A Trip-based Trajectory Matching Algorithm for Sparse GPS Data

Lipika Deka and Mohammed Quddus
School of Civil and Building Engineering
Loughborough University
Leicestershire LE11 3TU
United Kingdom

Introduction

Recent times have seen an increased penetration rate of mobile devices equipped with GPS. This has made it possible for the collection of trip-based sequences of spatio-temporal data and using it for a plethora of applications such as mobility modelling, traffic management and semantic applications such as finding point-of interests. Another emerging application scenario is the use of trip-based trajectory data to study the occurrences and magnitudes of speed limit violation by capturing trip-based driving speeds and comparing them with the corresponding posted speed limits with a view of understanding drivers' speed choice behaviours. Spatio-temporal variations in driving speeds have a significant impact on safety and mobility of a road network (Quddus, 2013).

All raw and hence imprecise GPS trajectories must first be mapped onto links of a road network using appropriate map-matching algorithms. This is challenging as there are inaccuracies in both GPS and road network data. GPS inaccuracies are further aggravated in urban areas due to Multi-Path Interference and as the need for energy efficiency in GPS devices lead to capturing low frequency GPS samples. Existing map-matching algorithms are mainly online and work with high frequency GPS data acquired from relatively precise vehicle-based navigation devices. Existing offline map-matching techniques, uses additional information such as driver's route choice, mobility patterns, assumptions regarding the nature of GPS error and caters to a fixed GPS sampling frequencies (Yuan et al., 2010; Zheng et al., 2011; Miwa et al., 2012) . Hence, we aim to develop an offline trajectory based map-matching algorithm that uses only spatio-temporal GPS location information and the underlying digital road map, and is independent of the type of road network, travel mode, sampling frequency of the GPS position recordings (may be variable even within the same trip). The proposed algorithm is based on the degree of closeness (in terms of proximity and heading) of the user trajectory as derived from every pair consecutive GPS position.

Methodology

The map-matching algorithm was developed based on two datasets. The first being the trip-level travelling characteristics dataset consisting of Person ID, Trip ID and Spatial-temporal GPS data at a regular interval (comprising of clock time and positioning fixes (a series of latitudes and longitudes associated with a trip)). The second dataset is the road network dataset containing the topological details of the underlying road network such as spatial information of link nodes, link length and grade separation. Given a GPS trajectory **T** comprising of chronologically ordered GPS positions and an underlying road network, the trajectory based map-matching algorithm identifies the actual path **P** in the road network comprising of a sequence of connected links, which closely matches the trajectory **T**.

The trajectory-matching process begins by identifying candidate links within an error circle of the first GPS position. These links are considered to be the initial candidate paths which then grow through the process of connected candidate link concatenation as candidate links are identified for each subsequent GPS positioning fix. Each growing candidate path comprising of connected links is associated with a similarity score which we term as the Total Pattern Matching Score or TPMS (the higher the score the higher is the probability that it is the correct path or trajectory).

$$TPMS = PMS_{01} + PMS_{12} + \cdots + PMS_{(n-1)n} = \sum_{i=1}^{N} PMS_{(i-1)i}$$

where, PMS_{01} is equal to 0 denoting the initial TPMS of each of the candidate paths, PMS_{12} is the pattern matching score derived from the matching of first and second GPS position and so on.

$$PMS = W_d\, f(\Delta d) + W_h\, g(\Delta\theta)$$

Where,

$$W_d + W_h = \mathbf{1}; \mathbf{0} < W_d < 1; 0 < W_h < 1 f(\Delta d) = \begin{cases} \frac{d_{th} - \Delta d}{d_{th}} & if\ \Delta d \le d_{th} \\ -1 & else \end{cases}$$

and $range(f) = [-1,1]$;

$$g(\Delta\theta) = \cos(\Delta\theta) \qquad \text{and } range(g) = [-1,1]$$

$W_d\ and\ W_h$ are the weighting parameters representing the relative importance of Δd and $(\Delta\theta)$ in selecting the correct segment, d_{th} is the threshold value for proximity of the GPS fixes to the correct candidate link, Δd is the average distance of consecutive GPS positions from respective candidate links and Δθ denotes the angle difference between the user heading and candidate path heading.

If none of the candidate paths is directly connected to any of the candidate links selected for the current GPS fix, the method then uses the Dijkstra's shortest path algorithm that identifies the shortest path (in terms of the link length) between each of the candidate paths and each of the identified candidate link of current GPS fix. Each candidate path is extended by concatenating the corresponding shortest path. Such a scenario arises if the time interval between two recorded consecutive GPS fixes is large enough for the user to have crossed more than one junction.

Result and Discussion

The values of parameters W_d and W_h can be obtained from an empirical analysis using a reference (true) dataset and we do so using a dataset from Washington DC, USA (5). The dataset was provided as part of the ACM SIGSPATIAL Cup 2012. The parameter values of $W_d = 0.7$ *and* $W_h = 0.3$ provide the highest accuracy (~97%) in the correct path identification for a dataset collected in a dense urban area. The accuracy is defined as the percentage of correctly matched road links as compared to the ground truth path.

On a further analysis it was seen that 74% of the mismatched links have been incorrectly map-matched because the GPS points erred by more than 300m from the given ground truth. The proposed algorithm has provided similar map-matching accuracy on datasets from Australia and England suggesting that the developed method is generic and transferable.

Reference:

1. J. Yuan, Y. Zheng, C. Zhang, X. Xie, G. Sun. "*An Interactive-Voting Based Map Matching Algorithm*". In Proceedings of the 2010 Eleventh International Conference on Mobile Data Management.
2 T. Miwa, D. Kiuchi, T. Yamamoto, T. Morikawa. "*Development of Map Matching Algorithm for Low Frequency Probe Data*". Journal of Transportation Research Part C. Vol. 22, 2012, pp. 132-145.
3 Y. Zheng, M. A. Quddus. *Weight-based shortest path aided map-matching algorithm for low frequency GPS data*. In Transportation Research Board, 2011
4 M. A. Quddus. *Exploring the Relationship Between Average Speed, Speed Variation, and Accident Rates Using Spatial Statistical Models and GIS.* Journal of Transport Safety and Security. Vol. 5, No. 1, pp. 27-45.
5 ACM SIGSPATIAL Cup 2012 (2012). (ACM SIGSPATIAL Cup 2012)