

Faculty of Engineering and Information Technology  
University of Technology Sydney

# **Non-IID Latent Variable Models**

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

by

Trong Dinh Thac Do

February 2019



## **CERTIFICATE OF ORIGINAL AUTHORSHIP**

I, Trong Dinh Thac Do declare that this thesis, is submitted in fulfilment of the requirements for the award of the Ph.D. degree, in the Faculty of Engineering and Information Technology at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise reference or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

Signature:

Date:



# Acknowledgments

Foremost, I would like to express the deepest appreciation to my supervisor, Professor Longbing Cao for the continuous support of my Ph.D. study and research, for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my Ph.D. study.

In addition, I would like to thank all colleagues in Advanced Analytics Institute: Guansong Pang, Shoujin Wang, Chengzhang Zhu, Songlei Jian, Liang Hu and Wei Wang for the stimulating discussions, for the sleepless nights we were working together before deadlines, and for all the fun we have had in the last three years and a half.

Last but not least, I would like to thank my family: my parents, my parents in law, my wife, my dear son, my younger sister, her husband and her two children for their unconditional support, both financially and emotionally throughout the Ph.D. study.

Trong Dinh Thac Do

February 2019 @ UTS



# Contents

Certificate . . . . .	i
Acknowledgment . . . . .	iii
List of Figures . . . . .	xiii
List of Tables . . . . .	xix
List of Publications . . . . .	xxiii
Abstract . . . . .	xxv
<b>Chapter 1 Introduction . . . . .</b>	<b>1</b>
1.1 Research Aims and Objectives . . . . .	3
1.2 Thesis Contributions . . . . .	4
1.3 Thesis Organization . . . . .	7
<b>Chapter 2 Preliminaries and Literature Survey . . . . .</b>	<b>15</b>
2.1 Bayesian Nonparametric (BNP) Methods . . . . .	16
2.1.1 Chinese Restaurant Process (CRP) . . . . .	16
2.1.2 Indian Buffet Process (IBP) . . . . .	17
2.1.3 Stick-breaking Construction . . . . .	18
2.2 Latent Variable Model (LVM) . . . . .	19
2.2.1 Latent Class Model (LCM) . . . . .	19
2.2.2 Latent Feature Model (LFM) . . . . .	22
2.2.3 Latent Factor Model - Matrix Factorization . . . . .	25

2.3	Preliminaries on Inference Method . . . . .	29
2.3.1	Markov Chain Monte Carlo (MCMC) . . . . .	30
2.3.2	Variational Inference (VI) . . . . .	32
2.4	Related Work on Non-IID Learning . . . . .	34
<b>I</b>	<b>Non-IID Latent Class Models</b>	<b>36</b>
<b>Chapter 3</b>	<b>Coupled Attributes-dependent Mondrian Process for Both Static and Dynamic Infinite Relational Learning . . . . .</b>	<b>37</b>
3.1	Introduction . . . . .	37
3.2	Coupled Attributes-dependent Mondrian Process . . . . .	42
3.2.1	Probabilistic Cutting Space . . . . .	42
3.2.2	Coupled Node Similarity . . . . .	44
3.2.3	The Static Model - sCAMP . . . . .	47
3.2.4	The Dynamic Model - dCAMP . . . . .	51
3.3	Experiments . . . . .	53
3.3.1	Datasets . . . . .	53
3.3.2	Baselines . . . . .	54
3.3.3	Evaluation Metrics . . . . .	54
3.3.4	Performance of Static Model sCAMP . . . . .	55
3.3.5	Performance of Dynamic Model dCAMP - The effi- ciency of modelling dynamic in the network data. . . . .	60
3.4	Summary . . . . .	61

---

<b>II</b>	<b>Non-IID Latent Feature Models</b>	<b>64</b>
<b>Chapter 4</b>	<b>Infinite Latent Feature Model by Learning Both Value-Attribute-Node Couplings and Latent Feature Relations</b>	<b>65</b>
4.1	Introduction	65
4.2	The VLI Model	73
4.2.1	Problem Statement	73
4.2.2	The VLI framework	74
4.2.3	The Technical Design of VLI	78
4.2.4	Inference	83
4.3	Experiments and Analysis	86
4.3.1	Datasets and Baseline Methods	86
4.3.2	Performance on Link Prediction with vs. without Node Metadata	89
4.3.3	Impact of Integrating Node Metadata with Latent Feature Interactions	91
4.3.4	Impact of Selecting Node Interactions in Node Metadata	92
4.4	Summary	94
<b>Chapter 5</b>	<b>HDIM: A Heterogeneous Data-driven Infinite Model for Learning Hierarchical Relations in Time-varying Attributed Networks</b>	<b>96</b>
5.1	Introduction	96
5.2	Motivation and Gap Analysis	100
5.2.1	Motivating Example	100
5.3	The HDIM Working Mechanism	101
5.3.1	The Framework	101

