

*C02029: Doctor of Philosophy*

*CRICOS Code: 009469A*

*Subject Code: 32903*

*October 2019*

# *Non-IID Representation Learning on Complex Categorical Data*

---

*Chengzhang Zhu*

School of Computer Science

Faculty of Engineering and Information Technology

University of Technology Sydney

NSW - 2007, Australia



---

---

# Non-IID Representation Learning on Complex Categorical Data

---

---

*A thesis submitted in partial fulfilment of the requirements  
for the degree of*

Doctor of Philosophy

*in*

Analytics

*by*

**Chengzhang Zhu**

*to*

School of Computer Science

Faculty of Engineering and Information Technology

University of Technology Sydney

NSW - 2007, Australia

October 2019



## ABSTRACT

Learning complex categorical data requires proper vector or metric representations of the intricate characteristics of that data. Existing methods for categorical data representation usually assume data is independent and identically distributed (IID). However, real-world data is often hierarchically associated with diverse couplings and heterogeneities (i.e., non-IIDness, e.g., various couplings such as value co-occurrences and attribute correlation and dependency, as well as heterogeneities such as heterogeneous distributions or complementary and inconsistent relations). Existing methods either capture only some of these couplings and heterogeneities or simply assume IID data in building their representations.

This thesis aims to deeply understand and effectively represent non-IIDness in categorical data. Specifically, it focuses on (1) modeling heterogeneous couplings within and between attributes in categorical data; (2) disentangling attribute couplings with a mixture of heterogeneous distributions; (3) hierarchically learning heterogeneous couplings; (4) integrating complementary and inconsistent heterogeneous couplings; and (5) adaptively identifying and learning dynamic couplings and heterogeneities.

Accordingly, this thesis proposes (1) a non-IID similarity metrics learning framework to model complex interactions within and between attributes in non-IID categorical data; (2) a decoupled non-IID learning framework to capture and embed heterogeneous distributions in non-IID categorical data with bounded information loss; (3) a heterogeneous metric learning method with hierarchical couplings to learn and integrate the heterogeneous dependencies and distributions in non-IID categorical data into a representation of a similarity metric; (4) an unsupervised heterogeneous coupling learning approach to integrate the complementary and inconsistent heterogeneous couplings in non-IID categorical data; and (5) an unsupervised hierarchical and heterogeneous coupling learning method to learn hierarchical and heterogeneous couplings on dynamic non-IID categorical data.

Theoretical analyses support the effectiveness of the proposed methods and bound the information loss in their generated high-quality representations. Extensive experiments demonstrate that the proposed non-IID representation methods for complex categorical data perform significantly better than state-of-the-art methods in terms of multiple downstream learning tasks and representation-quality evaluation metrics.



## AUTHOR'S DECLARATION

I, *Chengzhang Zhu* declare that this thesis, is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the *School of Computer Science, Faculty of Engineering and Information Technology* at the University of Technology Sydney. This thesis is wholly my own work unless otherwise referenced 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. This research is supported by the Australian Government Research Training Program.

Production Note:

SIGNATURE: Signature removed prior to publication.

Chengzhang Zhu

DATE: 22<sup>nd</sup> October, 2019

PLACE: Sydney, Australia





## DEDICATION

*To my son Jianwei Zhu...*



## ACKNOWLEDGMENTS

First and foremost I want to thank my supervisor Prof. Longbing Cao. It has been an honor to be his Ph.D. student. He has taught me the skills required by an independent researcher, especially critical thinking and academic writing. I appreciate all his contributions of time, ideas, discussions, and paper polishing to make my Ph.D. experience impressive and stimulating. I am deeply influenced by his professionalism, hard work, and profound ideas. I am also thankful for the excellent opportunity to serve in the research community and the industry internship he has provided.

I would also like to thank Prof. Jianping Yin for his insightful comments, encouragement, and support. His optimism and enthusiasm were contagious and motivational for me, even during the difficult times in my Ph.D. pursuit.

