

Faculty of Engineering and Information Technology  
University of Technology Sydney

# **Nonoccurring Sequential Behavior Analytics**

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

by

Wei Wang

July 2020



## CERTIFICATE OF ORIGINAL AUTHORSHIP

I, Wei Wang declare that this thesis, is submitted in fulfilment of the requirements for the award of the degree: Doctor of Philosophy, 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.

This research is supported by the Australian Government Research Training Program.

Signature:      Production Note:  
Signature removed prior to publication.

Date:      12 July 2020



# Acknowledgments

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

In addition, I would like to thank all my colleagues in Advanced Analytics Institute: Liang Hu, Guansong Pang, Thac Do, Chengzhang Zhu, and Shoujin Wang for their stimulating discussions and scientific advice. Without their generous support, this dissertation would not have been possible.

Last but not least, I would like to thank my family for their unconditional support, both financially and emotionally throughout my Ph.D. studies. Without their encouragement, finishing this dissertation would have been impossible; without them, nothing would have any value.

Wei Wang

July 2020 @ UTS



# Contents

Certificate . . . . .	<b>i</b>
Acknowledgment . . . . .	<b>iii</b>
List of Figures . . . . .	<b>xi</b>
List of Tables . . . . .	<b>xiii</b>
List of Publications . . . . .	<b>xv</b>
Abstract . . . . .	<b>xvii</b>
<b>Chapter 1 Introduction . . . . .</b>	<b>1</b>
1.1 Research Motivation and Challenges . . . . .	<b>1</b>
1.2 Research Aims and Objectives . . . . .	<b>8</b>
1.3 Thesis Contributions . . . . .	<b>9</b>
1.4 Thesis Organization . . . . .	<b>12</b>
<b>Chapter 2 Preliminaries and Literature Review . . . . .</b>	<b>19</b>
2.1 Negative Sequential Pattern Mining . . . . .	<b>20</b>
2.1.1 Categorization of Existing NSP Mining Methods . . . . .	<b>20</b>
2.1.2 Technical Challenges . . . . .	<b>24</b>
2.1.3 Gap Analysis in NSP Mining . . . . .	<b>27</b>
2.2 Representative Subset Selection . . . . .	<b>29</b>
2.2.1 Determinantal Point Processes . . . . .	<b>31</b>
2.2.2 Fixed-Size Determinantal Point Processes . . . . .	<b>33</b>

## CONTENTS

---

2.2.3	Structured Determinantal Point Processes . . . . .	34
2.2.4	Gap Analysis in Representative Subset Selection . . . . .	35
2.3	Sequential Basket Recommendation . . . . .	37
2.3.1	Collaborative Filtering Methods . . . . .	38
2.3.2	Sequential Recommendation Methods . . . . .	40
2.3.3	Session-based and Interactive Recommendation . . . . .	41
2.3.4	Hybrid Recommendation Methods . . . . .	42
2.3.5	Gap Analysis in Sequential Basket Recommendation . . . . .	44
2.4	Summary . . . . .	46
 <b>Chapter 3 Nonoccurring Sequential Behavior Analytics For-</b>		
	<b>malization . . . . .</b>	<b>47</b>
3.1	Basic Concepts . . . . .	47
3.1.1	Item and Its Properties . . . . .	49
3.1.2	Element and Its Properties . . . . .	50
3.1.3	Sequence and Its Properties . . . . .	54
3.2	Problem Statement of NSB Analytics . . . . .	56
3.2.1	Negative Sequential Pattern Mining . . . . .	56
3.2.2	Sequential Basket Recommendation . . . . .	57
3.3	Constraints and Negative Containment . . . . .	59
3.3.1	Constraint Settings . . . . .	59
3.3.2	Negative Containment . . . . .	61
3.4	Summary . . . . .	64
 <b>Chapter 4 Efficient Negative Sequential Pattern Mining . . .</b>		
4.1	Introduction . . . . .	65
4.1.1	Problem Statement . . . . .	65
4.1.2	Design and Contributions . . . . .	68



4.2	Loose Constraint-enabled Constraints and Containment . . . .	70
4.3	The VM-NSP Framework and bM-NSP Method . . . . .	73
4.3.1	Initial Seed Set Construction . . . . .	75
4.3.2	NSC Generation . . . . .	75
4.3.3	Vertical Representation Construction . . . . .	77
4.3.4	NSC Support Testing . . . . .	80
4.4	Theoretical Analysis . . . . .	81
4.4.1	Runtime Analysis of bM-NSP . . . . .	82
4.4.2	Runtime Analysis of NegGSP . . . . .	83
4.4.3	Runtime Comparison w.r.t. Data Factors . . . . .	84
4.5	Experiments and Evaluation . . . . .	85
4.5.1	Datasets and Baseline Methods . . . . .	85
4.5.2	Performance Evaluation . . . . .	87
4.5.3	Case Study . . . . .	97
4.6	Summary . . . . .	98

