INFO442-Exploraty Data Analysis-group2

Harry Zhao, Yi Pan, Yantian Ding

Problem Description

In this project, we will use uber movement speed data from New York City to try to build a prediction model for the average speed and congestion of road segments using a graph neural network approach.

File Included

- DSCI442_EDA_graph.ipynb: The notebook we used for the EDA of Graph data
- DSCI442_EDA_TS.ipynb: The notebook we used for the EDA of TimeSeries data
- README.md: The readme file describing our work
- DSCI442_EDA_centrality.ipynb: The notebook we used to calculate the centrality
- new_centrality.csv: A csv table strores the centrality data
- speed_data: The speed record data we use
- vis_OSM.png: The graph used for graph EDA

What we did

Time Series Data Analysis

- Time series decomposition. Decompose the time series into a trend cycle component, a seasonal component and a residual component (containing any other elements of the time series) and analyze each component
    - Hourly Average Speed Time Series Data decomposition.

- Daily Average Speed Time Series Data decomposition.
  - Daily Average Car Number (Volumn) Time Series Data decomposition.

- Stationarity test. The stationarity of the time-series data is tested.
- Traffic flow versus speed. The correlation between traffic flow and speed is predicted by observing the change of the two.
- Autocorrelation test. The autocorrelation was tested using ACF plots.
- Stochasticity test. The randomness of the data was tested using lag plot.

Road Network Graph Data Analysis

- Visualization. Visualize waypoint data on the actual map for analysis.
  - Using JOSM Map Editor to visualization.
- Centrality calculations. Discover the correlation by calculating centrality and examining the relationship between centrality and velocity.
  - Degree Centrality
  - Betweenness Centrality
  - Closeness Centrally
  - Eigenvector Centrality
- Classification. Records are classified into four categories for speed data values.

EDA Findings:

- Time Series Data
  - Traffic volume has a negative correlation with vehicle speed
  - Time-series data is seasonal
  - Time-series data is stable
  - Time-series data are lagged and autocorrelated

    - ARMA(3,0) model can be considered
- Graph Data
    - Degree Centrality has negative relationship with speed
    - Betweenness Centrality has positive relationship with speed
    - Closeness Centrality has negative relationship with speed
    - Eigenvector Centrality has little relationship with speed
    - GNN+LSTM related model can be considered
- Model Selection
    - We will consisder the GNN+LSTM related model

Bibliography

1. NIST. (n.d.). *Handbook of Statistical Methods*. NIST/SEMATECH e-Handbook of Statistical Methods. Retrieved February 12, 2022, from https://www.itl.nist.gov/div898/handbook/index.htm
2. Li, J. (n.d.). *timeseriesanalysis101*. 时间序列分析101： 序言 - timeseriesanalysis101. Retrieved February 12, 2022, from https://skywateryang.gitbook.io/timeseriesanalysis101/
3. Disney, A. (2022, January 21). *Social network analysis: Understanding centrality measures*. Cambridge Intelligence. Retrieved February 16, 2022, from https://cambridge-intelligence.com/keylines-faqs-social-network-analysis/

Reference

[1] Abaya, Ernesto B., et al. Instantaneous Fuel Consumption Models of Light Duty Vehicles and a Case Study on the Fuel Consumption at Different Traffic Conditions in Metro Manila Using Shepard’s Interpolation Method. No. 2018-01-0075. SAE Technical Paper, 2018.

[2] Tomko M, Winter S, Claramunt C. Experiential Hieranrchies of Streets [J]. Computers Environmnet and Urban Systems, 2008, 32(1): 41-52.

[3] Jayaweera, I. M. L. N., Perera, K. K. K. R., & Munasinghe, J. (2017). Centrality measures to identify traffic congestion on road networks: A case study of sri lanka. IOSR Journal of Mathematics (IOSRJM).

[4] Crucitti P, Latora V, Porta S. Centrality in networks of urban streets[J]. Chaos:An Interdisciplinary Journal of Nonlinear Science, 2006, 16(1): 015113 DOI:10.1063/1.2150162.

[5] Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Philip, S. Y. (2020). A comprehensive survey on graph neural networks. IEEE transactions on neural networks and learning systems, 32(1), 4-24.