5.3.2	Attribute-based Coupled Node Similarity . . . . .	104
5.3.3	Modeling Node-to-Community Interactions . . . . .	105
5.3.4	The HDIM Model . . . . .	107
5.4	Experiments and Analysis . . . . .	108
5.4.1	Datasets . . . . .	109
5.4.2	Accuracy Evaluation . . . . .	110
5.4.3	Case Study . . . . .	112
5.5	Summary . . . . .	114

### **III Non-IID Latent Factor Models - Matrix Factorization** **115**

<b>Chapter 6</b>	<b>Metadata-dependent Infinite Poisson Factorization for Efficiently Modelling Sparse and Large Matrices in Recommendation . . . . .</b>	<b>116</b>
6.1	Introduction . . . . .	116
6.2	The MPF/MIPF Models . . . . .	119
6.2.1	Integrating Metadata into PF - The MPF Model . . . . .	119
6.2.2	Taking the Infinite - The MIPF Model . . . . .	121
6.3	Inference . . . . .	123
6.3.1	Variational Inference for MPF . . . . .	123
6.3.2	Variational Inference for MIPF . . . . .	128
6.4	MPF/MIPF Properties . . . . .	130
6.5	Experiments . . . . .	131
6.5.1	Experimental Settings . . . . .	131
6.5.2	Result Evaluation . . . . .	133
6.6	Summary . . . . .	136

<b>Chapter 7 Bayesian Nonparametric Metadata-integrated Coupled Poisson Factorization for Scalable Recommendations . . . . .</b>	<b>137</b>
7.1 Introduction . . . . .	137
7.2 The CPF and mCPF Models . . . . .	141
7.2.1 Coupled User Poisson Factorization (CuPF) . . . . .	141
7.2.2 Coupled Item Poisson Factorization (CiPF) . . . . .	144
7.2.3 Integrating Metadata to Coupled User Poisson Factorization (mCuPF) . . . . .	146
7.2.4 Integrating Metadata into Coupled Item Poisson Factorization (mCiPF) . . . . .	147
7.3 Bayesian Nonparametric Metadata-integrated Coupled Poisson Factorization (BNPmCPF) . . . . .	148
7.4 Inference for CPF/mCPF/BNPmCPF . . . . .	150
7.4.1 Variational Inference for CuPF . . . . .	150
7.4.2 Integrating User/Item Metadata into Coupled User Poisson Factorization (mCuPF) . . . . .	152
7.4.3 Variational Inference for BNPmCPF . . . . .	154
7.5 Properties of CPF/mCPF/BNPmCPF . . . . .	158
7.6 Empirical Results . . . . .	160
7.7 Summary . . . . .	165
<b>Chapter 8 Gamma-Poisson Dynamic Matrix Factorization Embedded with Metadata Influence . . . . .</b>	<b>168</b>
8.1 Introduction . . . . .	168
8.2 The mGDMF Model . . . . .	170
8.3 Stochastic Variational Inference for mGDMF . . . . .	175

8.4	Experiments . . . . .	181
8.4.1	Datasets . . . . .	181
8.4.2	Baseline Methods . . . . .	183
8.4.3	Experiment Settings and Evaluation . . . . .	184
8.4.4	Results . . . . .	185
8.5	Summary . . . . .	192
<b>Chapter 9</b>	<b>Group-based Gamma-Poisson Matrix Factorization on Multi-source, Large and Sparse Recommendation Data . . . . .</b>	<b>193</b>
9.1	Introduction . . . . .	193
9.2	The $G^2MF$ Model . . . . .	198
9.3	$G^2MF$ Inference . . . . .	202
9.3.1	Batch Inference for $G^2MF$ . . . . .	202
9.3.2	Stochastic Variational Inference for $G^2MF$ . . . . .	205
9.3.3	Computational efficiency . . . . .	207
9.4	Empirical Results . . . . .	208
9.4.1	Baseline Methods . . . . .	208
9.4.2	Datasets . . . . .	209
9.4.3	Experimental Settings and Evaluation . . . . .	210
9.4.4	Results . . . . .	212
9.5	Summary . . . . .	215
<b>Chapter 10</b>	<b>Conclusions and Future Directions . . . . .</b>	<b>217</b>
10.1	Conclusions . . . . .	217
10.1.1	Non-IID Latent Class Models (non-IID LCM) . . . . .	218
10.1.2	Non-IID Latent Feature Models (non-IID LFM) . . . . .	218

