Generation of graph embedding vectors based on graph isomorphism problem

Yosuke Higuchi and Yoshimitsu Kuroki
National Institute of Technology, Kurume College, Fukuoka, Japan
Email: {a3922yh, kuroki}@kurume.kosen-ac.jp

***Abstract*—Graph theory is the main theory for handling data with complex relationships: social networks, protein structures, and web page links. In graph theory, an adjacency matrix represents a finite graph structure as a square matrix. Additionally, a graph embedding vector is a low-dimensional vector that extracts the features of a graph. However, adjacency matrices have a drawback of being sensitive to graph vertex ordering changes. In this study, we propose a neural network model using graph isomorphism problem to generate new graph embedding vectors. Experimental results showed that the embedding vectors are robust to vertex reordering.**

I. Introduction

Adjacency matrix is one of the methods in graph theory to represent a graph as a matrix, and has been used for spectral analysis of graphs. However, the drawback that vertex reordering changes the adjacency matrix can be problematic in specific situations where the same graph is observed multiple times. In this study, we propose a neural network model using graph isomorphism to solve this problem.

II. Proposed Method

Fig. 1 shows the architecture of the proposed model. The model takes two adjacency matrix inputs and outputs isomorphisms with a probability of 0 to 1. In front of the output layer, there are group of the layers in which two embedding vectors $z_1, z_2 \in \mathbb{R}^{256}$ are generated. The layers generating those vectors seem to be two independent models, but they share the same training parameters. The vector generation layers consist of Batch Normalization [1], feedforward layers, transposition layers, and skip connections.

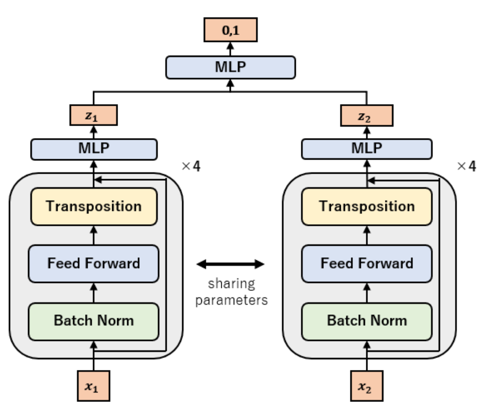

Figure 1. The proposed architecture

III. Experimental Results

The model shown in Fig. 1 was trained using 200,000 isomorphic and non-isomorphic dataset, which are a combination of the adjacency matrices of several random graphs with 10 to 50 vertices and the adjacency matrices of those graphs with vertex reordering. We found that the cosine similarity of the embeddings of the test dataset was independent of vertices ordering. The experiments compared classification accuracies of ENZYMES enzyme dataset from TUDataset [2] by two different MLP models: MLP1 uses the adjacency matrix as input, while MLP2 uses the generated embeddings of the dataset as input. The experimental results shown in Fig. 2 demonstrate that the classification accuracies of the two MLPs are comparable.

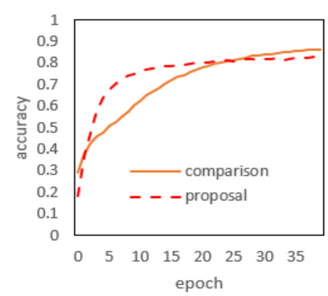

Figure 2. The classification accuracy

IV. Conclusion

By calculating graph isomorphism problem in the neural network model, we were able to generate graph embedding vectors that are robust to graph vertex reordering and preserve graph structure. Experimental results indicate the availability of the generated graph embedding vectors for graph classification. Future work includes improving the graph isomorphism model.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Numbers JP20K11878 and JP23K11159.

References

[1] L. Sergey and S. Christian. “Batch normalization: Accelerating deep network training by reducing internal covariate shift”, ICML, 2015.
[2] C. Morris, N.M. Kriege, F. Bause, K. Kersting, P. Mutzel, and M. Neumann. “Tudataset: A collection of benchmark datasets for learning with graphs” ICML, 2020.