	0.8279	0.8355	1.1382	0.8871	0.8381	0.8363	1.0826	0.8657	0.8176	0.8096	**0.7016**
80%	1.0547	1.0528	1.5785	1.1584	1.0429	1.0413	1.5692	1.1429	0.9657	0.9668	**0.9534**
60%	1.0541	1.0575	1.5398	1.1800	1.0483	1.0462	1.5698	1.1457	0.9681	0.9734	**0.9613**
40%	1.0510	1.0533	1.4824	1.2029	1.0497	1.0481	1.5549	1.1683	0.9754	0.9716	**0.9682**
20%	1.0429	1.0552	1.5129	1.2496	1.0415	1.0497	1.5116	1.1965	0.9943	0.9822	**0.9796**

3.3.5 Discussion

From the above experiments, we demonstrate the effectiveness of our proposed coupled neighbourhood-based CF methods over the classic CF methods which only utilize user-item ratings for recommendation, and their superiority over the hybrid CF methods which also consider users' and items' attributes in RS. Generally, we can conclude that the proposed coupled neighbourhood-CF methods CUCF, CICF and CCF are more effective than the benchmark CF and hybrid approaches regarding MAE and RMSE, due to the strength of coupling relations between users and between items. In addition, the results with a small portion of training data such as 20% or 40% also demonstrate that the proposed methods can partly help to overcome the cold-start challenges, resulting from the consideration of the couplings between users and between items.

3.4 Conclusion

In this chapter, we mainly studied the significant coupling relations between users and between items to improve recommendation qualities. The couplings deeply analysed the intrinsic non-IID relationships between users and between items. We also proposed three coupled neighbourhood-based collaborative filtering methods to integrate the user couplings, item couplings and the user-item ratings together. The experiments conducted on MovieLens100K data set demonstrated the superiority of the proposed CUCF, CICF and CCF methods with the consideration of couplings between users and items. In addition, the experiments also indicate that the user couplings and item couplings are helpful for solving the cold-start challenges for users and items. In future work, we need to explore the heterogeneity challenge of non-IID learning which is still not fully considered to further improve the recommendation quality.

Chapter 4

Coupled Matrix Factorization

4.1 Introduction

RS become increasingly important as they can impact our daily living, online, social, mobile and business activities. Typically, a set of users and items are involved, where each user rates various items according to his/her respective preferences (embodied by preference rates) (Melville & Sindhwani 2010). A new item is then recommended to a user based on the rating behaviors of similar users on existing items. As one of CF models, for example, MF is a latent factor model (Koren 2008) which predicts users' and items' overall structures in the form of latent factors. The preference of the active user to a specific item can be easily estimated by the multiplication of the decomposed user and item latent factors.

Often recommendation algorithms come up with the outcomes based on the aggregated understanding of individual commonality. A rate is then predicted for a new item to a given user or a new user for a given item. An essential problem of the existing recommendation algorithms is to compute the similarities between users/items which can be applied to recommend interesting items to users based on different data, such as historical ratings, users'/items' attributes, social networks, and textual information. For example, CF (Su & Khoshgoftaar 2009) utilizes rating information to compute the similarities between users/items, while social RS (Ma et al. 2008) (Ma et al. 2009) (Ma, Zhou, Liu, Lyu & King 2011) (Yang et al. 2012) (Ye, Liu & Lee 2012) applies users' social friendships to infer users' relations. It is commonly believed that only one type of data, such as rating data or attribute information of users/items, is not informative enough for making satisfactory recommendation. To overcome this problem, many hybrid methods have been proposed considering different types of data to improve recommendation quality. For example, many have tried to estimate the latent factors through incorporating the attributes (Rendle 2010) (Gantner et al. 2010) (Menon & Elkan 2010) (Agarwal & Chen 2009) or topic information of users and items (Agarwal & Chen 2010) (Wang & Blei 2011) (Hu, Koren & Volinsky 2008) for latent factor models.

Generally, leveraging other valuable information into RS is beneficial for enhancing recommendation performance. However, the performance of applying such algorithms for real-time recommendation for specific users and items is still often not very impressive. A very important reason for this is that the existing RS algorithms and systems, such as CF and MF, have been mainly built on the IID context, consequently they may overlook or may not fully capture the intrinsic heterogeneity and couplings. This is a fundamental and critical challenge for the RS community, as the big recommendation data in online, social, mobile and business applications is essentially non-IID. For example, a user's preference may influence his/her friends, users' attributes are often associated with each other via explicit or implicit relationships (Wang et al. 2013*a*) (Wang et al. 2013*b*) (Cao et al. 2012*b*). Actually, there are two important aspects that have not been considered thoroughly regarding non-IID issues, (1) the heterogeneity between users and between items, namely users and items are personalized and thus ratings need to be tailored according to individual characteristics; (2) the coupling relationships between users, between items, and between users and items, namely users and items are coupled and hence ratings need to capture the underlying interactions. These two aspects together essentially bring the recommendation problem to the non-IID context (Cao 2015) (Cao 2014). Therefore in this chapter, we incorporate the user couplings and item couplings as discussed in Chapter 3 into MF to improve recommendation quality.

The incorporation of such couplings into MF is motivated by analysing the interactions between different attributes which disclose the IID assumption. A complete consideration of couplings may provide a practical mean for enhancing the effectiveness of MF. Especially, if we do not have ample rating data, the objective user couplings and item couplings can be utilized to make recommendations. To date, our study is the first work to simultaneously consider user couplings and item couplings based on their objective attributes, and integrate them into the MF model.

The contributions of the chapter are summarized as follows:

- We propose a novel CUMF model by incorporating user couplings and user-item ratings into classic MF model.

- We propose a novel CIMF model with the integration of item couplings and user-item ratings into classic MF model.

- We propose a novel CMF framework by incorporating user couplings, item couplings and users' subjective rating preferences.

- We conduct experiments to evaluate the superiority of couplings and the effectiveness of proposed methods CUMF, CIMF and CMF.

The rest of the chapter is organized as follows. Section 2 details the proposed CUMF model integrating user couplings and rating preferences. Section 3 proposes the CIMF model incorporating item couplings and user-item ratings into MF. Section 4 details the CMF method which integrates user couplings, item couplings and user-item ratings. Experimental results and the analysis are presented in Section 5, followed by the conclusion.

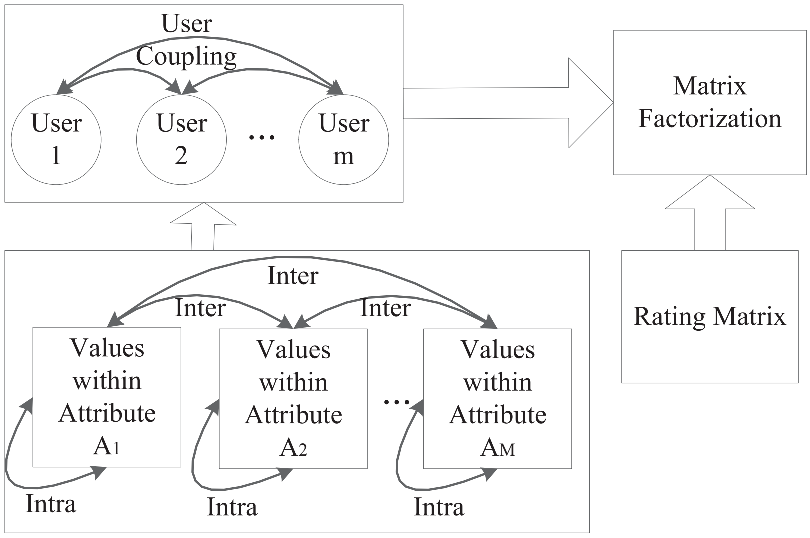

Figure 4.1: Coupled User-based MF Model

4.2 Coupled User-based MF Model

In this section, we introduce the CUMF model as in Fig. 4.1. We first construct the objective attribute spaces of users for computing user couplings which integrate $IaAVS$ and $IeAVS$, as shown in Definitions 2.2 2.6. Then, we incorporate user couplings and users' rating preferences into MF model. The CUMF model has the following characteristics: (1) the user coupling discloses the common IID assumption and considers the real world non-IID situations; (2) the user coupling reflects the relationships between users based on their attributes, which is able to partly overcome the problem of lacking informative rating knowledge; (3) users' rating preferences still play very important roles in the learning model.

4.2.1 Coupled User-based MF

On top of the traditional MF method, we propose a novel CUMF model which takes not only the rating preferences, but also the user couplings into account. The learning procedure is constrained two-fold: the learned rating values should be as close as possible to the observed rating values, and the user's preference should be similar to their neighbourhoods as well, which are derived from their couplings. More specifically, user couplings are incorporated to update the latent factors of users by considering their neighbours' preferences, as shown in Eqn. 4.1, which clearly shows that the latent factor vector P_i of user u_i is impacted by their neighbourhoods.

$$P_i' = \frac{P_i + \sum_{u_j \in N(u_i)} CUS(u_i, u_j) P_j}{\sum_{u_j \in (u_i + N(u_i))} CUS(u_i, u_j)} = \frac{\sum_{u_j \in (u_i + N(u_i))} CUS(u_i, u_j) P_j}{\sum_{u_j \in (u_i + N(u_i))} CUS(u_i, u_j)} \quad (4.1)$$

where $CUS(u_i, u_j)$ is the coupled similarity of users u_i and u_j, $N(u_i)$ represents the user neighbourhood.

After considering user couplings, the latent factors for users and items can be similarly optimized by minimizing the regularized squared error with the following amended objective function as Eqn. 4.2.

$$L = \frac{1}{2} \sum_{(u_i, o_j) \in E} \left(R_{ij} - P_i' Q_j^T \right)^2 + \frac{\lambda}{2} \left(\|P_i\|_F^2 + \|Q_j\|_F^2 \right) \quad (4.2)$$

In the objective function, the user coupling and the rating preferences are integrated. Accordingly, we optimize the above objective function by minimizing L through the gradient decent approach:

$$\frac{\partial L}{\partial P_i} = \sum_{o_j} I_{i,j} (P_i' {Q_j}^T - R_{ij}) Q_j + \lambda P_i \quad (4.3)$$

$$\frac{\partial L}{\partial Q_j} = \sum_{u_i} I_{i,j} (P_i' {Q_j}^T - R_{ij}) P_i' + \lambda Q_j \quad (4.4)$$

where $I_{i,j}$ is an logical function indicating whether the user u_i has rated the item o_j or not.

Furthermore, the updating rules for iteration are derived to learn the latent vectors P_i and Q_j:

$$P_i \leftarrow P_i + \eta((R_{ij} - P_i' {Q_j}^T) Q_j - \lambda P_i) \quad (4.5)$$

Algorithm 4.1 Coupled User-based Matrix Factorization Algorithm

Require: R: the user-item rating matrix.
A: the attributes information of users.
d: the dimension of latent factor vectors.
Z: the number of iterations.
λ: the regularization parameter for MF.
η: the learning rate.

Compute the coupled user similarity matrix using Eqn. 3.1 based on users' attributes
Initiate P^0 and Q^0 with random decimals and j=0
while $j < Z$ or $(L^j - L^{j+1} \leq \epsilon)$ **do**
  for all users and items
    if $R_{ij} \neq 0$ then,
      Compute the gradients P_i and Q_j using Eqn. 4.3 and 4.4
      Update P_u and Q_j by Eqn. 4.5 and 4.6
    end if
  end for
j++
end while
return the user latent factor matrix P and the item latent factor matrix Q

$$Q_j \leftarrow Q_j + \eta((R_{ij} - P_i' {Q_j}^T) P_i' - \lambda Q_j) \tag{4.6}$$

The optimum matrices P and Q can be computed by the above gradient descent approach. Generally, the CUMF model starts by computing user coupling based on the their attributes, then commences an iteration process for optimizing P and Q until convergence, according to Eqn. 4.3 and 4.4. Once P and Q are learned, the ratings for user-item pairs (u_i, o_j) can be easily predicted by Eqn. 2.7. Overall, the CUMF computation process can be described by Algorithm 4.1.

4.2.2 Complexity Analysis

The main computation cost of the CUMF mainly involves learning the latent factor vectors and computing the similarity between users based on their objective attributes. For learning the latent factor vectors, the main time cost is to evaluate the objective function L and the corresponding gradients for users and items. The computational complexity of evaluating objective function is $O(n\bar{r}d + n\bar{t}d)$, where $\bar{r}$ is the average number of ratings per item and $\bar{t}$ is the average number of most similar neighbours per item. The time complexities of evaluating the latent factor vectors for users and items are $O(m\bar{x}d)$ and $O(n\bar{x}d + n\bar{t}^2d)$, respectively, where $\bar{x}$ is the average number of ratings per user. We know that the value of $\bar{r}$ and $\bar{t}$ are usually small because the user-item rating matrix R is sparse, and only the most similar neighbours are selected for the target item. Therefore, the computation of L and latent factor vectors are fast and linear with respect to the number of items n and users m in the user-item rating matrix R. The time cost for computing the coupled similarities between users is $O(D^2w^3)$, where D is the number of user attributes, and w is the maximal number of attribute values for all the attributes in user-attribute information matrix.

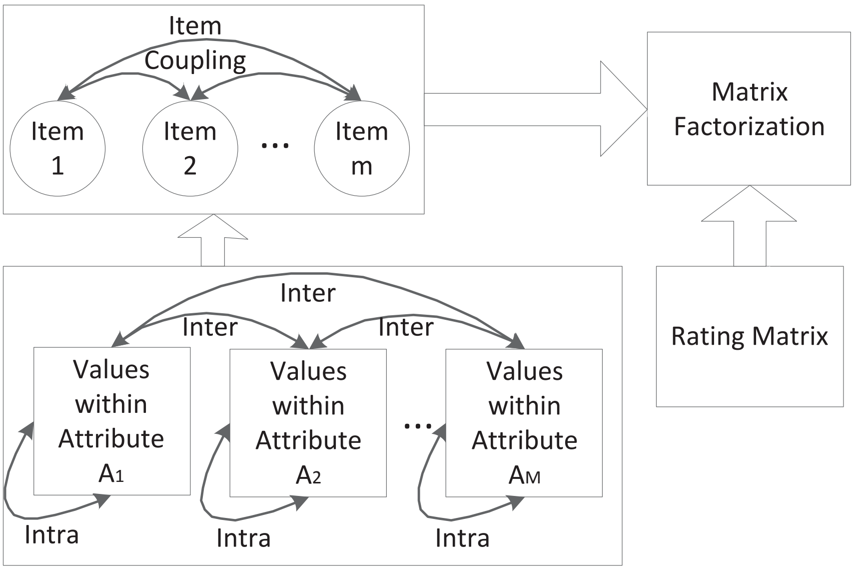

Figure 4.2: Coupled Item-based MF Model

4.3 Coupled Item-based MF Model

In this section, we introduce the CIMF model as in Fig. 4.2. We first construct the objective attribute spaces of items for computing item coupling. Then, we incorporate item coupling and users' rating preferences into MF model. The CIMF model has the following advantages: (1) the item coupling discloses the common IID assumption and consider the real world non-IID characteristic; (2) the item coupling reflects the relationships between items based on their attributes, which is able to partly overcome the problem of lacking informative rating knowledge; (3) users' subjective rating preference is taking the leading role in the learning model.

4.3.1 Coupled Item-based MF

On top of the traditional MF method, we propose a novel CIMF model which takes not only the rating preferences, but also item coupling into account. The learning procedure is constrained two-fold: the learned rating values should be as close as possible to the observed rating values, and the predicted item profiles should be similar to their neighbourhoods as well, which are derived from their coupling relations. More specifically, item coupling is incorporated by adding an additional regularization factor in the optimization step. Then, the computation of the mapping can be similarly optimized by minimizing the regularized squared error. The objective function is amended as Eqn. 4.7.

$$
\begin{aligned}
L = \frac{1}{2} \sum_{(u_i,o_j)\in E} \left(R_{ij} - \hat{R}_{u_i,o_j}\right)^2 + \frac{\lambda}{2}\left(\|Q_j\|^2 + \|P_i\|^2\right) \\
+ \frac{\alpha}{2} \sum_{all(o_j)} \left\| Q_j - \sum_{o_k\in N(o_j)} CIS(o_j,o_k) Q_k \right\|^2
\end{aligned}
\tag{4.7}
$$

In the objective function, the item coupling and users' rating preferences are integrated. Specifically, the first part reflects the subjective rating preferences and the last part reflects the item coupling. This means the users' rating preferences and item coupling act jointly to make recommendations. In addition, another distinct advantage is that, when we do not have ample rating data, it is still possible to make satisfactory recommendations via leveraging item couplings.

We optimize the above objective function by minimizing L through the gradient decent approach:

$$
\frac{\partial L}{\partial P_i} = \sum_{o_j} I_{i,j}(r_m + P_i {Q_j}^T - R_{ij}) Q_j + \lambda P_i \tag{4.8}
$$

$$
\begin{aligned}
&\frac{\partial L}{\partial Q_j} = \sum_{u_i} I_{i,j}(r_m + P_i {Q_j}^T - R_{ij}) P_i + \lambda Q_j + \\
&\alpha(Q_j - \sum_{o_i\in N(o_j)} CIS(o_i,o_j) Q_i) - \\
&\alpha \sum_{o_i : o_j \in N(o_i)} CIS(o_i,o_j)(Q_i - \sum_{o_k\in N(o_i)} CIS(o_i,o_k) Q_k)
\end{aligned}
\tag{4.9}
$$

where $I_{i,j}$ is an logical function indicating whether the user u_i has rated the item o_j or not, $CIS(o_i, o_j)$ is the coupled similarity of items o_i and o_j, $N(o_i)$ represent the item neighbourhood.

The optimum matrices P and Q can be computed by the above gradient descent approach. Generally, the CIMF model starts by computing item coupling based on the objective content, then commences an iteration process for optimizing P and Q until convergence, according to Eqn. 4.8 and 4.9. Once P and Q are learned, the ratings for user-item pairs (u_i, o_j) can be easily predicted by Eqn. 2.7.

4.4 Coupled MF Model

In this section, we first introduce the coupling relations between users and between items, then propose the coupled MF model which incorporates coupling relations and ratings into MF. In addition, we analyse the complexity of the proposed CMF method in the last part of this section.

4.4.1 Coupled MF

On top of the traditional MF method, we propose a novel CMF model, as shown in Fig. 4.3, which takes not only rating preferences, but also user couplings and item couplings into account. In Fig. 4.3, P_i and Q_j respectively represent the latent vector for user u_i and item o_j, and $P_{i_1}, ...P_{i_k}$, $Q_{j_1}, ...Q_{j_k}$ are the latent vectors of the user's and the item's neighbours. We first compute user couplings and item couplings based on their respective attributes. We then incorporate user couplings, item coupling, and users' rating preferences into MF model. Specifically, we employ users' neighbourhoods deprived from user couplings to update the latent factor vectors P, and use items' neighbourhoods deprived from item couplings to update the latent factor vectors Q. The CMF model has the following characteristics: (1) the user and item couplings disclose the common IID assumption and consider the real world non-IID situations; (2) the user and item couplings respec-

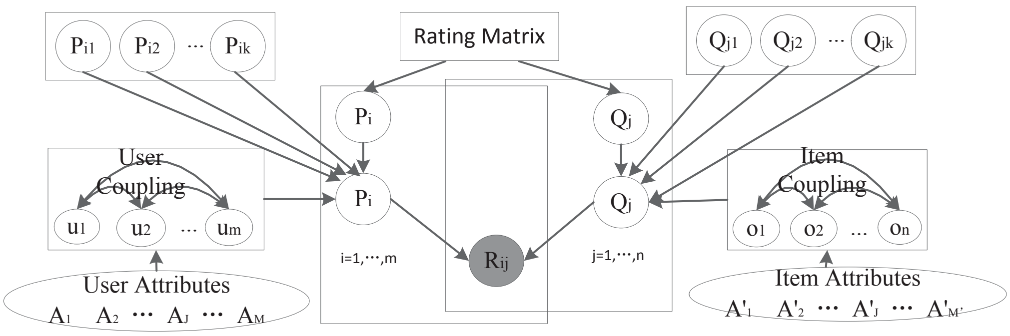

Figure 4.3: Coupled MF Model

tively reflect the coupling relationships between users and between items, which are able to partly overcome the problem of lacking informative rating knowledge; (3) users' rating preferences still play very important roles in the learning model.

