Designing Machine Learning Models for Graph Analytics

by

Hanchen Wang

A THESIS SUBMITTED IN FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Australian Artificial Intelligence Institute (AAII)
Faculty of Engineering and Information Technology (FEIT)
University of Technology Sydney (UTS)

June, 2021

CERTIFICATE OF ORIGINAL AUTHORSHIP

I, Hanchen Wang declare that this thesis, is submitted in fulfilment of the requirements for the award of 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 referenced 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: 15/06/2021

ACKNOWLEDGEMENTS

First of all, I would like to express my sincere gratitude to my principal supervisor Prof. Ying Zhang. I have benefited greatly from his constant support, knowledge, guidance and continuous encouragement. He is not only my valued mentor, but also a true friend. As a friend, he is always ready to help me. Whenever I feel frustrated, he is ready to offer advice and practical support. I cannot be more grateful for the incredible advice and support he has given me on my research and career.

Secondly, I would like to express my greatest gratitude to Prof. Defu Lian and my co-supervisor Lu Qin. They always provides insightful advice on my research topics. Their guidance significantly extended my knowledge and helped me pursue a correct research direction all the time. I would like to thank Prof. Xuemin Lin, Prof. Wenjie Zhang and Prof. Wei Wang for their support, they serve as examples of excellent researchers for me and share invaluable ideas and experiences with me.

Thirdly, I would like to acknowledge Prof. Xiangjian He, Prof. Xiaoyang Wang, Dr. Wanqi Liu, Dr. Dong Wen, Dr. Yiguang Lin, Dr. Chen Chen, Mr. Peilun Yang , Mr. Jianke Yu and Mr. Han Chen for supporting my works during my Ph.D. study. I thank Dr. Wanqi Liu for sharing her enjoyment of conducting research and her rich experience in research and coding. I also thank Mr. Peilun

Yang for the wonderful ideas and inspirations he gave me in our discussions.

My special thanks go to the following people: Prof. Long Yuan, Prof. Jianye Yang, Dr. Xin Cao, Prof. Fan Zhang, Prof. Yixiang Fang, Prof. Zengfeng Huang, Prof. Shiyu Yang, Prof. Peng Cheng, Prof. Weiwei Liu, Dr. Xubo Wang, Dr. Dian Ouyang, Dr. Longbin Lai, Dr. Xing Feng, Dr. Haida Zhang, Dr. Fei Bi, Dr. Shenlu Wang, Dr. Wei Li, Dr. Chen Zhang, Dr. Yang Yang, Dr. Kai Wang, Dr. You Peng, Dr. Boge Liu, Dr. Wentao Li, Dr. Conggai Li, Dr. Mingjie Li, Dr. Xuefeng Chen, Dr. Bing Li, Dr. Shuiqiao Yang, Dr. Jiaojiao Jiang, Mr. Yu Hao, Mr. Yuren Mao, Ms. Xiaoshuang Chen, Mr. Yuanhang Yu, Mr. Yixing Yang, Mr. Zhengyi Yang, Mr. Qingyuan Linghu, Mr. Michael Ruisi Yu, Mr. Chenji Huang, Mr. Bohua Yang, Mr. Jiahui Yang, Mr. Yilun Huang, Mr. Rong Hu, Mr. Junhua Zhang, Mr. Yuxuan Qiu, Mr. Kongzhang Hao, Mr. Yizhang He, Mr. Shunyang Li, Mr. Gengda Zhao, Mr. Qingshuai Feng, Mr. Zhuo Ma, Ms. Yuting Zhang. The days we spent together bear unforgettable memories.

Finally, I would also like to acknowledge my father Mr. Yongming Wang, and my mother Ms. Jing Chen, who provide me with selfless love and endless encouragement. I also thank all my relatives and friends for their love and support. Thanks to my wife Ms. Yufan Chen for her timely encouragement and inspiration in my hard times and bringing me joyful life.

Abstract

With growing popularity of the machine learning methods, there have been a great number of machine learning methods proposed for graph analytics. In this thesis, we design three machine learning based models for the popular graph analysis tasks such as node classification, graph representation learning, graph interaction prediction and subgraph matching.

