VE444 Project Proposal

Group 18

October 2020

Ge Jintian	517021911142
Ye Roushuang	516370910241
Xia Binyu	517370910012

1 Problem Description

In real-world networks, nodes often have attributes. For example, in a network of recommender system, a node may represents a person, and its attributes may be the click record of that person. From the attributes we can estimate useful information about the node and it's position in the whole graph. Attributes of a node and those of its local neighbourhood may contain information useful in downstream tasks. Therefore, it is very important to find embedding of nodes to lower dimensions so that "similar" nodes in the graph have embeddings that are close together. This introduces the purpose of our project: embedding similar nodes together.

In this project, we plan to use data collected by a Python package called 'Karate Club'. Their dataset come from facebook, twitch, github, wikipedia and some other online websites. They build a Class so that users can use these dataset easily.

2 Introduction to Multi-scale Attributed Embedding Method

In this project, we mainly focus on the study and application of Multi-scale Attributed Embedding (MUSAE) method, which supplement local network structure with attribute information. Compared with other node embedding methods, MUSAE method jointly learns distinct representations of nodes and features to have a better prediction on node attributes and links as well as a transfer learning between networks.

The Attributed Embedding (AE) algorithms are defined based on the structure and attributes of local neighborhoods, aiming at to learn similar embedding for nodes in neighborhoods of similar attributes and similar embedding for attributes in similar neighborhoods of nodes.

Since the AE algorithm pools features across neighborhoods of different proximity, multi-scale attributed embedding method is adapted based on AE to having a higher performance in a number of downstream settings without increasing the overall complexity or number of free parameters.

Detailed algorithms of attributed embedding will be presented in our project.

3 Evaluation and Task Distribution

We plan to evaluate our work by comparing the performance of our work to the original algorithms [1]. We will use three methods provided in the paper: node classification, regression and link prediction.

In all these methods, we will take the output of the embedding algorithms as new input, and then take them to predict the classification, regression or link. (The prediction algorithm will be a very basic one.) The test score will be compared to the paper's method as well as others' method. The improvement of the score will measure our success.

The final delivery will contain our code (embedding algorithms as well as verification), data set, parameter setting, output file and a report explaining our improvement and innovation point.

Ge Jintian and Ye Roushuang will be in charge of establishing and improving the embedding algorithms. Xia Binyu will focus on building the test and verification algorithms for the above algorithms.

References

[1] Benedek Rozemberczki, Carl Allen, and Rik Sarkar. Multi-scale attributed node embedding. *arXiv preprint arXiv:1909.13021*, 2019.