My sincere thanks also go to all of my colleagues in the Advanced Analytics Institute, for their enthusiastic help: Dr. Guansong Pang, Dr. Trong Dinh Thac Do, Dr. Shoujin Wang, Dr. Zhenjiang Lin, Dr. Liang Hu, Qi Zhang, Qing Liu, Zhilin Zhao, Wei Wang, and Jia Xu. I appreciate our intensive discussions, in-depth collaboration, and all the fun we have had in the last four years.

I thank the visiting scholars and students in our laboratory. They broadened the horizon of my research and shared unforgettable times with me at UTS. In particular, I am grateful to Prof. Wentao Zhao, Prof. Lizheng Wang, A/Prof. Quangui Zhang, A/Prof. Defu Lian, A/Prof. Lingyun Xiang, Dr. Wenpeng Lu, Dr. Xinwang Liu, Songlei Jian, Shangdong Yang, and Teng Li for their precious advice and encouragement in my Ph.D. research.

I gratefully acknowledge the funding from China Scholarship Council that made my Ph.D. work possible.

Lastly, I would like to thank my family for all their love and encouragement. Without their support, it would not be possible for me to conduct this research. In particular, I am grateful to my wife Qian Li whose spiritual and academic support are so appreciated. Thank you.

Chengzhang Zhu  
*University of Technology Sydney*  
October 2019



## LIST OF PUBLICATIONS

### RELATED TO THE THESIS :

#### Published Journal Papers :

1. **Zhu, C.**, Cao, L., Liu, Q., Yin, J. and Kumar, V., 2018. Heterogeneous metric learning of categorical data with hierarchical couplings. *IEEE Transactions on Knowledge and Data Engineering*, 30(7), pp.1254-1267.
2. Zhao, W., Li, Q., **Zhu, C.**, Song, J., Liu, X. and Yin, J., 2018. Model-aware categorical data embedding: a data-driven approach. *Soft Computing*, 22, pp.3603-3619.

#### Published Conference Papers :

3. Zhang, Q., Cao, L., **Zhu, C.**, Li, Z. and Sun, J., 2018, July. CoupledCF: Learning explicit and implicit user-item couplings in recommendation for deep collaborative filtering. *In International Joint Conference on Artificial Intelligence* (pp. 3662-3668).
4. Zhao, X., **Zhu, C.** and Cheng, L., 2017, October. Coupled Bayesian matrix factorization in recommender systems. *In 2017 IEEE International Conference on Data Science and Advanced Analytics* (pp. 522-528). IEEE.
5. Song, J., **Zhu, C.**, Zhao, W., Liu, W. and Liu, Q., 2017, September. Model-aware representation learning for categorical data with hierarchical couplings. *In International Conference on Artificial Neural Networks* (pp. 242-249). Springer, Cham.

#### Submitted Manuscripts :

6. **Zhu, C.**, Cao, L., and Yin, J., 2019. Unsupervised categorical representation with heterogeneous and inconsistent couplings. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

- 
7. **Zhu, C.**, Cao, L., Liu, Q., and Yin, J., 2019. Decoupled non-IID nominal data representation. *Journal of Machine Learning Research*.
  8. **Zhu, C.**, Cao, L., and Yin, J., 2018. Similarity metrics for non-IID categorical data. *Artificial Intelligence*.
  9. Li, Q., Wu, Q., **Zhu, C.**, Cao, L., and Zhang, J., 2019. Deep non-IID paragraph representation for sentiment analysis. *Pattern Recognition*. (Major Revision)

