

Multidimensional Analysis of Heterogeneous Relationships in Internet of Things



THE UNIVERSITY
of ADELAIDE

A dissertation submitted in fulfillment
of the requirements for the degree of

Doctor of Philosophy

in

Computer Science

Lina Yao

Supervisors: A/Prof. Michael Sheng

Prof. Anne H. H. Ngu

2014

© Copyright by

Lina Yao

2014

ORIGINALITY STATEMENT

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

I also give permission for the digital version of my thesis to be made available on the web, via the Universitys digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Lina Yao

January 10, 2014

*To my parents,
my husband and my daughter*

ACKNOWLEDGMENTS

I would like to thank all the people who have given me tremendous support and help to make this thesis happen.

First of all, I am deeply grateful to my supervisor, A/Prof. Michael Sheng, for his generous time and devotion on supervision and guidance during my PhD study. He instilled a thirst for excellence in me, taught me how to do high-quality research, and helped me think independently and creatively. In many ways, he has set up a great example for me: his passion on research and life, his patience in guiding students, and his diligence. I always feel lucky to have A/Prof. Michael Sheng as my supervisor, for all the earnest encouragements, and the strongest supports in every means.

I sincerely thank Prof. Anne H.H. Ngu and Assistant Professor Byron Gao, who have given me many valuable suggestions and insightful discussions in my research. I would like to thank Prof. David Suter who has been supporting me in many ways, from scholarship applications to fellowship application. I have also benefited a lot from his course on Computer Vision, which opens a fascinating research world for me.

I would like to thank my co-authors: Quan Z. Sheng, Anne H.H. Ngu, Byron Gao, Zakaria Marmarr, Xue Li, Talal Noor, Claudia Szabo, Aviv Segev, Rui Zeng for their productive and enjoyable collaborations.

I would like to also thank my friends for providing support and friendship whenever I needed. Especially thank my two dearest friends, Bo Zhao, for his unconditional support all the time, Lijuan Wang, who has kindly given me the precious help during my hardest time.

Finally and above all, I owe my deepest gratitude to my family. I want to thank my mother Junling Liu and my father Guangquan Yao for their endless and unreserved love, who have been encouraging and supporting me all the time. Their love has

provided my inspiration and has been my driving force. Without their patience and love I would have not survived the difficult times. I want to thank my husband, Tao Wang, for his love, understanding, patience, and support at every moment. I also want to thank my new born daughter Alicia, who is a miracle in my life and gives me ongoing courage to go through the challenging journey. This thesis is dedicated to them.

ABSTRACT OF THE DISSERTATION

Multidimensional Analysis of Heterogeneous Relationships in Internet of Things

by

Lina Yao

Doctor of Philosophy in Computer Science

The University of Adelaide, 2014

The Internet of Things (IoT) describes the evolution from systems linking digital information to systems relating digital information to real world physical items. In IoT, real-world objects are being connected to the Internet and interacted with traditional web entities (images, videos, texts and people in social networks), which complicates the relationships amongst heterogeneous entities. While it is well understood that IoT offers numerous opportunities and benefits, it also presents significant technical challenges. Among them, effective and efficient management of swift growing ubiquitous things in the boosting IoT is a fundamental challenge facing society and research community today. The essential prerequisite is efficient acknowledging and analysis of *heterogeneous relationships* existing in Internet of Things, and with it, many research challenges can be pushed forward.

In this thesis, we have systematically explored the principles and developed a series of methodologies of mining and modeling *three* types of heterogeneous relationships existing in Internet of Things.

Implicit mutual correlations of things. We study the problem of discovering implicit correlations among ubiquitous things, which have unique and challenging characteristics (e.g., no uniform features, diverse, and dynamic). We explore explicit user

interactions with things and develop two graph-based models, namely T-DisCor and T-DisCor+, for discovering latent correlations among things over graphs induced from thing usage events via modeling the contextual information and entity relationships respectively. We demonstrate how the discovered correlations of things can contribute to solving a number of important applications on things management. In particular, we develop an effective and flexible feature-based method for annotating things based on the outcome of our proposed approach.

