

Graph Convolutional Neural Networks with Negative Sampling

by **Wei Duan**

Thesis submitted in fulfilment of the requirements for
the degree of

Research Master of Science in Computing Science

under the supervision of Junyu Xuan, Jie Lu

University of Technology Sydney
Faculty of Engineering and Information Technology

January 2022

Certificate of Original Authorship

I, Wei Duan declare that this thesis, is submitted in fulfilment of the requirements for the award of master, in the faculty of engineering and information 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.

Production Note:

Signed: Signature removed prior to publication.

Date: 16 / 12 / 2021

Abstract

Convolutional neural networks (CNNs) can learn potential features from large amounts of Euclidean data, such as text, images, which produce a satisfactory performance on pattern recognition and data mining. Besides Euclidean data, graphs, as non-Euclidean data, are powerful structures for modeling molecules, social networks, citation networks, traffic networks, etc. However, to lend learning power from the Euclidean space to a graph is not trivial. Learning with graphs requires an effective representation of their data structure. Graph Convolutional Neural Networks (GCNs) have been generally accepted to be an effective tool for node representations learning. An interesting way to understand GCNs is to think of them as a message-passing mechanism where each node updates its representation by accepting information from its neighbors (also known as positive samples). However, beyond these neighboring nodes, graphs have a large, dark, all-but-forgotten world in which we find the non-neighboring nodes (negative samples).

In this thesis, we consider that this great dark world holds a substantial amount of information that might be useful for representation learning. Most specifically, it can provide negative information about the node representations. The key is how to select the appropriate negative samples for each node and incorporate the negative information contained in these samples into the representation update. We first propose a generalized method for fusing negative samples to graph convolutional neural networks. We next illustrate that the process of selecting the appropriate negative samples is not trivial. We proposed a determinantal point process algorithm-based method for efficiently obtaining samples. Finally, we use diverse negative samples to boost GCNs which jointly consider the positive and negative information when passing messages. The findings of this study not only have the development of the theory about negative samples in GCNs but also provide new ideas to alleviate the over-smoothing problem.

Acknowledgements

I would firstly like to express my earnest thanks to my principal supervisor Dr. Junyu Xuan, and my co-supervisor Prof. Jie Lu. Their comprehensive guidance has covered all aspects of my master study, including research topic selection, research methodology experiments, academic writing skills and thesis writing, even sentence structure and formulas. Their step-by-step approach taught me how to do scientific research. Their academic rigor and respectful personalities have benefited my master's research and will be a treasure in my life. Without their excellent guidance and constant encouragement, this research would not have been completed a semester early. Once again, I would like to thank my two supervisors for guiding me against numerous difficulties during the special time of the epidemic.

I am grateful to all members of the Decision Systems and e-Service Intelligent (DeSI) Lab in the Australian Artificial Intelligence Institute (AII) for their careful participation in my presentation and valuable comments on my research. I am especially grateful to Zihe Liu for staying in full communication with me, helping and supporting me with my coursework, papers, and thesis while we were stuck in China together by the epidemic. I would also like to thank Jemima for helping me to proofread my publications.

Last but not least, I would also like to thank my parents for their generous support of my study abroad, both financially and spiritually. I would also like to thank my friends for their support.

Contents

Declaration of Authorship	3
Abstract	5
Acknowledgements	7
List of Figures	11
List of Tables	13
1 Introduction	1
1.1 Background and motivation	1
1.2 Research questions	3
1.3 Research contributions	4
1.4 Research significance	4
1.5 Thesis structure	5
1.6 Publications	5
2 Literature review	7
2.1 Brief history of graph neural networks	7
2.2 Analytic tasks of graph neural networks	9
2.3 Categories of graph neural networks	9
2.3.1 Recurrent graph neural networks (RecGNNs)	9
2.3.2 Graph convolutional neural networks (GCNs)	10
2.3.2.1 Spectral-based graph neural networks	10
2.3.2.2 Spatial-based graph neural networks	15
2.3.3 Graph autoencoders (GAEs)	17
2.3.4 Spatial-temporal graph neural networks (STGNNs)	17
2.4 Main application areas of graph neural networks	17
2.4.1 Recommender systems	18
2.4.2 Computer vision	18
2.4.3 Natural language processing	19
2.4.4 Physics	20