## **OTHERS :**

### **Published Journal Papers :**

10. Yuan, L., Liu, Q., Lu, M., Zhou, S., **Zhu, C.** and Yin, J., 2017. A highly efficient human activity classification method using mobile data from wearable sensors. *International Journal of Sensor Networks*, 25(2), pp.86-92.
11. Zhao, L., Li, K., Wang, M., Yin, J., Zhu, E., Wu, C., Wang, S. and **Zhu, C.**, 2016. Automatic cytoplasm and nuclei segmentation for color cervical smear image using an efficient gap-search MRF. *Computers in biology and medicine*, 71, pp.46-56.
12. Zhou, S., Liu, X., Liu, Q., Wang, S., **Zhu, C.** and Yin, J., 2016. Random Fourier extreme learning machine with  $\ell_2$ , 1-norm regularization. *Neurocomputing*, 174, pp.143-153.
13. Liu, Q., Zhou, S., **Zhu, C.**, Liu, X. and Yin, J., 2016. MI-ELM: Highly efficient multi-instance learning based on hierarchical extreme learning machine. *Neurocomputing*, 173, pp.1044-1053.

### **Published Conference Papers :**

14. Li, Q., Wu, Q., **Zhu, C.**, and Zhang, J., 2019, September. Bi-level masked multi-scale CNN-RNN networks for short text representation *In International Conference on Document Analysis and Recognition*. IEEE.
15. Li, Q., Wu, Q., **Zhu, C.**, Zhang, J. and Zhao, W., 2019, July. An inferable representation learning for fraud review detection with cold-start problem. *In 2019 International Joint Conference on Neural Networks*. IEEE.
16. Li, Q., Wu, Q., **Zhu, C.**, Zhang, J. and Zhao, W., 2019, April. Unsupervised user behavior representation for fraud review detection with cold-start problem. *In*

---

*Pacific-Asia Conference on Knowledge Discovery and Data Mining* (pp. 222-236). Springer, Cham.

17. Wang, S., Zeng, Y., Liu, Q., **Zhu, C.**, Zhu, E. and Yin, J., 2018, October. Detecting abnormality without knowing normality: A two-stage approach for unsupervised video abnormal event detection. *In 2018 ACM Multimedia Conference on Multimedia Conference* (pp. 636-644). ACM.
18. Zhao, G., Xiang, L., **Zhu, C.** and Li, F., 2018, July. Two-stage unsupervised multiple kernel extreme learning machine. *In 2018 International Joint Conference on Neural Networks*. IEEE.





## TABLE OF CONTENTS

<b>List of Publications</b>	<b>ix</b>
<b>List of Figures</b>	<b>xix</b>
<b>List of Tables</b>	<b>xxiii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Background . . . . .	1
1.2 Current Work and Gap Analysis . . . . .	2
1.3 Research Problems and Objectives . . . . .	3
1.4 Thesis Contributions . . . . .	6
1.5 Thesis Organization . . . . .	7
<b>2 Literature Review</b>	<b>11</b>
2.1 Introduction . . . . .	11
2.2 Categorical Data Representation Paradigms . . . . .	12
2.2.1 Similarity-Based Representation Versus Vector-Based Representation . . . . .	12
2.2.2 IID Representation Versus Non-IID Representation . . . . .	13
2.3 Coupling Learning . . . . .	15
2.3.1 Learning Intra-Attribute Couplings . . . . .	16
2.3.2 Learning Inter-Attribute Couplings . . . . .	18
2.3.3 Learning Attribute-Label Couplings . . . . .	21
2.3.4 Discussion . . . . .	22
2.4 Heterogeneity Learning . . . . .	23
2.4.1 Learning Heterogeneous Distributions . . . . .	23
2.4.2 Learning Heterogeneous Dependencies . . . . .	24
2.4.3 Discussion . . . . .	26