10.1.3 Non-IID Latent Factor Model - Matrix Factorization (non-IID LFM-MF) . . . . .	219
10.2 Future Directions . . . . .	220
10.2.1 Apply to Other Applications . . . . .	221
10.2.2 Integrate More Sources of Side Information . . . . .	221
10.2.3 Online Applications . . . . .	221
<b>Chapter A Abbreviations . . . . .</b>	<b>222</b>
<b>Chapter B Notations . . . . .</b>	<b>224</b>
<b>Chapter C Basic Distributions . . . . .</b>	<b>225</b>
C.1 Gaussian Normal Distribution . . . . .	225
C.2 Dirichlet Distribution . . . . .	225
C.3 Beta Distribution . . . . .	226
C.4 Bernoulli Distribution . . . . .	226
C.5 Gamma Distribution . . . . .	226
C.6 Poisson Distribution . . . . .	226
C.7 Multinomial Distribution . . . . .	227
<b>Chapter D Derivation of mGDMF . . . . .</b>	<b>228</b>
D.1 Derivation of Metadata Integration . . . . .	228
D.2 Derivation of User Latent Preference and Item Latent Attractiveness . . . . .	230
D.3 Derivation of the Rating . . . . .	231
D.4 Derivation of the Global Static Factor of User/Item . . . . .	232
D.5 Derivation of the Local Dynamic Factor of User/Item . . . . .	233

CONTENTS

---

**Chapter E Derivations of  $G^2MF$  . . . . . 236**

E.1 Derivation of the Rating . . . . . 236

E.2 Derivation of User-group Preference . . . . . 237

E.3 Derivation of Item-group Preference . . . . . 238

E.4 Derivation of the Weight of Latent Group of User . . . . . 238

E.5 Derivation of the Weight of Latent Group of Item . . . . . 239

E.6 Derivation of Latent User Influence . . . . . 239

E.7 Derivation of Latent Item Influence . . . . . 240

E.8 Derivation of Latent User Preference . . . . . 240

E.9 Derivation of Latent Item Feature . . . . . 241

**Bibliography . . . . . 242**

# List of Figures

1.1	Thesis Structure . . . . .	14
2.1	Chinese Restaurant Process (CRP). Squares are the tables and Circles represent for the customers. . . . .	17
2.2	Indian Buffet Process (IBP). . . . .	18
2.3	Stick-breaking Construction. . . . .	19
2.4	Mondrian Process (MP) . . . . .	22
2.5	The Latent Feature Relational Model (LFRM). . . . .	24
2.6	Graphical Model of Poisson Factorization (PF). . . . .	27
3.1	Block modeling on a synthetic relational data set: (a) Original data, (b) regular block structure, (c) irregular block structure as in the Mondrian Process, and (d) irregular block structure as in our proposed model. . . . .	38
3.2	Illustration of block representation on the <i>probabilistic cutting space</i> (PCS) in sCAMP: (a) The blocks from MP, (b) the blocks in PCS transformed from MP's cutting space, (c) the node density based on coupled node similarities, and (d) the <i>coupled node similarity</i> -enhanced PCS blocks after rescaling. . . . .	43
3.3	Graphical Models of sCAMP and dCAMP. . . . .	47

*LIST OF FIGURES*

---

3.4 Illustration of PCS construction and CNS integration: (a) The blocks from MP, (b) the kd-tree of the MP blocks, (c) PCS-based blocks built on the MP’s blocks, (d) the MP blocks embedded with CNS, and (e) the refined PCS-based blocks after integrating CNS. . . . . 48

3.5 Block Structure on the Douban dataset. . . . . 56

3.6 Trace plot of Log-likelihood (LL) versus the number of iterations in MP, MDMP-G and sCAMP-GAL on the Lazaga dataset. 60

3.7 Trace plot of Log-likelihood (LL) versus the number of iterations in MP, MDMP and sCAMP on the Douban dataset. . . 61

3.8 Block structures on the Social Patterns dataset: (a) dIRM, (b) dCAMP without integrating attributes (dCAMP\_wo\_A), (c) CNS-based block weight, and (d) dCAMP integrated with attributes (dCAMP\_w\_A). Darker color means stronger relations in a block. . . . . 62

