

# GNN Research Summary

Chuyi Jiang      cj2792@columbia.edu

May 16, 2024

## 1 Introduction

This is a summary of the GNN research I did in this semester. The goal is to investigate the SHAP (Shapley Additive explanations) value, which is a measure used in explaining the output of machine learning models, and the modify the model through the generated values. The experiment model is EAST-Net[1].

The EAST-Net is a spatio-temporal graph neural network. The network design specifically for spatio-temporal data. It's used for tasks where both spatial and temporal dependencies are important, such as traffic prediction, weather forecasting, and urban sensing.

Totally, the work we finished is to train the network again and generate the SHAP values. Then we modify the dataset, to observe the influence on SHAP values, which is described in Section2. The uncompleted part is to modify the original model by the SHAP values, which would also be described in Section3.

## 2 Experiment

Based on the EAST-net, we tried to trained the model with two dataset, traffic demand and supply under winter storm Jonas in Washington DC and travel purpose (totally 10 categories) under COVID-19 pandemic in US. I implemented the code on Colab. Because I need to save the checking model and modify the dataset, I manually change the address of the files in the original codes.

Besides the original dataset, I manually modified it. The aim is the reduce the numbers of holiday during the time period. In each dataset, there is a category called meta\_onehot. Each row shows the time information, including hours, days, months and whether holiday. The holiday is the last column in this dataset. Due to a hour-based dataset, the holiday information is same for each day. For example, the Jonas-DC is collected with 1 hour interval. Each 24 rows in the meta\_onehot show the information of the same day and the holiday information. If a certain day is holiday(weekends), it will show '1', otherwise, it will be '0'. I randomly reduce the days of the holiday by 50% and train the two dataset again. All the result is stored in the SHAP value txt file in github.

As a result, the number of holidays could directly influences the SHAP value for the x sequence dramatically. With less holidays in the dataset, the SHAP value for x sequence will be decrease. That could explain the importance of holidays.

### **3 Further Work**

For the next step, after getting the SHAP values, we could put in the models. Personally, I supported that the SHAP values could be transformed to a ratio compared with different features. Using that ratio, we should add them to the model architecture and influence the generation of the prediction by models. The work should focus on modify part of the algorithm in the architecture.

### **References**

- [1] Z. Wang, R. Jiang, H. Xue, F. D. Salim, X. Song, and R. Shibasaki, “Event-aware multi-modal mobility nowcasting,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 4, 2022, pp. 4228–4236.