## Chapter 5 Representative Negative Sequential Pattern Dis-

	<b>covery . . . . .</b>	<b>99</b>
5.1	Introduction . . . . .	99
5.1.1	Problem Statement . . . . .	99
5.1.2	Design and Contributions . . . . .	100
5.2	The EINSF Method . . . . .	102
5.2.1	NSP Graph Construction . . . . .	103
5.2.2	Explicit Relation Modeling . . . . .	104
5.2.3	Implicit Relation Modeling . . . . .	107
5.2.4	Overall Relation-based Selection . . . . .	110
5.2.5	The EINSF Algorithm . . . . .	110
5.3	Experiments and Evaluation . . . . .	111

## CONTENTS

---

5.3.1	Experimental Setup . . . . .	113
5.3.2	Pattern Coverage Comparison . . . . .	116
5.3.3	Average Pattern Size Comparison . . . . .	123
5.3.4	Average Implicit Relation Strength Comparison . . . . .	125
5.3.5	Sensitivity Evaluation . . . . .	127
5.4	Summary . . . . .	129
<b>Chapter 6 Interactive Sequential Basket Recommendation . . . . .</b>		<b>131</b>
6.1	Introduction . . . . .	131
6.1.1	Problem Statement . . . . .	132
6.1.2	Design and Contributions . . . . .	137
6.2	The HAEM Model . . . . .	138
6.2.1	HAEM Architecture . . . . .	139
6.2.2	Basket Encoder . . . . .	141
6.2.3	Sequence Encoder . . . . .	145
6.2.4	Prediction Decoder . . . . .	146
6.2.5	The HAEM Learning and Prediction . . . . .	150
6.3	Experiments and Evaluation . . . . .	151
6.3.1	Datasets and Preparation . . . . .	151
6.3.2	Baseline Methods . . . . .	153
6.3.3	Experimental Settings . . . . .	155
6.3.4	Evaluation Measures . . . . .	156
6.3.5	Accuracy Evaluation . . . . .	158
6.3.6	Novelty Evaluation . . . . .	166
6.4	Summary . . . . .	168
<b>Chapter 7 Conclusions and Future Directions . . . . .</b>		<b>170</b>
7.1	Conclusions . . . . .	170

7.1.1	Efficient Negative Sequential Pattern Mining . . . . .	170
7.1.2	Representative Negative Sequential Pattern Discovery . . . . .	171
7.1.3	Interactive Sequential Basket Recommendation . . . . .	171
7.2	Future Directions . . . . .	172
7.2.1	Incremental NSP Mining over Streamed Behaviors . . . . .	172
7.2.2	Top-K NSP Mining . . . . .	173
7.2.3	Quantitative NSP Discovery . . . . .	173
7.2.4	NSP Discovery with Complex Hierarchical Structure . . . . .	174
7.2.5	Feature-aware Sequential Behavior Prediction . . . . .	175
	<b>Bibliography . . . . .</b>	<b>176</b>



# List of Figures

1.1	Thesis Structure . . . . .	13
4.1	Framework of Vertical Mining of Negative Sequential Patterns: VM-NSP . . . . .	74
4.2	Runtime Comparison on Datasets DS1-DS4. . . . .	88
4.3	Runtime Comparison on Datasets DS5-DS8. . . . .	89
4.4	Runtime Comparison on Data Factors (C and T). . . . .	90
4.5	Runtime Comparison on Data Factors (S, I and N). . . . .	91
4.6	Scalability Evaluation on Datasets DS6 and DS8. . . . .	95
4.7	Coverage Evaluation on Datasets DS3 and DS4. . . . .	96
5.1	Representative NSP Discovery for High-quality and Diverse NSP by Modeling Both Explicit and Implicit Element/Pattern- wise Relations . . . . .	101
5.2	Sequence Coverage Evaluation on Datasets DS1-DS8. . . . .	117
5.3	Item Coverage Evaluation on Datasets DS1-DS8. . . . .	120
5.4	Averagely Covered Item Frequency Evaluation on Datasets DS1-DS8. . . . .	122
5.5	Average Pattern Size Comparison on Datasets DS1-DS8. . . .	123
5.6	Comparison of Average Implicit Relation Strength (IRS) on Datasets DS1-DS8. . . . .	126

## LIST OF FIGURES

---

6.1	An Example of Interactive Sequential Basket Recommendation.	133
6.2	The HAEM Model for Sequential Basket Recommendation . . .	140
6.3	The Basket Encoder. It maps an input basket to a basket representation and models linear and pairwise attentive intra-basket relations. . . . .	141
6.4	The Sequence Encoder. It maps the basket representation sequence to an annotation representation sequence. . . . .	145
6.5	The Prediction Decoder. It makes predictions for sequential basket recommendations based on the annotation sequence learned. . . . .	147
6.6	Novelty Comparison of Methods on Two Datasets. . . . .	167

