Machine Learning on Graphs - Project Proposal

Mohammad Bahrami - Mostafa Mojtahedi

January 5, 2022

---

1 Sports Match Prediction Using GNNs

1.1 Problem definition in real life

Prediction of outcome of sports matches has always been a difficult task for both humans and machines. Some say that the maximum accuracy of predicting a football match for example, is 62%. Although this is not mathematically confirmed, it seems that the sports have their fans because their outcome is usually unpredictable.

1.2 How we are going to solve it

Using a heterogenous Spatio-Temporal graph of the teams and the players, along with a multi layer perceptron, in an end-to-end training method, we are going to train a model that will use the knowledge in the graph to obtain player and team embeddings. After that, using the MLP we will predict the outcome of a match as a 3-class classification problem.

2 Data

The data we use needs to have matches results in a weekly basis, and for each match, we need team names and team lineups of that match. We have found some data that has the capacity of being useful to us. The data contains matches of multiple leagues, containing a total of about 20000 matches.

3 Implementation

This project is a part of a much bigger project that I have been doing under supervision of Dr. Abdolreza Mirzaei for several months now. The project started with another type of graph and has evolved over time to many different type of models with supervision over match result, goal difference prediction and goal regression prediction. with different types of training methods, from real life progressive to random link prediction.

This work is the result of months of research in a team containing 3 bachelor students and Dr. Mirzaei himself. I will be presenting the PART THAT I HAVE DONE MYSELF. I have discussed this with Dr. Mirzaei in advance and he has accepted this in the terms that I note that this is a work of a team and not me alone, which I have.

The Model will be implemented using PyTorch and Torch Geometric Library for the neural network part. the preparation of data, cleaning, data partitioning and converting the data into the heterogenous graph will be implemented completely from scratch.

4 A second Idea?

I have also became interested in drug resistance prediction using graphs and have came across an interesting paper on it. But i have not found any public data or code base for it yet. in the case of the first project being denied, I may be able to implement the method of that paper (Or a part of it at least).