Utilizing Real-World Transportation Data for Accurate Traffic Prediction, ICDM 2012

Bei Pan, Ugur Demiryurek, Cyrus Shahabi.
Integrated Media System Center
University of Southern California
Los Angeles, United States

1. Summary
In this paper, the authors use the traffic data available from 8000 traffic loop detectors in Los Angeles, as well as two existing methods, to propose a new method that predicts the time and average speed to go from any place to any other in Los Angeles.

2. Positive/Strong Points

2.1 Good previous research and general description of the previously available methods ARIMA and HAM. The first one is useful for short-term predictions and the second one is better for long term predictions and rush hours.

2.2 By noticing the cons and pros of each method, they propose a new method, based on a decision tree that will choose either ARIMA or HAM depending on which one yields better accuracy.

2.3 Additionally, they give a lot of emphasis on special events that affect traffic, like accidents, sport events, road repairs, etc. Only ARIMA used to consider these events, but their method uses also online information to improve the accuracy of their predictions.

2.4 Their new method yielded far better results than any of the previous methods applied individually, and it is not much more costly since they are only adding a decision tree.

2.5 Their method can, under certain circumstances, improve the predictions by up to 67% and 78% in short-term and long-term predictions respectively. Similarly, by using online data, the prediction accuracy under special events can be increase by up to 91%.

2.6 It is a very promising method to reduce the time that people spend in traffic.

3. Negative/Weak Points

3.1 Not all cities have so many sensors (if any), so this method cannot be generally applied to any city.

3.2 For doing predictions without events, their training set was November and the testing set was the first week of December. However, HAM uses historical data and the traffic might be different from month to month, and from season to season.

3.3 Table I b) and c) has very little data from the events to be able to do inference. They do not specify how they treat the missing values.
3.4 For estimating what happens with the special events, they divide the day into time zones of 4 hours each. However, some rush hours are caught in-between two time zones, so the effect of special events during rush hours might not be captured correctly.

3.5 Predicts time and speed, but no alternate routes.

4. Research Questions and Points for Discussion

4.1 Could the method be generalized so that it can be applied to any city? E.g., they talked about the delays that could occur if an accident happened in CA-95. Would it be easy to generalize to any highway with so many lanes, at a specific time of the day and the week in any city? Or are there more parameters to be considered?

4.2 Can this information be used to obtain an optimal route from point A to point B, so that there is actually a saving in time and fuel?

4.3 Could there be a way to constantly incorporate the information of new special events, so that future predictions are more accurate?