Basically, the learned rating values should be as close as possible to the observed rating values for RS, which have been studied well in most existing recommendation algorithms. Besides this, users probably like the items that their neighbourhoods like, and users possibly like item's neighbourhoods as well. More specifically, we incorporate user couplings to derive users' neighbourhoods and update the latent factors of users by considering their neighbours' preferences, as shown in Eqn. 4.10, which clearly shows that the latent factor vector P_i of user u_i is impacted by their neighbourhoods. In addition, we also utilize item couplings to derive items' neighbourhoods and update the latent factors of items by considering items' neighbourhoods, as shown in Eqn. 4.11.

$$P'_i = \frac{P_i + \sum_{u_j \in N(u_i)} CUS(u_i, u_j) P_j}{\sum_{u_j \in (u_i + N(u_i))} CUS(u_i, u_j)} = \frac{\sum_{u_j \in (u_i + N(u_i))} CUS(u_i, u_j) P_j}{\sum_{u_j \in (u_i + N(u_i))} CUS(u_i, u_j)} \quad (4.10)$$

$$Q'_j = \frac{Q_j + \sum_{o_i \in N(o_j)} CIS(o_i, o_j) Q_i}{\sum_{o_i \in (o_j + N(o_j))} CIS(o_i, o_j)} = \frac{\sum_{o_i \in (o_j + N(o_j))} CIS(o_i, o_j) Q_i}{\sum_{o_i \in (o_j + N(o_j))} CIS(o_i, o_j)} \quad (4.11)$$

where $CUS(u_i, u_j)$ is the coupled similarity between users u_i and u_j, $N(u_i)$ represents the user neighbourhood, and $CIS(o_i, o_j)$ is the coupled similarity between items o_i and o_j, $N(o_j)$ represents the neighbours of item o_j.

After considering the user and item couplings, the latent factors for users and items can be similarly optimized by minimizing the regularized squared error with the following amended objective function as Eqn. 4.12.

$$L = \frac{1}{2} \sum_{(u_i,o_j)\in E} \left(R_{ij} - P_i' {Q_j'}^T\right)^2 + \frac{\lambda}{2} \left(\|P_i\|_F^2 + \|Q_j\|_F^2\right) \tag{4.12}$$

In the objective function, user couplings, item couplings and rating preferences are integrated. Accordingly, we optimize the above objective function by minimizing L through the gradient decent approach:

$$\frac{\partial L}{\partial P_i} = \sum_{o_j} I_{i,j}(P_i' {Q_j'}^T - R_{ij})Q_j' + \lambda P_i \tag{4.13}$$

$$\frac{\partial L}{\partial Q_j} = \sum_{u_i} I_{i,j}(P_i' {Q_j'}^T - R_{ij})P_i' + \lambda Q_j \tag{4.14}$$

where $I_{i,j}$ is an logical function indicating whether the user u_i has rated item o_j or not.

Furthermore, the updating rules for iteration are derived to learn the latent vectors P_i and Q_j:

$$P_i \leftarrow P_i + \eta((R_{ij} - P_i' {Q_j'}^T)Q_j' - \lambda P_i) \tag{4.15}$$

$$Q_j \leftarrow Q_j + \eta((R_{ij} - P_i' {Q_j'}^T)P_i' - \lambda Q_j) \tag{4.16}$$

The optimum matrices P and Q can be computed by the above gradient descent approach. Generally, the CMF model starts by computing user coupling and item coupling based on the their attributes, then commences an iteration process for optimizing P and Q until convergence with the consideration of user's and item's neighbourhoods, according to Eqn. 4.13 and 4.14. Once P and Q are learned, the ratings for user-item pairs (u_i, o_j) can be easily predicted by Eqn. 2.7. Overall, the CMF computation process can be described by Algorithm 4.2.

Algorithm 4.2 Coupled Matrix Factorization Algorithm

Require: R: the user-item rating matrix.
A: the attributes information of users.
B: the attributes information of items.
d: the dimension of latent factor vectors.
Z: the number of iterations.
λ: the regularization parameter for MF.
η: the learning rate.

Compute the coupled user similarity matrix using Eqn. 3.1 based on users' attributes
Compute the coupled item similarity matrix using Eqn. 3.2 based on items' attributes
Initiate P^0 and Q^0 with random decimals and z=0
while $z < Z$ or $(L^z - L^{z+1} \leq \epsilon)$ **do**
  for all users and items
    if $R_{ij} \neq 0$ then,
      Compute the gradients P_i and Q_j using Eqn. 4.13 and 4.14
      Update P_i and Q_j by Eqn. 4.15 and 4.16
    end if
  end for
z++
end while
return P: the user latent factor matrix.
Q: the item latent factor matrix.

4.4.2 Complexity Analysis

The main computation cost of the CMF mainly involves learning the latent factor vectors and computing the similarity between users and between items based on their attributes information. For learning the latent factor vectors, the main time cost is to evaluate the objective function L and the corresponding gradients for users and items. The computational complexity of evaluating objective function is $O(m\bar{t}_1 d + n\bar{t}_2 d)$, where $\bar{t}_1$ is the average number of most similar neighbours per user, and $\bar{t}_2$ is the average number of most similar neighbours per item. The time complexities of evaluating the latent factor vectors for users and items are $O(m\bar{x}d+m\bar{t}_1^2 d)$ and $O(n\bar{x}d+n\bar{t}_2^2 d)$, respectively, where $\bar{x}$ is the average number of ratings per user. We know that the value of $\bar{x}$, $\bar{t}_1$ and $\bar{t}_2$ are usually small because the user-item rating matrix R is sparse, and only the most similar neighbours for users and items are selected. Therefore, the computation of L and latent factor vectors are fast and linear with respect to the number of items n and users m in the user-item rating matrix R. The time cost for computing the coupled similarities between users/items is $O(D^2 w^3)$, where D is the number of user/item attributes, and w is the maximal number of attribute values for all the attributes in attribute information matrix for users and items. In practice, to speed up the online learning process, the computation of the coupled user or item similarity can be offline, which does not decrease the efficiency of online recommendation, as the user or item information space is usually fixed or stable for a given recommendation problem.

4.5 Experiments

In this section, we evaluate our proposed models and compare them to the existing approaches respectively, using the MovieLens100K, MovieLens1m and BookCrossing(Ziegler, McNee, Konstan & Lausen 2005) data sets.

4.5.1 Data Sets

The MovieLens data sets were collected by the GroupLens Research Project at the University of Minnesota and have been widely explored in RS research in the last decade. The MovieLens 100K data set consists of 100,000 ratings from 943 users on 1682 movies. The ratings are made on a 5-star scale and each user has rated at least 20 movies. In addition, simple demographic information for the users, such as "age", "gender", "occupation" and "zip code", are also included in the data, which can be used to compute user couplings. Similarly, the genre attributes of items, such as "Crime", "Documentary", "Drama", are used to infer item couplings.

In addition to MovieLens100K data set, the MovieLens1m data set consists of 1,000,209 anonymous ratings of approximately 3,900 movies made by 6,040 MovieLens users who joined MovieLens in 2000. The ratings are made on a 5-star scale and each user has at least 20 ratings. The movies have titles provided by the IMDB (including year of release) and a special "genre" attribute which is applied to compute the item couplings.

Similarly, collected by Cai-Nicolas Ziegler from the BookCrossing community, the BookCrossing data set involves 278,858 users with demographic information providing 1,149,780 ratings on 271,379 books. We remove users and books which have made/received less than 20 ratings, and finally there are 4,213 users and 2,869 books with 172,950 ratings left. The ratings range from 1 to 10 and the users' "country", "state", "city" and "age" attributes tare used to compute user couplings, items' "author", "publication date", "publisher","title" and "url" are used to infer item couplings.

4.5.2 Experimental Settings

We perform five-fold cross validation in our experiments for the MovieLens100K, MovieLens1m and BookCrossing data sets. We first split the original data into five equal sized samples, then we respectively keep a single sample of the five samples as the test set, and the remaining four samples are used as training set. In this way, the original data is converted to a five-fold data set with each fold 80% as the training set and 20% as the test set. Then the cross-validation process is repeated five times on the sampled data for each fold. Finally, the estimation on the whole data can be averaged from the five results for each fold. Here we use Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) as the evaluation metrics:

$$RMSE = \sqrt{\frac{\sum_{(u_i,o_j)|R_{test}} \left(R_{ij} - \hat{R}_{ij}\right)^2}{|R_{test}|}} \tag{4.17}$$

$$MAE = \frac{\sum_{(u_i,o_j)|R_{test}} \left|R_{ij} - \hat{R}_{ij}\right|}{|R_{test}|} \tag{4.18}$$

where R_{test} is the set of all pairs (u_i, o_j) in the test set.

4.5.3 Experimental Evaluation for CUMF

To evaluate the performance of our proposed CUMF, we consider six different models. These models include:

- Probabilistic matrix factorization (PMF) approach (Salakhutdinov & Mnih 2008). This model assume that users/items are independent with each other and ignore the similarities between users and between items.

- Singular value decomposition (RSVD) (Alter, Brown & Botstein 2000) model. This model also assume that users/items are independent with each other and ignore the similarities.

- SoRec model (Ma, King & Lyu 2011). This method takes users' similarities which are computed from users' ratings into consideration.

- Implicit social matrix factorization (ISMF) (Ma 2013) model. This model incorporates users' similarities and items' similarities which are computed from the rating preferences.

- PSMF model. This model incorporates the relations between users based on users' attributes instead of rating preferences. The relations between users are defined as the similarities between users computed by Pearson Correlation Coefficient with the same incorporation method as our proposed CUMF model.
- JSMF model. Similar to PSMF, this model also incorporates the similarities between users on attribute information instead of rating preferences but computed by Jaccard similarity .

Superiority over MF Baselines

We respectively evaluate the effectiveness of the proposed CUMF model by comparing it with the above methods. In Table 4.1, different latent dimensions regarding MAE and RMSE metrics on the MovieLens100K data are considered to evaluate the proposed CUMF model. In general, the experiments demonstrate that our proposed CUMF outperforms the methods which only utilize the user-item rating matrix such as PMF, RSVD, SoRec and ISMF. Specifically, when the latent dimension is set to 10 and 50, in terms of MAE, our proposed CUMF can reach improvements to 2.26% and 1.41% compared with the PMF method. Regarding RMSE, the improvements reach up to 5.19% and 5.70%. Similarly, CUMF can respectively improve by 3.32%, 2.42% regarding MAE, and 4.14%, 4.28% regarding RMSE over the RSVD approach. In addition, CUMF improves SoRec model by 1.12%, 0.44% regarding MAE, and 3.17%, 4.13% with regard to RMSE. In comparison with ISMF model, CUMF can also improve by 0.89%, 0.33% regarding MAE and 2.98%, 3.64% with regard to RMSE. In addition to the above comparisons, we also evaluate our model by comparing with another two models PSMF and JSMF which utilize two different data sources including rating preferences and users' attribute information. The experimental result clearly shows that CUMF can respectively improve PSMF by 1.02% and 1.33% regarding MAE, and 2.91%, 3.08% regarding RMSE for different dimensions 10 and 50. CUMF also improves JSMF by 1.13%, 0.2% with regard to MAE, and

2.96%, 3.59% regarding RMSE for dimensions 10 and 50.

Similarly, we depict the effectiveness comparisons with respect to different methods on the BookCrossing data set in Table 4.1. We can clearly see that our proposed CUMF method outperforms all its counterparts in terms of MAE and RMSE. Specifically, when the latent dimension is set to 10 and 50, in terms of MAE, our proposed CUMF can reach improvements to 3.60% compared with the PMF method. Regarding RMSE, CUMF also improves slightly by 1.12% and 2.11%. Similarly, CUMF can respectively improve by 3.59%, 3.63% regarding MAE, and 3.0%, 4.07% regarding RMSE over the RSVD approach. In addition, CUMF improves SoRec model by 3.53%, 3.87% regarding MAE, and 2.07%, 1.87% with regard to RMSE. In comparison with ISMF model, CUMF can also improve by 3.32% regarding MAE and 0.81%, 1.74% with regard to RMSE for dimensions 10 and 50. In addition, the comparison results with PSMF and JSMF show that CUMF can respectively improve PSMF by 3.42% and 2.17% regarding MAE, and 0.89%, 2.4% regarding RMSE for different dimensions 10 and 50. CUMF also improves JSMF by 2.58%, 2.09% with regard to MAE, and 0.93%, 2.81% regarding RMSE for dimensions 10 and 50.

Based on the experimental results on the MovieLens100K and BookCrossing data sets, we can conclude that our CUMF method not only outperforms the traditional MF methods which only utilize user-item ratings, such as PMF, SVD, SoRec, ISMF, but also performs better than the hybrid methods utilizing user-item ratings and users' attributes, such as PSMF and JSMF in terms of MAE and RMSE metrics.

Superiority for Cold-start Users

Besides the above experiments, we also compared the effectiveness of solving the problem of cold-start users. To select the cold-start users, we filter the data sets by selecting the users who have rated items less than 50 times as cold-start users. To evaluate the effectiveness of our proposed approach on cold-start users, we compare the RMSE and MAE results with other

Table 4.1: Performance of CUMF on MovieLens100K and BookCrossing Data Sets

Data Set	Dim	Metrics	PMF	ISMF	RSVD	SoRec	PSMF	JSMF	CUMF
MovieLens100K	50D	MAE	0.7536	0.7428	0.7642	0.7439	0.7528	0.7415	**0.7395**
		RMSE	0.9789	0.9583	0.9647	0.9632	0.9527	0.9578	**0.9219**
	10D	MAE	0.7662	0.7525	0.7768	0.7548	0.7538	0.7549	**0.7436**
		RMSE	0.9857	0.9636	0.9752	0.9655	0.9629	0.9634	**0.9338**
BookCrossing	50D	MAE	1.5128	1.5100	1.5131	1.5155	1.4985	1.4977	**1.4768**
		RMSE	3.7452	3.7415	3.7648	3.7428	3.7481	3.7522	**3.7241**
	10D	MAE	1.5135	1.5107	1.5134	1.5128	1.5117	1.5033	**1.4775**
		RMSE	3.7483	3.7440	3.7659	3.7566	3.7448	3.7452	**3.7359**

Table 4.2: Performance of CUMF for Cold-start Users

Data Set	Metrics	PMF	ISMF	RSVD	SoRec	PSMF	JSMF	CUMF
MovieLens100K	MAE	0.9746	0.9578	0.9684	0.9598	0.9476	0.9593	**0.9372**
	RMSE	1.2385	1.1437	1.2258	1.1697	1.1329	1.1286	**1.1139**
BookCrossing	MAE	1.5459	1.5410	1.5447	1.5428	1.5539	1.5426	**1.4981**
	RMSE	3.7862	3.7824	3.7893	3.7857	3.8122	3.7944	**3.7624**

benchmark methods as showed in Table 4.2. The dimension parameter is fixed as d=50 for the MovieLens100K and BookCrossing data. The comparison results clearly demonstrate that the proposed CUMF method not only outperforms the benchmark methods which only utilize user-item ratings, but also performs better than the models which also consider the users' attributes. The improvements are resulted from considering the coupling relations between users.

Discussion

From the above experiments, we demonstrate the effectiveness of our proposed CUMF model over the MF methods which only utilize user-item ratings for recommendation, and the superiority over the MF methods which also consider users' attributes in RS. Generally, we can conclude that the proposed CUMF is more effective than the benchmark MF approaches regarding MAE and RMSE for different latent dimensions, due to the strength of coupling relations between users.

4.5.4 Experimental Evaluation for CIMF

To evaluate the performance of our proposed CIMF, we consider five baseline approaches based on a user-item rating matrix: (1) the basic probabilistic matrix factorization (PMF) approach (Salakhutdinov & Mnih 2008); (2) the singular value decomposition (RSVD) (Alter et al. 2000) method; (3) the implicit social matrix factorization (ISMF) (Ma 2013) model which incorporates implicit social relationships between users and between items; (4) user-based CF (UBCF) (Su & Khoshgoftaar 2009); and (5) item-based CF (IBCF) (Deshpande & Karypis 2004).

The above five baselines only consider users' rating preferences on items but ignore item coupling. In order to completely evaluate our method, we also compare the following three models PSMF, CSMF and JSMF, which respectively augment MF using the same strategy as in Eqn. 4.7 with the

Pearson Correlation Coefficient, and the Cosine and Jaccard similarity measures to compute item coupling relationships, based on the objective attribute information.

Superiority over MF Methods

It is well known that MF methods are popular and successful in RS, hence, in this experiment, we compare our proposed CIMF model with the existing MF methods. In Table 4.3, different latent dimensions regarding MAE and RMSE metrics on the MovieLens1m data are considered to evaluate the proposed CIMF model. In general, the experiments demonstrate that our proposed CIMF outperforms the other three MF baselines. Specifically, when the latent dimension is set to 10, 50 and 100, in terms of MAE, our proposed CIMF can reach improvements of 16.33%, 17.53% and 27.85% compared to the PMF method. Regarding RMSE, the improvements reach up to 48.35%, 55.00% and 70.53%. Similarly, CIMF can respectively improve by 6.02%, 9.89%, 20.74% regarding MAE, and 26.76%, 32.84%, 57.76% regarding RMSE over the RSVD approach. In addition to basic comparisons, we also compare our CIMF model with the latest research outcome ISMF which utilizes the implicit relationships between users and items based on the rating matrix by Pearson similarity. From the experimental result, we can see that CIMF can respectively improve by 11.55%, 10.89% and 21.23% regarding MAE, and by 41.07%, 35.52%, 58.60% regarding RMSE for different dimensions, 10, 50 and 100.

Similarly, we depict the effectiveness comparisons with respect to different methods on the BookCrossing data set in Table 4.3. We can clearly see that our proposed CIMF method outperforms all its counterparts in terms of MAE and RMSE. Specifically, when the latent dimension is set to 10, 50 and 100, in terms of MAE, our proposed CIMF can reach improvements of 3.72%, 3.65% and 3.64% compared to the PMF method. Regarding RMSE, CIMF also improves slightly by 0.85%, 0.80% and 0.69%. Similarly, CIMF can respectively improve by 3.71%, 3.68%, 3.68% regarding MAE, and 2.61%,

Table 4.3: MF Comparisons on MovieLens1m and BookCrossing Data Sets for CIMF

Data	Dim	Metrics	PMF (Improve)	ISMF (Improve)	RSVD (Improve)	CIMF
MovieLens1m	100D	MAE	1.1787(27.85%)	1.1125 (21.23%)	1.1076 (20.74%)	**0.9002**
		RMSE	1.7111 (70.53%)	1.5918 (58.60%)	1.5834 (57.76%)	**1.0058**
	50D	MAE	1.1852 (17.53%)	1.1188 (10.89%)	1.1088 (9.89%)	**1.0099**
		RMSE	1.8051 (55.00%)	1.6103 (35.52%)	1.5835 (32.84%)	**1.2551**
	10D	MAE	1.2129 (16.33%)	1.1651 (11.55%)	1.1098 (6.02%)	**1.0496**
		RMSE	1.8022 (48.35%)	1.7294 (41.07%)	1.5863 (26.76%)	**1.3187**
BookCrossing	100D	MAE	1.5127 (3.64%)	1.5102 (3.39%)	1.5131 (3.68%)	**1.4763**
		RMSE	3.7455 (0.69%)	3.7397 (0.11%)	3.7646 (2.60%)	**3.7386**
	50D	MAE	1.5128 (3.65%)	1.5100 (3.37%)	1.5131 (3.68%)	**1.4763**
		RMSE	3.7452 (0.80%)	3.7415 (0.43%)	3.7648 (2.76%)	**3.7372**
	10D	MAE	1.5135 (3.72%)	1.5107 (3.44%)	1.5134 (3.71%)	**1.4763**
		RMSE	3.7483 (0.85%)	3.7440 (0.42%)	3.7659 (2.61%)	**3.7398**

2.76%, 2.60% regarding RMSE over the RSVD approach. In addition, from the experimental results, compared to the latest research outcome ISMF, we can see that CIMF can respectively improve by 3.44%, 3.37%, 3.39% regarding MAE, and by 0.42%, 0.43%, 0.11% regarding RMSE for different dimensions, 10, 50 and 100.

Based on the experimental results on the MovieLens1m and BookCrossing data sets, we can conclude that our CIMF method not only outperforms the traditional MF methods PMF and SVD, but also performs better than the state-of-the-art model ISMF in terms of MAE and RMSE metrics. Furthermore, the prominent improvements are the result of considering item couplings.

Table 4.4: CF Comparisons on MovieLens1m and BookCrossing Data Sets for CIMF

Data Set	Metrics	UBCF (Improve)	IBCF (Improve)	CIMF
MovieLens	MAE	0.9027 (0.25%)	0.9220 (2.18%)	**0.9002**
	RMSE	1.0022 (-0.36%)	1.1958 (19.00%)	**1.0058**
BookCrossing	MAE	1.8064 (33.01%)	1.7865 (31.02%)	**1.4763**
	RMSE	3.9847 (24.61%)	3.9283 (18.97%)	**3.7386**

Superiority over CF Methods

In addition to the MF methods, we also compare our proposed CIMF model with two different CF methods, UBCF and IBCF. In this experiment, we fix the latent dimension to 100 for our proposed CIMF model. On the MovieLens1m data set, the results in Table 4.4 indicate that CIMF can respectively improve by 0.25% and 2.18% regarding MAE. Regarding RMSE, CIMF can improve by 19.00% compared with IBCF, and slightly decreases by 0.36% compared with UBCF but it is still comparable. Similarly, on the BookCrossing data set, the results show that CIMF can respectively reach improvements of 33.01%, 31.02% regarding MAE, and 24.61%, 18.97% regarding RMSE compared to UBCF and IBCF. Therefore, this experiment clearly demonstrates that our proposed CIMF performs better than the traditional CF methods, UBCF and IBCF. The consideration of item couplings in RS contribute to these improvements.

Superiority over Hybrid Methods

In order to demonstrate the effectiveness of our proposed model, we compare it with three different hybrid methods, PSMF, CSMF and JSMF, which respectively augment MF with the Pearson Correlation Coefficient, and the Cosine and Jaccard similarity measures.

From the results shown in Fig. 4.4 on the MovieLens1m data set, generally we can clearly see that the coupled similarity method CIMF largely outperforms the three different comparisons with PSMF, CSMF and JSMF in terms of MAE and RMSE. Specifically, on the MovieLens1m data set for three different dimensions 10, 50 and 100, CIMF can respectively reach an improvement of 47.87%, 44.72% and 68.72% regarding RMSE compared to PSMF. In terms of MAE, CIMF also can increase by 15.79%, 16.14% and 26.48% compared to PSMF. Similarly, compared to CSMF on the MovieLens1m data set, CIMF can improve by 47.22%, 43.57% and 67.14% regarding RMSE, while for MAE, the improvement can reach up to 15.49%, 15.38% and 26.03%. Additionally, CIMF also performs better than JSMF, the respective improvements regarding RMSE being 74.18%, 70.02% and 93.47%, while regarding MAE, CIMF also improves by 23.23%, 23.63% and 34.48%.

On the BookCrossing data set, the results in Fig. 4.5 also indicate that the coupled similarity method CIMF constantly performs better than corresponding comparison methods regarding RMSE and MAE. Specifically, for three different dimensions 10, 50 and 100, CIMF can respectively reach an improvement of 7.91%, 8.10% and 8.01% regarding RMSE compared to PSMF. In terms of MAE, CIMF also can slightly improve by 2.25%, 2.21% and 2.22%, compared to PSMF. Similarly, compared to CSMF, on the MovieLens data set, CIMF can improve by 44.82%, 44.24% and 43.13% regarding RMSE, while for MAE, the improvement can reach up to 19.31%, 19.62%

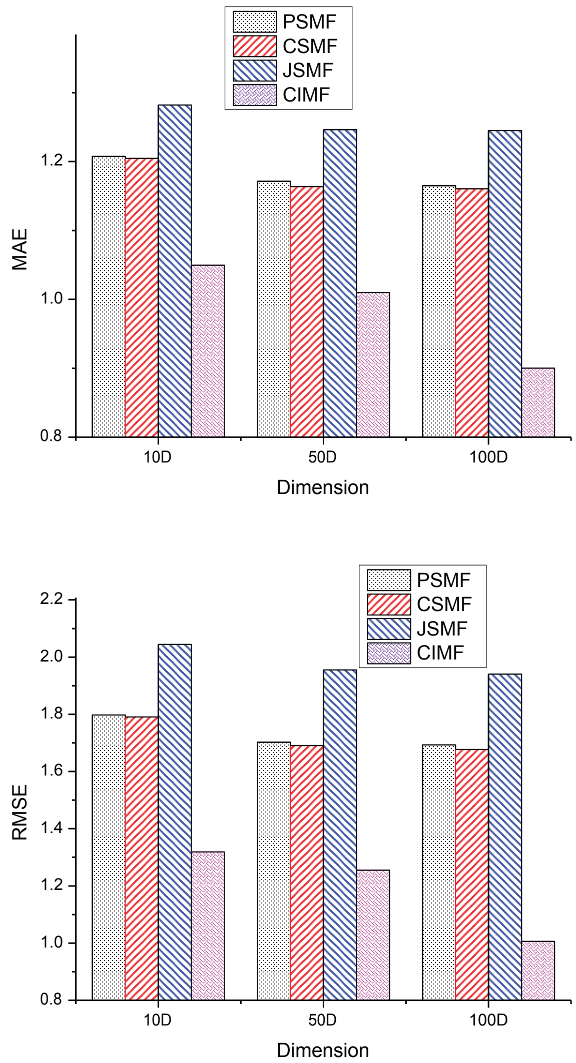

Figure 4.4: Performance of CIMF over Hybrid Methods on MovieLens1m

and 19.75%. Additionally, CIMF also performs better than JSMF, the respective improvements regarding RMSE being 8.15%, 8.4% and 8.22%, while regarding MAE, CIMF also slightly improves by 2.22%, 2.18% and 2.16%.

From these comparisons, we can conclude that our proposed CIMF model is more effective than the three different hybrid methods, PSMF, CSMF and JSMF.

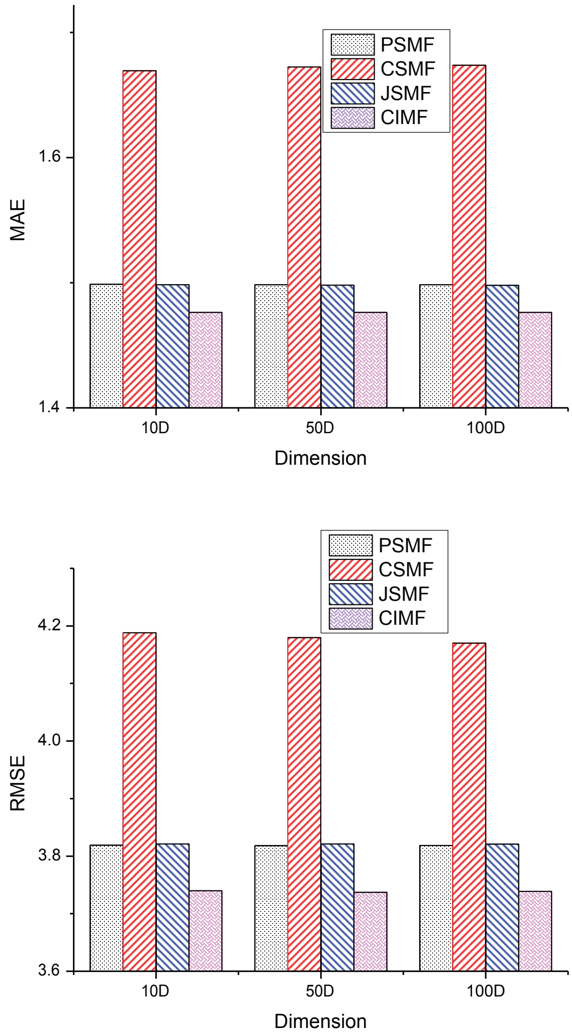

Figure 4.5: Performance of CIMF over Hybrid Methods on BookCrossing

Impact of Parameters

Parameter α controls the influence of the item couplings. Bigger value of α in the objective function of Eqn. 4.7 indicates higher impact of the item coupling. To select the optimum parameters, we depict the MAE changing trends of CIMF methods when α is ranged in [0,1]. Fig. 4.6 and 4.7 shows the impacts of parameter α when neighbourhood size for items is respectively set

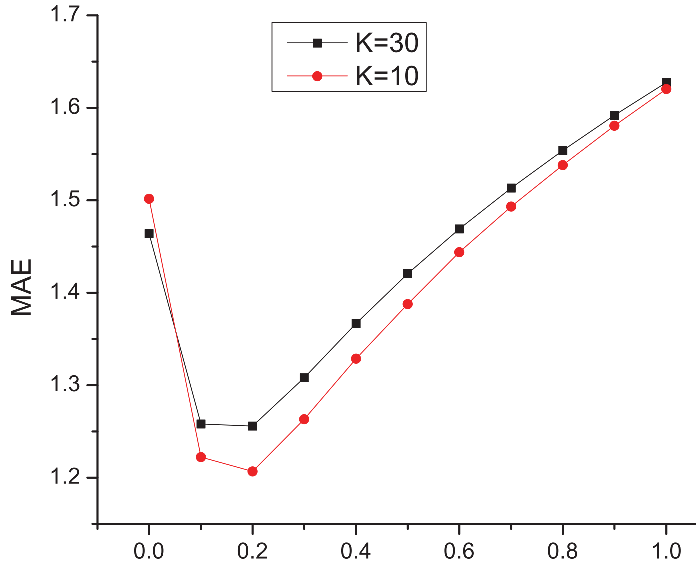

Figure 4.6: Parameter α of CIMF on MovieLens1m

to 10 or 30 on the Movielens1m and BookCrossing data sets. Experimental results show that α=0.2 are proper values for Movielens1m, while α=1.0 are more suitable for the BookCrossing data set. Additionally, in this chapter, for computing item coupling, we set the parameter $\gamma_k = \frac{1}{n-1}$ which controls the weight of attribute A_k for items, n is the number of attributes.

Discussion

From the above experiments, we demonstrate the impact of our coupled similarity and the superiority over MF and CF methods. Also, we notice that sometimes on the MovieLens1m CF or SVD-based approaches perform more satisfactorily than PMF method, which violates the well-known research findings from others that PMF usually performs better than SVD or CF methods. This result inspires us to meditate on the data characteristics, specifically we believe that the performance of the recommendation methods might be closely dependent on the data characteristics in addition to approaches themselves. For example, when the data sets largely follow

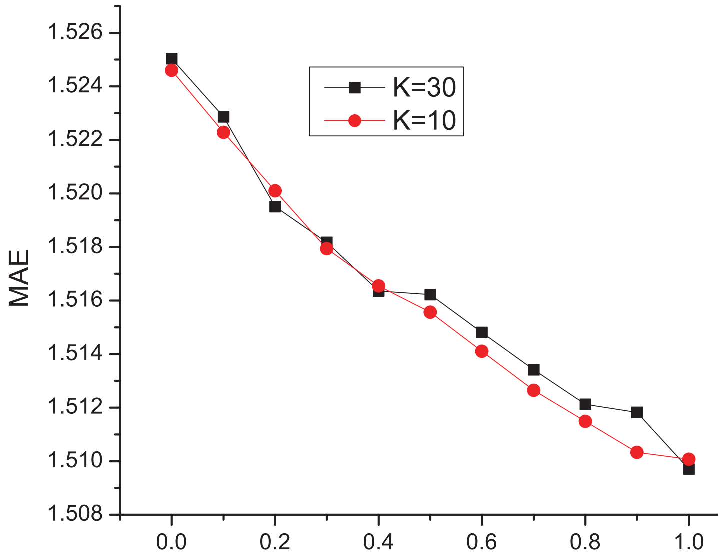

Figure 4.7: Parameter α of CIMF on BookCrossing

Gaussian distribution, PMF method which assumes that the rating matrix is Gaussian distribution may be more suitable for recommendation. Otherwise, other methods may get a better result. Therefore, we may need to incorporate the data characteristics into our model, we believe that the performance can be further enhanced through this consideration.

4.5.5 Experimental Evaluation for CMF

In this section, we evaluate our proposed CMF model and compare it to the existing approaches respectively, using the MovieLens100K and BookCrossing data sets. To evaluate the performance of our proposed CMF, we consider six different models. These models include:

- Probabilistic matrix factorization (PMF) approach (Salakhutdinov & Mnih 2008). This model assume that users/items are independent with each other and ignore the similarities between users and between items.

- Singular value decomposition (RSVD) (Alter et al. 2000) model. This model also assume that users/items are independent with each other and ignore the similarities.

- SoRec model (Ma, King & Lyu 2011). This method takes users' similarities which are computed from users' ratings into consideration.

- Implicit social matrix factorization (ISMF) (Ma 2013) model. This model incorporates users' similarities and items' similarities which are computed from the rating preferences.

- PSMF model. This model incorporates the relations between users and between items based on users'/items' attributes instead of rating preferences. The relations between users/items are defined as the similarities between users/items computed by Pearson Correlation Coefficient with the same incorporation method as our proposed CMF model.

- JSMF model. Similar to PSMF, this model also incorporates the similarities between users and between items on attribute information instead of rating preferences but computed by Jaccard similarity .

Superiority over MF Baselines

We respectively evaluate the effectiveness of the proposed CMF model by comparing it with the above methods. In Table 4.5, different latent dimensions regarding MAE and RMSE metrics on the MovieLens100K are considered to evaluate the proposed CMF model. Generally, the experiments demonstrate that our proposed CMF outperforms the methods which only utilize the user-item rating matrix, such as PMF, RSVD, SoRec and ISMF. Specifically, when the latent dimension is set to 10 and 50, in terms of MAE, our proposed CMF can reach improvements to 4.33% and 4.10% compared to the PMF method. Regarding RMSE, the improvements reach up to 7.30% and 7.55%. Similarly, CMF can respectively improve by 5.39%, 5.16% regarding MAE, and 6.25%, 6.13% regarding RMSE over the RSVD approach. Besides, CMF improves SoRec model by 3.19%, 3.13% regarding MAE, and 5.28%, 5.98% with regard to RMSE. In comparison with ISMF model, CMF can also improve by 2.96%, 3.02% regarding MAE and 5.09%, 5.49% with regard to RMSE. In addition to the above comparisons, we also evaluate our model compared to another two models PSMF and JSMF which utilize three different data sources including rating preferences, users' attributes and items' attributes. The experimental result clearly shows that CMF can respectively improve PSMF by 2.17% and 2.89% regarding MAE, and 4.11%, 4.24% regarding RMSE for different dimensions 10 and 50. CMF also improves JSMF by 2.22%, 1.82% with regard to MAE, and 4.5%, 4.31% regarding RMSE for dimensions 10 and 50.

Similarly, we depict the effectiveness comparisons with respect to different methods on the BookCrossing data set in Table 4.5. We can clearly see that our proposed CMF method outperforms all its counterparts in terms of MAE and RMSE. Specifically, when the latent dimension is set to 10 and 50, in terms of MAE, our proposed CMF can reach improvements to 5.77%, 5.89% compared to the PMF method. Regarding RMSE, CMF also improves by 3.35% and 3.96%. Similarly, CMF can respectively improve by 5.76%, 5.92% regarding MAE, and 5.11%, 5.92% regarding RMSE over the RSVD approach. Besides, CMF improves SoRec model by 5.7%, 6.16% regarding MAE, and 4.18%, 3.72% with regard to RMSE. In comparison with ISMF model, CMF can also improve by 5.49%, 5.61% regarding MAE and 2.92%, 3.59% with regard to RMSE for dimensions 10 and 50. In addition, the comparison results with PSMF and JSMF show that CMF can respectively improve PSMF by 4.74% and 3.18% regarding MAE, and 1.97%, 3.40% regarding RMSE for different dimensions 10 and 50. CMF also improves JSMF by 4.04%, 3.23% with regard to MAE, and 1.74%, 3.85% regarding RMSE for dimensions 10 and 50.

Based on the experimental results on the MovieLens100K and BookCrossing data sets, we can conclude that our CMF method not only outperforms the traditional MF methods which only utilize the user-item ratings, such as PMF, SVD, SoRec, ISMF, but also performs better than the hybrid methods utilizing the user-item ratings, users' attributes and items' attributes, such

Table 4.5: Performance of CMF on MovieLens100K and BookCrossing Data Sets

Data Set	Dim	Metrics	PMF	ISMF	RSVD	SoRec	PSMF	JSMF	CMF
MovieLens100K	50D	MAE	0.7536	0.7428	0.7642	0.7439	0.7415	0.7308	**0.7126**
		RMSE	0.9789	0.9583	0.9647	0.9632	0.9458	0.9465	**0.9034**
	10D	MAE	0.7662	0.7525	0.7768	0.7548	0.7446	0.7451	**0.7229**
		RMSE	0.9857	0.9636	0.9752	0.9655	0.9538	0.9577	**0.9127**
BookCrossing	50D	MAE	1.5128	1.5100	1.5131	1.5155	1.4857	1.4862	**1.4539**
		RMSE	3.7452	3.7415	3.7648	3.7428	3.7396	3.7441	**3.7056**
	10D	MAE	1.5135	1.5107	1.5134	1.5128	1.5032	1.4962	**1.4558**
		RMSE	3.7483	3.7440	3.7659	3.7566	3.7345	3.7322	**3.7148**

as PSMF and JSMF, in terms of MAE and RMSE metrics.

Superiority for Cold-start Users and Items

Besides the above experiments, we also evaluate the effectiveness of solving the cold-start problem for users and items. To select the cold-start users, we filter the data sets by selecting the users who have rated items less than 50 times as cold-start users. Similarly, the items that are rated less than 50 times as cold-start items. To evaluate the effectiveness of our proposed approach on cold-start users and items, we compare the RMSE and MAE results with other benchmark methods as showed in Table 4.6 and Table 4.7. The dimension parameter is fixed as d=50 for the MovieLens100K and BookCrossing data. The comparison results for cold-start users and items clearly demonstrate that the proposed CMF method not only outperforms the benchmark methods which only utilize the user-item ratings, but also performs better than the models which also consider the users' attributes and items' attributes. The improvements are resulted from considering the coupling relations between users and between items.

Discussion

From the above experiments, we demonstrate the effectiveness of our proposed CMF model over the MF methods which only utilize the user-item ratings for recommendation, and the superiority over the MF methods which also consider users' and items' attributes in RS. Generally, we can conclude that the proposed CMF is more effective than the benchmark MF approaches regarding MAE and RMSE for different latent dimensions, due to the strength of coupling relations between users and between items. In addition, the proposed CMF method also outperforms over the baselines for cold-start users and items, which indicate that the couplings between users and between items are beneficial for solving the cold-start challenge.

