

Dynamic Implicit Social Recommendation



Qin Zhang

Faculty of Engineering and Information Technology
University of Technology Sydney

A thesis submitted for the degree of

Doctor of Philosophy

January 2018

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 Student

Production Note:

Signature removed prior to publication.

Acknowledgements

I would like to express my earnest thanks to my supervisor, Professor Chengqi Zhang, and co-supervisors, Doctor Peng Zhang, Doctor Guodong Long and Doctor Jing Jiang. They have provided tremendous support and guidance for my research and my life in the past three and half years.

I would like to express my sincere appreciation to Professor Chengqi Zhang, for his wise guidance and supervisions during my Ph.D study. Professor Chengqi Zhang provided me an opportunity to study in the Center for Artificial Intelligence (which used to be the Center for Quantum Computation and Intelligent Systems), where I met and learnt a lot from many smart and sharp people. I benefited significantly from his unselfish help and valuable suggestion on my research career.

I would like to thank Doctor Peng Zhang who encouraged me staying in academia after I got my Master degree from the University of Chinese Academy of Sciences. Doctor Peng Zhang got me start to love research and taught me how to write an academic paper step by step with his endless patience. His great passion on research also impacted me so much and made me longing to learn new technologies. I could remember the time we discuss the research details, mainly because his broad knowledge background often motivated me to target some interesting fields, which were always exciting experiences.

I also would like to thank Doctor Guodong Long, who gave me a different point of view on how to observe the world, how to know more about myself and how to plan my career. The conversations with him always benefited me so much and gave me a new perspective

on the problems. His vision, creativeness and enthusiasm in solving challenging problems both in research and in projects have greatly encouraged me and inspired my works. I also want to give thanks to him as a friend of mine, for giving me invaluable instructions and suggestions during my life.

I would like to thank Doctor Jing Jiang who gave me a lot of help even before I came to UTS. From her, I learnt a lot on time and energy management. I benefited so much from various discussions and communications with her. She gave me advices, encouragement on the difficult problems in research and in life.

I also wish to express my sincere appreciation to Professor Wei Wang for all the support and help during my visiting in University of California, Los Angeles (UCLA). The guidance and support from her and her team benefited me so much. Her persistent enthusiasm on research and life impressed me so much. I learnt a lot from her, not only on research, but also on how to explore the real-world problem for great ideas, on how to keep firm and tenacious on what we love, and on how to manage well a big team.

I would like to give special thanks to Professor Ivor Tsang. I was so fortunate to spend time and cooperated with him. Discussing a problem and writing a paper with him have been always a pleasure and eye-opening experience. He gave me sufficient freedom and encouragement to think and explore my research interests.

I would also like to thank all the people that had a positive influence on my day-to-day enjoyment of the job: Jia Wu, Shuirui Pan, Haishuai Wang, Maoying Qiao, Tongliang Liu, Zhibin Hong, Shaoli Huang, Sujuan Hou, Zhiguo Long, Zhe Xu, Bozhong Liu, Bo Du, Daokun Zhang, Chaoyue Wang, Shan Xue, Xiaolin Zhang, Ruizhi Zhou, Yue Ma, Li Gao, Guoqing Xu, Stephanie Li, Grace Lin, Christina Tong, FuChien Tseng, Paulina Rivera, Teigan Valenzuela. They are the ones who have given me support during both joyful and stressful times, to whom I will always be thankful. Without their endless patience,

generous support, and constant guidance, this thesis could not have been accomplished.

Finally, and above all, I want to thank my family for their continuous support. I especially thank my parents for their unconditional encouragement and support, both emotionally and financially. No words could possibly express my deepest gratitude for their endless love, self-sacrifice and unwavering help. To them I dedicate this dissertation.

