

DOCTORAL THESIS

Detecting stochastic motifs in network and sequence data for human behavior analysis

Liu, Kai

Date of Award:
2014

[Link to publication](#)

General rights

Copyright and intellectual property rights for the publications made accessible in HKBU Scholars are retained by the authors and/or other copyright owners. In addition to the restrictions prescribed by the Copyright Ordinance of Hong Kong, all users and readers must also observe the following terms of use:

- Users may download and print one copy of any publication from HKBU Scholars for the purpose of private study or research
- Users cannot further distribute the material or use it for any profit-making activity or commercial gain
- To share publications in HKBU Scholars with others, users are welcome to freely distribute the permanent URL assigned to the publication

Abstract

With the recent advent of Web 2.0, mobile computing, and pervasive sensing technologies, human activities can readily be logged, leaving digital traces of different forms. For instance, human communication activities recorded in online social networks allow user interactions to be represented as “network” data. Also, human daily activities can be tracked in a smart house, where the log of sensor triggering events can be represented as “sequence” data. This thesis research aims to develop computational data mining algorithms using the generative modeling approach to extract salient patterns (*motifs*) embedded in such *network* and *sequence* data, and to apply them for human behavior analysis.

Motifs are defined as the recurrent over-represented patterns embedded in the data, and have been known to be effective for characterizing complex networks. Many motif extraction methods found in the literature assume that a motif is either present or absent. In real practice, such salient patterns can appear partially due to their stochastic nature and/or the presence of noise. Thus, the probabilistic approach is adopted in this thesis to model motifs.

For network data, we use a probability matrix to represent a network motif and propose a mixture model to extract network motifs. A component-wise EM algorithm is adopted where the optimal number of stochastic motifs is automatically determined with the help of a minimum message length criterion. Considering also the edge occurrence ordering within a motif, we model a motif as a mixture of first-order Markov chains for the extraction. Using a probabilistic approach similar to the one for network motif, an optimal set of stochastic temporal network motifs are

extracted. We carried out rigorous experiments to evaluate the performance of the proposed motif extraction algorithms using both synthetic data sets and real-world social network data sets and mobile phone usage data sets, and obtained promising results. Also, we found that some of the results can be interpreted using the social balance and social status theories which are well-known in social network analysis.

To evaluate the effectiveness of adopting stochastic temporal network motifs for not only characterizing human behaviors, we incorporate stochastic temporal network motifs as local structural features into a factor graph model for followee recommendation prediction (essentially a link prediction problem) in online social networks. The proposed motif-based factor graph model is found to outperform significantly the existing state-of-the-art methods for the prediction task.

For extract motifs from sequence data, the probabilistic framework proposed for the stochastic temporal network motif extraction is also applicable. One possible way is to make use of the edit distance in the probabilistic framework so that the subsequences with minor ordering variations can first be grouped to form the initial set of motif candidates. A mixture model can then be used to determine the optimal set of temporal motifs. We applied this approach to extract sequence motifs from a smart home data set which contains sensor triggering events corresponding to some activities performed by residents in the smart home. The unique behavior extracted for each resident based on the detected motifs is also discussed.

Keywords: Stochastic network motifs, finite mixture models, expectation maximization algorithms, social networks, stochastic temporal network motifs, mixture of Markov chains, human behavior analysis, followee recommendation, signed social networks, activity of daily living, smart environments

Acknowledgements

First of all, I would like to thank my supervisor, Dr. William CHEUNG and my co-supervisor, Prof. Jiming LIU. They provided guidance, support and a lot of patience over the past four years. Their collaboration was essential and we always had fruitful discussions leading to new ideas of my work. When I first came to Hong Kong Baptist University in 2010 as a new graduate student, I knew little about research. Dr. William Cheung kindly showed me the ways of doing research, patiently went through my shaky first few papers and kept encouraging me. He lets me understand that research is not only to read like what I did in undergraduate years, but also to deliberate and criticize the underlying ideas. Understanding only the existing ideas is not enough, and it is important to come up with new ideas for my own research. To this date, he has constantly helped me in every aspect of my studies. He taught me how to communicate with others and how to present my ideas to those with even very little background of my work. I also learnt from him how to be respectful towards others and how to be understanding. I am grateful for having Dr. Cheung and Prof. Liu as my supervisor and co-supervisor, and for all of your endless support during my studies.