Relationship between things and its attributes. We investigate the problem of exploiting relationships between ubiquitous things and their different attributes, we propose multiple approaches:

- Thing’s semantic labels. We propose a semi-supervised learning framework for multi-label classification, which fully considers objects mutual affinity, semantic label correlation, coupled by thing-label assignment relationships for effective semi-supervised classification.
- Thing’s social tags. We explore social tags for web object classification by developing a tag-centric, unified and discriminative classification framework. We not only use social tags conveying partial and latent information about the web objects as a novel evidence to facilitate classifying objects on the web, but also exploit the relative information among tags.
- Thing’s availability. We propose an algorithm to estimate the availability of things so that those with higher availability can be selected and delivered in applications such as things composition.

Interactive relationship between people and things. Things recommendation is a crucial step for promoting and taking full advantage of Internet of Things, where it benefits the individuals, businesses and society on a daily basis in terms of two main

aspects. On the one hand, it can deliver relevant things (things have similar functionalities that users might need) to users based on users preference and interest. On the other hand, it can also serve to optimize the time and cost of using IoT in a particular situation. Physical things in reality have multiple unique attributes. We propose two novel approaches to explore things recommendation problem.

- We propose a probabilistic matrix factorization based joint model to address things recommendation problem in Internet of Things. We fuse information from users social networks and things correlation networks, by learning shared latent factors stemming from the probabilistic matrix factorization on three matrices, namely users relations, things correlations, and observable things usage interactions.
- We propose a novel approach that unifies collaborative filtering and content-based recommendations. In particular, our approach considers simultaneously both rating data (e.g., QoS) and semantic content data (e.g., functionalities) of services offered by things using a three-way aspect model. Unobservable user preferences are represented by introducing a set of latent variables, which is statistically estimated.

For each of model and approach, we have conducted extensive experiments to validate and evaluate it using either public data set or datasets generated from an environment we built.

The studies presented in this dissertation lead to a series of models and real-world applications, including things classification, things clustering and things recommendation. Some open research directions are also discussed in the dissertation.

TABLE OF CONTENTS

1	Introduction	1
1.1	Motivations and Overview	1
1.1.1	Internet of Things	3
1.1.2	Heterogeneous Relationships in Internet of Things Network	6
1.1.3	Why Mining Heterogeneous Relationships with Internet of Things	9
1.2	Dissertation Organization	10
I	Learning Things Mutual Relationships	15
2	Correlation Discovery via Fusing Contextual Information in Things Usage Events	16
2.1	Motivations and Challenges	17
2.2	Problem Formulation	18
2.3	T-DisCor	20
2.3.1	Approach Overview	20
2.3.2	Social Dimension Information Modeling	21
2.3.3	Spatial Dimension Information Modeling	23
2.3.4	Temporal Dimension Information Modeling	25
2.3.5	Correlation Inference Algorithm	25
2.4	T-DisCor+	28
2.4.1	Approach Overview	28

2.4.2	Spatio-temporal Fusion Graph	29
2.4.3	Periodical Pattern Detection	32
2.4.4	Inference Algorithm	33
2.5	Related Work	35
2.6	Conclusion	37
3	Things Annotation by Exploiting Correlations of Things	38
3.1	Algorithm Overview	38
3.2	Structural Feature Space	40
3.2.1	Spectral Bisection Partitioning	40
3.2.2	Modularity Optimization	42
3.2.3	Structural Features Extraction	45
3.3	Posterior Probability Feature Space	46
3.4	Explicit Feature Space	47
3.5	Conclusion	48
4	Implementation and Experiments	50
4.1	T-Mine Platform	50
4.2	Implementation Details	53
4.2.1	Sensors Hive	53
4.2.2	Virtual Resource	54
4.2.3	Event Processing	54
4.2.4	Web API/Services	56
4.2.5	Rules Engine	57

4.3	A Demonstration	58
4.4	Experiments for T-DisCor	62
4.4.1	Overall Performance	62
4.4.2	Parameters Tuning	63
4.4.3	Effects on Time and Location Smoothing	66
4.5	Experiments for T-DisCor+	68
4.5.1	Overall Performance	69
4.5.2	Impact of Integrating Spatio-Temporal Information	70
4.6	Conclusion	72
 II Mining Things and Their Attribute Relationships		73
 5 Things Classification Using Bi-Relational Graph of Things and Semantic Labels		74
5.1	Motivations and Challenges	75
5.2	Problem Formulation	78
5.3	Bi-relational Graph Construction	80
5.3.1	Affinity Subgraph	80
5.3.2	Categorical Correlation Graph	85
5.4	Learning Algorithm	86
5.5	Algorithm Analysis	90
5.6	Experiments	92
5.6.1	Experimental Settings	92

