CROSS-DOMAIN LEARNING FOR NETWORK REPRESENTATIONS

Shan Xue

Faculty of Engineering and Information Technology

University of Technology Sydney

A thesis submitted for the Degree of

Doctor of Philosophy

February 2019

CERTIFICATE OF AUTHORSHIP/ORIGINALITY

This thesis is the result of a research candidate conducted jointly with Shanghai University as part of a collaborative Doctoral degree. I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of the requirements for a degree except as part of the collaborative doctoral degree and/or 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.

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

Production Note:
Signature removed prior to publication.

Shan Xue

February 2019

Acknowledgements

On having completed this thesis, I would like to express my sincere gratitude to my principal supervisor Distinguished Professor Jie Lu. Without her patience and encouragement, I would not have been able to complete this PhD program. Professor Lu led me in a new academic research direction, giving me as much guidance as possible and trusted me to pursue my own research interests. She taught me how to think and study independently, and how to solve difficult scientific problems in a flexible and rigorous way. She instilled me with confidence and motivated me when I felt lost or afraid about the future. Her wisdom, decisiveness and sharp insights inspired me to keep going regardless of any difficulty I encountered. The help that Professor Lu gave me will continue to benefit me throughout my life.

I would also to express my thankfulness to my co-supervisor A./Professor Guangquan Zhang, who has been like a father to me throughout the duration of my four-year PhD candidature. I always knew that, in him, I had someone to whom I could talk about any of my concerns and he would always lead me to a solution.

I would like to express my thankfulness to every member of the Decision Systems & e-Service Intelligence Lab (DeSI) in the Centre for Artificial Intelligence (CAI). It was a wonderful experience to spend four years with these dedicated researchers. I will also have fond memories of UTS and Sydney. I especially thank Dr. Junyu

Xuan who helped me greatly during my PhD candidature, Dr. Vahid Behbood who strictly helped me solve several problems related to my research, Dr. Hua Zuo who was like an older sister to me and Dr. Qian Zhang who studied with me throughout the four years. I would also thank Dr. Fan Dong, Yiliao Song and Feng Liu who helped me and shared my joys and sadness.

I am incredibly grateful to my parents for their generosity and encouragement. This thesis would never have been completed without their constant support and understanding. I am also thankful to my friends who accompanied me and were by my side, listening to me and supporting me through the arduous journey of four years.

Abstract

Network representation aims to learn a latent feature space so that artificial intelligent algorithms can be applied based on the latent features. The set of latent features is obtained from the information hidden behind network structures, which is learned to provide knowledge for traditional machine learning tasks, such as node classification, recommendation and data visualization. Networks, which are a kind of structured data, limit the representation performance in the structure searching process. Therefore, a good node sampling strategy plays an important role in network representation. Recent research has driven significant progress in network representation by employing random walk as the network sampling strategy. However, real-world large-scale information networks naturally have structural sparsity. The existing approaches to random walk-based network representations are in the domain-specific view to represent the nodes in a vector format, which cannot guarantee a good representation by one network knowledge learning.

To address these gaps, this research proposes a framework and develops two algorithms to adapt useful information across relational large-scale information networks and allows the information of the network structure to be transferred from one network to another network to improve the performance of network representation. First, a novel framework of transferring structures across large-scale information

networks (FTLSIN) is proposed. FTLSIN consists of a two-layer random walk to measure the relations between two networks and predict the links across them. Second, a cross-domain network representation algorithm (CDNR) is proposed to demonstrate the knowledge which transfers across domains. CDNR learns the structural information from dense networks to sparse networks and further defines the two-layer random walk in unsupervised feature learning with a cross-domain node mapping procedure and a cross-domain walk mapping procedure. Thirdly, a cross-domain similarity learning algorithm (CDSL) is proposed to acquire the most relevant knowledge from the external network. CDSL is nested in the biased random walk-based node sampling and targets the minimum cost of searching the neighborhood in the biased random walk that considers the first-order and second-order walking; and the neighborhood is described by a dual centrality indicator which consists of closeness centrality and betweenness centrality. The developed framework and the two algorithms are very innovative and significantly contribute to both fields of transfer learning and network representation.

Table of Contents

List of Figures

List of Tables