Inferring Information from mobility traces

Daniel A. Desmond (Supervisor: Prof. Kenneth N. Brown)

Insight
Centre for Data Analytics

Thesis Statement

The integration of shortest path planning with mobility data mining to infer further information.

Previous work - Waypoint Inference

Where we have a single GPS trace from a vehicle, it is possible to infer waypoints when we know the driving style but having no other information about the individual except their driving style preference and using only a map of the area concerned[1]. We achieved a recall of up to 97% on estimating waypoints, with a precision of 93% and an accuracy of 95%.

Because drivers do not reliably choose shortest paths[2] this method was then expanded to infer waypoints where the drivers have different preferences for driving styles.[3] We achieved over 95% recall on estimating waypoints, with a precision of 93% and an accuracy of 94%.

Current work – Time, Distance and Route Inference

Question

Where we have historical trip data of New York taxi trips[4] consisting of the start and end points along with the distance travelled and time taken for the trip, we can use this information to create a more accurate map by inferring the delays due to junctions.

Our hypothesis is that in urban driving conditions a vehicle in isolation will travel at the posted speed limit. Due to the effects of other vehicles and junctions the vehicle will travel at less than the posted limit and the effects of the of other vehicles and junctions on the travel time between two points can be simulated by setting the time to travel between junctions as a function of the posted speed limit and then estimating the time to pass through a junction.

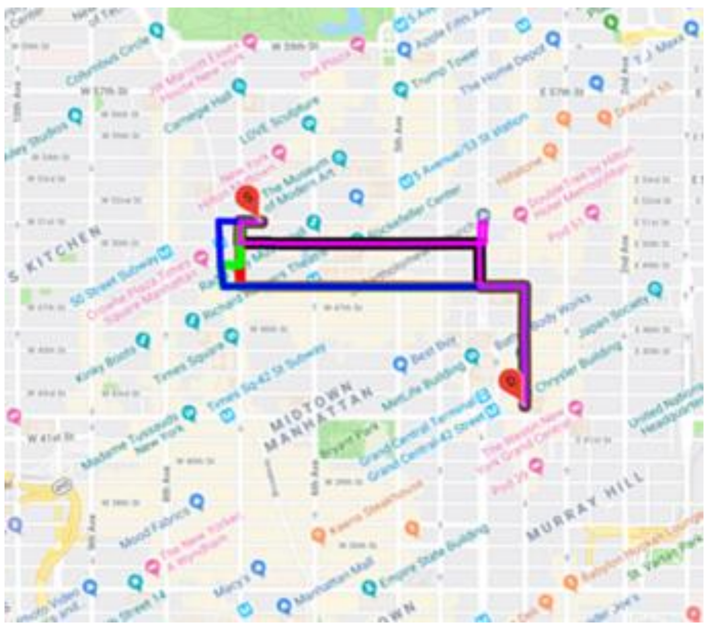

Figure 1

The actual route taken from the start to the end of the trip is not known but a set of possible routes for the trip can be obtained as we know the distance of the trip. The data consists of trips which began and ended on Manhattan island in a one hour time slice. An Example of this is shown in Figure 1.

Method

Each trip has a recorded distance (d). For each trip in our training set we create a set of possible routes where the distance travelled (x) is $d \leq x \leq d + 0.11$.

Each feasible trip is broken into two parts

- Travelling between junctions
- Waiting at and travelling through junctions.

We set the time to travel between junctions as the time to cover the distance at the posted speed limit.

To infer the time required to pass through each junction we look to adjust the junction time based on the errors occurring to the possible routes, and minimise the total absolute error on the possible routes that have the minimum absolute error for each trip using the following algorithm. An initial value for each junction is set.

```
S ← set of feasible routes for each trip in training set
J ← initial values for each junction
LE ← feasible route with least error for each trip
a ← maximum allowed percentage difference in absolute errors
previous ← absolute error of feasible routes in LE
check ← 0
required ← number of successive iterations that meet the stopping criteria
While check < required
    J ← adjusted values for each junction
    LE ← feasible route with least error for each trip
    current ← absolute error of feasible routes IN LE
    If current < previous and difference between both is < a then
         check++
    Else
         check ← 0
    previous ← current
Create graph G with inferred times
```

Testing

The trips in our testing set were then processed using graph G searching for the quickest path, and the inferred time were compared to the recorded values as shown in Table 1. These are an improvement on giving each junction an initial value which id shown in brackets

Initialization	Mean Trip Error (secs)	Mean Abs. Percentage Error (%)	RMSLE	Trips within 10% recorded time (%)	Coefficient of Determination (R^2)
Zeros	-38.49 (-483.10)	21.07 (64.91)	0.2832 (1.1624)	40.45 (0.46)	0.7605 (0.1580)
Average	-35.75 (-109.18)	21.07 (30.03)	0.2832 (0.4083)	40.82 (20.76)	0.7602 (0.6147)
Average with junction diff	-35.73 (-110.96)	21.07 (29.94)	0.2832 (0.4085)	40.81 (20.73)	0.7603 (0.6162)
Traffic lights with junction diff	-35.65 (-149.63)	21.08 (31.11)	0.2831 (0.4520)	40.70 (17.78)	0.7597 (0.6023)
Average closest with junction diff	-38.68 (-45.80)	21.06 (31.26)	0.2830 (0.3896)	40.73 (22.71)	0.7630 (0.6668)

Table 1

Future Work

Further examine the effects of different adjustment methods and termination values.

References

[1] D. A. Desmond, K. N. Brown.: *Inferring Waypoint Using Shortest Paths.* In: Proceedings of the 24th Irish Conference on Artificial Intelligence and Cognitive Science (AICS 2016): 45-56 (2016)

[2] S. Zhu, D. Levinson. *Do People Use the Shortest Path? An Empirical Test of Wardrop's First Principle.* Proc. of the 91st Annual Meeting of the Transportation Research Board, Washington, DC, 2012.

[3] D. A. Desmond, K. N. Brown.: *Inferring Waypoints in the Absence of Knowledge of Driving Style.* In: Proceedings of the 25th Irish Conference on Artificial Intelligence and Cognitive Science (AICS 2017): 166-178 (2017)

[4] http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml

A World Leading SFI Research Centre

This project has been funded by Science Foundation Ireland under Grant Number SFI/12/RC/2289