CS 5785 AML: Final Project - Project Proposal

Project Title: Spatio-temporal prediction of road conditions based on machine learning

Project Category: Application of machine learning to a practical problem or a dataset

Team Information

Scarlett Huang (sh2557)

Zihan Zhang (zz698)

Panda Xu (px48)

Project Proposal

With the advent of the mobile Internet era, all mobile device holders can become portrayers of traffic conditions and road capacity. If it is possible to have a more accurate prediction of traffic conditions based on real-time and historical traffic information, it will undoubtedly play a vital role in travel decision-making and alleviating urban congestion.

However, it is very difficult to predict future road conditions, which are affected by many factors such as time period, road capacity, the downstream topology of road networks, navigation traffic, and sudden road conditions.

The goal of our project is to accurately predict the traffic conditions (ie, unblocked, eased, and congested) in a certain period of time in the future based on real-time and historical road condition characteristics of road segments, basic road attributes, and road network topology relationship diagrams to help drivers to optimize route plans as well as city planning.

Most of the prior works use the probabilistic models to predict the traffic conditions, which are mainly suitable for traditional data with small sample size and simple data structure. However, the performances of traditional prediction methods for big data problems with large amounts of data and high complexity are severely restricted by data noise and influences of emergencies.

Our project plans to apply machine learning methods, to conduct feature screening and dimensionality reduction on-road sections to be predicted, use a variety of machine learning classification algorithms (such as SVM, Bayesian Network, Decision Tree) and neural network models to build classifiers separately, and apply ensemble learning methods to traverse different

model combinations in turn to find the best ensemble model to achieve accurate short-term prediction of the road network traffic status.

We are planning on finding a high-quality travel tracking dataset driver and passenger that is suitable to model real-time traffic conditions. We plan to classify the traffic state of the road based on the average speed of the road, road grade, and other information. Use discrete time and the time unit is 2 minutes. The road conditions are considered uniform within 2 minutes. According to the historical speed and status of the road conditions in the same period, predict the target road conditions. At the same time, we need to know the specific information of the road, including speed limit, length, number and width of lanes, etc.