5.6.2	Comparison Partners	94
5.6.3	Experimental Results	96
5.7	Related Work	99
5.8	Conclusion	101
6	A Discriminative Model for Things Classification by Learning Social Tags	103
6.1	Problem Formulation	103
6.2	A Discriminative Model	106
6.2.1	Modeling Functions	106
6.2.2	Max-Margin Learning Algorithm	109
6.2.3	Convex Optimization	110
6.2.4	Finding Optimal Latent Variable t'	111
6.3	Loss Function	112
6.4	Tag Tree Generation	113
6.5	Discussion: Learning with Invisible Side Information	115
6.6	Experiments	116
6.6.1	Data Preparation	116
6.6.2	Tag Preprocessing	117
6.6.3	Performance Comparison	119
6.7	Related Work	121
6.8	Conclusion	122
7	Particle Filtering based Services Availability Estimation	123
7.1	Motivations and Challenges	124

7.2	Model Overview	125
7.3	Availability Modeling	126
7.4	The Particle Filter Based Approach	128
7.4.1	Particle Filtering	128
7.4.2	Particle Filtering Based Estimation	132
7.4.3	The Dynamic Tracking Algorithm	134
7.5	Experiments	136
7.6	Related Work	140
7.7	Conclusion	142
III	Analysis on Things Recommendation Relationship	143
8	Exploring Things Recommendation in Internet of Things	144
8.1	Motivations and Challenges	145
8.2	Problem Description	147
8.3	The Model	148
8.3.1	Encoding Users Friendship	149
8.3.2	Encoding Things Correlations	150
8.3.3	Encoding User-Things Interactions	151
8.3.4	Model Learning	153
8.4	Experiments	154
8.4.1	Dataset Description	154
8.4.2	Performance Metrics	155

8.4.3	Performance Comparison	156
8.4.4	Impact of Things Correlations and Social Influence	158
8.5	Related Work	159
8.6	Conclusion	161
9	Unified Collaborative Service Recommendation Combined with Content-based Features	163
9.1	Motivations and Challenges	164
9.2	Requirements in Service Recommendation	166
9.2.1	Problem Description	167
9.3	The Unified Service Recommendation Model	168
9.3.1	Model Description	168
9.3.2	Model Learning	171
9.4	Dealing with Data Sparsity	174
9.4.1	Data Smoothing	175
9.4.2	Implicit User-Descriptor Aspect Model	175
9.5	Performance Evaluation	176
9.5.1	Dataset Setup	177
9.5.2	Metrics	180
9.5.3	Performance Comparison	181
9.5.4	Impact of Markoff Ratio and Latent Variables	185
9.6	Related Work	189
9.6.1	CF Methods for Service Recommendation	189

9.6.2	Content-based Methods for Service Recommendation	191
9.6.3	Hybrid Methods for Service Recommendation	192
9.7	Conclusion	193
10	Conclusion and Future Work	196
10.1	Conclusion	196
10.2	Future Directions	198
	Bibliography	203
A	Curriculum Vitae	218

LIST OF FIGURES

1.1	An illustrative scenario of Internet of Things: Tony’s one day	4
1.2	Three relationships and corresponding schemas studied in this thesis. The upper part shows three summarized relationships with Internet of Things: Thing-to-Thing (R1:T2T); Thing-to-Attribute (R2:T2A); People-to-Thing (R3:P2T). The lower part demonstrates three schemas of T2T, T2A and P2T studied in this thesis: (a) learning mutual relationship between things (Part I) (b) mining relationship between things and its attributes (Part II) (c) predicting dyadic relationship between people and things (Part III)	7
2.1	Illustration of problem description: explicit things usage events vs. implicit correlation things	19
2.2	Three dimensional graph extracted from the things usage events: (a) social graph G_f , (b) spatial graph G_s and temporal graph G_t	21
2.3	Illustration of two graphs induced from thing usage events: (a) spatio-temporal graph G_m (b) social graph G_f	29
3.1	Using the first two eigenvectors of the modularity matrix, vertices can be represented as points on a plane. By cutting the plane with a line passing through the origin (like the dashed line in the figure) one obtains bipartition’s of the graph with possibly high modularity values [89].	44
4.1	Some devices used in T-Mine platform	51
4.2	The architecture of T-Mine platform	53
4.3	Web-based user interface	59