Publications

In Journals

1. **Qin Zhang**, Peng Zhang, Guodong Long, Wei Ding, Chengqi Zhang and Xindong Wu. Online learning from trapezoidal data streams. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, vol. 28(10), pp. 2709-2723, 2016. (Australia ERA Ranked **A**, Core ranked **A**)
2. **Qin Zhang**, Jia Wu, Peng Zhang, Guodong Long and Chengqi Zhang. Discriminative subsequence mining for time series clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, **Under third-round review**, 2018. (Australia ERA Ranked **A***, Core ranked **A***)
3. Qinzhe Zhang, Jia Wu, **Qin Zhang**, Peng Zhang, Guodong Long and Chengqi Zhang. Dual influence embedded social recommendation. *World Wide Web Journal (WWWJ)*, pp. 1-26, 2017. (Australia ERA Ranked **A**, Core ranked **A**)

In Conferences

4. **Qin Zhang**, Jia Wu, Hong Yang, Yingjie Tian and Chengqi Zhang. Unsupervised feature learning from time series. *26rd International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 2322-2328, 2016. (Australia ERA Ranked **A**, Core ranked **A***)
5. **Qin Zhang**, Jia Wu, Peng Zhang, Guodong Long, Ivor W. Tsang and Chengqi Zhang. Inferring latent network from cascade data for dynamic social recommendation. *International Conference on Data Mining (ICDM)*, pp. 669-678, 2016. (Australia ERA Ranked **A**, Core ranked **A***)
6. **Qin Zhang**, Peng Zhang, Guodong Long, Wei Ding, Chengqi Zhang and Xindong Wu. Towards mining trapezoidal data streams. *International Conference on Data Mining (ICDM)*, pp. 1111-1116, 2015. (Australia ERA Ranked **A**, Core ranked **A***)

-
7. Qinzhe Zhang, **Qin Zhang**, Guodong Long, Peng Zhang and Chengqi Zhang. Exploring heterogeneous product networks for discovering collective marketing hyping behavior. *Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*, pp. 40-51, 2016. (Australia ERA Ranked **A**, Core ranked **A**)
 8. Li Gao, Yao Lu, **Qin Zhang**, Weixue Lu and Yue Hu. Query expansion for exploratory search with subtopic discovery in community question answering. *International Joint Conference on Neural Networks IJCNN*, pp. 4715-4720, 2016. (Australia ERA Ranked **A**, Core ranked **A**)

Abstract

Due to the potential value of social relations [125] [34], social recommendation has attracted a lot of attention recently in the research communities.

Modelling time drifting data is a central problem in this area. Temporal changes, such as the emergence of new products or services and the changes in users' purchase behaviour pattern, bring unique challenges. The need for modelling time changes at the level of each individual significantly reduces the amount of available data for detecting such changes. Thus we should resort to more accurate techniques that suffice the dynamic nature of social recommendations.

Besides, the essence of social recommendation methods is utilizing users explicit social connections to improve recommendation results. However, explicit social connection information is not always available in real-world recommender systems. Only few Web sites have implemented the social or trust mechanisms, like Epinions and Douban. Lacking social recommendation data greatly limits the impact and utilization of social recommendation methods. Fortunately, in case that we do not have explicit social information, we can always compute a set of implicit social information to improve the recommendation performance.

To this end, we propose a new Implicit Social Recommendation(ISR) model from cascade data (a variant of time series) in Chapter 3, which makes recommendations based on the inferred latent social network. It can sufficiently mine the information contained in time by mining the cascade data and identify the dynamic changes in the users in time by using the latest updated social network to make recommendations.

Further, since the temporal information is crucial in social recommendation, we explore the temporal behaviour pattern from users' time series data in Chapter 4. The main challenge is finding the discriminative and explainable features (shapelets) that can best represent the raw time series data. We build an economical shapelet learning model that can automatically-learn shapelets without labels. Specifically, a novel Unsupervised Discriminative Subsequence Mining (UDSM) model that automatically-learn shapelets from unlabelled time series data.

Another challenge in social recommendation is the continually appeared new users and new items. In real-world application, this problem is unavoidable. How to involve this kind of trapezoidal data streams and how to handle this problem smoothly must be considered. Most of the existing research either builds a general model in a sample scenario such as ignoring the dynamic nature of the system and the problem of new users and items, or only focuses on some specific fields, such as dynamic restaurant recommendation and timely news recommendation. We propose a new Online Learning with Streaming Features (OLSF) algorithm and its two variants OLSF-I and OLSF-II for mining trapezoidal data streams in Chapter 5. OLSF and its variants combine online learning and streaming feature (i.e. continuously appeared new items) together to handle the double-streaming data.