Table 4.6: Performance of CMF for Cold-start Users

Data Set	Metrics	PMF	ISMF	RSVD	SoRec	PSMF	JSMF	CMF
MovieLens100K	MAE	0.9746	0.9578	0.9684	0.9598	0.9468	0.9452	**0.9264**
	RMSE	1.2385	1.1437	1.2258	1.1697	1.1329	1.1286	**1.1139**
BookCrossing	MAE	1.5459	1.5410	1.5447	1.5428	1.5481	1.5344	**1.4839**
	RMSE	3.7862	3.7824	3.7893	3.7857	3.8122	3.7944	**3.7624**

Table 4.7: Performance of CMF for Cold-start Items

Data Set	Metrics	PMF	ISMF	RSVD	SoRec	PSMF	JSMF	CMF
MovieLens100K	MAE	0.9574	0.9381	0.9437	0.9398	0.9376	0.9393	**0.9172**
	RMSE	1.2037	1.1193	1.1945	1.1438	1.1146	1.1013	**1.0944**
BookCrossing	MAE	1.5394	1.5285	1.5352	1.5243	1.5160	1.5148	**1.4732**
	RMSE	3.7637	3.7619	3.7648	3.7672	3.7848	3.7754	**3.7487**

4.6 Conclusion

In this chapter, we mainly studied the significant coupling relations between users and between items to improve recommendation quality. The couplings deeply analysed the intrinsic non-IID relationships between users and between items. Based on the coupling analysis, we first proposed a novel CUMF model integrating user couplings and user-item ratings into the MF model, then proposed the CIMF model incorporating item couplings and user-item ratings. Besides, we also proposed a CMF model to consider user couplings, item couplings and user-item ratings into MF. The experiments conducted on open data sets demonstrated the superiority of the proposed method with the consideration of user couplings and item couplings. In addition, the experiments also indicate that user couplings and item couplings are helpful for solving the cold-start challenges for users and items.

Chapter 5

Two-level Matrix Factorization

5.1 Introduction

RS become increasingly important as they deeply involve our daily living, online, social, mobile and business activities. Typically, a set of users and items are involved, where each user u rates various items according to his/her respective preferences (embodied by preference rates) (Melville & Sindhwani 2010). A new rate or item is then recommended to a user based on the rating behaviors of similar users on existing items.

Many researchers tried to incorporate side-information, such as social relationships (Ma et al. 2008) (Yang et al. 2012), item relations (Rendle 2010) (Gantner et al. 2010) (Menon & Elkan 2010) (Agarwal & Chen 2009), topic distribution (Agarwal & Chen 2010) (Wang & Blei 2011), coupling relations (Li et al. 2013) (Li et al. 2014) (Li et al. 2015a) into MF, for more precisely predicting the latent factors of users and items. However, the existing MF algorithms have still not fully captured the intrinsic relations between items and users, especially the semantic couplings. Therefore, we integrate the semantic couplings between items into MF in this chapter to improve the performance of RS. Specifically, we first deeply analyse items' semantic couplings based on textual MF, then incorporate the semantic couplings into the upper level of MF on rating matrix.

Table 5.1: A Toy Example

	i_1	i_2	i_3	i_4
u_1	1	3	5	4
u_2	4	2	1	5
u_3	-	2	-	4

Item	Title	Introduction
i_1	The Godfather	The aging patriarch of an organized ...
i_2	Goodfellas	Henry Hill and his friends work their way ...
i_3	Vertigo	A retired San Francisco detective ...
i_4	N or NW	Correspondence between young lovers ...

To illustrate the semantic couplings between items in RS, we give a toy example in Table 5.1. There is a rating matrix consisting of three users and four movies with textual information. Most existing CF methods utilize the rating matrix for recommendation but ignore the textual context of items. However, when the rating matrix is very sparse, the textual titles and introductions of items may also contribute to improve recommendation quality. Specifically, we can infer the movies' semantic coupling relations based on their textual introductions by text mining techniques. Intuitively, the existing text mining technique, such as TF*IDF (Wu et al. 2008), can be applied to infer the relationships between items. However this method ignores the semantic couplings between items. As we know that latent factor models, such as Latent Semantic Analysis (LSA) (Landauer et al. 1998), Probabilistic Latent Semantic Analysis (PLSA) (Hofmann 1999), Latent Dirichlet Allocation (LDA) (Blei et al. 2003), are beneficial for identifying textual semantic relations. However, a fact that missing words of a sentence are irrelevant to the sentence was not considered in these latent factor models. Recently, a novel Weighted Textual Matrix Factorization (WTMF) method (Guo & Diab 2012*b*) was proposed to compute the semantic similarities between sentences and achieved better performance. However, this valuable semantic analysis had not been applied in RS, which greatly motivates us to model the textual semantic couplings into MF approach of RS.

The contributions of the chapter are summarized as follows:

- We apply the novel WTMF model to infer the semantic couplings between items based on the textual context.
- We propose a two-level Matrix Factorization (TLMF) recommendation model by accommodating item textual semantic couplings and users' subjective rating preferences.
- We conduct experiments to evaluate the superiority of the proposed TLMF model.

The rest of the chapter is organized as follows. Section 2 first states the notations, then introduces the WTMF method and semantic couplings for items, followed by the details and complexity analysis of the proposed TLMF which incorporates the textual semantic couplings into MF. In Section 3, experimental results and analysis are presented, followed by the conclusion in the last Section.

5.2 Two-level Matrix Factorization

In this section, we mainly introduce the two-level MF method, as shown in Fig. 5.1. TLMF first computes the semantic couplings between items according to textual MF. Then items' semantic couplings and users' rating preferences are incorporated into TLMF model. The TLMF model is advantageous in the following aspects: (1) the semantic couplings between items are analysed by the lower level MF, which are able to remedy the problem of lacking informative rating knowledge, further to improve the quality of recommendation; (2) user's subjective rating preference is also incorporated in the learning model by the upper level MF. To help understand the proposed TLMF model, we below respectively introduce the notations, WTMF method, and items' semantic couplings, followed by the integration of the semantic couplings into the proposed TLMF model and complexity analysis.

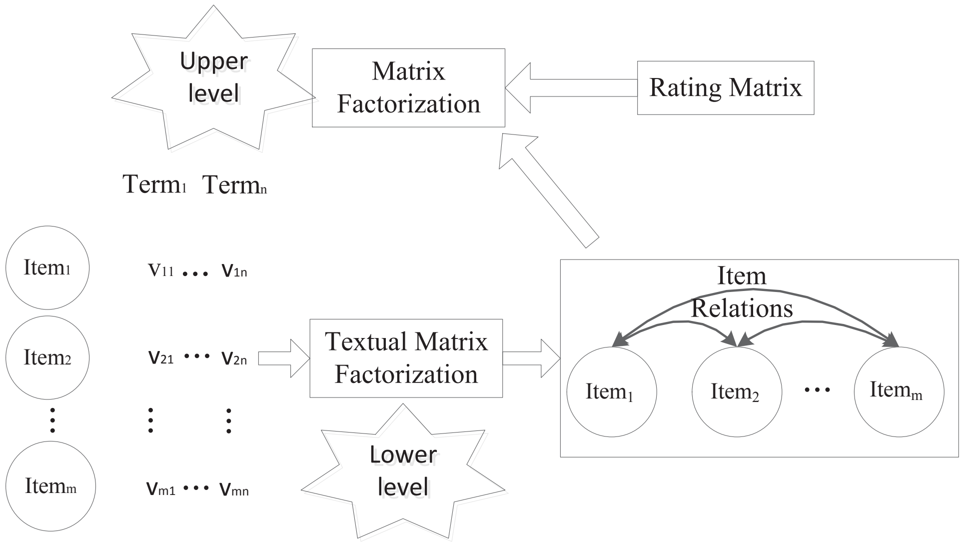

Figure 5.1: Two-level Matrix Factorization Model

5.2.1 Notations

A large number of user and item sets with textual information can be organized by a triple $C =< C_U, C_O, h >$, where $C_U = U = \{u_1, u_2, ..., u_m\}$ is a nonempty finite set of users, $C_O =< O, X >$, $O = \{o_1, o_2, ..., o_n\}$ is a nonempty finite set of items, X is the term document matrix for item's textual information. In the triple $C =< C_U, C_O, h >$, $h(u_i, o_j) = R_{ij}$ expresses the subjective rating preference on item o_j for user u_i. Through the mapping function h, user rating preferences on items are then converted into a user-item matrix R, with m rows and n columns. Each element R_{ij} of R represents the rating given by user u_i on item o_j. And the term document matrix X derived from textual information of items can be decomposed into two latent factor matrices $A^{M \times K}$ and $B^{N \times K}$ which respectively represent word and sentence, where K is the dimension of the latent factors for words and sentences, M is the number of terms, and N is the number of sentences in the textual corpus. Then the semantic similarities between items can be computed from the decomposed latent factor vectors of sentences, which are further incorporated into MF model of RS.

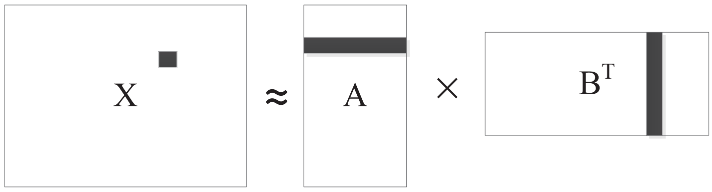

Figure 5.2: Weighted Textual Matrix Factorization

5.2.2 Weighted Textual Matrix Factorization

The WTMF (Guo & Diab 2012*b*) method has been successfully applied in many natural language processing tasks, such as sentence similarity computation (Guo & Diab 2012*a*) and linkage analysis between tweets and news (Guo, Li, Ji & Diab 2013), achieving state-of-the-art unsupervised performance. A big contribution of WTMF is that the missing words are modelled to greatly enrich the features of sentences and short text. According to Guo and Diab, the missing words of a sentence are defined as all the vocabulary of the training corpus minus the observed words in a sentence. Intuitively, we know that a sentence should not be related to the missing words of the sentence (we call it missing words principle), which is often ignored by most sentence similarity measures. WTMF thus models sentences by textual MF method which considers missing words as additional constraints for the semantics of sentences.

Similar to the above traditional MF based on rating matrix in RS, textual MF is the applied on term document matrix. Specifically, the textual corpus is first represented by a term document matrix X with the TF*IDF values of words in each cell, where the rows of X are words and columns are sentences. Matrix X is then similarly approximated by the product of a $M \times K$ matrix A and a $N \times K$ matrix B. Accordingly, each sentence s_j is represented by a K dimensional latent factor vector B_j , and each word w_i is generalized by latent factor vector A_i. Therefore, the inner product of a word vector A_i and a sentence vector B_j is to approximate the cell X_{ij} (shaded part in Fig. 5.2). Thereby, the constraint of the inner product of A_i and B_j to be close to 0 ensures the missing words are modelled in line with the above principle.

To overcome the over influence of missing words, a small weight w_m is assigned for 0 cells in matrix X, since 99% of the cells are 0 values, possibly diminishing the effect of observed words. The model parameters (vectors in A and B) are optimized by minimizing the objective function:

$$L_l = \min_{A,B} \frac{1}{2} \sum_i \sum_j W_{ij} \big(X_{i,j} - A_i B_j^T \big)^2 + \frac{\gamma}{2} \left(\|A\|_F^2 + \|B\|_F^2 \right) \tag{5.1}$$

$$W_{ij} = \begin{cases} 1, & if \ X_{ij} \neq 0, \\ w_m, & if \ X_{ij} = 0. \end{cases} \tag{5.2}$$

where γ is a free regularization factor, and the weight matrix W defines a weight for each cell in X.

5.2.3 Semantic Couplings for Items

The semantic couplings between items are derived from their textual information by the mentioned WTMF approach. According to Eqn. 5.1, the best latent factor matrices A and B need to be optimized first. Then the semantic couplings between items o_i and o_j can be computed from that between sentences s_i and s_j by cosine similarity of vectors B_i and B_j, which is given in Eqn. 5.3.

$$S(o_i, o_j) = S(s_i, s_j) = cos < B_i, B_j >= \frac{B_i . B_j}{\|B_i\| \|B_j\|} \tag{5.3}$$

To optimize the latent vectors, A and B are first randomly initialized, then can be computed iteratively by the following equations (Nati & Jaakkola 2003):

$$A_i = (B \tilde{W}^{(i)} B^T + \lambda I)^{-1} B \tilde{W}^{(i)} X_{i,.}^T \tag{5.4}$$

$$B_j = (A \tilde{W}^{(j)} A^T + \lambda I)^{-1} A \tilde{W}^{(i)} X_{.,j}^T \tag{5.5}$$

where A_i is a K-dimension latent semantic vector profile for word w_i; similarly, B_j is the K-dimension vector profile for sentence s_j.

5.2.4 Two-level MF Model

MF approaches have been recognized as the main stream in RS through a latent topic projection learning model. In this work, we attempt to incorporate the discussed semantic couplings between items into a MF scheme.

As shown in Fig. 5.1, TLMF not only takes the rating matrix, but also the semantic couplings between items into account. All these aspects should be accommodated into a unified learning model. The learning procedure is constrained by three-fold: the learned rating values should be as close as possible to the observed rating values, the predicted item profiles should be similar to their neighbourhoods as well, which are derived from their semantic couplings. In addition, the lower level textual MF models relations between words and sentences by accommodating the impact of missing words, which improves the performance of semantic analysis. Specifically, in order to incorporate the couplings between items, we add two additional regularization factors in the optimization step. Then the computation of the mapping can be similarly optimized by minimizing the regularized squared error. The objective function is given as Eqn. 5.6.

$$\begin{aligned}L_u = \frac{1}{2}\sum_{(u_i,o_j)\in E}\left(R_{ij}-\hat{R}_{ij}\right)^2 + \frac{\lambda}{2}\left(\|Q\|_F^2+\|P\|_F^2\right) + \\ \frac{\alpha}{2}\sum_{o_j}\left(Q_j-\sum_{o_i}S(o_i,o_j)Q_i\right)\left(Q_j-\sum_{o_i}S(o_i,o_j)Q_i\right)^T\end{aligned} \tag{5.6}$$

In the objective function, semantic couplings between items and users' rating preferences are integrated. This means the users' rating preferences and items' semantic couplings act together to make recommendations. In addition, another distinct advantage is that, when we do not have ample rating data, it is still possible to make satisfactory recommendations via leveraging the semantic couplings between items. Similar to Eqn. 2.10 and 2.11, we optimize the above objective function by minimizing L_u through the gradient decent approach:

$$\frac{\partial L_u}{\partial P_i} = \sum_{o_j} I_{i,j}(P_i {Q_j}^T - R_{ij})Q_j + \lambda P_i \tag{5.7}$$

$$\begin{aligned}
\frac{\partial L_u}{\partial Q_j} &= \sum_{u_i} I_{i,j}(P_i {Q_j}^T - R_{ij})P_i + \lambda Q_j + \\
&\alpha(Q_j - \sum_{o_i \in \mathbb{N}(o_j)} S(o_i, o_j)Q_i) - \\
&\alpha \sum_{o_i} S(o_i, o_j)(Q_i - \sum_{o_k \in \mathbb{N}(o_i)} S(o_i, o_k)Q_k)
\end{aligned} \tag{5.8}$$

where $I_{i,j}$ is an logical function indicating whether the user u_i has rated item o_i or not. $S(o_i, o_j)$ is the semantic similarity between items o_i and o_j. $\mathbb{N}(o_i)$ represents the item neighbourhood.

Furthermore, the updating rules for iteration are derived to learn the latent vectors P_i and Q_j:

$$P_i \leftarrow P_i + \eta((R_{ij} - P_i {Q_j}^T)Q_j - \lambda P_i) \tag{5.9}$$

$$\begin{aligned}
&Q_j \leftarrow Q_j + \eta((R_{ij} - P_i {Q_j}^T)P_i - \lambda Q_j - \\
&\alpha(Q_j - \sum_{o_i \in \mathbb{N}(o_j)} S(o_i, o_j)Q_i) + \\
&\alpha \sum_{o_i} S(o_i, o_j)(Q_i - \sum_{o_k \in \mathbb{N}(o_i)} S(o_i, o_k)Q_k))
\end{aligned} \tag{5.10}$$

The optimum matrices P and Q can be computed by the above gradient descent approach. Generally, the TLMF model starts by computing item couplings based on the textual content by the lower level TMF method, then commences an iteration process for optimizing P and Q until convergence, according to Eqn. 5.9 and 5.10. Once P and Q are learned, the ratings for user-item pairs (u_i, o_j) can be easily predicted by Eqn. 2.7. Overall, the TLMF computation process can be described by Algorithm 5.1.

5.2.5 Complexity Analysis

The main computation cost of the TLMF mainly involves learning the latent factor vectors and computing the similarity between items with the textual

Algorithm 5.1 Two-level Matrix Factorization Algorithm

Require: R: the user-item rating matrix.
d: the dimension of latent feature vector on upper level.
K: the dimension of latent feature vector on lower level.
T: the textual corpus of items.
Z: the number of iterations.
λ: the regularization parameter for upper level MF.
γ: the regularization parameter for lower level MF.
η: the learning rate.
w_m: the parameter of missing words.

Build term document matrix with TF*IDF values based on textual corpus of items T
Initiate A^0 and B^0 with random decimals and j=0
while $j < Z$ or $(L_l^j - L_l^{j+1} \leq \epsilon)$ **do**
  for all words and sentences
    Compute and update A and B by Eqn. 5.4 and 5.5
  end for
j++
end while
Compute similarity matrix S by Eqn. 5.3
Initiate P^0 and Q^0 with random decimals and j=0
while $j < Z$ or $(L_u^j - L_u^{j+1} \leq \epsilon)$ **do**
  for all users and items
    Compute and update P and Q by Eqn. 5.7 to 5.10
  end for
j++
end while
return the user latent feature matrix P and the item latent feature matrix Q

MF method. For learning the latent factor vectors, the main time cost is to evaluate the objective function L_u and the corresponding gradients for users and items. The computational complexity of evaluating objective function is $O(n\bar{r}d + n\bar{t}d)$, where $\bar{r}$ is the average number of ratings per item and $\bar{t}$ is the average number of most similar neighbours per item. The time complexities of evaluating the latent factor vectors for users and items are $O(m\bar{x}d)$ and $O(n\bar{x}d + n\bar{t}^2d)$, respectively, where $\bar{x}$ is the average number of ratings per user. We know that the value of $\bar{r}$ and $\bar{t}$ are usually small because the user-item rating matrix R is sparse, and only the most similar neighbours are selected for the target item. Therefore, the computation of L_u and latent factor vectors are fast and linear with respect to the number of items n and users m in the user-item rating matrix R.

For computing the textual similarity by the textual MF, the main time cost is to learn the latent factor vectors A and B for textual terms and sentences. Similar to the complexity analysis for the upper level MF, the computation complexity for computing objective function L_l is $O(NaK)$, where N is the number of sentences in the textual corpus, K is the dimension of the latent vectors, a is the average number of terms per sentence. The time complexity for evaluating the latent factor vectors A and B are respectively $O(MaK)$ and $O(NaK)$. Therefore, the evaluation of the lower level MF is dependent on the number of terms for sentences which is impacted by the sentence length. That means the longer of the textual sentences the more time consuming for inferring the similarities between items. In practice, to speed up the online learning process, the computation process of textual similarity between items can be offline, which can also be very beneficial for the recommendation community. In addition, online recommendation strategy can be practically implemented on big data platforms such as Hadoop or Spark.

Table 5.2: Basic Statistics for Data Sets

Statistics	MovieLens1m	BookCrossing
Num. of Ratings	1,000,209	1,149,780
Num. of Users	6040	278,858
Num. of Items	3076	271,379
Sparsity	94.62%	99.99%

5.3 Experiments and Results

In this section, we evaluate our proposed model and compare it with the existing approaches respectively using MovieLens[1] and BookCrossing[2] data sets.