## TABLE OF CONTENTS

---

2.5	Non-IID-Completeness and Non-IID-Hardness Learning . . . . .	27
2.5.1	Learning Non-IID-Completeness . . . . .	27
2.5.2	Learning Non-IID-Hardness . . . . .	28
<b>3</b>	<b>Preliminaries</b>	<b>31</b>
3.1	Introduction . . . . .	31
3.2	Symbol Styles and Key Notations . . . . .	31
3.3	Basic Information Functions . . . . .	33
3.4	Data Sets for Representation Performance Evaluation . . . . .	35
3.5	Metrics for Representation Performance Evaluation . . . . .	37
3.5.1	Representation-Quality-Based Evaluation Metrics . . . . .	37
3.5.2	Down-Stream-Task-Based Evaluation Metrics . . . . .	38
<b>I</b>	<b>Coupling Learning</b>	<b>41</b>
<b>4</b>	<b>Non-IID Similarity Metrics Learning on Categorical Data</b>	<b>43</b>
4.1	Introduction . . . . .	43
4.2	The Non-IID Similarity Metrics Learning Framework . . . . .	45
4.2.1	Coupled Kernel Metric for Categorical Data . . . . .	46
4.2.2	Intra-Attribute Coupling Metric . . . . .	47
4.2.3	Inter-Attribute Coupling Metric . . . . .	48
4.2.4	Object Coupling Metric . . . . .	51
4.2.5	Non-IID Distance Metric . . . . .	55
4.3	Theoretical Analysis of Non-IID Similarity Metrics Learning Properties . . . . .	55
4.3.1	The Positive Semi-Definite Property of Coupled Kernels . . . . .	55
4.3.2	Stationary and Metric Properties of Coupled Kernels . . . . .	57
4.3.3	Time Complexity of Coupled Kernel Metrics Learning . . . . .	58
4.4	Experiments and Evaluation of Non-IID Similarity Metrics Learning Performance . . . . .	58
4.4.1	Testing Non-IID Similarity Metrics-Enabled Learning Performance . . . . .	59
4.4.2	Testing Non-IID Similarity Metrics Learning Quality . . . . .	68
4.4.3	Testing Non-IID Similarity Metrics Learning Efficiency . . . . .	70
4.4.4	Discussion . . . . .	71
4.5	Summary . . . . .	73

<b>II</b>	<b>Heterogeneity Learning</b>	<b>75</b>
<b>5</b>	<b>Decoupled Non-IID Categorical Data Representation</b>	<b>77</b>
5.1	Introduction . . . . .	77
5.2	The Decoupled Non-IID Learning Framework . . . . .	79
5.3	Non-BEND: DNL-Based Nonparametric Bayesian Embedding . . . . .	82
5.3.1	The Prior Distribution of Non-IID Categorical Data . . . . .	82
5.3.2	The Posterior Probability of Non-IID Categorical Data . . . . .	83
5.3.3	The Non-BEND Representation . . . . .	88
5.4	Theoretical Analysis of Non-BEND Properties . . . . .	91
5.4.1	Information Loss Bound of Non-BEND . . . . .	91
5.4.2	Computational Complexity of Non-BEND . . . . .	93
5.5	Connections between Non-BEND and the Existing Representation Methods	94
5.6	Experiments and Evaluation of Non-BEND Performance . . . . .	94
5.6.1	Testing Non-BEND Effectiveness . . . . .	95
5.6.2	Testing Non-BEND-Enabled Classification Performance . . . . .	97
5.6.3	Testing Non-BEND Representation Quality . . . . .	99
5.6.4	Testing Non-BEND Flexibility . . . . .	101
5.6.5	Testing Non-BEND Efficiency . . . . .	102
5.6.6	Ablation Study of Non-BEND Design . . . . .	103
5.7	Summary . . . . .	105
<b>III</b>	<b>Non-IID-Completeness Learning</b>	<b>107</b>
<b>6</b>	<b>Heterogeneous Metric Learning of Categorical Data with Hierarchical Couplings</b>	<b>109</b>
6.1	Introduction . . . . .	109
6.2	The HELIC Design . . . . .	111
6.2.1	Problem Statement of Metric Learning . . . . .	111
6.2.2	The HELIC Framework . . . . .	111
6.2.3	Learning Value-to-Class Couplings . . . . .	112
6.2.4	Learning Heterogeneity in Heterogeneous Couplings . . . . .	115
6.2.5	Learning HELIC Representation . . . . .	116
6.3	Theoretical Analysis of HELIC Properties . . . . .	119
6.3.1	HELIC Effectiveness . . . . .	119