Finally, we introduce a novel General Online Dynamic Social Recommendation(GODSR) model in Chapter 6, which combines network inference from cascade data, double-streaming users and items, and collaborative filtering together in an iterative process. By inferring the latent dynamic core social networks from cascade data, identifying the drift of a users preferences and involving new users and items in the online learning process, GODSR method reacts rapidly and makes accurate recommendations to users when new data arrive.

Contents

Contents	ix
List of Figures	xiv
List of Tables	xix
1 Introduction	1
1.1 Background	1
1.2 Motivations and Significances	4
1.2.1 Implicit networks in social recommendation	4
1.2.2 Temporal Pattern learning from User behaviours	6
1.2.3 Dynamics in doubly-streaming Data	7
1.2.4 All-in-One: Online dynamic implicit social recommendation	9
1.3 Research Problems	10
1.3.1 Implicit social recommendation with cascade data	10
1.3.2 Unsupervised user behaviour pattern learning	11
1.3.3 Online learning with streaming features	11
1.3.4 Generalized online dynamic implicit social recommendation	12
1.4 Thesis Contributions	12
1.4.1 Implicit social recommendation with cascade data	12
1.4.2 Feature learning from unsupervised time series	13
1.4.3 Online learning with trapezoidal data streams	13
1.4.4 Online dynamic implicit social recommendation	14
1.5 Thesis Overview	15

Nomenclature		1
2 Literature review		17
2.1 Implicit social recommendation		17
2.1.1 Traditional recommender systems		17
2.1.2 Explicit social recommenders		19
2.1.3 Implicit social recommenders		19
2.2 Feature learning from unsupervised time series		20
2.2.1 Shapelets selection		20
2.2.2 Shapelet learning		21
2.2.3 Unsupervised seature selection		21
2.2.4 Shapelets for clustering		22
2.3 Online learning from trapezoidal data streams		22
2.3.1 Traditional online learning		23
2.3.2 Traditional feature selection		23
2.3.3 Online feature selection		24
2.3.4 Online streaming feature selection		25
2.4 Online dynamic social recommendation		25
2.4.1 Offline recommenders		25
2.4.2 Online recommenders		26
I Implicit Social Recommendation with cascade data		27
3 Inferring Latent Network from Cascade Data for Social Recommendation		28
3.1 Introduction		28
3.2 Problem definition		31
3.3 The proposed ISR model		32
3.3.1 Dynamic social network inference		33
3.3.2 Social recommendation regularization term		35
3.3.3 Low-rank matrix factorization.		36
3.4 Algorithm		37
3.4.1 Implicit social recommendation algorithm		37

3.4.2	Convergence analysis	40
3.4.3	Complexity analysis	40
3.5	Experiments	40
3.5.1	Datasets	41
3.5.2	Metrics	43
3.5.3	Experimental setup	44
3.5.4	Comparisons	45
3.5.5	Performance in terms of data freshness	46
3.5.6	Performance in terms of observation window	47
3.6	Conclusion	47
II Exploration on users' temporal behaviours		51
4	Discriminative Subsequence Mining for Unsupervised Time Series	52
4.1	Introduction	52
4.2	Preliminaries	54
4.3	Unsupervised shapelet learning	56
4.3.1	Shapelet-transformed representation	56
4.3.2	Pseudo-class labels	57
4.3.3	Spectral analysis	57
4.3.4	Least-squares minimization	58
4.3.5	Shapelet similarity minimization	58
4.3.6	Unsupervised discriminative subsequence mining	59
4.4	The algorithm	59
4.4.1	Learning algorithm	59
4.4.2	Convergence	64
4.4.3	Initialization	64
4.4.4	Complexity analysis	65
4.5	Experiments	65
4.5.1	Data sets	65
4.5.2	Measures	66
4.5.3	Baseline methods	71