2.4.5	Chemistry and biology	20
2.4.6	Others	21
2.5	Negative sampling in graph neural networks	21
2.6	Summary	22
3	Graph convolutional neural networks with Monte-Carlo negative sampling	25
3.1	Introduction to negative sampling in Graph convolutional neural networks .	25
3.2	Graph convolutional neural networks with Monte-Carlo negative sampling .	26
3.2.1	Negative sampling	27
3.2.2	Graph convolution with negative sampling	28
3.3	Experiments	30
3.3.1	Datasets	30
3.3.2	Experimental setup	31
3.3.3	Baselines	31
3.3.4	Experimental results	31
3.4	Summary	34
4	Graph convolutional neural networks with DPP negative sampling	37
4.1	Introduction to determinant point process (DPP)	37
4.2	Graph convolutional neural networks with DPP negative sampling	39
4.2.1	DPP-based negative sampling	40
4.2.2	GCN boosted by diverse negative samples	43
4.3	Experiments and results analysis	45
4.3.1	Baselines	45
4.3.2	Setup	46
4.3.3	Metrics	46
4.3.4	Results analysis	46
4.4	Summary	48
5	Conclusion and Further Study	51
5.1	Conclusions	51
5.2	Further study	52
	Appendices	53
	Bibliography	53

List of Figures

1.1	Illustration of the motivation of this work, including the dark world (gray shadow), different semantic clusters (green, yellow, purple, blue nodes), positive samples (nodes 2, 3, 4), and selected diverse negative samples (nodes 6, 11, and 18) of a given node (node 1).	2
3.1	Mechanism of the negative sampling graph convolution. The central node is $v = 5$ and $f(\cdot)$ is graph convolution layer [1]. Node 4, 7 are directly linked with node 5 by real positive edges, thus positive sampling convolution is performed by $x_{pos} = f(4, 7, 5)$. Node 3, 8 are negative sampled using MCNS methods [2], which are based on Markov Chain Monte-Carlo methods and DFS, message passing to central node $v = 5$ along virtual imaginary edges, then negative sampling convolution is performed by $x_{neg} = f(3, 8)$. Given a certain negative rate β , we get negative sampling graph convolution result of this layer, i.e. $x' = x_{pos} - \beta x_{neg}$	26
3.2	Performance of the three models on the training set in the Cora dataset. We choose the NegGCNs model with the highest accuracy for comparison, when $\beta = 1.25$	32
3.3	Performance of the three models on the training set in the Citeseer dataset. We choose the NegGCNs model with the highest accuracy for comparison, when $\beta = 1.25$	33
3.4	Performance of the three models on the training set in the Pubmed dataset. We choose the NegGCNs model with the highest accuracy for comparison, when $\beta = 1.00$	33
3.5	Performance of a different number of negative sampling nodes on Cora training set using NegGCNs with $\beta = 1.25$. We tested sampling 10%, 30%, 50%, 70%, and 90% of the total negative sampling points to calculate x_{neg}	34
4.1	The concept of DPP-based negative sampling. The target node is Node 1. Nodes 2, 3 and 4 are positive samples. Nodes 5-18 are the dark world of Node 1. The 4-length DFS path of Node 1 is $\{3, 5, 11, 13\}$, where $\{5, 11, 13\}$ are the central nodes on the path in the dark world. With their first-order neighbouring nodes, they form the candidate set of DPPs, i.e. $\{5, 6, 7, 11, 12, 13, 14, 18\}$. The selected negative samples from this set are 6, 11, and 18, which can be seen as virtual negative links to Node 1.	42

- 4.2 The Accuracy and MAD of five models on three datasets with a varying number of layers from 2 to 6. The x-axis denotes the layer number, and the mean and standard deviation of 10 runs are given for each model with each layer number. 47
- 4.3 The Accuracy and MAD of D2GCN with 5 layers on Citeseer datasets in the varying length of DFS and scale of negative rate. The mean and standard deviation of 10 runs are given for each setting. 47

List of Tables

3.1	Statistics of the dataset	30
3.2	Experimental results	32

