Using Graph Embedding with Machine Learning on Simulation Models

Rajpal Singh, Master of Science in Computer Science
University of Dublin, Trinity College, 2022

Supervisor: Goetz Botterweck

Graphs are the general language for describing and analyzing entities with relations and interactions. Graphs provide the optimal structure for the representation and manipulation of non-euclidean data. In this research, simulation models are structured in graphs and are used for cost prediction through machine learning models. In the comparison of simulation, graph neural networks with machine learning help to work seamlessly and takes much fewer resources and time for structuring and training the models.

This dissertation aims to apply tools for converting JSON data to graphs. With the help of graph embedding techniques and methods, graphs are converted into vectors which machine learning models can train the data with the fan cost value provided as the target value for prediction.

We carried out an experiment with the dataset that consist of 10000 graphs, of which 1200 were taken with the target value (fan cost) associated with each graph for prediction. NetworkX, a python framework, was used to structure graphs consisting of rooms as nodes and links between them as edges. Random Walks methods such as DeepWalk and Node2Vec were used to transform graphs focusing on node-level structuring. Moreover, Weisfeiler-Lehman methods such as Graph2Vec and GL2Vec, along with Laplacian matrix with eigenvalues methods such as NetLSD, were used as whole graph level structuring for graphs into vectors. Vectors were then trained with machine learning regression models like Linear Regression, Lasso/Ridge Regression, AdaBoost, Random Forest and Multi-Layer Perceptron (MLP). MLP with K Fold cross-validation has proven to be the best fitting model with an accuracy score of 21%, an error rate of 0.7 among whole-graph level methods and an accuracy score of 16% error rate of 0.14 for node-level methods.