TABLE OF CONTENTS

---

6.3.2	The Generalization Error Bound of HELIC . . . . .	121
6.3.3	Computational Complexity of HELIC . . . . .	124
6.4	Experiments and Evaluation of HELIC Performance . . . . .	125
6.4.1	Parameter Settings of HELIC . . . . .	125
6.4.2	Testing HELIC Representation Performance . . . . .	126
6.4.3	Testing HELIC Representation Quality . . . . .	129
6.4.4	Testing the Effect of Learning Couplings and Heterogeneity . . . . .	130
6.4.5	Testing HELIC Scalability . . . . .	132
6.4.6	Testing HELIC Stability . . . . .	134
6.5	Summary . . . . .	135
<b>7</b>	<b>Unsupervised Categorical Representation with Heterogeneous and In-</b>	
	<b>consistent Couplings</b>	<b>137</b>
7.1	Introduction . . . . .	137
7.2	The UNTIE Design . . . . .	139
7.2.1	The UNTIE Framework . . . . .	139
7.2.2	Heterogeneous Coupling Learning . . . . .	141
7.2.3	Heterogeneity Learning in Kernel Spaces . . . . .	143
7.2.4	Kernel $K$ -Means-Based Representation Learning . . . . .	145
7.2.5	The UNTIE Algorithm . . . . .	147
7.3	Theoretical Analysis of UNTIE Properties . . . . .	148
7.3.1	The Fitness of Heterogeneity Hypotheses . . . . .	148
7.3.2	The Positive Semi-Definite Property of UNTIE Wrapper Kernel . . . . .	149
7.3.3	The Separability of UNTIE-Represented Data . . . . .	150
7.3.4	Convergence of the UNTIE Algorithm . . . . .	152
7.3.5	Computational Complexity of UNTIE . . . . .	152
7.4	Experiments and Evaluation of UNTIE Performance . . . . .	153
7.4.1	Parameter Settings of UNTIE . . . . .	153
7.4.2	Testing UNTIE Effectiveness . . . . .	154
7.4.3	Testing UNTIE Representation Quality . . . . .	157
7.4.4	Testing UNTIE Efficiency . . . . .	163
7.4.5	Testing UNTIE Flexibility . . . . .	168
7.4.6	Testing UNTIE Stability . . . . .	168
7.5	Summary . . . . .	169

<b>IV Non-IID-Hardness Learning</b>	<b>171</b>
<b>8 Unsupervised Coupling Learning on Dynamic Categorical Data</b>	<b>173</b>
8.1 Introduction . . . . .	173
8.2 The UNICORN Method . . . . .	174
8.2.1 The UNICORN Architecture . . . . .	174
8.2.2 Unsupervised Heterogeneous Dynamic Coupling Learning . . . . .	177
8.2.3 Transforming Heterogeneous Couplings to Kernelized Spaces . . . . .	179
8.2.4 Building a Unified Global Representation . . . . .	180
8.2.5 The Learning Objective Function of UNICORN . . . . .	182
8.3 Theoretical Analysis of UNICORN Properties . . . . .	185
8.3.1 The Positive Semi-Definite Property of UNICORN Wrapper Kernel	185
8.4 Experiments and Evaluation of UNICORN Performance . . . . .	186
8.4.1 Dynamic Categorical Data Sets . . . . .	186
8.4.2 Parameter Settings of UNICORN . . . . .	186
8.4.3 Testing Effectiveness of Learning Dynamic Categorical Data . . . . .	186
8.4.4 Comparison to State-of-the-Art Static Categorical Data Representation Methods . . . . .	187
8.4.5 Ablation Study of UNICORN Design . . . . .	191
8.4.6 Testing UNICORN Stability . . . . .	193
8.5 Summary . . . . .	193
<b>9 Conclusions and Future Directions</b>	<b>195</b>
9.1 Conclusions . . . . .	195
9.1.1 Coupling Learning . . . . .	195
9.1.2 Heterogeneity Learning . . . . .	196
9.1.3 Non-IID-Complete Learning . . . . .	196
9.1.4 Non-IID-Hard Learning . . . . .	196
9.2 Future Directions . . . . .	197
9.2.1 Exploiting Other Non-IID Representation Methods for Categorical Data . . . . .	197
9.2.2 Studying Non-IID Representation on More Types of Data . . . . .	198
9.2.3 Quantifying the Non-IID Data Complexities . . . . .	198
<b>A Appendix</b>	<b>199</b>
A.1 List of Notations for Static Data . . . . .	199