5.3.1 Data Sets

The MovieLens data set has been widely explored in RS research in the last decade. The MovieLens 1M data set consists of 1,000,209 anonymous ratings of approximately 3,076 movies made by 6,040 MovieLens users who joined MovieLens in 2000. The ratings are made on a 5-star scale and each user has at least 20 ratings. The movies have titles provided by the IMDB (including year of release), we also extract relevant descriptions about the movies from wikipedia.

Similarly, collected by Cai-Nicolas Ziegler from the BookCrossing community, the BookCrossing data set involves 278,858 users with demographic information providing 1,149,780 ratings on 271,379 books. The ratings range from 1 to 10 and the books' titles are used to form the item semantic couplings. The basic statistics of the MovieLens1m and BookCrossing data sets are shown in Table 5.2.

[1] www.movielens.org

[2] www.bookcrossing.com

5.3.2 Experimental Settings

We perform five-fold cross validation in our experiments for both MovieLens1m and BookCrossing data sets. We first split the original data into five equal sized samples, then we respectively keep a single sample of the five samples as the test set, and the remaining four samples are used as training set. In this way, the original data is converted to a five-fold data set with each fold 80% as the training set and 20% as the test set. Then the cross-validation process is repeated five times on the sampled data for each fold. Finally, the estimation on the whole data can be averaged from the five results for each fold. Here we use Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) as the evaluation metrics, which are defined as follow:

$$RMSE = \sqrt{\frac{\sum_{(u_i,o_j)|R_{test}} \left(R_{ij} - \hat{R}_{ij}\right)^2}{|R_{test}|}} \tag{5.11}$$

$$MAE = \frac{\sum_{(u_i,o_j)|R_{test}} \left|R_{ij} - \hat{R}_{ij}\right|}{|R_{test}|} \tag{5.12}$$

where R_{test} is the set of all pairs (u_i, o_j) in the test set.

To evaluate the performance of our proposed TLMF we consider eight baseline approaches:

- The basic probabilistic matrix factorization (PMF) approach (Salakhutdinov & Mnih 2008);

- Singular value decomposition (RSVD) (Alter et al. 2000) is a factorization method to decompose the rating matrix;

- Implicit social matrix factorization (ISMF) (Ma 2013) is an unified model which incorporates implicit social relationships between users and between items computed by Pearson similarity based on the user-item rating matrix;

- User-based CF (UBCF) (Su & Khoshgoftaar 2009) first computes users' similarity by Pearson Correlation on the rating matrix, then recommends relevant items to the given user according to the users who have strong relationships;

- Item-based CF (IBCF) (Deshpande & Karypis 2004) first considers items' similarity by Pearson Correlation on the rating matrix, then recommends relevant items which have strong relationships with the given user's interested items

- MF model with edit distance (EDMF) applies edit distance measure to compute the similarities between items based on textual information, with the incorporation of item relations into MF as objective function shown in Enq. 5.6

- MF model with term frequency (TFMF) first directly computes the similarities between items using term frequency vectors after transforming items' textual information into a term-document matrix with the value of term frequency, then integrates the item relations into MF as Enq. 5.6;

- MF model with TFIDF (TIMF) directly calculates the similarities between items using term TFIDF vectors, followed by the same integration approach with TFMF.

5.3.3 Experiments and Discussions

We respectively evaluate the effectiveness of our TLMF model in comparison with the above baselines on the MovieLens1m and BookCrossing data sets.

Superiority over CF Methods

We first compare our proposed TLMF model with two different CF methods UBCF and IBCF. In this experiment, we fix the latent dimension to 50 for TLMF model. On the MovieLens1m, the results in Table 5.3 indicate

Table 5.3: CF Comparisons on MovieLens1m and BookCrossing for TLMF

Data Set	Metrics	UBCF (Improve)	IBCF (Improve)	TLMF
MovieLens1m	MAE	0.9027 (23.47%)	0.9220 (25.4%)	**0.668**
	RMSE	1.0022 (14.92%)	1.1958 (34.28%)	**0.853**
BookCrossing	MAE	1.8064 (35.25%)	1.7865 (33.26%)	**1.4539**
	RMSE	3.9847 (25.82%)	3.9283 (20.21%)	**3.7262**

that TLMF can respectively improve 23.47% and 25.4% regarding MAE, and 14.92% and 34.28% in terms of RMSE. Similarly compared to UBCF and IBCF, on the BookCrossing data set, the results show that the TLMF can reach improvements respectively 35.25% and 33.26% regarding MAE, and 25.82% and 20.21% regarding RMSE. Therefore, this experiment clearly demonstrates that our proposed TLMF performs better than UBCF and IBCF methods on both data sets. The improvements are contributed by the consideration of item semantic couplings in RS.

Superiority over MF Methods

In addition to the comparisons with pure CF methods, we also evaluate our proposed two-level MF model compared to the above MF series, with the results shown in Table 5.4, which clearly demonstrate our proposed TLMF outperforms other MF methods which only utilize user item rating matrix such as PMF, RSVD and ISMF. Specifically, TLMF respectively improves PMF by 1.7% and 5.89% on the MovieLens1m and BookCrossing data sets with MAE metric and latent dimension as 50. With the same setting, the MAE performance of ISMF can be improved by 1.1% and 5.61% respectively on the MovieLens1m and BookCrossing data. TLMF achieves better than RSVD as well, which is improved by 1.0% and 5.92%, respectively for the MovieLens1m and BookCrossing data.

We then compare the proposed TLMF model with other three MF models EDMF, TFMF and TIMF which also consider item couplings based on their textual information. From Table 5.4, we can clearly see that TLMF achieves

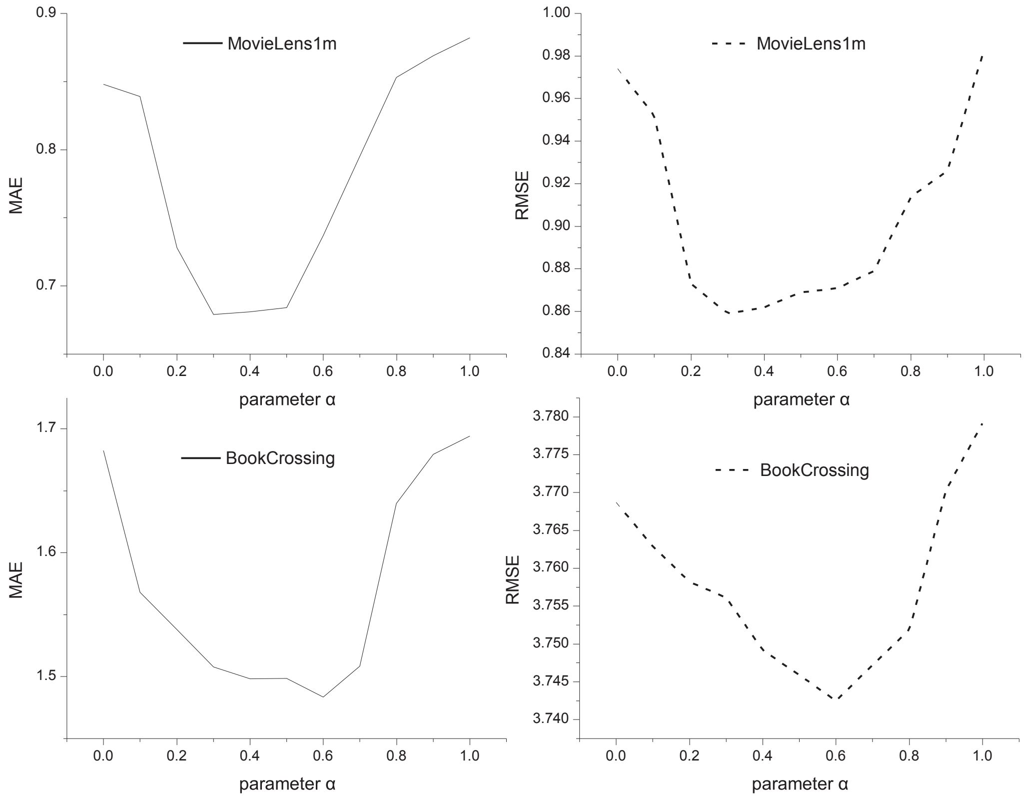

Figure 5.3: Impact of Parameter α on MovieLens1m and BookCrossing for TLMF

better performance than the three models. In detail, TLMF respectively improves EDMF by 1.5% and 7% regarding MAE evaluation metric on the MovieLens1m and BookCrossing data sets when latent dimension is 50. In the same setting, TFMF can also be improved by 0.9%, 6.44%, as well for TIMF by 0.7%, 2.34%.

Impact of Parameter α

Parameter α controls the influence of semantic couplings between items in TLMF model. A bigger value of α in the objective function of Eqn. 5.6 indicates a higher impact of the items' couplings. To select the optimum

Table 5.4: MF Comparisons on MovieLens1m and BookCrossing for TLMF

Data Set	Dim	Metrics	PMF	ISMF	RSVD	EDMF	TFMF	TIMF	TLMF
MovieLens1m	50D	MAE	0.685	0.679	0.678	0.683	0.677	0.675	**0.668**
		RMSE	0.865	0.859	0.862	0.863	0.861	0.864	**0.853**
	10D	MAE	0.688	0.685	0.683	0.697	0.689	0.684	**0.676**
		RMSE	0.871	0.864	0.867	0.869	0.863	0.862	**0.859**
BookCrossing	50D	MAE	1.5128	1.5100	1.5131	1.5239	1.5183	1.4773	**1.4539**
		RMSE	3.7452	3.7415	3.7648	3.7894	3.7850	3.7426	**3.7262**
	10D	MAE	1.5135	1.5107	1.5134	1.5164	1.5128	1.5085	**1.4836**
		RMSE	3.7483	3.7440	3.7659	3.7591	3.7483	3.7426	**3.7425**

Table 5.5: MF Comparisons on Cold-start Items for TLMF

Data Set	Metrics	PMF	ISMF	RSVD	EDMF	TFMF	TIMF	TLMF
MovieLens1m	MAE	0.697	0.693	0.692	0.694	0.687	0.689	**0.671**
	RMSE	0.878	0.874	0.876	0.878	0.875	0.872	**0.868**
BookCrossing	MAE	1.5332	1.5318	1.5338	1.5445	1.5385	1.4978	**1.4751**
	RMSE	3.7655	3.7618	3.7852	3.8098	3.8055	3.7632	**3.7470**

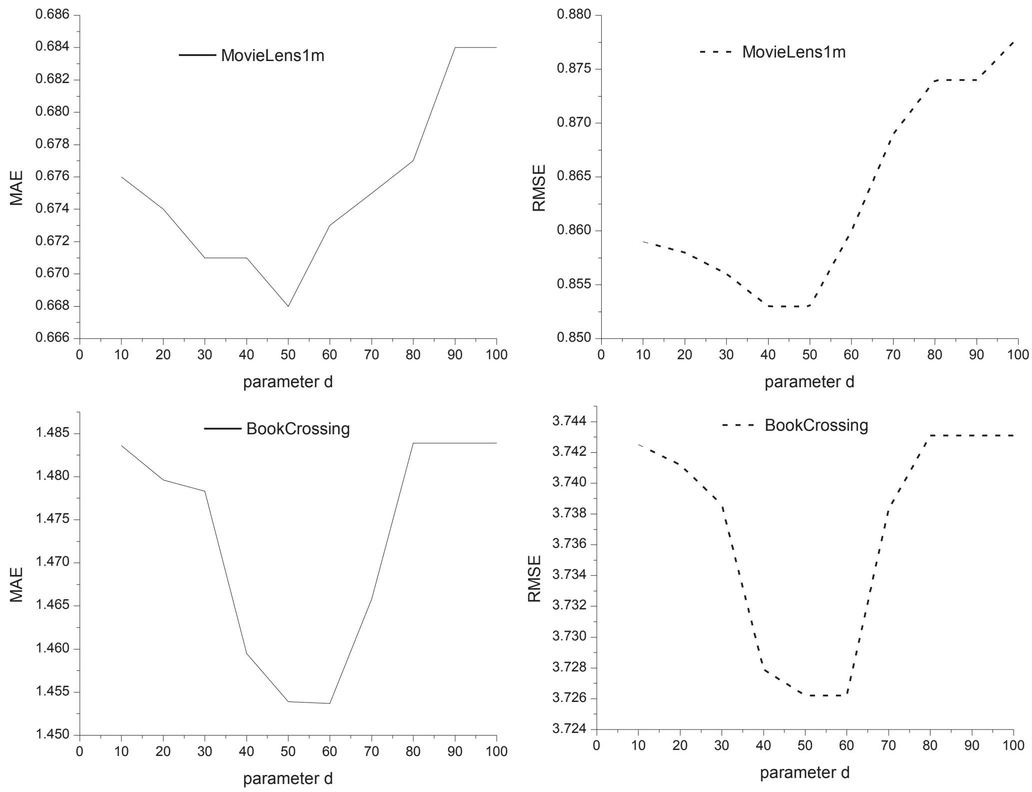

Figure 5.4: Impact of Parameter d on MovieLens1m and BookCrossing for TLMF

parameter α, we depict the MAE and RMSE changing trends of TLMF model when α ranges in [0,1]. Fig. 5.3 shows the impacts of parameter α with latent dimension d=50 on the MovieLens1m and BookCrossing data sets. Experimental results show that the proper values of α for the MovieLens1m and BookCrossing are respectively 0.3 and 0.6.

Impact of Dimension of Latent Vectors

Parameters d and K control the dimension of latent vectors respectively for the upper and lower level MF. In this Chapter, we also investigate the impact of the dimension parameters by fixing parameter α to 0.3 and 0.6 for the MovieLens1m and BookCrossing. Fig. 5.4 clearly shows the impact of

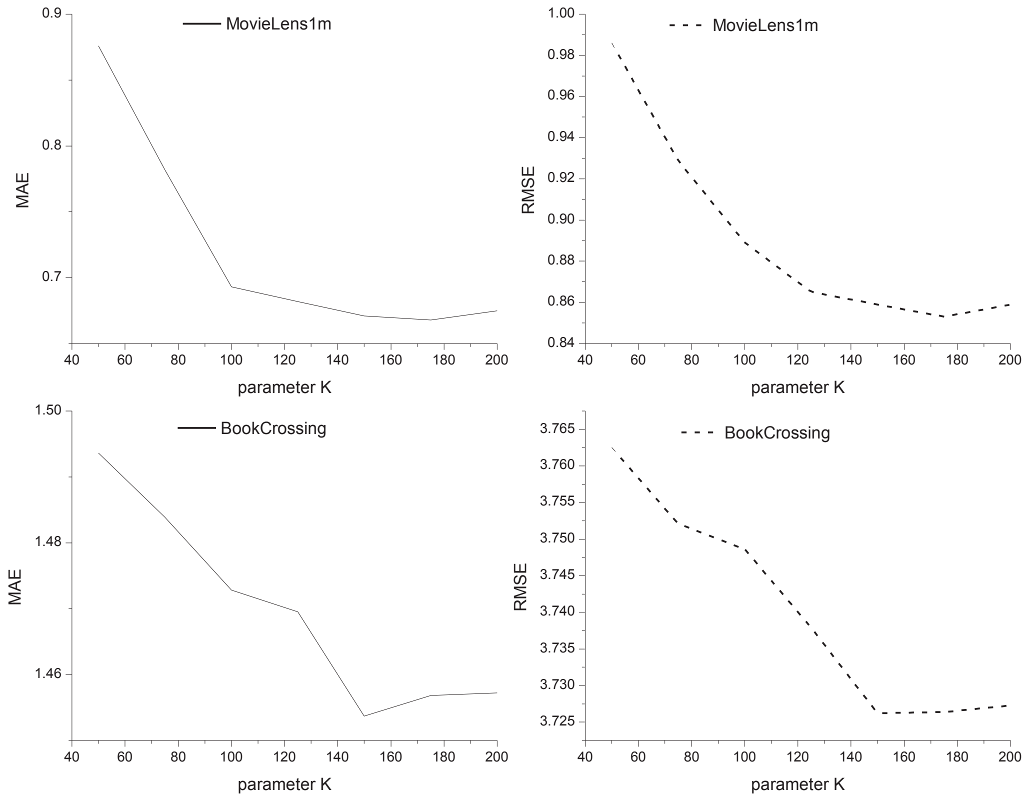

Figure 5.5: Impact of Parameter K on MovieLens1m and BookCrossing for TLMF

parameter d, indicating that the performance regarding MAE and RMSE would be decreased after the parameter d reaching 50 and 60 for the MovieLens1m and BookCrossing, respectively. Similarly, Fig. 5.5 shows the MAE and RMSE changing trends for different parameter K of lower level MF with w_m=0.01, indicating that optimum values are 175 and 150 respectively for the MovieLens1m and BookCrossing data. It is also noticed that the optimum parameter value of K is bigger than parameter d, which is possibly caused by big textual corpus having much more implicit features to depict.

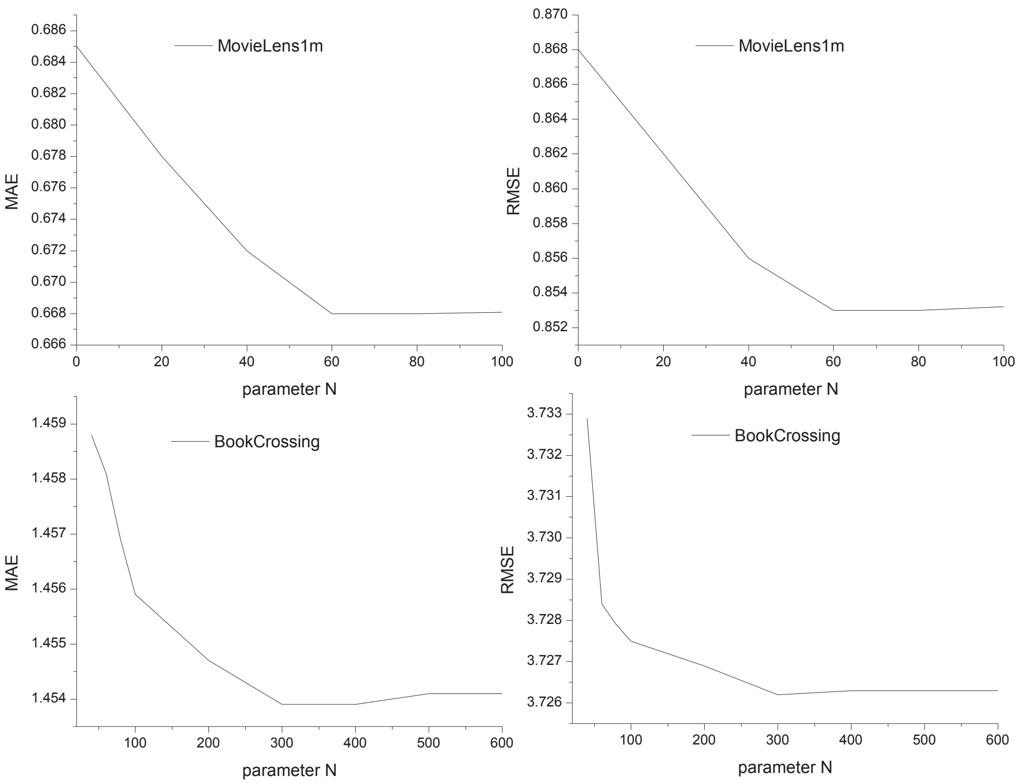

Figure 5.6: Impact of Neighbourhood Size of Items on MovieLens1m and BookCrossing for TLMF

Impact of Neighbourhood Size

In addition to the above parameters, the neighbourhood size of items also influence the optimization process of the objective function. Fig. 5.6 shows the effect of the neighbourhood size of items for the MovieLens1m and BookCrossing data sets. The experimental results indicate that the MAE and RMSE decrease sharply with the increase of the neighbourhood size of items until reaching a steady point. Continually increasing the neighbourhood size would not improve the performance after the steady point. From the experiments, we can see that the best neighbourhood size of items for the MovieLens1m and BookCrossing is respectively 60 and 300.