There are also a number of great people who have been influential in my PhD studies. I would like to thank Dr. Yu ZHANG, Dr. Benyun SHI, Dr. Xiang WAN, Dr. Li TAO, Dr. Xia SHANG for their advices about my research project. And Dr. Pang-Ning Tan provided me great help during my visit to his research lab in College of Engineering, Michigan State University. He brought me many good ideas, and we made one to be a paper. I am very happy to have the opportunity to collaborate

with him. I also would like to thank the Computer Science Department for providing three years studentship and the support staffs in our department for their help. I learnt from some professors how to communicate with students when I worked as a teaching assistant with them. My dear friends also provide support to my study. The friendship established also made my life happy and fruitful for the passed four years.

I would like to thank also my parents and my brother for their unconditional support. I specially like to thank my wife Haiyan WU. It was impossible to finish my studies abroad without her endless love and support. I am so lucky to have her as my wife. Thank you for being there for me!

Table of Contents

Declaration	i
Abstract	ii
Acknowledgements	iv
Table of Contents	vi
List of Tables	ix
List of Figures	xi
Chapter 1 Introduction	1
1.1 From Motifs to Stochastic Motifs	3
1.2 Our Contributions	6
1.3 Organization of the Thesis	9
Chapter 2 Related Work	11
2.1 Network Motifs and Their Applications	11
2.2 Network Motif Detection	15
2.3 Probabilistic Models for Social Network Analysis	19
2.4 Probabilistic Models for Activity Recognition in Smart Environments	21
Chapter 3 Detecting Multiple Stochastic Network Motifs in Network	
Data	23
3.1 Problem Formulation and Methodology	25

3.1.1	A finite mixture model for stochastic network motif detection	26
3.2	Parameter Estimation Using the EM Algorithm	28
3.2.1	The basic EM algorithm	28
3.2.2	The CEM algorithm for estimating the optimal set of motifs	31
3.3	Experimental Results	33
3.3.1	Evaluation based on synthetic networks	33
3.3.2	Evaluation based on social network datasets	38
3.4	Computational Complexity and Runtime	50
3.4.1	Runtime	51
3.4.2	On initialization of k_{max} in CEM algorithm	52
3.5	Summary	55

Chapter 4 Detecting Stochastic Temporal Network Motifs for Human

	Communication Patterns Analysis	56
4.1	Problem Formulations	59
4.2	Methodology	61
4.2.1	Sampling temporal subgraphs	61
4.2.2	Canonical labeling and isomorphism	63
4.2.3	Inferring a mixture of STNMs	65
4.2.4	Inferring motif evolution using HMM	66
4.3	Experimental Results	68
4.3.1	Performance evaluation on synthetic networks	68
4.3.2	Application to mobile phone usage analysis	71
4.4	Runtime Evaluation	77
4.5	Summary	79

Chapter 5 Temporal Network Motif Factor Graph Model for Followee

	Recommendation in Social Media	80
5.1	Problem Formulation	82
5.2	Modeling	83

5.2.1	Motif-based factor graph model	84
5.2.2	Parameter estimation of MotifGM	88
5.3	Experimental Results	89
5.3.1	Performance on followee recommendation accuracy	94
5.3.2	Runtime	96
5.4	Summary	99
Chapter 6 Extracting Behavioral Motifs for Characterizing Human Daily Activities in Smart Environments		100
6.1	Problem Formulation	102
6.1.1	Step 1: Sampling event patterns	103
6.1.2	Step 2: Motif candidates based on Z-score	104
6.1.3	Step 3: Behavioral motifs as components of mixture models	105
6.1.4	The EM algorithm for mixture model	105
6.2	Experimental Results and Discussion	107
6.2.1	Results of motif detection	107
6.2.2	Sensitivity analysis on parameter settings	110
6.3	Summary	113
Chapter 7 Conclusions and Future Work		117
7.1	Concluding Remarks	117
7.2	Future Directions	118
7.2.1	Scalable subgraph sampling algorithms	119
7.2.2	Online learning algorithms for stream data	120
7.2.3	Other applications of motifs	124
Bibliography		125
Curriculum Vitae		141