## TABLE OF CONTENTS

---

A.2 List of Notations for Dynamic Data . . . . .	201
A.3 List of Abbreviations . . . . .	202
<b>Bibliography</b>	<b>205</b>

## LIST OF FIGURES

FIGURE	Page
1.1 Non-IIDness in complex categorical data. . . . .	4
1.2 The research problems and their relations. . . . .	5
2.1 Hierarchical and heterogeneous couplings in dynamic categorical data. . . . .	28
4.1 The non-IID similarity metrics learning framework . . . . .	46
4.2 The precision@ $k$ -curve and recall@ $k$ -curve of different categorical data representation methods: A better metric yields a higher curve in this figure. . . . .	68
4.3 The visualization of (dis-)similarity representation of categorical data on Wc-s data set. These figures illustrate the data distribution in the cKML learned metric space has clearer boundaries between different clusters. The plotted two-dimensional embedding is converted from the metric space by multidimensional scaling (Borg & Groenen 2005). Different symbols refer to different data clusters per ground truth. . . . .	69
4.4 Time cost of cKML on synthetic data sets with different data factors. . . . .	70
5.1 Categorical data representation framework with decoupled non-IID learning.	81
5.2 The graphical model of non-BEND. . . . .	83
5.3 Comparison of clustering performance enabled by non-BEND against the other representation methods per the Bonferroni–Dunn test. All representation methods with ARs outside the marked interval are significantly different ( $p < 0.1$ ) from non-BEND. . . . .	97
5.4 Comparison of classification performance enabled by non-BEND against the other representation methods per the Bonferroni–Dunn test. All representation methods with ARs outside the marked interval are significantly different ( $p < 0.1$ ) from non-BEND. . . . .	99

5.5	The visualization of different representation methods on data set Tr. These results illustrate the non-BEND-represented data shows clearer boundaries between different clusters. The plotted two-dimensional embedding is converted from high-dimensional representation by $t$ -SNE. Different colors refer to different data categories per the ground truth. . . . .	100
5.6	The non-BEND time cost with respect to data factors: object number $n_o$ , attribute number $n_a$ , and maximum number of attribute values $n_{mv}$ . . . . .	103
6.1	The HELIC framework: The coupling learning first represents the couplings in categorical data; then, the heterogeneity learning reveals the heterogeneous distributions of categorical values in the coupling spaces and feeds them into metric learning. . . . .	112
6.2	Comparison of HELIC against the other distance measures as per the Bonferroni–Dunn test. All distance measures with ranks outside the marked interval are significantly different ( $p < 0.05$ ) from HELIC. . . . .	128
6.3	The precision@ $k$ -curve of different distance measures: A better metric yields a higher curve in this figure. . . . .	128
6.4	The $(\epsilon, \gamma)$ -curve of different transformed similarity measures: A better metric would yield a curve with higher $y$ -axis values. . . . .	129
6.5	Comparison of HELIC against its variants per the Bonferroni–Dunn test. All distance measures with ranks outside the marked interval are significantly different ( $p < 0.05$ ) from HELIC. . . . .	130
6.6	Comparison of HC against the other distance measures per the Bonferroni–Dunn test. All distance measures with ranks outside the marked interval are significantly different ( $p < 0.05$ ) from HC. . . . .	132
6.7	The HELIC training loss on different data sets. The stochastic optimization method for HELIC is Adam (Kingma & Ba 2014), the initial learning rate is $10^{-3}$ , and the batch size is 20. The $x$ -axis refers to the number of iterations, and the $y$ -axis refers to the loss value of HELIC metric learning objective function Equation (6.19). . . . .	133
6.8	The HELIC time cost with respect to data factors: object number $n_o$ , attribute number $n_a$ , and maximum number of attribute values $n_{mv}$ . . . . .	134
6.9	The HELIC time cost with respect to number of kernels. . . . .	134
6.10	The HELIC-enabled KNN classification $F$ -score with respect to $\lambda$ . . . . .	135