Cold-start Recommendation

In addition to the above experiments, we also compared the effectiveness of solving the problem of cold-start items. To select the cold-start items, we filter the data sets by selecting the items which are rated less than 20 times as cold-start items. To evaluate the effectiveness of our proposed approach on cold-start items, we compare the RMSE and MAE results with other benchmark methods as showed in Table 5.5. The parameters are respectively fixed as α=0.3, d=50, K=175, $\mathbb{N}$=60 for the MovieLens1m data, and α=0.6, d=60, K=150, $\mathbb{N}$=300 for the BookCrossing data. The comparison results clearly demonstrate that the proposed TLMF method outperforms other benchmark methods, which is resulted from considering the textual semantic couplings between items.

Discussion

From the above experiments, we demonstrate the effectiveness of our proposed TLMF model and the superiority over MF and CF methods. Generally, we can conclude that TLMF is more effective than the benchmark MF and CF approaches regarding MAE and RMSE for different latent dimensions, due to the strength of semantic couplings between items. In addition, the comparisons with EDMF, TFMF and TIMF show that the consideration of missing words in lower level MF is beneficial for computing item semantic couplings and further helpful for improving recommendation quality.

5.4 Conclusion

In this chapter, we first studied semantic couplings between items based on textual information to improve recommendation quality. Actually, the semantic couplings were captured by modelling the term document matrix with TF*IDF values by textual MF approach which considered missing words to enhance the capability of semantic analysis. A two-level MF model was then proposed to incorporate the semantic couplings between items and the rating matrix. The proposed two-level MF model, on one hand, took the advantage of latent semantic analysis of textual MF. On the other hand, it also balanced the traditional rating matrix based MF model. The experiments conducted on the real data sets demonstrated the superiority of the proposed TLMF model and suggested that semantic couplings could be effectively applied in RS.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

In this dissertation, we studied the non-IID RS by incorporating the coupling relations into CF methods. Specifically, the attribute-based couplings between users and between items were integrated into CF methods to improve recommendation quality. By modelling this coupling relation, we proposed several novel CF recommendation methods such as CCF and CMF. In addition, we also analysed the textual semantic couplings between items based on the textual context, and proposed an innovative TLMF model integrating the semantic couplings and the user-item ratings. The proposed methods demonstrate that modelling different coupling relations into non-IID RS is indeed beneficial for enhancing recommendation quality and solving the well-known cold-start problems. The details are concluded as follows.

1. In Chapter 3, we proposed three novel recommendation methods CUCF, CICF and CCF incorporating user couplings and item couplings into the classic neighbourhood-based CF method. Specifically, we first analysed the user couplings and item couplings derived from a novel non-IID based coupled object similarity measure. The user couplings and item couplings disclosed the traditional IID assumption and deeply analysed the intrinsic relations between users and between items. We then in-

tegrated the user couplings, item couplings, and user-item ratings to generate two hybrid similarity measures for users and items, respectively, which were further utilized to propose the CUCF, CICF and CCF methods. Lastly, the experimental evaluation demonstrated that incorporating the coupling relations into neighbourhood-based CF can largely improve the recommendation quality.

2. In Chapter 4, we proposed three novel recommendation methods CUMF, CIMF, and CMF which incorporate the user couplings and item couplings into the classic model-based CF method. Specifically, CUMF integrated the user couplings and the user-item ratings together, CIMF considered the item couplings and the user-item ratings together, and CMF balanced the user couplings, item couplings and the user-item ratings. In these three models, the coupling relations disclosed the traditional IID assumption and deeply analysed the intrinsic relationships between users and between items. In addition, the performance evaluation of the proposed models on different open data sets demonstrated that the coupling relations are helpful for improving the recommendation quality of the MF model.

3. In Chapter 5, we proposed a two-level MF model TLMF considering the impact of textual semantic couplings between items for RS. The semantic couplings were modelled based on the term-document matrix with TF*IDF values by a textual MF approach which considered missing words to enhance the capability of semantic analysis. A two-level MF model was then proposed to integrate the semantic couplings between items and the rating matrix. The proposed TLMF model, on one hand, took advantage of the latent semantic analysis of textual MF. On the other hand, it also balanced the traditional rating-based MF model. The experiments demonstrated the superiority of the proposed TLMF model and suggested that semantic couplings could be effectively applied in RS.

6.2 Future Work

To better research the non-IID RS, this work can be further explored by understanding non-IID data characteristics, goals, learning tasks, deliverables and evaluation, which are summarized as follows:

(1) Non-IID data characteristic understanding: To better explore the complicated data for new theoretical and practical innovation, we need to deeply understand the non-IID data characteristics and their influence, and the reflection of business on data. For example, in this dissertation, the studied coupling relations for users and items are mainly explicit. A future direction is to deeply analyse the non-IID data characteristics, such as the implicit coupling relations between users and items for better modelling.

(2) Non-IID data analytical goal: After gaining an understanding of the non-IID characteristics, non-IID RS needs to address the two essential problems of heterogeneity and couplings. In this dissertation, we mainly focused on analysing the coupling relations for users and items in RS, however a future direction is to thoroughly address the heterogeneity challenge by innovative modelling techniques.

(3) Non-IID RS learning task: Non-IID recommendation algorithms need to consider non-IID characteristics to train an innovative theoretical model. For example, this dissertation incorporated the user couplings and item couplings into MF by adding regularization factors into the objective function. However, this learning approach still did not consider the heterogeneity problem. Hence, a possible direction is to learn and model the heterogeneity problem into recommendation algorithms.

(4) Non-IID RS learning deliverable: After the non-IID RS algorithms are learned, the learning outcomes should hopefully solve the fundamental problems of non-IID RS with a thorough and comprehensive viewpoint. In this dissertation, the outcomes clearly demonstrated that modelling

couplings into RS is beneficial for improving recommendation quality. However, it is still very costly to compute the coupling relations. Thus, the efficient implementation of the algorithms needs to be explored using some big data techniques such as Apache Hadoop and Spark, in the future.

(5) Non-IID RS evaluation: Finally, new evaluation systems probably need to be invented in order to assess the proposed non-IID theories and tools. In this dissertation, RMSE and MAE metrics were mainly applied to evaluate the recommendation quality. However, these two evaluation measures did not consider the non-IID characteristics, and probably could not completely reflect the effectiveness of non-IID RS. Therefore, innovative evaluation methods considering the non-IID data characteristics should be designed in future work.

Bibliography

Adali, S. (2013), *Modeling Trust Context in Networks*, Springer Publishing Company, Incorporated.

Adomavicius, G. & Tuzhilin, A. (2005), 'Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions', *IEEE Transactions on Knowledge and Data Engineering* **17**(6), 734–749.

Agarwal, A. & Duchi, J. C. (2013), 'The generalization ability of online algorithms for dependent data.', *IEEE Transactions on Information Theory* **59**(1), 573–587.

Agarwal, D. & Chen, B. (2009), Regression-based latent factor models, *in* J. F. E. IV, F. Fogelman-Soulié, P. A. Flach & M. J. Zaki, eds, 'Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Paris, France, June 28 - July 1, 2009', ACM, pp. 19–28.

Agarwal, D. & Chen, B.-C. (2010), flda: matrix factorization through latent dirichlet allocation, *in* B. D. Davison, T. Suel, N. Craswell & B. Liu, eds, 'Proceedings of the Third International Conference on Web Search and Web Data Mining, WSDM 2010', ACM, pp. 91–100.

Agarwal, N., Liu, H., Tang, L. & Yu, P. S. (2008), Identifying the influential bloggers in a community, *in* 'Proceedings of the 2008 International

Conference on Web Search and Data Mining', WSDM '08, ACM, New York, NY, USA, pp. 207–218.

Aggarwal, C. C. (2011), *Social Network Data Analytics*, 1st edn, Springer Publishing Company, Incorporated.

Ahmad, A. & Dey, L. (2007), 'A k-mean clustering algorithm for mixed numeric and categorical data', *Data and Knowledge Engineering* **63**(2), 503–527.

Alter, O., Brown, P. O. & Botstein, D. (2000), 'Singular value decomposition for genome-wide expression data processing and modeling', *Proceedings of the National Academy of Sciences (PNAS)* **97**(18).

Amaral, L. A., Scala, A., Barthelemy, M. & Stanley, H. E. (2000), 'Classes of small-world networks.', *Proceedings of the National Academy of Sciences* **97**(21), 11149–11152.

Andritsos, P., Tsaparas, P., Miller, R. J. & Sevcik, K. C. (2004), Limbo: Scalable clustering of categorical data, *in* 'Advanced Database Technology - EDBT 2004', pp. 531–532.

Balabanović, M. & Shoham, Y. (1997), 'Fab: Content-based, collaborative recommendation', *Communications of the ACM* **40**(3), 66–72.

Basu, C., Hirsh, H. & Cohen, W. (1998), Recommendation as classification: Using social and content-based information in recommendation, *in* 'Proceedings of the Fifteenth National/Tenth Conference on Artificial Intelligence/Innovative Applications of Artificial Intelligence', AAAI '98/IAAI '98, American Association for Artificial Intelligence, Menlo Park, CA, USA, pp. 714–720.

Bertsekas, D. P., Nedić, A. & Ozdaglar, A. E. (2003), *Convex Analysis and Optimization*, Athena Scientific.

Billsus, D. & Pazzani, M. J. (1998), Learning collaborative information filters, *in* 'Proceedings of the Fifteenth International Conference on Machine Learning', ICML '98, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp. 46–54.

Blei, D. M., Ng, A. Y., Jordan, M. I. & Lafferty, J. (2003), 'Latent dirichlet allocation', *Journal of Machine Learning Research* **3**, 2003.

Bobadilla, J., Hernando, A., Ortega, F. & Bernal, J. (2011), 'A framework for collaborative filtering recommender systems', *Expert Systems with Applications* **38**(12), 14609–14623.

Boden, C., Karnstedt, M., Fernández, M. & Markl, V. (2013), Large-scale social-media analytics on stratosphere, *in* L. Carr, A. H. F. Laender, B. F. Lóscio, I. King, M. Fontoura, D. Vrandecic, L. Aroyo, J. P. M. de Oliveira, F. Lima & E. Wilde, eds, '22nd International World Wide Web Conference, WWW '13, Rio de Janeiro, Brazil, May 13-17, 2013, Companion Volume', International World Wide Web Conferences Steering Committee / ACM, pp. 257–260.

Boriah, S., Chandola, V. & Kumar, V. (2008), Similarity measures for categorical data: A comparative evaluation, *in* 'Proceedings of the SIAM International Conference on Data Mining, SDM 2008, April 24-26, 2008, Atlanta, Georgia, USA', SIAM, pp. 243–254.

Breese, J. S., Heckerman, D. & Kadie, C. (1998), Empirical analysis of predictive algorithms for collaborative filtering, *in* 'Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence', UAI'98, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp. 43–52.

Cacheda, F., Carneiro, V., Fernández, D. & Formoso, V. (2011), 'Comparison of collaborative filtering algorithms: Limitations of current techniques and proposals for scalable, high-performance recommender systems', *ACM Transactions on the Web* **5**(1), 2:1–2:33.

Cai, D., He, X. & Han, J. (2007), Spectral regression for efficient regularized subspace learning, *in* 'IEEE 11th International Conference on Computer Vision, ICCV 2007, Rio de Janeiro, Brazil, October 14-20, 2007', IEEE, pp. 1–8.

Candillier, L., Meyer, F. & Boullé, M. (2007), Comparing state-of-the-art collaborative filtering systems, *in* P. Perner, ed., 'Machine Learning and Data Mining in Pattern Recognition, 5th International Conference, MLDM 2007, Leipzig, Germany, July 18-20, 2007, Proceedings', Vol. 4571 of *Lecture Notes in Computer Science*, Springer, pp. 548–562.

Cao, L. (2013*a*), 'Combined mining: Analyzing object and pattern relations for discovering and constructing complex yet actionable patterns', *Wiley Interdisc. Rew.: Data Mining and Knowledge Discovery* **3**(2), 140–155.

Cao, L. (2013*b*), 'Combined mining: Analyzing object and pattern relations for discovering and constructing complex yet actionable patterns.', *Wiley Interdisc. Rew.: Data Mining and Knowledge Discovery* **3**(2), 140–155.

Cao, L. (2014), 'Non-iidness learning in behavioral and social data', *The Computer Journal* **57**(9), 1358–1370.

Cao, L. (2015), 'Coupling learning of complex interactions', *Information Processing & Management* **51**(2), 167 – 186.

Cao, L., Dai, R. & Zhou, M. (2009), 'Metasynthesis: M-space, m-interaction, and m-computing for open complex giant systems.', *IEEE Transactions on Systems, Man, and Cybernetics, Part A* **39**(5), 1007–1021.

Cao, L., Li, J., Wang, C. & Yu, P. S. (2013), Efficient selection of globally optimal rules on large imbalanced data based on rule coverage relationship analysis, *in* 'Proceedings of the 13th SIAM International Conference on Data Mining, May 2-4, 2013. Austin, Texas, USA.', SIAM, pp. 216–224.

Cao, L., Ou, Y. & Yu, P. S. (2012*a*), 'Coupled behavior analysis with applications.', *IEEE Transactions on Knowledge and Data Engineering* **24**(8), 1378–1392.

Cao, L., Ou, Y. & Yu, P. S. (2012*b*), 'Coupled behavior analysis with applications', *IEEE Transactions on Knowledge and Data Engineering* **24**(8), 1378–1392.

Cao, L. & Yu, P. S. (2015), 'Non-iid recommendation theories and systems', *IEEE Intelligent System* .

Cao, L., Zhang, H., Zhao, Y., Luo, D. & Zhang, C. (2011), 'Combined mining: Discovering informative knowledge in complex data', *IEEE Transactions on Systems, Man, and Cybernetics, Part B* **41**(3), 699–712.

Cao, L., Zhao, Y. & Zhang, C. (2008), 'Mining impact-targeted activity patterns in imbalanced data.', *IEEE Transactions on Knowledge and Data Engineering* **20**(8), 1053–1066.

Cao, W., Cao, L. & Song, Y. (2013), Coupled market behavior based financial crisis detection., in *The 2013 International Joint Conference on Neural Networks, IJCNN 2013, Dallas, TX, USA, August 4-9, 2013* (*The 2013 International Joint Conference on Neural Networks, IJCNN 2013, Dallas, TX, USA, August 4-9, 2013* 2013), pp. 1–8.

Caruana, R. (1993), Multitask learning: A knowledge-based source of inductive bias, *in* 'Machine Learning, Proceedings of the Tenth International Conference, University of Massachusetts, Amherst, MA, USA, June 27-29, 1993', Morgan Kaufmann, pp. 41–48.

Chaturvedi, A., Green, P. E. & Caroll, J. D. (2001), 'K-modes clustering', *Journal of Classification* **18**(1), 35–55.

Cheng, X., Miao, D., Wang, C. & Cao, L. (2013), Coupled term-term relation analysis for document clustering., in *IJCNN* (*The 2013 International*

Joint Conference on Neural Networks, IJCNN 2013, Dallas, TX, USA, August 4-9, 2013 2013), pp. 1–8.

Cho, E., Myers, S. A. & Leskovec, J. (2011), Friendship and mobility: User movement in location-based social networks, *in* 'Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining', KDD '11, ACM, New York, NY, USA, pp. 1082–1090.

Cinbis, R. G., Verbeek, J. J. & Schmid, C. (2012), Image categorization using fisher kernels of non-iid image models., *in* '2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, June 16-21, 2012', IEEE Computer Society, pp. 2184–2191.

Claypool, M., Gokhale, A., Miranda, T., Murnikov, P., Netes, D. & Sartin, M. (1999), Combining content-based and collaborative filters in an online newspaper, *in* 'Proceedings of the ACM SIGIR '99 Workshop on Recommender Systems: Algorithms and Evaluation', ACM, Berkeley, California.

Cohn, D., Atlas, L. & Ladner, R. (1994), 'Improving generalization with active learning', *Machine Learning* **15**(2), 201–221.

Cook, D. J. & Holder, L. B. (2006), *Mining Graph Data*, John Wiley & Sons.

Cotter, P. & Smyth, B. (2000), Ptv: Intelligent personalised tv guides, *in* 'Proceedings of the Seventeenth National Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence', AAAI Press, pp. 957–964.

Crandall, D., Cosley, D., Huttenlocher, D., Kleinberg, J. & Suri, S. (2008), Feedback effects between similarity and social influence in online communities, *in* 'Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining', KDD '08, ACM, New York, NY, USA, pp. 160–168.

Das, G. & Mannila, H. (2000), Context-based similarity measures for categorical databases, *in* D. A. Zighed, H. J. Komorowski & J. M. Zytkow, eds, 'Principles of Data Mining and Knowledge Discovery, 4th European Conference, PKDD 2000, Lyon, France, September 13-16, 2000, Proceedings', Vol. 1910 of *Lecture Notes in Computer Science*, Springer, pp. 201–210.

Davis, D. A., Lichtenwalter, R. & Chawla, N. V. (2013), 'Supervised methods for multi-relational link prediction.', *Social Network Analysis and Mining* **3**(2), 127–141.

Dellarocas, C., Zhang, X. M. & Awad, N. F. (2007), 'Exploring the value of online product reviews in forecasting sales: The case of motion pictures', *Journal of Interactive Marketing* **21**(4), 23–45.

Deshpande, M. & Karypis, G. (2004), 'Item-based top-n recommendation algorithms', *ACM Transactions on Information Systems* **22**(1), 143–177.

Dunbar, R. (2010), *How Many Friends Does One Person Need?: Dunbar's Number and Other Evolutionary Quirks*, Faber and Faber.

Dundar, M., Krishnapuram, B., Bi, J. & Rao, R. B. (2007), Learning classifiers when the training data is not iid., *in* M. M. Veloso, ed., 'Proceedings of the 20th International Joint Conference on Artificial Intelligence IJCAI 2007', pp. 756–761.

Falcone, R. & Castelfranchi, C. (2010), Transitivity in trust. A discussed property, *in* A. Omicini & M. Viroli, eds, 'Proceedings of the 11th WOA 2010 Workshop, Dagli Oggetti Agli Agenti, Rimini, Italy, September 5-7, 2010.', Vol. 621 of *CEUR Workshop Proceedings*, CEUR-WS.org.

Fan, W. & Gordon, M. D. (2014), 'The power of social media analytics', *Communications of the ACM* **57**(6), 74–81.

Feng, W. & Wang, J. (2012), Incorporating heterogeneous information for personalized tag recommendation in social tagging systems, *in* Q. Yang, D. Agarwal & J. Pei, eds, 'The 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '12', ACM, pp. 1276–1284.

Figueiredo, F., Rocha, L., Couto, T., Salles, T., Gonçalves, M. A. & Meira Jr., W. (2011), 'Word co-occurrence features for text classification', *Information Systems* **36**(5), 843–858.

Fortunato, S. (2010), 'Community detection in graphs', *Physics Reports* **486**(3-5), 75 – 174.

Friedman, N., Getoor, L., Koller, D. & Pfeffer, A. (1999), Learning probabilistic relational models, *in* 'Proceedings of the 16th International Joint Conference on Artificial Intelligence - Volume 2', IJCAI'99, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp. 1300–1307.

Gama, J. (2010), *Knowledge Discovery from Data Streams*, 1st edn, Chapman & Hall/CRC.

