Minh-Le Nguyen
School of Information Science,
Japan Advance Institute of Science and Technology.
1-1 Asahidai, Nomi-city, Ishikawa, 923-1292, JAPAN.
☎ *+81 (0761) 51 1221*
FAX *+81 (0761) 51 1149*
✉ *nguyenml@jaist.ac.jp*

Prof. Kenji Doya and Prof. DeLiang Wang
Co-Editors-in-Chief
Journal of Neural Networks

February 15, 2018

Dear Prof. Kenji Doya and Prof. DeLiang Wang:

I am writing to submit our manuscript entitled: DGCNN: A Convolutional Neural Network over Large-scale Labeled Graphs, which is an improved and extended version of the paper: Convolutional Neural Networks over Control Flow Graphs for Software Defect Prediction [1], presented at *International Conference on Tools with Artificial Intelligence* (ICTAI) 2017, for the consideration of publication in Neural Networks. We confirm that this manuscript has not been published elsewhere and is not under consideration by another journal.

Labeled graph representations are available in many real applications. The traditional learning techniques for labeled graphs are based on frequent subgraph mining and graph-based kernels. However, these approaches suffer from large-scale graphs due to computational complexity. Recently, although many graph-based neural networks have been developed, these models are impractical to adapt to dynamic graphs. In this research, we proposed a deep neural network that can handle large-scale and dynamic structures of graphs.

Experimental results on two tasks of software defect prediction and malware analysis show three important points: (1) the network achieves high performance on classifying labeled graphs, (2) the network can handle large-scale graphs with hundred thousand nodes, and does not need any padding or alignment for dynamic structures, (3) Our design can easily adapt to other tasks using labeled graph representations. We believe our findings are likely to be of great interests to deep learning, software engineering and data mining scientists who read your journal.

Comparing to the original paper, this manuscript makes eight new and significant improvements as follows.

- We clearly analyze the impossibility to adapt graph kernels as well as other graph-based networks to labeled graphs of large-scale.
- We clearly describe the proposed neural network which is not sufficiently mentioned in the original paper.
- We formulate a new approach that applies deep learning on control flow graphs for program analysis.
- We also validate the ability to process huge graphs of the proposed network on a malware analysis task.
- We collect and preprocess a dataset for the malware analysis task.
- We apply a tool namely BE-PUM to generate control flow graphs for executable files.
- We deeply observe and analyze the performance of the approaches according to various criteria such as the performance measures, and the convergence of the learning process.
- Our implementation is released to motivate further research.

All authors have approved the manuscript and agree with its submission to the journal of Neural Networks.

Thank you very much for receiving our manuscript and considering it for review. We appreciate your time and look forward to your response.

Sincerely,

Minh-Le Nguyen

[1] All the necessary documents can be accessed at https://github.com/nguyenlab/DGCNN