7.1	The UNTIE framework: It first transforms the coupling spaces into multiple kernel spaces and then learns the heterogeneity within and between couplings in these kernel spaces by solving a kernel $k$ -means objective. . . . .	140
7.2	Comparison of UNTIE against the other representation methods per the Bonferroni–Dunn test. All representation methods with ranks outside the marked interval differ significantly ( $p < 0.1$ ) from UNTIE. . . . .	156
7.3	The precision@ $k$ of different categorical data representation methods: A better metric yields a higher value. . . . .	157
7.4	The probability density of intra-coupling heterogeneity indicator per kernel density estimation. . . . .	158
7.5	The probability density of inter-coupling heterogeneity indicator per kernel density estimation. . . . .	160
7.6	The UNTIE-enabled clustering performance on data sets with different inconsistency levels per the intra-heterogeneity indicator. . . . .	161
7.7	The UNTIE-enabled clustering performance on data sets with different inconsistency levels per the inter-heterogeneity indicator. . . . .	162
7.8	The $(\epsilon, \gamma)$ -curves of different transformed similarity measures: A better metric yields a better result. . . . .	163
7.9	The visualization of different representation methods on Dmg. The UNTIE-represented data shows clearer boundaries between different clusters. The plotted two-dimensional embedding is converted from high-dimensional representation by $t$ -SNE. Different symbols refer to different data clusters, per the ground truth. . . . .	164
7.10	The UNTIE’s training loss on different data sets. The stochastic optimization method for UNTIE is Adam (Kingma & Ba 2014), with an initial learning rate of $10^{-3}$ and a batch size of 20. The $x$ -axis refers to the number of iterations, and the $y$ -axis refers to the loss value of UNTIE’s objective function Equation (7.24). . . . .	165
7.11	The UNTIE time cost with respect to data factors: object number $n_o$ , attribute number $n_a$ , and maximum number of attribute values $n_{mv}$ . The solid line refers to the total time cost of UNTIE. The dotted line refers to the time cost of building the coupling spaces. The star line refers to the time cost of the heterogeneity learning. . . . .	166
7.12	The clustering $F$ -score (%) with UNTIE with respect to different kernel function sets: The same color indicates the same kernel function group. . . . .	169

8.1	The UNICORN architecture: A multi-step dynamic learning process to capture and convert local heterogeneous couplings to unified global representations on dynamic categorical data. . . . .	175
8.2	Comparison of UNICORN-enabled $k$ -means clustering against categorical data stream clustering methods per the Bonferroni–Dunn test. All clustering methods with ranks outside the marked interval are significantly different ( $p < 0.05$ ) from the UNICORN-enabled $k$ -mean. . . . .	187
8.3	Comparison of UNICORN against other representation methods per the Bonferroni–Dunn test. All representation methods with ranks outside the marked interval differ significantly ( $p < 0.05$ ) from UNICORN. . . . .	189
8.4	The precision@ $k$ -curve of different representations: A better representation yields a higher curve. . . . .	190
8.5	The Visualization of different representation methods: A better representation yields closer grouping results. . . . .	191
8.6	Averaged ranks over three iterations of data dynamics: Each iteration averages the ARs of four data partitions of all data sets. A better method incurs a lower AR in each iteration. The methods that can capture dynamic couplings will generate non-increasing ARs over iterations. . . . .	192
8.7	The UNICORN-enabled clustering $F$ -score with respect to hyper-parameter $\delta$ . . . . .	192