4.5.4	Comparison	74
4.5.5	Time series with unequal length	75
4.5.6	Parameter study	78
4.5.7	Running time and convergence curve	80
4.6	Conclusions	81
III Online Learning with double-streaming data		83
5	Online Learning from Trapezoidal Data Streams	84
5.1	Introduction	84
5.2	Problem setting	87
5.3	Online learning with trapezoidal data streams	89
5.4	Theoretical analysis	94
5.5	Experiments	100
5.5.1	Experiment I: Comparisons between OL_{SF} and its two variants	101
5.5.2	Experiment II: Comparisons with benchmarks	107
5.5.3	Experiment III: Comparisons with the state-of-the-art online feature selection algorithms	109
5.5.4	Experiment IV: Applications to real-world trapezoidal data streams	113
5.5.5	Discussions	114
5.6	Conclusions	114
IV Dynamic implicit social recommendation		116
6	Online dynamic implicit social recommendation	117
6.1	Introduction	117
6.2	Problem setting	120
6.3	General online dynamic social recommendation	121
6.3.1	Online update strategy	123
6.3.2	The scenario with newly appearing users and items	126
6.3.3	Complexity analysis	127

6.4 Experiments	127
6.5 Conclusion	131
V Conclusions and future works	134
7 Conclusions and Future Works	135
7.1 Summary	135
7.2 Future Work	137
References	139

List of Figures

3.1	An example of Implicit Social Recommendation (ISR) framework. We infer both the structure and the strength of latent dynamic social network A from cascade data C to catch the dynamic changes of the users by using a generative probabilistic model. Based on the inferred latent dynamic social network A , ISR model learns the low-rank item matrix V and user matrix U to predict the missing values in rating data R . By jointly learning the social relationships and missing ratings from both rating and cascade data, ISR can absorb the potential damage brought up by noisy explicit social network and capture the most recent preferences of the users and give the exact recommendations.	29
3.2	MAEs and RMSEs obtained by ISR algorithm in terms of different data freshness (time distances between the target ratings and historical ratings) values on the three datasets.	49
3.3	MAEs and RMSEs obtained by ISR algorithm in terms of observation window on the three datasets.	50
4.1	The top left blue curve is a rectangular signal with Gaussian noise while the bottom left blue curve is a sinusoidal signal with Gaussian noise. The short curves on the right, marked in bold red, are learned shapelets. We observe that the learned shapelets differ from all the candidate segments and are robust to noise.	53

4.2	The framework of the proposed Unsupervised Discriminative Subsequence Mining (UDSM) model. From the original time series ①, we first learn shapelets using a shapelet similarity minimization principle ②. Then we calculate the minimum distance between the time series and the learned shapelets ③. We transform the time series into a shapelet-based space ④. In low-dimensional shapelet-space, the UDSM learns the pseudo-class labels and a pseudo classifier using spectral analysis and regularized least-squares minimization⑤. Then, we update the shapelet with the new learned pseudo-class labels and the pseudo classifier ②. This process is repeated until convergence. Finally, the UDSM ensures the optimal shapelets and the pseudo-class labels.	55
4.3	An illustration of UDSM on the four data sets: Coffee, DiatomSizeReduction, ECG200 and SonyAIBoRobotSurfaceII. The left part of each figure shows the two shapelets S1 and S2 learned from the original time series. The middle shows the closest match of the learned shapelets to the original time series examples, where T_1 and T_2 are drawn from one of the classes and T_3 and T_4 are drawn from the other class. We can see that the learned shapelets are discriminative features. The right shows the shapelet-transformed plots of the original time series data set. Different colors show different class labels. The dots in red circles and blue squares are the labels obtained by UDSM, which closely matches the original class labels.	73
4.4	An example of the generalizability of UDSM on sinusoidal signals and rectangular signals. The top left blue curves are rectangular signals with Gaussian noise of length 100 while the bottom black curves are sinusoidal signals with Gaussian noise of length 200. The top right short curves marked in bold red are two learned shapelets of lengths 12 and 30 respectively. We see that UDSM can generally handle both the time series and shapelets of different length. The RI and NMI obtained by UDSM for this data set are both 1, which means perfect clustering.	75