# List of Tables

1.1	Thesis Research Overview (Main Body) . . . . .	18
2.1	Categorization Overview of Existing NSP Mining Methods . .	25
3.1	Main Notations in NSB Analytics (Part I) . . . . .	47
3.2	Main Notations in NSB Analytics (Part II) . . . . .	48
3.3	Main Notations in NSB Analytics (Part III) . . . . .	49
4.1	Toy Example . . . . .	78
4.2	The Constructed BHTs of the 1-length Sequences in Table 4.1	78
4.3	Runtime Sensitivity against Data Factors on DS1.X (Part I) .	92
4.4	Runtime Sensitivity against Data Factors on DS1.X (Part II) .	93
5.1	Sequence Coverage Sensitivity of EINSF and Baselines w.r.t. Data Factors for $k = 150$ . . . . .	128
6.1	Notations in Sequential Basket Recommendation (Part I) . . .	138
6.2	Notations in Sequential Basket Recommendation (Part II) . .	139
6.3	Statistics of Experimental Datasets . . . . .	153
6.4	Accuracy of HAEM against Baselines for Next-basket Recom- mendation on the Ta-Feng Dataset . . . . .	159

## LIST OF TABLES

---

6.5	Accuracy of HAEM against Baselines for Next-basket Recommendation on the IJCAI-15 Dataset . . . . .	160
6.6	Accuracy Effect of Negative Feedback-based Refinement on the Ta-Feng Dataset . . . . .	164
6.7	Accuracy Effect of Negative Feedback-based Refinement on the IJCAI-15 Dataset . . . . .	165



# List of Publications

## Published Papers

1. **Wei Wang**, and Longbing Cao, 2019. “Negative Sequence Analysis: A Review,” *ACM Computing Surveys*, 52(2), pp. 32:1-32:39.

## Papers under Revision

1. **Wei Wang**, and Longbing Cao, “VM-NSP: An Efficient Vertical Negative Sequential Pattern Mining Framework,” *ACM Transactions on Information Systems*, Under review (Minor revision, Submission ID: TOIS-2020-0018).
2. **Wei Wang**, and Longbing Cao, “Determinantal Point Process-based Relation Modeling in Negative Sequence Analysis,” *IEEE Transactions on Knowledge and Data Engineering*, Under review (Major revision, Submission ID: TKDE-2020-03-0252).
3. **Wei Wang**, and Longbing Cao, “Sequential Basket Recommendation by Iteratively Learning Basket Relations and Interactive Feedbacks,” *ACM Transactions on Information Systems*, Under review (Minor revision, Submission ID: TOIS-2019-0030.R1).



# Abstract

Behavior analytics has attracted increasing attention in broad communities as a major research area in understanding and managing the dynamics of complex systems and problems such as series of medical treatments, interactions between customers and service providers, and online communications. Sequential behavior analytics aims to understand, analyze, detect, and predict existing or future behaviors and behavior sequences. Existing methods for sequential behavior analytics only focus on occurred or to-occur behaviors (also called positive behaviors), while ignoring nonoccurring behaviors (also called negative behaviors), which are often useful for understanding, managing and predicting hidden or unseen yet important behaviors that differ from and typically mix with occurred ones. Nonoccurring behaviors complement occurring ones for complete and deep behavior analytics, while very limited theoretical progress has been made.

This thesis studied the theory to comprehensively model the complex relations within and between behaviors and to effectively discover and predict interesting sequential occurring and nonoccurring behaviors. Specifically, it focused on (1) forming a comprehensive and systematic representation, formalization, and theoretical system for defining and representing the concepts, problems, constraint settings, and negative containment of nonoccurring sequential behavior (NSB) analytics; (2) efficiently discovering the

## *ABSTRACT*

---

high-frequency negative sequential patterns (NSP) composed of both occurring/nonoccurring behaviors; (3) discovering the representative NSP subset by exploring the complicated explicit/implicit behavior relations; and (4) enabling the sequential basket recommendation system (SBRS) through learning behavior relations and interactive feedback. Accordingly, this thesis proposed (1) a vertical NSP mining framework and its instantiation for the efficient discovery of the complete set of NSP with the loose negative element constraint via the vertical representation of each sequence, which guarantees the coverage of flexible patterns with complicated behavior relations; (2) a determinantal point processes-based (DPP-based) representative NSP discovery approach for the selection of a representative subset of the high-quality and diverse patterns by jointly modeling explicit and implicit sequential element/pattern relations; and (3) a hierarchical attentive encoder-decoder model for interactive sequential basket recommendation, which jointly models both intra-/inter-basket relations in sequential user basket behaviors as well as incorporates positive/negative feedback to enable negative feedback-based refinement.

The extensive empirical analysis of the proposed methods demonstrated that our methods performed significantly better than the state-of-the-art methods in the same domain in terms of multiple evaluation metrics.