Calibrating Large Scale Vehicle Trajectory Data

Siyuan Liu, Carnegie Mellon University
pkv@cs.cmu.edu

Ce Liu, University of Pittsburgh
pkv@cs.cmu.edu

Qiong Luo, Hong Kong University of Science & Technology
pkv@cs.cmu.edu

Lionel Ni, Hong Kong University of Science & Technology
pkv@cs.cmu.edu

Ramayya Krishnan, Carnegie Mellon University
@cs.cmu.edu

13th IEEE International Conference on Mobile Data Management (MDM'12), July 2012

LARC Accepted Conference Papers: http://smu.edu.sg/centres/larc/conference-papers/
ABSTRACT

An accurate and sufficient vehicle trajectory data set is the basis to many trajectory-based data mining tasks and applications. However, vehicle trajectories sampled by GPS devices are usually at a relatively low sampling rate and contain notable location errors. To address these two problems in GPS trajectory data, we propose WI-matching, the first vehicle trajectory calibration framework to take advantage of road networks topology and geometry information and trajectory historical information in large scale. WI-matching consists of a Weighting-based map matching algorithm and a trajectory Interpolation-based matching algorithm. In our WI-matching framework, we first integrate the vehicle GPS data with digital road networks data, to identify the roads where a vehicle traveled and the vehicle locations along the roads. Then our weighting-based map matching algorithm considers (1) the geometric and topological information of the road networks and (2) the spatiotemporal trajectory information to efficiently and effectively calibrate the GPS data points. Finally, our interpolation algorithm identifies paths between consecutive GPS points, and adds points with estimated vehicle status (location and time stamp) along the paths to construct sufficient vehicle trajectories. We have evaluated our algorithms on a large-scale real life data set in comparison with the state of the art. Our extensive and empirical results indicate that our WI-matching achieves a high accuracy as well as a high efficiency on real-world data which beats the state of the art.