4.5	The performance of UDSM with respect to the number and the length of the learned shapelets (measured by the Rand index, the NMI and the running time). In terms of RI and NMI, UDSM reaches its best performance at the top-middle or top-right corners. And the running time continuously increases with an increase in the number and the length of the shapelets. This shows that the UDSM does not need to learn long shapelets since short shapelets achieve sufficient or better clustering performance and also save time. In terms of the number of shapelets, it is not necessary to set a large value. In most of the data sets too many shapelets would lead to repeating or overlapping. For the time series sets with complex lines, like SonyAIBORobotSurfaceII, a reasonably large number of shapelets can be set to learn the numerous distinct shapes of the original time series. In summary, the length and the number of shapelets, in practice, does not need to be large values.	76
4.6	An illustration of the shapelets learned by UDSM on the four real world data sets. The left shows time series examples of the original data sets. Time series from the same class are shown together. The right shows the learned shapelets. Shapelets in blue are learned by increasing the length from 10 to 60 with a step 10. Shapelets in red are learned by increasing the number from 1 to 6 with a step of 1. We see when increasing the length of shapelets, there is a heavy overlap in learned shapelets. When varying the shapelet number, the learned shapelets change but they still overlap. These observations confirm that in practice the length and the number of shapelets does not need to be large values.	77
4.7	Running time with respect to the length and number of time series.	79
4.8	Convergence curves of UDSM over Coffee, DiatomSizeReduction, ECG200 and SonyAIBORobotSurfaceII data sets.	81

5.1	Each column is a document set. We observe document sets continuously arrive as a stream. In each column, words in colored boxes are new words introduced by document sets and the number associates with each word is the importance rank for classification. For example, in document set 16, the word "wolverin" in the blue box was first observed and then became one of the most important words for classification in Document set 39.	85
5.2	Comparison among the proposed three algorithms OL_{SF} , OL_{SF} -I and OL_{SF} -II on the 12 UCI data sets. We can observe that OL_{SF} -I and OL_{SF} -II outperform OL_{SF} because their "soft" update strategies can avoid overfitting to noise.	103
5.3	Performance comparison with respect to the projected feature space B . The results show that OL_{SF} -I and OL_{SF} -II are robust algorithms under different projected subspaces B	104
5.4	The average number of error predictions with respect to parameter C . We can choose the best parameter for the algorithms on the 12 UCI data sets.	106
5.5	Comparison of the four algorithms under online learning setting. We can observe that OL_{SF} -I obtains the best results on all the 12 data sets because its sparsity strategy can significantly improve the performance and outperforms the Perceptron update strategy.	108
5.6	Online classification accuracy with respect to the parameter B . We can observe that OL_{SF} -I performs the best especially when the feature space is sparse, i.e., B is very small.	110
5.7	Comparison with respect to the average number of error predictions. We can observe that OL_{SF} -I and OL_{SF} -II performs better than OFS and OFS_P by adding a flexible learning rate τ_t	111
5.8	Comparison with respect to online prediction. We can observe that the curves of OL_{SF} -I and OL_{SF} -II descend much faster than those of OFS and OFS_P and eventually become stable with lower error rates.	112
5.9	Performance on real trapezoidal data streams($B = 0.01$).	115

6.1	The framework of the General Online Dynamic Social Recommendation method. When a new record $(user_n, item_n, t_n, r_n)$ arrives, we first find out whether a new user or/and new item has appeared or not. If the answer is yes, we go to the extension and initiation step, otherwise we go to the loss calculation step directly. Then, if it suffers loss in this current record, we go to the update step to improve the model, otherwise we skip this record and go to the next loop with the following new record.	118
6.2	Examples for influence paths on the three data sets	132

List of Tables

1.1	Structure of the thesis with references to the chapters.	15
3.1	Statistics of Datasets	42
3.2	Performance Comparisons with 80% of data as training set	42
3.3	Performance Comparisons with 60% of data as training set	43
4.1	Statistics of the benchmark time series data sets	63
4.2	Rand Index (RI) comparisons on 36 time series data sets.	67
4.3	Rand Index (RI) comparisons on 36 time series data sets (continued with Table 4.2).	68
4.4	Normalized Mutual Information (NMI) comparisons on 36 time series data sets.	69
4.5	Normalized Mutual Information (NMI) comparisons on 36 time series data sets (continued with Table 4.4)	70
5.1	Symbols and Notations.	88
5.2	The data sets used in the experiments	101
5.3	The average number of prediction errors on the 12 UCI data sets .	102
5.4	The average number of error predictions on the 12 UCI data sets with respect to parameter B	105
5.5	Comparison with respect to the average number of error predictions.	111
5.6	Comparison with respect to the average number of error predictions ($B = 0.001$).	113
6.1	Performance Comparisons in terms of MAE	128
6.2	Performance Comparisons in terms of RMSE	129