VTrack: Accurate, Energy-aware Road Traffic Delay Estimation Using Mobile Phones

Arvind Thiagarajan, Lenin Ravindranath, Katrina LaCurts, Samuel Madden, Hari Balakrishnan, Sivan Toledo, Jakob Eriksson

Presented By: Rajarshi Bhowmik
January 26, 2017
Outline

• Brief Overview of Trajectory Data Mining
  • Trajectory Data Sources
  • Trajectory Data Preprocessing
  • VTrack
Brief Overview of Trajectory Data Mining

• What is a trajectory?
  – A trace generated by a moving object in geographical spaces
  – Usually represented as a series of chronologically ordered points
  – Each point consists of geo-spatial coordinate + time stamp

• Why trajectory data mining is important?
  - Advancement in location acquisition technologies generate huge amount of spatial trajectory data from various moving objects
  - Novel applications in location based social networks, intelligent transport system and urban computing
Outline

• Brief Overview of Trajectory Data Mining
• Trajectory Data Sources
• Trajectory Data Preprocessing
• VTrack
Trajectory Data Sources

- Human Mobility
  - Active recording
  - Passive recording

- Mobility of Transportation Vehicles
- Mobility of Animals
- Mobility of Natural Phenomena
Outline

• Brief Overview of Trajectory Data Mining
• Trajectory Data Sources
• Trajectory Data Preprocessing
• VTrack
Trajectory Data Preprocessing

• Noise Filtering
  – Mean (or Median) Filter
  – Kalman and Particle Filter
  – Heuristics based Outlier Detection

• Stay Point Detection

• Map Matching
Trajectory Data Preprocessing (continued)

- Trajectory Compression
  - Distance Metric
  - Offline Compression
  - Online Compression
Trajectory Data Preprocessing (continued)

• Trajectory Segmentation
  – Time Interval based Segmentation
  – Turning Point based Segmentation
  – Stay Point based Segmentation
  – Semantic Meaning based Segmentation
Outline

• Brief Overview of Trajectory Mining
• Trajectory Data Sources
• Trajectory Data Preprocessing
• VTrack
  ❖ Motivation
  ❖ Problem Specification
  ❖ Key Challenges
  ❖ Architecture
  ❖ Requirements
  ❖ Solution
  ❖ Evaluation
Motivation

• Traffic Congestion in Cities
  - Wastage of time and money
  - Impact on environment
  - Frustration for commuters

• Trends
  - 1.2 Billion cars on road (2014). ~2B cars by 2035.
  - 4.2B hours spent in traffic in 2007 (Source: US Bureau of Transportation Statistics)

• Use smartphone capabilities
  - Leverage GPS and WiFi to gather traffic data
Problem Specification

• Detecting and visualizing hotspots
  - Goal: low *miss rate* and *false positive rate*

• Real-time route planning
  - Goal: provide users routes with minimum travel time
Key Challenges

• Energy Consumption
  - GPS accurate but power hungry

• Outage
  - GPS is not available in urban canyons and tunnels

• Possible solution: use WiFi as a location sensor
  - Energy efficient but noisy
  - Availability may not be uniform

• Privacy and security concerns (not considered in this paper)
Architecture
Requirements

• Accuracy
  - Route planning error up to 10-15% is tolerable.

• Efficient enough to run in real-time
  - A*-style shortest path algorithms fail to achieve this

• Energy efficient
  - Should have an energy-accuracy trade-off
Solution : High Level Concepts

• Map Matching: matching a sequence of noisy and sparse position samples to a sequence of road segments

• Question: How to obtain position samples?
  - For GPS, just report the position directly
  - For WiFi, it can only report the observed access points (APs)
  - Convert the APs to position samples using a wardriving database

• Naïve method for map matching: Assign each position samples to nearest road segments

• VTrack uses HMM and Viterbi Decoding to perform map matching.
Solution: HMM and Viterbi Algorithm

- HMM consists of hidden and observed variables.
- Observed variables: Position samples
- Hidden variables: Road segments
- HMM is described by a combination of two probabilistic models: State Transition Model and Observation Model
- Viterbi Algorithm: Given a sequence of positions, determine the most likely sequence of segments
Map Matching

RAW TRACE

OUTLIER REMOVAL & INTERPOLATION

VITERBI MATCHING

BAD ZONE REMOVAL
Travel Time Estimation

• Travel time for any segment $S$ is given by:

$$T(S) = T_{left}(S) + T_{matched}(S) + T_{right}(S)$$

$T_{left}(S)$ – time between the entry point for $S$ and the first observed point matched to $S$

$T_{matched}(S)$ – time between first and last observed point matched to $S$

$T_{right}(S)$ – time between the last point matched to $S$ and the exit point from $S$
Evaluation

• Obtaining the ground truth
  - uses aggressive data cleaning
  - For each GPS point $g$, consider a set of segments $S_g$ within 15m radius
  - Search the space of segments to match sequence of $g$’s to continuous sequence of segments $X_g$
  - Project each $g$ to the closest point on $X_g$ to obtain ground truth point $g'$

• Sensor sampling strategies
  - Continuous WiFi
  - GPS every $k$ seconds
  - GPS every $k$ seconds + WiFi in between
  - Scaled speed limit
Route Planning

- CDF of Optimality Gap

Optimality Gap = \[ \frac{\text{Time}(P_{\text{noisy}}) - \text{Time}(P_{gt})}{\text{Time}(P_{gt})} \]
PER and SER

- **Point Error Rate** (PER) is the fraction of position samples assigned to the wrong segment.
- **Segment Error Rate** (SER) is edit distance between the map-matched output and the ground truth trajectory.
Hotspot Detection

- **Success Rate** – fraction of ground truth hotspots that VTrack detected.
Hotspot Detection

• **False Positive Rate** – fraction of hotspots detected that are not actually hotspots
HMM vs Nearest Segment Matching
Energy vs Accuracy

Power Budget $p/w$

$\frac{p}{w} = 1$

$\frac{p}{g} = \frac{6}{k}$

Critical

GPS $k$

WiFi + GPS $k'$

GPS Cost $g/w$

WiFi only
Conclusions and Follow-up Works

• Three key contributions
  1) Reduced energy consumption using WiFi
  2) Accurate travel time estimate from inaccurate positions
  3) Accurate detection of hotspots

• Follow-up Works
  1) T-Drive by Yuan and Zheng et al. 2011 (personalized driving direction that adapts to weather, traffic condition and persons own driving habit)
  2) Wang et al. 2014 (citywide real-time model)
Questions?
Thank You