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

List of Figures

List of Tables

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 nonocurring 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

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.