4.1 Latent Class Model - Non-Overlapping classes . . . . . 68

4.2 Latent Class Model - Overlapping classes . . . . . 69

4.3 Latent Feature Model . . . . . 70

4.4 The state-of-the-art models. . . . . 71

4.5 VLI: Integrating Value Coupling, Attribute Coupling and Metadata-based Node Relations to Latent Feature Interactions. . . . . 72

4.6 VLI Framework . . . . . 75

4.7 The mVLI main process. . . . . 77

4.8 The graphical model of mVLI . . . . . 78

4.9 The aVLI main process. . . . . 79

4.10 The graphical model of aVLI . . . . . 80

4.11 Log-likelihood and AUC by LFRM, niLF, mVLI and aVLI w.r.t. network density on Synthetic_2 data. . . . .	92
4.12 Log-likelihood and AUC by LFRM, niLF and VLI w.r.t. net- work density on Sampson data. . . . .	93
5.1 Detecting communities in the toy data: (a) using node behav- iors at time $t_2$ (e.g., by LFRM); (b) combining node behaviors at $t_2$ with behaviors at the past time $t_1$ (e.g., by DRIFT); and (c) integrating node attributes and social relationships with dynamic behaviors (our model HDIM). . . . .	101
5.2 The HDIM framework: Dynamic infinite models for learning hierarchical relations in heterogeneous data sources. . . . .	103
5.3 Link prediction with sample data from Twitter API data: (a) - (c) the observed data represented by LFRM from time pe- riod $t_1$ to $t_3$ (white dots represent the linkage between nodes while black dots indicate no linkage); (d) social relationships data; (e) ACNS based on node attributes; (f) the link predic- tion result using LFRM (green dots are the correctly predicted linkage, red dots are the incorrectly predicted linkage); (g) the result of DRIFT; (h) the result of HDIM. . . . .	113
6.1 The graphical models of MPF and MIPF. . . . .	122
6.2 Top-20 recommendations compared with baselines. . . . .	133
6.3 Performance of top-20 recommendations made by finite model MPF and infinite model MIPF. . . . .	134
6.4 Example of MIPF in handling sparse items in comparison with HCPF. . . . .	135

*LIST OF FIGURES*

---

7.1	Graphical Models of (a) Coupled User Poisson Factorization and (b) Coupled Item Poisson Factorization. . . . .	145
7.2	Graphical Models of (a) Metadata-integrated Coupled User Poisson Factorization and (b) Metadata-integrated Coupled Item Poisson Factorization. . . . .	147
7.3	Graphical Models of (a) Bayesian Nonparametric Metadata-integrated Coupled User Poisson Factorization and (b) Bayesian Nonparametric Metadata-integrated Coupled Item Poisson Factorization. . . . .	148
7.4	Predictive Performance w.r.t. Top $N$ Recommendations on Four Datasets ( $N = 10, 20, 30, 40$ ). . . . .	163
7.5	Rating Popularity and Sparsity Test. . . . .	164
7.6	Top-10 Recommendations for User 184. . . . .	165
7.7	Performance of top-20 recommendations made by finite model mCPF and infinite model BNPmCPF. . . . .	167
8.1	Graphical Model of mGDMF . . . . .	174
8.2	Top-50 Recommendations Compared with Baselines. . . . .	185
8.3	Percentage (%) of Sparse Items Recommended Precisely for 10 Users by mGDMF and DCPF. . . . .	189
8.4	Percentage (%) of sparse items recommended precisely for 10 users by mGDMF and DCPF. . . . .	190
8.5	Analysis on two users ‘ $U270$ ’ and ‘ $U437$ ’ with the same metadata in Last.fm. The number of times that users listened to two ‘rock’ and ‘pop’ tracks with 16 time slices is shown on the left. The distribution of the number of times that $U270$ and $U437$ listened to top 10 ‘rock’ and ‘pop’ tracks and the top 10 precise recommendations by mGDMF are shown on the right. . . . .	191

8.6 The number of times listened to ‘Zombie’ track by two users ‘U270’ and ‘U437’ through 16 time slices. . . . . 191

9.1 A Toy Example on Multi-source, Sparse Recommendation Data:  
 (a) Sparse items (e.g., ‘Goodfellas’), Cold-start items (e.g., ‘Alien’), Cold-start users (e.g., user  $u_8$ ); (b) User friendship; (c) and User metadata. . . . . 195