## LIST OF TABLES

TABLE	Page
2.1 Toy Example: The Watermelon Information Table. Each Watermelon with a Different Sweetness is Described with respect to Three Attributes: <i>Texture</i> , <i>Color</i> , and <i>Root Shape</i> . . . . .	14
3.1 List of Symbol Styles . . . . .	32
3.2 Characteristics of Benchmark Data Sets . . . . .	36
4.1 <i>K</i> -modes Clustering <i>F</i> -score with Different Similarity Measures . . . . .	61
4.2 <i>K</i> -modes Clustering NMI with Different Similarity Measures . . . . .	62
4.3 <i>F</i> -score for <i>K</i> -modes Clustering with Low-level cKML and COS Measures . .	63
4.4 NMI for <i>K</i> -modes Clustering with Low-level cKML and COS Measures . . .	64
4.5 <i>F</i> -score for KNN Classification with Different State-of-the-art Similarity Measures . . . . .	66
4.6 <i>F</i> -score for KNN Classification with Low-level cKML and COS Measures . .	67
5.1 <i>F</i> -score of <i>K</i> -means and <i>K</i> -modes Enabled by Different Categorical Data Representation Methods . . . . .	96
5.2 <i>F</i> -score of KNN Enabled by Distance and Similarity Learning Methods . . . .	98
5.3 The Accuracy of Bayesian Matrix Factorization Enabled by Different Representation Methods . . . . .	102
5.4 <i>F</i> -score of <i>K</i> -means Enabled by Non-BEND and Its Variants . . . . .	104

6.1	KNN Classification $F$ -score (%) with Different Distance Measures. The Monte Carlo cross-validation results are reported with respect to <i>mean <math>\pm</math> standard deviation</i> . The best results are highlighted in bold, the results without a significant difference from the best results for a data set under the <i>student t-test</i> ( $p$ -value $> 0.05$ ) are labelled by *, and $\Delta$ is the HELIC's improvement over the best results of the other measures. The AR of a method over all data sets with significant difference from others with respect to the <i>Bonferroni–Dunn test</i> ( $p$ -value $< 0.05$ ) is labelled by ** . . . . .	127
6.2	KNN Classification $F$ -score (%) with HELIC Variants. The Monte Carlo cross-validation results are reported as <i>mean <math>\pm</math> standard deviation</i> . $\Delta$ shows the HELIC improvement over the best results of its variants. . . . .	131
7.1	Clustering $F$ -score (%) with Different Embedding Methods: The value-based representations are fed into $k$ -means and the similarity-based representations are fed into $k$ -modes to get the clustering results. The best results are highlighted in bold. $\Delta$ indicates the UNTIE's improvement over the best results of the other measures. The AR of a method over all data sets with significant difference from others with respect to the Bonferroni-Dunn test ( $p$ -value $< 0.1$ ) is labelled by * . . . . .	155
7.2	KNN, SVM, RF and LR Classification $F$ -score (%) with UNTIE and CDE. The best results are highlighted in bold typeface. . . . .	167
7.3	Three Groups of Kernel Function Sets for UNTIE Stability Evaluation . . . .	169
8.1	Clustering $F$ -score of UNICORN-enabled $K$ -Means vs. Different Categorical Data Stream Clustering Methods. The experiment shows the mean value of results on random partitions of a data set. The best results are highlighted in bold, and $\Delta$ shows the performance lift of UNICORN-enabled $k$ -means over the best results of the baselines. The AR cross different data sets is reported to show the overall performance. . . . .	188
8.2	Clustering $F$ -score of $K$ -Means with Different Representations. The best results are highlighted in bold typeface, and $\Delta$ shows the UNICORN's improvement (lift) over the best results of the baselines. The AR across different data sets shows the overall performance. . . . .	189