Developing a Benchmark for Using Trajectories of Moving Objects in Traffic Prediction and Management

Győző Gidófalvi, Ehsan Saqib

The Royal Institute of Technology (KTH), Geoinformatics, Drottning Kristinas väg 30, 100 44 Stockholm, Sweden
Email: gyozo.gidofalvi@abe.kth.se, esaqib@kth.se

1. Application Scenario

Several trends in urban mobility put the development of effective traffic prediction and management systems in high demand. Early systems have primarily used punctuated speed and flow measurements from fixed location sensors in conjunction with traffic models to tackle the prediction and management tasks. More recently, fuelled by the wide-spread adoption of GPS-based on-board navigation systems and location-aware mobile devices, to improve accuracy the use of moving object trajectories in such systems have been proposed. In such proposals it is assumed that vehicles periodically submit their location to a central server. In turn the server extracts mobility patterns from the submitted locations. The extracted mobility patterns, together with the current locations of the vehicles, are both used in short- and long term traffic prediction, management and planning tasks. Additionally, the current and near-future traffic conditions are sent in real-time to the likely-to-be-affected vehicles. To aid the development and assessment of such systems, a benchmark for traffic prediction and management is needed.

2. Data Description

One possible real world moving object trajectory data set (stream) for developing such a traffic prediction and management benchmark is provided by Trafik Stockholm and is available at the Transport and Logistic Division of the Department of Urban Planning and Environment, Royal Institute of Technology (KTH), Sweden (collaborating partner). The trajectory data set contains the GPS readings of 1500 taxis and 400 trucks travelling on the streets of Stockholm. Each taxi produces a reading once every 60 seconds approximately. This reading includes only taxi identification and location information. Taxis produce readings less frequently when they are not carrying any passengers. Trucks use more recent and more accurate GPS devices that produce readings once every 30 seconds and include identification, location, speed and heading information. The peak data rate for the whole city is over 1000 readings per minute, and there are approximately 170 million readings during the course of a year. Two notable characteristics of the described trajectory data set are that it represents a relatively large sample of the moving object population, but locations of samples are obtained relatively infrequently.

3. Benchmark Tests

Benchmark tests are to be developed with clearly defined criteria for evaluating and comparing the *performance* and *accuracy* of systems for the traffic prediction and management tasks outlined in Section 1. Benchmark test that evaluate the *scalability* of systems should also be designed. The development of benchmark tests, given the data and the application scenario, should carefully examine and address the following issues:

- **Trajectory sample bias**: Taxi trajectories have a different spatial and temporal distribution from that of private vehicle trajectories.
- **Absence of individual mobility patterns**: Taxi trajectories do not contain *periodic* patterns that pertain to the mobility of individual persons. Consequently, benchmark tests developed based on the taxi trajectories cannot be used to adequately assess the performance of systems that exploit such patterns.
- **Need for privacy**: Privacy is a major issue in the application setting, therefore benchmark tests should be able to assess the performance and accuracy of systems for different privacy levels, which are provided by some location anonymization framework for example by blurring/cloaking locations of objects.
- **Realistic scalability tests**: Replaying data at several times the actual data rate can only partially simulate higher input data rates. However, given this data scaling method, the accurate evaluation of scalability is questionable, because the number of patterns in the data does not increase, i.e., the processing time for pattern extraction and management remains constant or only increases linearly with the scaling factor.

4. Raw Trajectory Data and Derived Traffic Patterns

To gain a better understanding of the nature of the proposed trajectory data to be used in the development of the proposed benchmark, Figure 1 visualizes a *small subset* of the raw data and the extracted mobility patterns. In particular, the subset is spatially restricted to the inner city of Stockholm; the area of the full spatial extent of the complete trajectory data is approximately 25 times larger than the area of the extent that is visualized in Figure 1. Furthermore, the subset is also temporally restricted and only includes data for the course of a single day. Given this subset, the data visualized in Figure 1(left) is further restricted to only contain the raw GPS readings of 100 taxis. Subsequently, the data visualized in Figure 1(center) is even further restricted to only contain the raw 2D trajectories of 10 taxis. Figure 1(center) clearly illustrates one consequence of the low sampling rate and visualizes the *linear interpolation* of the trajectories. Finally, Figure 1(right) shows the extracted mobility patterns, i.e., frequent routes with speed profiles, which, given the previously described spatio-temporal extent, are extracted from the raw data of *all* 1500 taxis using *road network based interpolation* of the trajectories and frequent pattern mining techniques.

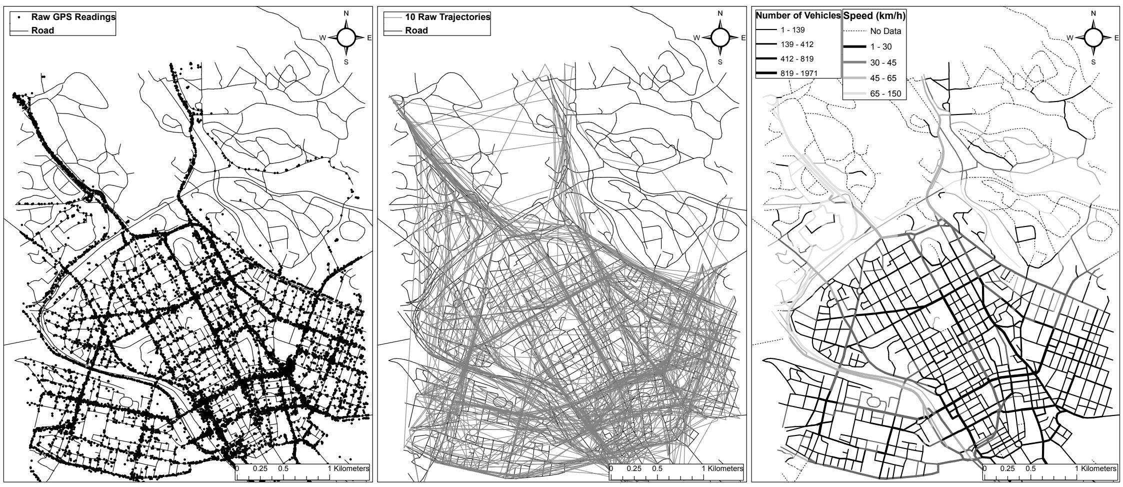

Figure 1: Raw Trajectory Data and Derived Traffic Patterns. Raw GPS readings of 100 taxis (left). Raw trajectories of 10 taxis (center). Frequent routes of all taxis (right).