Doctor of Philosophy

Diversified Probabilistic Graphical Models

by

Maoying Qiao

supervised

by

Prof. Dacheng Tao

the Centre for Quantum Computation and Intelligent Systems (QCIS)

the Faculty of Engineering and Information Technology (FEIT)

the University of Technology Sydney (UTS)

July, 2016

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 Candidate

UNIVERSITY OF TECHNOLOGY SYDNEY

ACKNOWLEDGEMENTS

Firstly, I would like to express my sincere gratitude to my supervisor Prof. Dacheng Tao for his continuous support of my Ph.D study. Thanks for his consistent patience and motivation, for his encouraging attitude and expert knowledage for my research. His strict academic attitude and deligent work style have played a model role for me and will continue to benefit me through my life. It is no exaggeration to say without his help steering my research direction, I would not have finished this thesis so smoothly and on time.

Besides my principal supervisor, I would like to thank Dr. Wei Bian. I want to thanks for his never bored discussion with my seemingly endless simple questions from research motivation, model development, algorithm implementation, and paper drafting. With these specific and detailed technical discussion, I have gained practical skills to effectively and efficiently develop and implement my research problems. I also gain deep understandings for my research areas. In addition, these discussion moments between us and with Qiang Li have been an inseparable part of my Ph.D life, and will always be cherished by myself.

I would like to thank Dr. Richard Yi da Xu for his advices on my study of Monte Carlo sampling methods. His help undoubtably has widened my research area. I also want to thank my colleages in Prof. Tao's group for their academic discussion and help. My special thanks goes to Ms. Jemima Moore for helping improve my English skills in both academic publications and presentations.

My sincere thanks also goes to China Scholarship Council (CSC), University of Technology Sydney (UTS) and the centre for Quantum Computation & Information System (QCIS). With the financial support from CSC-UTS joint scholarship, I could concentrate on my research study and did not have to worry about my living. QCIS has provided a excellent and supportive research environment for academic-related activities.

Last but not the least, my gratitude extends to my family who have been patiently encouraging and waiting for the finish of this thesis.

ABSTRACT

Probabilistic graphical models (PGMs) as diverse as Bayesian networks and Markov random fields have provided a fundamental framework to learn and reason using limited and noisy observations. Examples include, but are not limited to, hidden Markov models (HMMs), sequential graphical models, and probabilistic principal component analysis mixture models (PPCA-MM). PGMs have been used in a wide variety of applications such as speech recognition, natural language processing, web searching, and image understanding. However, one potential drawback of using PGMs with traditional learning and inference methods is that the learned parameters or inferred variables are easily trapped within local, clustered optima rather than distributed evenly across the whole space. Taking mixture models as an example, the learned mixing components might overlap. Consequently, the resulting models might show ambiguity when clustering is performed based on these overlapping mixing components. This phenomenon might limit PGM performance.

Although efforts have been made to explore a variety of priors to alleviate this potential drawback and to enhance PGM performance, diverse priors have yet to be fully explored and utilized. Diversity is a concept that encourages counterpart model parameters and variables to repel as much as possible and, in doing so, spread out model components and decrease overlapping. However, how to explicitly encode these priors into a PGM and how to solve the resulting diversified PGMs are two critical problems that must be solved. This thesis proposes a unified framework to constrain PGMs with diverse priors. Three different PGMs - HMMs, time-varying determinantal point processes (TV-DPPs), and PCA-MMs - are elaborated to demonstrate the proposed diversified PGM framework. For each PGM, three basic constituent framework elements are examined: which part of the traditional PGM is diversified, how to formulate the diversity, and how to solve the diversified version, e.g., parameter learning and inference. In addition, experiments are conducted using various application scenarios to verify the effectiveness of the proposed diversified PGMs.

***Keywords*:** Probabilistic graphical models (PGMs), diversity prior, determinantal point processes (DPPs), hidden Markov models (HMMs), time-varying DPPs, probabilistic principal component analysis (PPCA).

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES