Validating the use of probe data for highway traffic estimation in the Mobile Millennium highway model using Bluetooth measurements

1 Introduction

This document explains the process used for validating the effectiveness of GPS probe data for traffic velocity estimation in the Mobile Millennium highway model using Bluetooth measurements. The Mobile Millennium highway model is a flow based traffic estimation engine that uses traffic counts from loop detectors and instantaneous velocities from probe data to create a velocity map for a highway network. Traditionally count data from loop detectors has been used as the primary input to the system, but due to the high cost of maintaining these fixed sensors we wish to explore the effectiveness of using GPS probe measurements that can potentially be obtained at a cheaper cost.

GPS probe data measurements are obtained from third party vendors using a bidding process. Determining the validity of these measurements is a hard problem in general as there is no measurable ground truth value for instantaneous point velocities. Thus, the best we can do is to determine the validity of these measurements for the specific application of traffic velocity estimation. Again, since velocity ground truth measurements are not available, travel-times obtained via Bluetooth sensors are used as a proxy. The purpose of this document is to explain this validation process in terms of experimental setup and validation procedure, while also pointing out the limitations and errors involved.

2 Experiment dates and locations

The experiment dates are yet to be determined. There will be three experiments in total for three highway sections.

1. Highway I-880 between Winton Avenue and Stevenson Blvd.
2. Highway I-15 between TBD and TBD.
3. Highway I-15 between Los Angeles and Victorville.

These three segments provide areas with varying probe data and sensor penetration rates allowing us to validate the performance of the velocity estimation under these different conditions. More details about the experimental setup are provided in a separate document.

3 GPS probe data

Probe data is procured from third party vendors by CCIT using a bidding process. The data will be supplied by the vendors in a format specified by CCIT during the RFP process. Initially, the data will be obtained in batched for dates of the experiment and inserted to the Mobile Millennium database. Eventually, the data will be obtained via an input stream and updated in the database on line for real-time traffic estimation.

The accuracy and density of the probe data will determine the performance of the traffic estimation model. Therefore, we wish to use this study as a means to determining the adequate accuracy levels and penetration rates required for meaningful traffic estimation using probe data.

4 Bluetooth data

Vehicle travel times measured with Bluetooth readers will be used to evaluate the accuracy of the highway model traffic estimate. A series of ten Bluetooth readers will be placed on each of the three highway segments described above. The vehicle travel times are determined by capturing the times at which vehicles that contain a discoverable Bluetooth device pass by these readers. The travel time is computed by looking at the time taken for the same Bluetooth device to be detected at different readers. The goal is to accurately estimate the average travel times between consecutive Bluetooth reader locations at a given time interval using individual travel-times. In this experiment, Traffax BluFax Bluetooth readers are used to collect these travel times by measurements. These devices are being used due to the current availability at CCIT, the familiarity of operating them, existing data collection scripts and their general appropriateness for this experiment as elaborated below.

4.1 Reader requirements

The selection of Bluetooth readers is guided by several requirements. The units need to be weatherproof and robust to the external environment, since they will be deployed on California highways for a two week period without the ability to monitor them during this period. We need the flexibility to be able to mount the readers on poles, cabinets, or secure them on the roadside. We also need the ability to be able to run the detectors without access to external power for a period of at least 10-14 days. A live data connection from the sensor to receive data from the units during deployment is desired, but a nightly download of the data is acceptable. The readers should be able to synchronize their data with GPS time. The Bluetooth sensors need to be able to detect Bluetooth devices across all lanes of the highway in a given direction.

BluFax units are battery powered in weatherproof cases and come with brackets suitable for mounting on poles. The Bluetooth antenna is sufficient to cover all lanes (in fact, they have a range up to approximately 100 meters). They are able to send email bundles of the days' observations via cell modem once a day.

4.2 Deployment

The deployment of Bluetooth readers in governed by two requirements.

- Finding appropriate locations with good directional visibility of the highway traffic so that the antenna of the reader has the best chance of capturing as many vehicular Bluetooth devices as possible. From previous experience this includes:
  - Installing the reader roughly 15 feet above the road to give a direct line of sight through the windows of a car where the Bluetooth signal is least obstructed.
  - Keeping the reader away from other secondary roads to avoid spurious readings.
- Working with Caltrans to determine locations that are easily accessible and satisfy all the Caltrans safety requirements such as not distracting drivers.

4.3 Error profile

The BluFax readers, like all Class 1 Bluetooth devices (and therefore any other Bluetooth readers we might use), have a practical reception range of about 100 meters, with no way to detect where in the reception area a detected device is. Even in a highway environment, 100 meters can be a long distance that unavoidably covers cars queuing at intersections, vehicles on adjacent or crossing streets, pedestrians on overpasses or adjacent streets, etc. This can have several side-effects on the data that have to be managed in the analysis. As we plan to deploy 10 readers over a roughly 10 mile section of road, readers should try to be placed 1 mile apart, such that the 100 meter error range does not make up a significant portion of the distance between the two readers. These factors need to be managed carefully in the data analysis, and have implications for how best to deploy Bluetooth readers in this environment.

The Bluetooth readers scan for Bluetooth devices every 5 seconds within the detection range of 100 meters. This means that they detect passing vehicles for up to 200 meters, which leads to imprecision in the collected travel times. The Bluetooth readers check for nearby devices every 5 seconds, which means that they detect the same vehicle multiple times as it passes. Multiple readings at each Bluetooth reader make identifying the "true" travel time a non-trivial task. The post processing section describes the approaches taken to minimize this error and determine the best estimate for travel-time.

4.4 Post processing

A post processing algorithm is used to eliminate duplicate entries from the same vehicle (within a single trip) at a given reader. Multiple readings can occur due to a number of reasons such as high congestion, nearby ramps, overpasses, parallel roads or Bluetooth devices that are not in a vehicle. Listed below are three basic strategies we consider with their advantages and disadvantages.

- Average time: Picking the average time of the measurements should give an estimate of the time at which that vehicle was closest to the sensor if the multiple readings were simply due to congestion. This would be the strategy of choice in a case where there are no nearby ramps.
- Last time: Picking the last time of measurement is robust to multiple readings due to on-ramps. By picking the last time we eliminate the time spent on the ramp and get a more accurate estimate. This is especially important at metered lights.
- First time: Picking the first time of measurement is robust to multiple readings due to off-ramps. This is important because an off-ramp can slow down traffic in the rightmost lane of the highway and what we're really interested in is the travel-time for mainline flow along the highway.

5 Validation method

The validity of probe data collected from the vendors can not be measured directly, since there is no such thing as a ground truth measurement for point speeds. Thus, validation requires some model (from as simple as linear interpolation to the flow model which we use) to convert the point speed measurements to travel-times. We determine the validity of this data in relation to the specific application it is used for, the Mobile Millennium highway velocity model.

We assume that the Bluetooth data provides link travel-times for a sample of individual vehicles. The goal is to determine how accurately the highway model (with the purchased probe data) can estimate these travel times.

5.1 Computing the model error

For any given vehicle traveling from reader a to reader b the model error is estimated as follows.

- Let t_a, t_b be the times at which a given vehicle was sampled at readers a and b.
- Let the travel-time from a to b for the vehicle as given by the Bluetooth readings be $\beta_{ab} = t_b - t_a$.
- Estimate the model travel-time for the vehicle by integrating the model velocity along the vehicle trajectory starting at time t_a at location a. This is done as follows:

- The highway model provides a velocity $v_{(s,l)}$ for each time-space cell (s, l), where s is the space discretization index and l is the time discretization index.
- Let location a at time t_a be in cell $c_1 = (s_1, l_1)$. Thus, the vehicle starts in cell c_1.
- Simulate the vehicle trajectory by using the velocity $v_{(s_1,t_1)} = v_{c_1}$ while the vehicle is in cell c_1. Once the vehicle exits cell c_1 use the velocity for the next cell and continue until location b is reached[1].
- Call this travel time obtained using this trajectory similation τ_{ab}.

- The estimation error is given by $\epsilon = |\beta_{ab} - \tau_{ab}|$.

It should be noted that this estimate is more sensitive to model errors at the start of the trajectory as opposed to at the end of it. This is due to the fact that any initial errors can propagate by moving into incorrect time-space discretization cells. This needs to be explained more carefully by including a diagram.

It is important to note that the highway model provides an estimate for the velocity of an average vehicle. Assuming that the Bluetooth travel-times are uniform independent samples, a large number of samples should give us an average travel-time. Unfortunately, we will not have a large number of samples for any given time, so there will be some inherent estimation error. We will need to consider an appropriate averaging of the Bluetooth travel-times based on the number of samples we get to minimize this error.

6 Limitations

As with all experiments, there are a number of limitations that could effect the accuracy of the measurements and process described above. In this section we wish to identify these limitations so that the reader is aware of them and can take any necessary action to minimize their impact.

6.1 Sources of error

The probe data measurements can have multiple sources of error. Given below are some of the errors we expect to see.

- **GPS error:** Each probe device will have a GPS location measurement error. This error is a function of the GPS hardware used in the device, the weather conditions at the time of measurement, the number of visible satellites and the spatial distribution of these satellites. Some GPS devices will provide an estimate of the measurement error with the measurement. Typically this is in the form of a Dilution of Precision (DOP) value.

- **Map matching error:** Once a GPS coordinate for a probe is obtained it needs to be mapped to a location on a road. The highway model uses a representation of the road network as given by Navteq maps. This representation with is a collection of geo-lines is an approximation of the actual road topology. Furthermore, the maps are only updated every six months and the version we use could be many years old. Therefore, the GPS measurements (even if they had zero measurement error) might not map directly to a location on the road network. Map matching algorithms are used to make the best guess as to which location on the road network each GPS measurement corresponds to. These algorithms work best by considering multiple consecutive readings and assignment locations to a series of measurements.

[1]There are two general models in which to do this simulation. Namely, the frozen travel time model and the elastic travel time model. In the frozen travel time model, the link (spatial dimension of the cell) travel time is assumed to be fixed apon entering the link and will not change even if the time discretization changes before exiting the link. This is the most commonly used model in most travel time estimations due to its simplicity. In contrast, the elastic travel time model adjusts the velocity of the vehicle as it moves from one time discretization to another while within a link. This corresponds to changing the velocity when changing cells

- **Model error:** The highway traffic estimation algorithm uses a model based on hydrodynamic flow theory for velocity estimation. As with all models, there are a number of simplifying assumptions made in theory that generally do not hold in reality. While the performance of the model has been shown to be quite accurate, it is important to realize that the model estimate will by no means be perfect even with perfect measurements and map matching.

- **Bluetooth measurement error:** The highway model output is validated using Bluetooth readings as explained previously. As alluded to in that discussion, there are many factors that can add errors to the estimation of the travel times with Bluetooth readings.

- **Variability due to driver behavior:** The highway model estimates the velocity of an average driver while the Bluetooth measurements estimate travel times across the distribution of all drivers. This adds another source of error to the comparison. As mentioned above, to obtain a meaningful comparison we would need to average Bluetooth travel times over a reasonable number of vehicles.

- **Bias in Bluetooth equipped vehicles:** The driving behavior of the vehicles that are equipped with a Bluetooth device might not be representative of the entire population of vehicles on a road segment. We are assuming there is no correlation between driver behaviour and having a Bluetooth device in the car.