9.2 Graphical Model of G<sup>2</sup>MF.  $\Gamma$  and  $P$  represent Gamma distribution and Poisson distribution. . . . . 200

9.3 User/item-group Modeling Effect of G<sup>2</sup>MF Compared to Baselines on 11 Datasets w.r.t. *Recall-at-20* ( $R@20$ ). rG<sup>2</sup>MF only models groups of ratings. uG<sup>2</sup>MF combines user/item group modeling with integrating user relations. iG<sup>2</sup>MF combines user/item group modeling with integrating item relations. mG<sup>2</sup>MF combines user/item group modeling with integrating user/item metadata. . . . . 213

9.4 Percentage (%) of Sparse Items Precisely Recommended for 15 Users by SPF, MIPF, mCiPF and G<sup>2</sup>MF on Last.fm Dataset. 214



# List of Tables

3.1	A Toy Example . . . . .	45
3.2	Performance on the Douban dataset w.r.t. Log-likelihood (LL), Bayesian Information Criterion (BIC), and Area Under ROC Curve (AUC). . . . .	55
3.3	Performance on the Lazega dataset w.r.t. Log-likelihood (LL). There are seven different attributes: ‘status’ (S), ‘gender’ (G), ‘office’ (O), ‘years with firm’ (Y), ‘age’ (A), ‘practice’ (P) and ‘law school’ (L). . . . .	57
3.4	Performance on the Lazega dataset w.r.t. Bayesian Information Criterion (BIC). There are seven different attributes: ‘status’ (S), ‘gender’ (G), ‘office’ (O), ‘years with firm’ (Y), ‘age’ (A), ‘practice’ (P) and ‘law school’ (L). . . . .	58
3.5	Performance on the Lazega dataset w.r.t. Area Under ROC Curve (AUC). There are seven different attributes: ‘status’ (S), ‘gender’ (G), ‘office’ (O), ‘years with firm’ (Y), ‘age’ (A), ‘practice’ (P) and ‘law school’ (L). . . . .	59
3.6	Performance on the Social Patterns dataset w.r.t. Log-likelihood (LL), Bayesian Information Criterion (BIC), and Area Under ROC Curve (AUC). . . . .	62

LIST OF TABLES

---

4.1 A toy example of node linkage and metadata. . . . . 66

4.2 Performance on link prediction w.r.t. log-likelihood of node linkage. . . . . 90

4.3 Performance on link prediction w.r.t. AUC of node linkage . . 91

4.4 The impact of selecting node attributes and interactions (AUC of mVLI/aVLI) on Lazega dataset. The first seven lines are the AUC of mVLI with each attribute of metadata. The eighth line is the AUC of mVLI with all attributes, and the ninth line is AUC with chosen attributes. The last line is the AUC of aVLI together with the weight of each attribute learned from the model at the last iteration. . . . . 94

5.1 Toy example: Multiple heterogeneous aspects of data in time-varying social networks at two time points  $t_1$  and  $t_2$  (Note: Values in bold face show the changes). . . . . 99

5.2 Performance of link prediction for Synthetic, Enron, Social Patterns (SP), Higgs Twitter (HT) and Twitter API (TA) w.r.t. log-likelihood (LL) and AUC. . . . . 111

7.1 Examples of Item-oriented Rating Popularity and Sparsity . . 139

8.1 Latent Variables, Type, Variational Variables and Variational Update for Users (i.e.,  $hu_m, \xi_u, \bar{\theta}_{uk}, \theta_{uk}, \lambda_{uk,t}$ .  $\aleph_m$  is the number of users having the  $m^{th}$  attribute,  $K$  is the number of latent components, and  $\Psi(\cdot)$  is the *digamma* function. The Gamma distribution is parameterized by *shape* (*shp*) and *rate* (*rte*). . . . . 177

