电子信息与电气工程学院
School of Electronic, Information and Electrical Engineering

Traffic Prediction for Urban Planning

许方正

计算机科学与工程系

引言介绍

Traffic problems are becoming more and more common in urban areas, and there are many reasons for causing it, including population density, capacity of road networks, and urban planning.

Traffic prediction is useful for travel planning, transport distribution, urban planning, etc. Many other companies and researchers have done such things, like Google, Microsoft, Baidu, etc. But they focus more on the traffic prediction of the fixed location on existing map data based on moving average of historical data.

主要研究目标

Build a system to predict traffic conditions on any roads given a map formatted in OpenStreetMap data.

- ***Input***: road network map, points of interests, time, and the weather condition, etc.
- ***Output***: 4 classes at each spot: green (good), yellow (slow), red (congested) and deep red (extremely congested).

主要方法

Step 1 - Collect data

Road network data from OpenStreetMap, open map data source represented in graphs as nodes, ways and relations.

Points of interest data from Baidu Map. The coordinate system is different between OpenStreetMap (WGS-84) and Baidu Map (BD-09), so we used some approximation algorithms to transform the coordinate of Baidu Map's POI into WGS-84 within an accuracy of 0.5m.

Traffic data from Baidu Map. Traffic tile images are crawled every 30 minutes. Parsing real-time traffic situation at each node of OSM is done by intercepting the color of pixels labeling the traffic situation on two directions nearby.

Step 2 – Feature engineering

Local spatial features include road type, number of POIs of each type exist near the current node, as well as those to the left side, right side, forward and backward both up to three crossings, and the average distance between road crossings both forward and backward to the current node.

Global spatial features are overview of traffic flows in large urban perspective and reflection of commuters and driver's thoughts and behaviors. By clustering using Affinity Propagation on the most-visited POIs we find out those functional areas Three types of functional areas: residence, workplace, commercial. Then optimal routings were made between pairs of the areas. This feature reflects the popularity of the way.

Other implicit features like time of the day, day of the week, weather conditions and house rental price reflecting popularity density also plays a role.

Step 3 – Model training & Testing

We formatted categorical features to binary and scaled features to same range to normalize. As a 4-class classification problem with many features and massive data instances, we used L2-regularized L2-loss linear SVC (primal) as model. We also added weights to classes to mitigate data imbalance.

结论

Conclusion

We solved many challenges in data collection and feature engineering, and have developed a working demo capable of predicting traffic situation in user given area and input, and achieved an overall accuracy of 0.87 by feature engineering and using linear weighted SVC. It outperforms Baidu Map prediction accuracy with side-by-side comparison divided into hourly periods of several days.

Issues and future work

Data imbalance problem still affects the result, as the prediction rate of heavy traffic classes are relatively low. We are to try new techniques and models like ensemble methods and decision trees to tackle the problem, since it is hard for linear SVM to classify such extreme imbalances and may also consider from other perspectives of the problem like anomaly detection.

参考文献

Coast S. OpenStreetMap, www.openstreetmap.org

Çolak, S., Lima, A., & González, M. C. (2016). Understanding congested travel in urban areas. Nature Communications, 7, 10793.

Ren, Y., Ercsey-Ravasz, M., Wang, P., González, M. C., & Toroczkai, Z. (2014). Predicting commuter flows in spatial networks using a radiation model based on temporal ranges. Nature Communications, 5, 5347.