AUTOMATING GRAPH REPRESENTATION LEARNING (AUTOGRL)

Sanjula Madurapperuma

A dissertation submitted in partial fulfilment of the requirement for

Bachelor of Engineering (Honours) degree in Software Engineering

Department of Computing

Informatics Institute of Technology, Sri Lanka
in Collaboration with
University of Westminster, UK

2021

Abstract

Graph-structured data is omnipresent through a staggering amount of industries, with usage ranging from telecommunication networks to 3D-vision and quantum chemistry. However, in order to take advantage of all that data and gain insights in order to solve real-world problems, Graph Representation Learning (GRL) has to be brought into play. This area has been surging in activity, most of which are of tremendous value in solving pressing problems. Even though the new innovations that have come about due to this research activity allow for several important downstream tasks to be performed, a significant amount of computational and expert resources are still required to conduct GRL. Not everyone may have access to such computational resources, be it financially or physically, nor have the level of expertise required to achieve the best possible performance. With the opportunity for non-technical users or domain experts to conduct GRL without the need for extensive programming knowledge, there will be a significant increase in the pace at which real-world problems are solved. This dissertation is about building an automated Graph Representation Learning system called AutoGRL. It aims to abstract away the complex nature of GRL and utilizes an intelligent way to identify even edge-case scenarios in graph data and make decisions regarding the feature extraction, algorithms, hyperparameters, the training and evaluation process. At the end of the training, the user is presented with a summary of the results including the best performing model and the decisions made by the system. AutoGRL consists of a novel design and architecture, including the extensive and intelligent decision-making operations, pipelines and input graph data standardization. Compared to existing similar systems, it performs equally or better regardless of the size of the graph dataset. For Link Prediction downstream task, it can be observed that the performance improves the larger and more complex the data gets. It is the first of its kind to offer support for node classification and link prediction downstream tasks along with end-to-end automation of the GRL process.

Keywords: AutoGRL, Graph Representation Learning, node classification, link prediction, Python