Gantner, Z., Drumond, L., Freudenthaler, C., Rendle, S. & Schmidt-Thieme, L. (2010), Learning attribute-to-feature mappings for cold-start recommendations, *in* Webb, Liu, Zhang, Gunopulos & Wu (2010), pp. 176–185.

Gao, H., Tang, J. & Liu, H. (2012*a*), Exploring social-historical ties on location-based social networks., *in* J. G. Breslin, N. B. Ellison, J. G. Shanahan & Z. Tufekci, eds, 'Proceedings of the Sixth International Conference on Weblogs and Social Media, Dublin, Ireland, June 4-7, 2012', The AAAI Press.

Gao, H., Tang, J. & Liu, H. (2012*b*), gscorr: Modeling geo-social correlations for new check-ins on location-based social networks, *in* 'Proceedings of the 21st ACM International Conference on Information and Knowledge Management', CIKM '12, ACM, New York, NY, USA, pp. 1582–1586.

Gefen, D., Karahanna, E. & Straub, D. W. (2003), 'Trust and tam in online shopping: An integrated model.', *MIS Quarterly* **27**(1), 51–90.

George, T. & Merugu, S. (2005), A scalable collaborative filtering framework based on co-clustering, *in* 'Proceedings of the Fifth IEEE International Conference on Data Mining', ICDM '05, IEEE Computer Society, Washington, DC, USA, pp. 625–628.

Getoor, L. & Taskar, B. (2007), *Introduction to Statistical Relational Learning (Adaptive Computation and Machine Learning)*, The MIT Press.

Gibson, D., Kleinberg, J. M. & Raghavan, P. (2000), 'Clustering categorical data: An approach based on dynamical systems', *VLDB Journal: Very Large Data Bases* **8**(3–4), 222–236.

Girvan, M. & Newman, M. E. J. (2002), 'Community structure in social and biological networks', *Proceedings of the National Academy of Sciences* **99**(12), 7821–7826.

Golbeck, J. (2006), Generating predictive movie recommendations from trust in social networks, *in* 'Proceedings of the 4th International Conference on Trust Management', iTrust'06, Springer-Verlag, Berlin, Heidelberg, pp. 93–104.

Golbeck, J. (2009), 'Trust and nuanced profile similarity in online social networks', *ACM Transactions on the Web* **3**(4), 12:1–12:33.

Good, N., Schafer, J. B., Konstan, J. A., Borchers, A., Sarwar, B., Herlocker, J. & Riedl, J. (1999), Combining collaborative filtering with personal agents for better recommendations, *in* 'Proceedings of the Sixteenth National Conference on Artificial Intelligence and the Eleventh Innovative Applications of Artificial Intelligence Conference Innovative

Applications of Artificial Intelligence', AAAI '99/IAAI '99, American Association for Artificial Intelligence, Menlo Park, CA, USA, pp. 439–446.

Guha, S., Rastogi, R. & Shim, K. (1999), Rock: A robust clustering algorithm for categorical attributes, *in* 'Proceedings of the 15th International Conference on Data Engineering, 23-26 March 1999, Sydney, Austrialia', IEEE Computer Society, pp. 512–521.

Guo, W. & Diab, M. (2012*a*), A simple unsupervised latent semantics based approach for sentence similarity, *in* 'Proceedings of the First Joint Conference on Lexical and Computational Semantics', SemEval '12, Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 586–590.

Guo, W. & Diab, M. T. (2012*b*), Modeling sentences in the latent space, *in* 'The 50th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference, July 8-14, 2012, Jeju Island, Korea - Volume 1: Long Papers', pp. 864–872.

Guo, W., Li, H., Ji, H. & Diab, M. T. (2013), Linking tweets to news: A framework to enrich short text data in social media, *in* 'Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, ACL 2013, 4-9 August 2013, Sofia, Bulgaria, Volume 1: Long Papers', pp. 239–249.

Guo, Z.-C. & Shi, L. (2011), 'Classification with non-i.i.d. sampling.', *Mathematical and Computer Modelling* **54**(5-6), 1347–1364.

Guy, I. & Carmel, D. (2011), Social recommender systems, *in* 'Proceedings of the 20th International Conference Companion on World Wide Web', WWW '11, ACM, New York, NY, USA, pp. 283–284.

Herlocker, J., Konstan, J. A. & Riedl, J. (2002), 'An empirical analysis of design choices in neighborhood-based collaborative filtering algorithms', *Information Retrieval* **5**(4), 287–310.

Herlocker, J. L., Konstan, J. A., Borchers, A. & Riedl, J. (1999), An algorithmic framework for performing collaborative filtering, *in* 'Proceedings of the 22Nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval', SIGIR '99, ACM, New York, NY, USA, pp. 230–237.

Ho, Y., Fong, S. & Yan, Z. (2007), A hybrid ga-based collaborative filtering model for online recommenders., *in* J. Filipe, D. A. Marca, B. Shishkov & M. van Sinderen, eds, 'ICE-B 2007 - Proceedings of the International Conference on e-Business, Barcelona, Spain, July 28-31, 2007, ICE-B is part of ICETE - The International Joint Conference on e-Business and Telecommunications', INSTICC Press, pp. 200–203.

Hofmann, T. (1999), Probabilistic latent semantic indexing, *in* 'Proceedings of the 22Nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval', SIGIR '99, ACM, New York, NY, USA, pp. 50–57.

Hofmann, T. (2004), 'Latent semantic models for collaborative filtering', *ACM Transactions on Information Systems* **22**(1), 89–115.

Houle, M. E., Oria, V. & Qasim, U. (2010), Active caching for similarity queries based on shared-neighbor information, *in* 'Proceedings of the 19th ACM International Conference on Information and Knowledge Management', CIKM '10, ACM, New York, NY, USA, pp. 669–678.

Hu, L., Cao, J., Xu, G., Wang, J., Gu, Z. & Cao, L. (2013), Cross-domain collaborative filtering via bilinear multilevel analysis., *in* Rossi (2013).

Hu, Y., Koren, Y. & Volinsky, C. (2008), Collaborative filtering for implicit feedback datasets, *in* 'Proceedings of the 2008 Eighth IEEE International Conference on Data Mining', ICDM '08, IEEE Computer Society, Washington, DC, USA, pp. 263–272.

Huopaniemi, I., Suvitaival, T., Nikkila, J., Oresic, M. & Kaski, S. (2010), 'Multivariate multi-way analysis of multi-source data.', *Bioinformatics [ISMB]* **26**(12), 391–398.

Itskov, M. (2009), *Tensor Algebra and Tensor Analysis for Engineers: With Applications to Continuum Mechanics*, 2nd edn, Springer Publishing Company, Incorporated.

Jacob, L., Obozinski, G. & Vert, J.-P. (2009), Group lasso with overlap and graph lasso, *in* 'Proceedings of the 26th Annual International Conference on Machine Learning', ICML '09, ACM, New York, NY, USA, pp. 433–440.

Jaeger, M. (1997), Relational bayesian networks, *in* 'Proceedings of the 13th Annual Conference on Uncertainty in AI (UAI)', Morgan Kaufmann, pp. 266–273.

Jamali, M. & Ester, M. (2009), Trustwalker: A random walk model for combining trust-based and item-based recommendation, *in* 'Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining', KDD '09, ACM, New York, NY, USA, pp. 397–406.

Jamali, M. & Ester, M. (2010*a*), A matrix factorization technique with trust propagation for recommendation in social networks, *in* X. Amatriain, M. Torrens, P. Resnick & M. Zanker, eds, 'Proceedings of the 2010 ACM Conference on Recommender Systems, RecSys 2010', ACM, pp. 135–142.

Jamali, M. & Ester, M. (2010*b*), A matrix factorization technique with trust propagation for recommendation in social networks, *in* 'Proceedings of the Fourth ACM Conference on Recommender Systems', RecSys '10, ACM, New York, NY, USA, pp. 135–142.

Jordan, M. I., ed. (1999), *Learning in Graphical Models*, MIT Press, Cambridge, MA, USA.

Jøsang, A., Ismail, R. & Boyd, C. (2007), 'A survey of trust and reputation systems for online service provision', *Decision Support Systems* **43**(2), 618–644.

Kaufman, L. & Rousseeuw, P. J. (1990), *Finding Groups in Data: An Introduction to Cluster Analysis.*, John Wiley.

Kautz, H., Selman, B. & Shah, M. (1997), 'Referral web: Combining social networks and collaborative filtering', *Communications of the ACM* **40**(3), 63–65.

Kersting, K. & Raedt, L. D. (2008), Basic principles of learning bayesian logic programs, *in* K. K. S. M. L. De Raedt, P. Frasconi, ed., 'Probabilistic Inductive Logic Programming: Theory and Applications', Vol. 4911 of *LNAI*, Springer, pp. 193–226.

King, I., Lyu, M. R. & Ma, H. (2010), Introduction to social recommendation., *in* M. Rappa, P. Jones, J. Freire & S. Chakrabarti, eds, 'Proceedings of the 19th International Conference on World Wide Web, WWW 2010', ACM, pp. 1355–1356.

Kleinberg, J. M. (1999), 'Authoritative sources in a hyperlinked environment', *Journal of the ACM* **46**(5), 604–632.

Kok, S. & Domingos, P. (2007), Statistical predicate invention, *in* 'In Z. Ghahramani (Ed.), Proceedings of the 24th annual international conference on machine learning (ICML-2007', pp. 433–440.

Koller, D., Schuurmans, D., Bengio, Y. & Bottou, L., eds (2008), *Advances in Neural Information Processing Systems 21, Proceedings of the Twenty-*

Second Annual Conference on Neural Information Processing Systems, Vancouver, British Columbia, Canada, December 8-11, 2008, Curran Associates, Inc.

Kontorovich, L. A. & Ramanan, K. (2008), 'Concentration inequalities for dependent random variables via the martingale method', *ANNALS OF PROBABILITY* **36**(6), 2126–2158.

Kontorovich, L. A., Ramanan, K. et al. (2008), 'Concentration inequalities for dependent random variables via the martingale method', *The Annals of Probability* **36**(6), 2126–2158.

Koren, Y. (2008), Factorization meets the neighborhood: a multifaceted collaborative filtering model, *in* Y. Li, B. Liu & S. Sarawagi, eds, 'Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining', ACM, pp. 426–434.

Koren, Y., Bell, R. M. & Volinsky, C. (2009), 'Matrix factorization techniques for recommender systems', *IEEE Computer* **42**(8), 30–37.

Krulwich, B. (1997), 'Lifestyle finder: Intelligent user profiling using large-scale demographic data', *AI Magazine* **18**(2), 37–45.

Kumar, A. & III, H. D. (2012), Learning task grouping and overlap in multi-task learning., *in* 'Proceedings of the 29th International Conference on Machine Learning, ICML 2012, Edinburgh, Scotland, UK, June 26 - July 1, 2012', icml.cc / Omnipress.

Kumar, S. (2013), 'Review of: Kadushin, charles. understanding social networks: theories, concepts and findings. new york, ny: Oxford university press, 2012.', *Information Research* **18**(1).

L, L., Medo, M., Yeung, C. H., Zhang, Y.-C., Zhang, Z.-K. & Zhou, T. (2012), 'Recommender systems', *CoRR* **abs/1202.1112**.

Lacoste-Julien, S., Sha, F. & Jordan, M. I. (2008), Disclda: Discriminative learning for dimensionality reduction and classification., *in* Koller, Schuurmans, Bengio & Bottou (2008), pp. 897–904.

Lafferty, J. D., McCallum, A. & Pereira, F. C. N. (2001), Conditional random fields: Probabilistic models for segmenting and labeling sequence data, *in* 'Proceedings of the Eighteenth International Conference on Machine Learning', ICML '01, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp. 282–289.

Landauer, T. K., Foltz, P. W. & Laham, D. (1998), 'An introduction to latent semantic analysis', *Discourse Processes* (25), 259–284.

Lang, K. (1995), Newsweeder: Learning to filter netnews, *in* 'in Proceedings of the 12th International Machine Learning Conference (ML95'.

Laskey, K. B. (2008), 'Mebn: A language for first-order bayesian knowledge bases', *Artificial Intelligence* **172**(2-3).

Lauritzen, S. L. (1996), *Graphical Models*, Oxford University Press.

Lee, D. D. & Seung, H. S. (1999), 'Learning the parts of objects by nonnegative matrix factorization', *Nature* **401**, 788–791.

Lehmann, E. L. & Romano, J. P. (2005), *Testing statistical hypotheses*, Springer Texts in Statistics, third edn, Springer, New York.

Leskovec, J., Adamic, L. A. & Huberman, B. A. (2007), 'The dynamics of viral marketing', *ACM Transactions on the Web* **1**(1).

Leskovec, J., Huttenlocher, D. & Kleinberg, J. (2010), Predicting positive and negative links in online social networks, *in* 'Proceedings of the 19th International Conference on World Wide Web', WWW '10, ACM, New York, NY, USA, pp. 641–650.

Li, F., Xu, G. & Cao, L. (2014), Coupled item-based matrix factorization, *in* B. Benatallah, A. Bestavros, Y. Manolopoulos, A. Vakali & Y. Zhang, eds, 'Web Information Systems Engineering - WISE (1) 2014', Vol. 8786 of *Lecture Notes in Computer Science*, Springer, pp. 1–14.

Li, F., Xu, G. & Cao, L. (2015*a*), Coupled matrix factorization within non-iid context, *in* T. Cao, E. Lim, Z. Zhou, T. Ho, D. Cheung & H. Motoda, eds, 'Advances in Knowledge Discovery and Data Mining - 19th Pacific-Asia Conference, PAKDD 2015, Ho Chi Minh City, Vietnam, May 19-22, 2015, Proceedings, Part II', Vol. 9078 of *Lecture Notes in Computer Science*, Springer, pp. 707–719.

Li, F., Xu, G. & Cao, L. (2015*b*), 'Two-level matrix factorization for recommender systems', *Neural Computing and Applications* **26**(7), 1–12.

Li, F., Xu, G., Cao, L., Fan, X. & Niu, Z. (2013), Cgmf: Coupled group-based matrix factorization for recommender system, *in* X. Lin, Y. Manolopoulos, D. Srivastava & G. Huang, eds, 'Web Information Systems Engineering - WISE(1) 2013', Vol. 8180 of *Lecture Notes in Computer Science*, Springer, pp. 189–198.

Littlestone, N. (1988), 'Learning quickly when irrelevant attributes abound: A new linear-threshold algorithm', *Machine Learning* **2**(4), 285–318.

Loomes, G. & Sugden, R. (1982), 'Regret theory: An alternative theory of rational choice under uncertainty', *Economic Journal* **92**(368), 805–24.

Losee, R. M. (2006), 'Browsing mixed structured and unstructured data.', *Information Processing & Management* **42**(2), 440–452.

Luo, X., Xia, Y. & Zhu, Q. (2012), 'Incremental collaborative filtering recommender based on regularized matrix factorization', *Knowledge-Based Systems* **27**, 271–280.

Luo, X., Xia, Y. & Zhu, Q. (2013), 'Applying the learning rate adaptation to the matrix factorization based collaborative filtering', *Knowledge-Based Systems* **37**, 154–164.

Ma, H. (2013), An experimental study on implicit social recommendation, *in* G. J. F. Jones, P. Sheridan, D. Kelly, M. de Rijke & T. Sakai, eds, 'The 36th International ACM SIGIR conference on research and development in Information Retrieval, SIGIR '13', ACM, pp. 73–82.

Ma, H., King, I. & Lyu, M. R. (2009), Learning to recommend with social trust ensemble, *in* J. Allan, J. A. Aslam, M. Sanderson, C. Zhai & J. Zobel, eds, 'Proceedings of the 32nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval', ACM, pp. 203–210.

Ma, H., King, I. & Lyu, M. R. (2011), 'Learning to recommend with explicit and implicit social relations', *ACM Transactions on Intelligent Systems and Technology* **2**(3), 29:1–29:19.

Ma, H., Yang, H., Lyu, M. R. & King, I. (2008), Sorec: social recommendation using probabilistic matrix factorization, *in* J. G. Shanahan, S. Amer-Yahia, I. Manolescu, Y. Zhang, D. A. Evans, A. Kolcz, K. Choi & A. Chowdhury, eds, 'Proceedings of the 17th ACM Conference on Information and Knowledge Management, CIKM 2008', ACM, pp. 931–940.

Ma, H., Zhou, D., Liu, C., Lyu, M. R. & King, I. (2011), Recommender systems with social regularization, *in* 'Proceedings of the Forth International Conference on Web Search and Web Data Mining, WSDM 2011', pp. 287–296.

Ma, H., Zhou, T. C., Lyu, M. R. & King, I. (2011), 'Improving recommender systems by incorporating social contextual information', *ACM Transactions on Information Systems* **29**(2), 9:1–9:23.

man Au Yeung, C. & Iwata, T. (2011), Strength of social influence in trust networks in product review sites., *in* I. King, W. Nejdl & H. Li, eds, 'Proceedings of the Forth International Conference on Web Search and Web Data Mining, WSDM 2011, Hong Kong, China, February 9-12, 2011', ACM, pp. 495–504.

Manning, C. D. & Schütze, H. (1999), *Foundations of Statistical Natural Language Processing*, MIT Press, Cambridge, MA, USA.

Marsden, P. V. & Friedkin, N. E. (1993), 'Network studies of social influence', *Sociological Methods Research* **22**(1), 127–151.

Massa, P. (2006), A survey of trust use and modeling in current real systems, *in* R. Song, L. Korba & G. Yee, eds, 'Trust in E-Services: Technologies, Practices and Challenges', Idea Group Publishing.

Massa, P. & Avesani, P. (2007), Trust-aware recommender systems, *in* 'Proceedings of the 2007 ACM Conference on Recommender Systems', RecSys '07, ACM, New York, NY, USA, pp. 17–24.

McKnight, D. H., Choudhury, V. & Kacmar, C. J. (2002), 'Developing and validating trust measures for e-commerce: An integrative typology.', *Information Systems Research* **13**(3), 334–359.

McPherson, M., Smith-Lovin, L. & Cook, J. M. (2001), 'Birds of a feather: Homophily in social networks', *Annual Review of Sociology* **27**(1), 415–444.

Meier, L., van de Geer, S. & Bühlmann, P. (2008), 'The group lasso for logistic regression', *Journal of the Royal Statistical Society. Series B* **70**(1), 53–71.

Melville, P., Mooney, R. J. & Nagarajan, R. (2002), Content-boosted collaborative filtering for improved recommendations, *in* 'Eighteenth National Conference on Artificial Intelligence', American Association for Artificial Intelligence, Menlo Park, CA, USA, pp. 187–192.

Melville, P. & Sindhwani, V. (2010), Recommender systems, *in* 'Encyclopedia of Machine Learning', pp. 829–838.

Menon, A. K. & Elkan, C. (2010), A log-linear model with latent features for dyadic prediction, *in* Webb et al. (2010), pp. 364–373.

Milch, B. & Russell, S. (2007), 'First-Order probabilistic languages: Into the unknown', *Inductive Logic Programming* pp. 10–24.

Mohri, M. & Rostamizadeh, A. (2008), Rademacher complexity bounds for non-i.i.d. processes., *in* Koller et al. (2008), pp. 1097–1104.

Mohri, M. & Rostamizadeh, A. (2010), 'Stability bounds for stationary phi-mixing and beta-mixing processes', *Journal of Machine Learning Research* **11**, 789–814.

Mooney, R. J. & Roy, L. (2000), Content-based book recommending using learning for text categorization, *in* 'Proceedings of the fifth ACM conference on Digital libraries', DL '00, ACM, New York, NY, USA, pp. 195–204.

Nati, N. S. & Jaakkola, T. (2003), Weighted low-rank approximations, *in* 'In 20th International Conference on Machine Learning', AAAI Press, pp. 720–727.