4.4	Screenshot of Rule Composer: a user can conveniently drag any available thing's widget panel (left) to the Rules Editor panel (right) and start to create a new rule or modify the old rules.	61
4.5	Overall performance comparison on six categories	64
4.6	Parameters tuning on six categories	65
4.7	Impact of smoothed time	67
4.8	Impact of clustered location	68
4.9	Overall performance comparison: (a) Micro-F1, (b) Macro-F1	70
5.1	Illustration of a mixed graph. The lower component is the text-affinity subgraph, where the edges indicate text-text affinity. The upper component is the label-correlation subgraph, where the edges indicate label-label correlation. The dashed edges across the two components indicate text-label assignment relationships. For example, text 17 is labeled with <i>Celebrities</i> , <i>Art</i> and <i>Entertainment</i> . We aim to infer the most appropriate labels for the unlabeled texts, which are the ones with question marks in the figure, for example, text 25.	79
5.2	One-to-one pairwise affinity graph vs. sparse graph. We randomly selected 100 vertexes from the Yahoo! Arts dataset and applied both traditional similarity graph construction (left) and sparse graph reconstruction (right). The thickness of lines is proportional to the edge weight.	82
5.3	Micro-F1 and Macro-F1 comparison on SVM, GH, wvRN, wvRN+RL and RMG	97
5.4	Micro-F1 and Macro-F1 of SVM, GH, wvRn, wvRN+RL and RMG on each dataset	98
5.5	Comparing subgraph construction methods on ℓ_1 sparse graph and k -nn graph	99
6.1	Tags as Puzzles: top-5 association tags assigned by users for one store selling Lady Gaga's album: the Fame Monster	105

6.2	Tags as Bridge: transferring category information between different domains	105
6.3	Graphical representation of the decomposition of the model	107
6.4	Our Tag-Centric Model with Different Base Feature Settings: Partial Textual Features (Title Only) vs. Full Textual Feature (Title+Description)	120
7.1	Services selection based on availability prediction	126
7.2	The main processes of the particle filter technique	129
7.3	Availability of a service: actual availability vs. estimated availability	137
7.4	Availability of a composite service	138
7.5	The impact of σ on estimated availability error of a component service	139
7.6	The impact of σ on the availability of a composite service	140
7.7	Time for composing a composite service	141
8.1	The graph induced from users' social networks and things correlations. The connections consist of three types of links, including users' friendships within <i>users' social network</i> , things correlations with correlation graph of things and things usage links within <i>user-thing interactions</i> . Our task is to predict the dyadic relationship between users and things, in other words, recommending certain things to users.	148
8.2	Graphical representation of our approach.	149
8.3	Impact of social influence and things correlations	159
9.1	Graphical representation of our probabilistic model	169
9.2	Distribution of Web services and users (computers) of the WS-DREAM dataset: the circles denote Web services at different locations and the stars represent the users.	178
9.3	QoS data collection in the WS-DREAM dataset (a) response time and (b) throughput	180

9.4	Illustration of experiment implementation: we divided rating matrix \mathcal{R} into the training matrix \mathcal{R}_t , in which partial rating scores are masked and the rest ratings are used to train our model, and evaluation matrix \mathcal{R}_e , in which the masked rating value can be predicted using our validated model.	183
9.5	Recommendation performance with different markoff ratios for top 1, 5, 10 Web services	186
9.6	Recommendation performance with different number of latent variables \mathcal{N}_z for top 1, 5, 10 Web services	187

LIST OF TABLES

4.1	Data collection from T-Mine platform	52
4.2	Performance comparison with STI and without STI	71
5.1	Yahoo! datasets	93
5.2	Overall performance comparison	94
5.3	Performance comparison on each dataset	97
6.1	Data statistics	117
6.2	Classification performance comparison	119
8.1	Dataset statistics	155
8.2	Comparison with PMF, SVD++, SoRec on all categories	157
9.1	Dataset Statistics	177
9.2	Performance Comparison with Other Approaches	184