8.2	Latent Variables, Type, Variational Variables and Variational Update for Items (i.e., $h_{i_n}$ , $\eta_i$ , $\bar{\beta}_{ik}$ , $\beta_{ik}$ , $\iota_{ik,t}$ ). $\chi_n$ is the number of items having the $n^{th}$ attribute and $K$ is the number of latent components. The Gamma distribution is parameterized by <i>shape</i> ( <i>shp</i> ) and <i>rate</i> ( <i>rte</i> ). . . . .	178
8.3	Predictive Performance on Five Datasets w.r.t. NDCG and AUC. . . . .	187
9.1	User group’s preferences based on the multi-source data in Figure 9.1: rating table, user friendship and user metadata. . .	196
9.2	Predictive Performance on Last.fm w.r.t. <i>Precision-at-20</i> ( $P@20$ ), <i>Recall-at-20</i> ( $R@20$ ), AUC and NDCG. SPF only integrates user friendship. MPF, MIPF, mCuPF and mCiPF only integrate item metadata. $G^2$ MF integrates both item metadata and user friendship. . . . .	214
9.3	Predictive Performance on Four Amazon Datasets w.r.t. <i>Precision-at-20</i> ( $P@20$ ), <i>Recall-at-20</i> ( $R@20$ ), AUC and NDCG. $C^2$ PF only integrates item relations. MPF, MIPF, mCuPF and mCiPF only integrate item metadata. $G^2$ MF integrates both item metadata and item relations. . . . .	216



# List of Publications

## Published Papers

- C-1. **Trong Dinh Thac Do**, and Longbing Cao, “Gamma-Poisson Dynamic Matrix Factorization Embedded with Metadata Influence,” *The Thirty-second Conference on Neural Information Processing Systems (NIPS-18)*, 2018. Accepted.
- C-2. **Trong Dinh Thac Do**, and Longbing Cao, “Coupled Poisson Factorization Integrated with User/Item Metadata for Modeling Popular and Sparse Ratings in Scalable Recommendation,” *The Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18)*, 2018.
- C-3. **Trong Dinh Thac Do**, and Longbing Cao, “Metadata-dependent Infinite Poisson Factorization for Efficiently Modelling Sparse and Large Matrices in Recommendation,” *The 27th International Joint Conference on Artificial Intelligence (IJCAI-18)*, 2018.

## To be Submitted/Under Reviewed Papers

- J-1. **Trong Dinh Thac Do**, and Longbing Cao, “Coupled Attributes-dependent Mondrian Process for Both Static and Dynamic Infinite Relational Learning,” *Submitted to Machine Learning Journal*. Under Reviewed.

*LIST OF PUBLICATIONS*

---

- J-2. **Trong Dinh Thac Do**, and Longbing Cao, “Bayesian Nonparametric Metadata-integrated Coupled Poisson Factorization for Scalable Recommendations.” To be Submitted.
- J-3. **Trong Dinh Thac Do**, and Longbing Cao, “Explicit and Implicit Relations-based Infinite Latent Feature Learning for Link Prediction”. To be Submitted.
- C-4. **Trong Dinh Thac Do**, and Longbing Cao, “Group-based Gamma-Poisson Matrix Factorization on Multi-source, Large and Sparse Recommendation Data.” To be Submitted.
- C-5. **Trong Dinh Thac Do**, and Longbing Cao, “HDIM: A Heterogeneous Data-driven Infinite Model for Learning Hierarchical Relations in Time-varying Attributed Networks”. To be Submitted.

# Abstract

Latent Variable Model (LVM) is the statistical model that aims to uncover hidden information behind data. These models have been widely used for real-world applications such as community detection, link prediction or recommender systems. However, LVM faces significant challenges in modeling complex relations since LVM assumes that the data are independent and identically distributed (IID). However, real-world data are often coupled in terms of object attributes, object relations, or even hidden variable relations. For example, in social networks, users that indicate a similar ‘age’, ‘location’ and ‘high school’ are often friends. To this end, non-IID learning has the potential to describe the above hierarchical relations in real-world data which are typically not independent or identically distributed (non-IID).

In this thesis, we are interested in determining the relations behind observations and hidden variables in LVM. More specifically, we focus on coupling relations in non-IID data in terms of various LVM, including Latent Class Model (LCM), Latent Feature Model (LFM), and Latent Factor Model-Matrix Factorization (LFM-MF). In particular, we aim to model the following relations: (1) relations between attributes in observed data (e.g., user/item metadata such as ‘location’ of a user or ‘genre’ of a movie); (2) relations between different sources of observed data (e.g., metadata and user’s friendships); and (3) relations between latent variables in LVM. We also apply

## *ABSTRACT*

---

Bayesian Nonparametric (BNP) techniques to the proposed LVM models to automatically tune the number of latent variables in LVM for efficient computation. Furthermore, to work with large and sparse data, we introduce several methods for better inference of the proposed LVM models.

The empirical analysis of both proposed models reveals that our models significantly outperform state-of-the-art models in the same family. Together with improved optimization techniques (i.e., BNP and inference methods), our proposed models indicate their potential for online modeling of large, sparse data.