Nelsen, R. B. (2006), *An Introduction to Copulas (Springer Series in Statistics)*, Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Neville, J. & Jensen, D. (2007), 'Relational dependency networks', *Journal of Machine Learning Research* **8**, 653–692.

Newman, M. (2003), 'The structure and function of complex networks', *SIAM review* **45**(2), 167–256.

Newman, M. E. J. (2005), 'Power laws, pareto distributions and zipf's law', *Contemporary Physics* **46**, 323–351.

Nicholas, I. S. C. & Nicholas, C. K. (1999), Combining content and collaboration in text filtering, *in* 'In Proceedings of the IJCAI99 Workshop on Machine Learning for Information Filtering', pp. 86–91.

Olsson, F. (2009), A literature survey of active machine learning in the context of natural language processing, Technical Report 06, Box 1263, SE-164 29 Kista, Sweden.

Page, L., Brin, S., Motwani, R. & Winograd, T. (1998), The pagerank citation ranking: Bringing order to the web, *in* 'Proceedings of the 7th International World Wide Web Conference', Brisbane, Australia, pp. 161–172.

Palla, G., Dernyi, I., Farkas, I. & Vicsek, T. (2005), 'Uncovering the overlapping community structure of complex networks in nature and society', *Nature* **435**(7043), 814–818.

Pan, R. & Scholz, M. (2009), Mind the gaps: Weighting the unknown in large-scale one-class collaborative filtering, *in* 'Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining', KDD '09, ACM, New York, NY, USA, pp. 667–676.

Pan, S. J. & Yang, Q. (2010), 'A survey on transfer learning', *IEEE Transactions on Knowledge and Data Engineering* **22**(10), 1345–1359.

Pan, Z.-W. & Xiao, Q.-W. (2009), 'Least-square regularized regression with non-iid sampling', *Journal of Statistical Planning and Inference* **139**(10), 3579–3587.

Park, M.-H., Hong, J.-H. & Cho, S.-B. (2007), Location-based recommendation system using bayesian user's preference model in mobile devices, *in* 'Proceedings of the 4th International Conference on Ubiquitous Intelligence and Computing', UIC'07, Springer-Verlag, Berlin, Heidelberg, pp. 1130–1139.

Pazzani, M. & Billsus, D. (1997), 'Learning and revising user profiles: The identification ofinteresting web sites', *Machine Learning* **27**(3), 313–331.

Pazzani, M. J. (1999), 'A framework for collaborative, content-based and demographic filtering', *Artificial Intelligence Review* **13**(5-6), 393–408.

Pazzani, M. J. & Billsus, D. (2007), Content-based recommendation systems, *in* 'THE ADAPTIVE WEB: METHODS AND STRATEGIES OF WEB PERSONALIZATION. VOLUME 4321 OF LECTURE NOTES IN COMPUTER SCIENCE', Springer-Verlag, pp. 325–341.

Peddle, D. R. & Ferguson, D. T. (2002), 'Optimisation of multisource data analysis: An example using evidential reasoning for gis data classification', *Comput. Geosci.* **28**(1), 45–52.

Pham, M. C., Cao, Y., Klamma, R. & Jarke, M. (2011), 'A clustering approach for collaborative filtering recommendation using social network analysis.', *Journal of Universal Computer Science* **17**(4), 583–604.

Ping, W., Xu, Y., Ren, K., Chi, C. & Furao, S. (2010), Non-i.i.d. multi-instance dimensionality reduction by learning a maximum bag margin subspace, *in* M. Fox & D. Poole, eds, 'Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2010, Atlanta, Georgia, USA, July 11-15, 2010', AAAI Press.

Pitsilis, G., Zhang, X. & Wang, W. (2011), Clustering recommenders in collaborative filtering using explicit trust information., *in* I. Wakeman, E. Gudes, C. D. Jensen & J. Crampton, eds, 'Trust Management V, 5th IFIP WG 11.11 International Conference, IFIPTM 2011, Copenhagen, Denmark, June 29-July 1, 2011, Proceedings', Vol. 358 of *IFIP Advances in Information and Communication Technology*, Springer, pp. 82–97.

Popescul, A., Ungar, L. H., Pennock, D. M. & Lawrence, S. (2001), Probabilistic models for unified collaborative and content-based recommendation in sparse-data environments, *in* 'Proceedings of the 17th Conference in Uncertainty in Artificial Intelligence', UAI '01, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp. 437–444.

Porcel, C., Tejeda-Lorente, A., Martnez, M. A. & Herrera-Viedma, E. (2012), 'A hybrid recommender system for the selective dissemination of research resources in a technology transfer office.', *Information Sciences* **184**(1), 1–19.

Porter, M. A., Onnela, J.-P. & Mucha, P. J. (2009), 'Communities in networks', *Notices of the American Mathematical Society* **56**(9), 1082–1097.

Prez-cruz, F. (2008), Kullback-leibler divergence estimation of continuous distributions, *in* 'Proceedings of IEEE International Symposium on Information Theory', pp. 1666–1670.

Ralaivola, L., Szafranski, M. & Stempfel, G. (2010), 'Chromatic pac-bayes bounds for non-iid data: Applications to ranking and stationary -mixing processes.', *Journal of Machine Learning Research* **11**, 1927–1956.

Rendle, S. (2010), Factorization machines, *in* Webb et al. (2010), pp. 995–1000.

Rennie, J. D. M. & Srebro, N. (2005), Fast maximum margin matrix factorization for collaborative prediction, *in* 'Proceedings of the 22Nd International Conference on Machine Learning', ICML '05, ACM, New York, NY, USA, pp. 713–719.

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P. & Riedl, J. (1994), Grouplens: An open architecture for collaborative filtering of netnews, *in* 'Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work', CSCW '94, ACM, New York, NY, USA, pp. 175–186.

Ribeiro, L. A. & Harder, T. (2011), 'Generalizing prefix filtering to improve set similarity joins.', *Inf. Syst.* **36**(1), 62–78.

Richardson, M. & Domingos, P. (2006), 'Markov logic networks', *Machine Learning* **62**(1-2), 107–136.

Roh, T. H., Oh, K. J. & Han, I. (2003), 'The collaborative filtering recommendation based on som cluster-indexing cbr', *Expert Systems with Applications* **25**(3), 413–423.

Rossi, F., ed. (2013), *IJCAI 2013, Proceedings of the 23rd International Joint Conference on Artificial Intelligence, Beijing, China, August 3-9, 2013*, IJCAI/AAAI.

Roussas, G. G. (1990), 'Nonparametric regression estimation under mixing conditions', *Stochastic Processes and their Applications* **36**(1), 107 – 116.

Rubens, N., Kaplan, D. & Sugiyama, M. (2011), Active learning in recommender systems, *in* F. Ricci, L. Rokach, B. Shapira & P. B. Kantor, eds, 'Recommender Systems Handbook', Springer US, chapter 23, pp. 735–767.

Salakhutdinov, R. & Mnih, A. (2008), Probabilistic matrix factorization, *in* 'Advances in Neural Information Processing Systems', Vol. 20.

Salter, J. & Antonopoulos, N. (2006), 'Cinemascreen recommender agent: Combining collaborative and content-based filtering', *IEEE Intelligent Systems* **21**(1), 35–41.

Sarwar, B., Karypis, G., Konstan, J. & Riedl, J. (2001*a*), Item-based collaborative filtering recommendation algorithms, *in* 'Proceedings of the 10th international conference on World Wide Web', WWW '01, ACM, New York, NY, USA, pp. 285–295.

Sarwar, B., Karypis, G., Konstan, J. & Riedl, J. (2001*b*), Item-based collaborative filtering recommendation algorithms, *in* 'Proceedings of the 10th International Conference on World Wide Web', WWW '01, ACM, New York, NY, USA, pp. 285–295.

Sarwar, B. M., Karypis, G., Konstan, J. A. & Riedl, J. T. (2000), Application of dimensionality reduction in recommender systems: A case study, *in* 'WebKDD Workshop at the ACM SIGKKD'.

Schafer, J. B., Frankowski, D., Herlocker, J. & Sen, S. (2007), The adaptive web, Springer-Verlag, Berlin, Heidelberg, chapter Collaborative Filtering Recommender Systems, pp. 291–324.

Schein, A. I., Popescul, A., Ungar, L. H. & Pennock, D. M. (2002), Methods and metrics for cold-start recommendations, *in* 'Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval', SIGIR '02, ACM, New York, NY, USA, pp. 253–260.

Scott, J. (2000), *Social Network Analysis: A Handbook*, SAGE Publications.

Scott, J. (2011), 'Social network analysis: developments, advances, and prospects', *Social Network Analysis and Mining* **1**(1), 21–26.

Settles, B. (2009), Active learning literature survey, Computer Sciences Technical Report 1648, University of Wisconsin–Madison.

Shepitsen, A., Gemmell, J., Mobasher, B. & Burke, R. (2008), Personalized recommendation in social tagging systems using hierarchical clustering, *in* 'Proceedings of the 2008 ACM Conference on Recommender Systems', RecSys '08, ACM, New York, NY, USA, pp. 259–266.

Sherchan, W., Nepal, S. & Paris, C. (2013), 'A survey of trust in social networks', *ACM Computing Surveys* **45**(4), 47:1–47:33.

Shinde, S. K. & Kulkarni, U. (2012), 'Hybrid personalized recommender system using centering-bunching based clustering algorithm', *Expert Systems with Applications* **39**(1), 1381–1387.

Sinha, R. R. & Swearingen, K. (2001), Comparing recommendations made by online systems and friends., *in* 'DELOS Workshop: Personalisation and Recommender Systems in Digital Libraries'.

Song, L., Smola, A., Gretton, A., Bedo, J. & Borgwardt, K. (2012), 'Feature selection via dependence maximization', *Journal of Machine Learning Research* **13**(1), 1393–1434.

Su, X., Greiner, R., Khoshgoftaar, T. M. & Zhu, X. (2007), Hybrid collaborative filtering algorithms using a mixture of experts, *in* 'Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence', WI '07, IEEE Computer Society, Washington, DC, USA, pp. 645–649.

Su, X. & Khoshgoftaar, T. M. (2009), 'A survey of collaborative filtering techniques', *Advances in Artificial Intelligence* **2009**, 421425:1–421425:19.

Sun, Y. & Han, J. (2012), *Mining Heterogeneous Information Networks: Principles and Methodologies*, Morgan & Claypool Publishers.

Takács, G., Pilászy, I., Németh, B. & Tikk, D. (2009), 'Scalable collaborative filtering approaches for large recommender systems', *Journal of Machine Learning Research* **10**, 623–656.

Tang, J., Gao, H., Hu, X. & Liu, H. (2013), Exploiting homophily effect for trust prediction, *in* 'Proceedings of the Sixth ACM International Conference on Web Search and Data Mining', WSDM '13, ACM, New York, NY, USA, pp. 53–62.

Tang, J., Gao, H. & Liu, H. (2012), mtrust: discerning multi-faceted trust in a connected world., *in* E. Adar, J. Teevan, E. Agichtein & Y. Maarek, eds, 'Proceedings of the Fifth International Conference on Web Search and Web Data Mining, WSDM 2012, Seattle, WA, USA, February 8-12, 2012', ACM, pp. 93–102.

Tang, J., Hu, X. & Liu, H. (2013), 'Social recommendation: a review', *Social Network Analysis and Mining* **3**(4), 1113–1133.

Tang, L. & Liu, H. (2010), *Community Detection and Mining in Social Media*, Synthesis Lectures on Data Mining and Knowledge Discovery, Morgan & Claypool Publishers.

Taskar, B., Abbeel, P., Wong, M.-F. & Koller, D. (2007), Relational markov networks, *in* L. Getoor & B. Taskar, eds, 'Introduction to Statistical Relational Learning', MIT Press.

The 2013 International Joint Conference on Neural Networks, IJCNN 2013, Dallas, TX, USA, August 4-9, 2013 (2013), IEEE.

Tillman, R. E. (2009), Structure learning with independent non-identically distributed data, *in* 'Proceedings of the 26th Annual International Conference on Machine Learning', ICML '09, ACM, New York, NY, USA, pp. 1041–1048.

Troncy, R., Huet, B. & Schenk, S. (2011), *Multimedia semantics : metadata, analysis and interaction*, Wiley, Chichester, West Sussex, U.K.

Tse, Y. K. & Tsui, A. (2002), 'A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations', *Journal of Business & Economic Statistics* **20**(3), 351–62.

Vozalis, M. G. & Margaritis, K. G. (2007), 'Using svd and demographic data for the enhancement of generalized collaborative filtering', *Information Sciences* **177**(15), 3017–3037.

Wang, C. & Blei, D. M. (2011), Collaborative topic modeling for recommending scientific articles, *in* C. Apté, J. Ghosh & P. Smyth, eds, 'Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA, USA, August 21-24, 2011', ACM, pp. 448–456.

Wang, C. & Cao, L. (2012), Modeling and analysis of social activity process, *in* L. Cao & P. S. Yu, eds, 'Behavior Computing', Springer London, pp. 21–35.

Wang, C., Cao, L., Wang, M., Li, J., Wei, W. & Ou, Y. (2011), Coupled nominal similarity in unsupervised learning, *in* C. Macdonald, I. Ounis & I. Ruthven, eds, 'Proceedings of the 20th ACM Conference on Information and Knowledge Management, CIKM 2011', ACM, pp. 973–978.

Wang, C., She, Z. & Cao, L. (2013*a*), Coupled attribute analysis on numerical data., *in* Rossi (2013).

Wang, C., She, Z. & Cao, L. (2013*b*), Coupled clustering ensemble: Incorporating coupling relationships both between base clusterings and objects, *in* C. S. Jensen, C. M. Jermaine & X. Zhou, eds, '29th IEEE International Conference on Data Engineering, ICDE 2013', IEEE Computer Society, pp. 374–385.

Wasserman, S. & Faust, K. (1994), *Social network analysis: Methods and applications*, Vol. 8, Cambridge university press.

Webb, G. I., Liu, B., Zhang, C., Gunopulos, D. & Wu, X., eds (2010), *ICDM 2010, The 10th IEEE International Conference on Data Mining, Sydney, Australia, 14-17 December 2010*, IEEE Computer Society.

Wei, W., Fan, X., Li, J. & Cao, L. (2012), Model the complex dependence structures of financial variables by using canonical vine, *in* X. Chen, G. Lebanon, H. Wang & M. J. Zaki, eds, '21st ACM International Conference on Information and Knowledge Management, CIKM'12, Maui, HI, USA, October 29 - November 02, 2012', ACM, pp. 1382–1391.

Weng, J., Lim, E.-P., Jiang, J. & He, Q. (2010), Twitterrank: Finding topic-sensitive influential twitterers, *in* 'Proceedings of the Third ACM International Conference on Web Search and Data Mining', WSDM '10, ACM, New York, NY, USA, pp. 261–270.

Wu, H. C., Luk, R. W. P., Wong, K. F. & Kwok, K. L. (2008), 'Interpreting tf-idf term weights as making relevance decisions', *ACM Transactions on Information Systems* **26**(3), 13:1–13:37.

Wu, L., Chen, E., Liu, Q., Xu, L., Bao, T. & Zhang, L. (2012), Leveraging tagging for neighborhood-aware probabilistic matrix factorization, *in* 'Proceedings of the 21st ACM International Conference on Information and Knowledge Management', CIKM '12, ACM, New York, NY, USA, pp. 1854–1858.

Xu, Z., Tresp, V., Yu, K. & Kriegel, H. P. (2006), Infinite hidden relational models, *in* 'Proceedings of the 22nd International Conference on Uncertainity in Artificial Intelligence (UAI 2006)', AUAI Press, Cambridge, MA, USA.

Yager, R. R. (2003), 'Fuzzy logic methods in recommender systems', *Fuzzy Sets and Systems* **136**(2), 133–149.

Yang, X., Steck, H. & Liu, Y. (2012), Circle-based recommendation in online social networks, *in* Q. Yang, D. Agarwal & J. Pei, eds, 'The 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining', ACM, pp. 1267–1275.

Yao, Z. & Zhang, Q. (2009), Item-based clustering collaborative filtering algorithm under high-dimensional sparse data, *in* 'Proceedings of the 2009 International Joint Conference on Computational Sciences and Optimization - Volume 01', CSO '09, IEEE Computer Society, Washington, DC, USA, pp. 787–790.

Ye, M., Liu, X. & Lee, W.-C. (2012), Exploring social influence for recommendation: a generative model approach, *in* W. R. Hersh, J. Callan, Y. Maarek & M. Sanderson, eds, 'The 35th International ACM SIGIR conference on research and development in Information Retrieval, SIGIR '12', ACM, pp. 671–680.

Yu, Y., Wang, C., Gao, Y., Cao, L. & Chen, Q. (2013), A coupled clustering approach for items recommendation, *in* J. Pei, V. S. Tseng, L. Cao, H. Motoda & G. Xu, eds, 'Advances in Knowledge Discovery and Data

Mining, 17th Pacific-Asia Conference, PAKDD 2013', Vol. 7819 of *Lecture Notes in Computer Science*, Springer, pp. 365–376.

Yuan, Q., Chen, L. & Zhao, S. (2011), Factorization vs. regularization: Fusing heterogeneous social relationships in top-n recommendation, *in* 'Proceedings of the Fifth ACM Conference on Recommender Systems', RecSys '11, ACM, New York, NY, USA, pp. 245–252.

Zhang, H., Zhao, Y., Cao, L. & Zhang, C. (2008), Combined association rule mining., *in* T. Washio, E. Suzuki, K. M. Ting & A. Inokuchi, eds, 'Advances in Knowledge Discovery and Data Mining, 12th Pacific-Asia Conference, PAKDD 2008, Osaka, Japan, May 20-23, 2008 Proceedings', Vol. 5012 of *Lecture Notes in Computer Science*, Springer, pp. 1069–1074.

Zhang, S., Wang, W., Ford, J., Makedon, F. & Pearlman, J. (2005), Using singular value decomposition approximation for collaborative filtering, *in* 'In Proceedings of the 7th IEEE Conference on E-Commerce', IEEE Computer Society, pp. 257–264.

Zhang, X., Song, L., Gretton, A. & Smola, A. (2009), Kernel measures of independence for non-iid data, *in* 'Advances in neural information processing systems 21', Max-Planck-Gesellschaft, Curran, Red Hook, NY, USA, pp. 1937–1944.

Zhao, Y., Zhang, H., Figueiredo, F., Cao, L. & Zhang, C. (2007), Mining for combined association rules on multiple datasets, *in* 'Proceedings of the 2007 International Workshop on Domain Driven Data Mining', DDDM '07, ACM, New York, NY, USA, pp. 18–23.

Zhao, Y., Zhang, H., Wu, S., Pei, J., Cao, L., Zhang, C. & Bohlscheid, H. (2009), Debt detection in social security by sequence classification using both positive and negative patterns., *in* W. L. Buntine, M. Grobelnik, D. Mladenic & J. Shawe-Taylor, eds, 'Machine Learning and Knowledge Discovery in Databases, European Conference, ECML PKDD 2009, Bled, Slovenia, September 7-11, 2009, Proceedings, Part II', Vol. 5782 of *Lecture Notes in Computer Science*, Springer, pp. 648–663.

Zhong, J. & Li, X. (2010), 'Unified collaborative filtering model based on combination of latent features', *Expert Systems with Applications* **37**(8), 5666–5672.

Zhou, Z.-H., Sun, Y.-Y. & Li, Y.-F. (2009), Multi-instance learning by treating instances as non-i.i.d. samples, *in* 'Proceedings of the 26th Annual International Conference on Machine Learning', ICML '09, ACM, New York, NY, USA, pp. 1249–1256.

Ziegler, C.-N., McNee, S. M., Konstan, J. A. & Lausen, G. (2005), Improving recommendation lists through topic diversification, *in* 'Proceedings of the 14th international conference on World Wide Web', WWW '05, ACM, New York, NY, USA, pp. 22–32.