Abstract

Team- DC20036

* Raw Data Observation
* Technologies utilized:
* Data Analysis

Exploratory analysis

We first used the graphs in EXCEL to get a rough idea of the dataset. We found there were fluctuations, and especially there were peaks in a couple of days. A few factors could affect this were weather and events that took place on campus. Thus, naturally the next step was to mine the related data of the factors.

* New variable

In order to figure out which factors may affect the traffic count, our team added some related variables: weekday/weekend, event, temperature(15 min interval), precipitation (15 min interval), wind speed (15 min interval)

* Data clean

Remove error records e.g., rows where each mode counts 0 at the same time.

* Predictive Model

Our team uses the M5P tree to build a traffic count predictive model. M5P combines a conventional decision tree with the possibility of linear regression at the nodes

* Expected outcome
* Data visualization of the temporal dynamic trend for each traffic mode on each location, observing the time trend of each traffic mode to find whether it exists fixed pattern, peak value at each location, comparing traffic trends of weekday and weekday
* Combine predictive model and visualization tools
* Limitations
* Overlapped data, the same car/bus/pedestrian/truck/bicycle may be counted multiple times at the same location by different sensors.

-----

Team - DC20036

Our team worked on the Campus Traffic Count Sensor Data provided by DOTS. We’ve finished our mining, analysis and modeling prediction.

Through the data preprocess, we found that campus traffic is highly correlated with multiple factors: schedules, weather and events. We tried multiple models to make predictions and finally found that one model took into account many characteristics of this problem. To get sustainable optimization, we will also create a tool combining visualization and data prediction, for the benefit of future use. Please see more details below.

* Technologies utilized
* Excel for initial data observation
* Weka for model evaluation
* JS/HTML for visualization and tool implementation.
* Expected outcome
  + Predictive Model:  
    With the input of factors, the model outputs the prediction of traffic count for the selected location. Compared with the actual count, we can conclude a reduction in carbon footprint on campus during that time.
  + Data visualization:  
    still working on it.
* Limitations
  + Overlapped data, the same car/bus/pedestrian/truck/bicycle may be counted multiple times at the same location by different sensors, which makes some values outliers and cause the predictive model to be less accurate;
  + Missing data: some sensors reported 0 counts in a certain period, which is obviously inaccurate;
  + Limited samples, only 2-week span, allowing less space for model optimization.