Firstly, we design a binarized graph neural network to efficiently obtain the vector representations for vertices and graphs. Recently, there have been some breakthroughs in graph analysis by applying the Graph Neural Networks (GNNs) following a neighborhood aggregation scheme, which demonstrate outstanding performance in many tasks. However, we observe that the parameters of the network and the embedding of nodes are represented in real-valued matrices in existing GNN-based graph embedding approaches which may limit the efficiency and scalability of these models. It is well-known that binary vector is usually much more space and time efficient than the real-valued vector. This motivates us to develop a binarized graph neural network to learn the binary representations of the nodes with binary network parameters following the GNN-based paradigm. Our proposed method can be seamlessly integrated into the existing GNN-based embedding approaches to binarize the model parameters and learn the compact embedding. Extensive experiments indicate that the proposed binarized graph neural network, namely BGN, is orders of magnitude more efficient in terms of both time and space while matching the state-of-the-art performance.

Secondly, we first design a graph of graphs neural network for entity interaction prediction, and then extend the model to support the graph classification task with more expressive representations. Entity interaction prediction is essential in many important applications such as chemistry, biology, material science, and medical science. The problem becomes quite challenging when each entity is

represented by a complex structure, namely structured entity, because two types of graphs are involved: local graphs for structured entities and a global graph to capture the interactions between structured entities. We observe that existing works on structured entity interaction prediction cannot properly exploit the unique graph of graphs structure. In this thesis, we propose a Graph of Graphs Neural Network, namely GoGNN, which extracts the features in both structured entity graphs and the entity interaction graph in a hierarchical way. We also propose the dual-attention mechanism that enables the model to preserve the neighbor importance in both levels of graphs. Based on GoGNN, we further propose a Powerful Graph Of graphs neural Network, namely PGON, which has 3-Weisfeiler-Lehman expressive power and captures the attributes and structural information from both structured entity graphs and entity interaction graph hierarchically. Extensive experiments are conducted on real-world datasets, which show the superior performance of GoGNN and PGON compared to other state-of-the-art methods on both graph classification and graph interaction prediction tasks.

Thirdly, we design a reinforcement learning based query vertex ordering model for subgraph matching. Subgraph matching is a fundamental problem in various fields that use graph structured data. Subgraph matching algorithms enumerate all isomorphic embeddings of a query graph q in a data graph G. We apply the Reinforcement Learning (RL) and Graph Neural Networks (GNNs) techniques to generate the high-quality matching order for subgraph matching algorithms. Instead of using the fixed heuristics to generate the matching order, our model could capture and make full use of the graph information, and thus determine the query vertex order with the adaptive learning-based rule that could significantly reduce the number of redundant enumerations. With the help of the reinforcement learning framework, our model is able to consider the long-term

benefits rather than only consider the local information at current step.

PUBLICATIONS

- **Hanchen Wang**, Defu Lian, Ying Zhang, Lu Qin, Xiangjian He, Yiguang Lin, and Xuemin Lin."Binarized graph neural network." *World Wide Web Journal (2021).* (**Chapter 3**)
- **Hanchen Wang**, Defu Lian, Ying Zhang, Lu Qin, Xuemin Lin. "GoGNN: Graph of Graphs Neural Network for Predicting Structured Entity Interactions." *In the Proceedings of 29th International Joint Conference on Artificial Intelligence. (IJCAI 2020)* (**Chapter 4**)
- **Hanchen Wang**, Defu Lian, Wanqi Liu, Dong Wen, Chen Chen, Xiaoyang Wang. "Powerful Graph of Graphs Neural Network for Graph Classification" *World Wide Web Journal (2021).* (**Chapter 4**)
- **Hanchen Wang**, Ying Zhang, Lu Qin, Wei Wang, Wenjie Zhang. "Reinforcement Learning Based Query Vertex Ordering Model for Backtracking Based Subgraph Matching" *In Submission* (**Chapter 5**)

Contents

List of Figures