A Maximum Likelihood Map-Matching Algorithm

Nanne J. van der Zijpp
Delft University of Technology
Faculty of Civil Engineering and Geosciences

-extended abstract-
submitted to the 15th International Symposium on Transportation and Traffic Theory

Introduction

With the development of increasingly advanced traveller information and traffic control systems the capabilities to observe travel speeds, occupancies and traffic flows are more and more becoming a limiting factor in the application of this technology. Technically the ultimate traffic surveillance system is already feasible, for example by equipping all cars with a GPS device and a wireless connection to a central computer. However, for the majority of consumers, the individual advantages of such a system do not outweigh its costs of hardware and communication.

This opens up the market for surveillance systems that are at first sight less ideal but do not require any hardware investments on the part of consumers. One possibility is to use anonymous cell phone data. When a cell phone is actively used, its location can be determined by different techniques. The accuracy of these techniques varies widely, but is generally less then the accuracy that is reached with GPS. However, this is compensated by the high number of measurement reports that are available, due to the widespread use of cellular phones. Therefore the use of appropriate statistical techniques may enable the extraction of valuable information on traffic conditions from these data.

This paper will focus on the problem of matching a sequence of measurements to a sequence of locations on a network. Because of the lack of accuracy, this is not a trivial task. The problem will be solved in a Maximum Likelihood manner: each sequence of measurements will be matched in such a way that the probability of observing these measurements is maximised. The algorithm must ensure that the location estimates satisfy speed constraints and constraints posed by the topography of the network.

The novelty of the paper is that the Maximum Likelihood map-matching problem is shown to be equivalent to a shortest path search problem, which is efficiently solved with existing algorithms.

Mathematical formulation of the map matching problem

We will now describe the problem in mathematical terms. For this purpose we introduce the following notation:

$\{t_k, x_k, y_k\}$ A series of probe location messages, k=1,2, …K, with t_k the time instant of the kth message and $\{x_k, y_k\}$ the co-ordinates of the kth message.

$\{d_{ar}, x_{ar}, y_{ar}\}$ A series of points along route r, r=1,2,..R, a=1,2,..A, with d_{ar} the distance of the ath point from the Origin along route r, and $\{x_{ar}, y_{ar}\}$ the co-ordinates of this point. These points can be chosen arbitrarily close to each other.

π A route matching map with:

$$\pi^r_{ka} = 1: \text{ message } k \text{ originates from point } a \text{ along route } r$$
$$\pi^r_{ka} = 0: \text{ otherwise} \qquad (1)$$

We will formulate the constraints that apply to the route matching map and define the map matching problem as an optimisation problem. We assume that the location in the measurement reports $\{t_k, x_k, y_k\}$ satisfy the following equation:

$$\begin{pmatrix} x_k \\ y_k \end{pmatrix} = \begin{pmatrix} x(t_k) \\ y(t_k) \end{pmatrix} + \begin{pmatrix} e_k^x \\ e_k^y \end{pmatrix} \qquad (2)$$

with $[x(t_k), y(t_k)]$ the exact position of the vehicle at instant t_k, and $[e_k^x, e_k^y]$ the observation error of the kth observation for which the distribution $p^k[e_k^x, e_k^y]$ is known and independent from observation errors at all other periods.

We will now formulate the constraints that apply to the route matching map π**.** Obviously each message originates from exactly one point hence:

$$\sum_{r=1}^{R} \sum_{a=1}^{A} p^r{}_{ka} = 1\ , \qquad \forall k \qquad (3)$$

Also a vehicle can only travel at one route at a time, hence:

$$p^r{}_{ka} = 1 \Rightarrow p^s{}_{pb} = 0,\ \forall s \neq r,\ \forall p, b \qquad (4)$$

Finally there are a number of speed constraints to be taken into account. Vehicles have a positive speed, hence:

$$p^r{}_{ka} = 1 \Rightarrow p^r{}_{k+1,b} = 0,\ \forall b, a \,|\, (d_{br} - d_{ar}) < 0 \qquad (5)$$

The speed of vehicles does not exceed a given maximum speed $v^{\max}$, hence:

$$p^{r}{}_{ka} = 1 \Rightarrow p^{r}{}_{k+1,b} = 0, \ \forall b \ \Big| \ (d_{br} - d_{ar}) > v^{\max}.(t_{k+1} - t_k) + \Delta \qquad (6)$$

where ? is the distance between two consecutive points along a route.

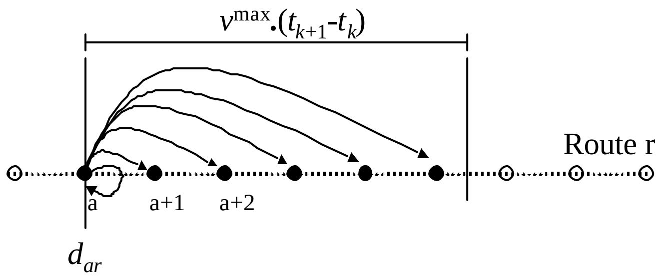
The objective is to find a map p that satisfies these conditions and maximises the likelihood of observing the messages $\{t_k, x_k, y_k\}$. This likelihood can be written as:

$$L(p) = \prod_{r=1}^{R} \prod_{a=1}^{A} \prod_{k=1}^{K} \left(\mathrm{p}^k[x_k - x_{ar}, y_k - y_{ar}]\right)^{p^{r}{}_{ka}} \qquad (7)$$

A Maximum Likelihood map matching algorithm

It is clear that an effective search strategy is needed to maximise the likelihood specified in (7): for a brute force method the number of matching maps π that are feasible in terms of constraints (3), (4), (5), (6) is simply too large. In the present paper we will define such a strategy.

Suppose that we match message k with point a at route r then the constraint (3) implies that message k+1 should also be matched to a point along route r. Moreover constraints (4) en (5) imply that this point should be somewhere in the interval $[d_{ar}, d_{ar} + v^{\max}.(t_{k+1}-t_k)+?]$. In other words: the vehicle can only travel *forward* over a route and can not exceed its maximum speed. Graphically this is represented in Figure 1.



● : Feasible ○ : Infeasible

Figure 1: *If point a is the point to which message k is matched, the arrows in above figure indicate the feasible matching points for message k+1.*

We will now construct a so-called hypernetwork consisting of nodes and links to which we will give specific interpretations. These interpretations will enable us to formulate the route matching problem as a shortest path search problem.

We start with defining the nodes of the hypernetwork. Of each node on each route we make K (=the number of messages) copies, see Figure 2. Each node represents a physical node in a specific time period. Subsequently we add the links. To the links we give the following interpretation:

“The existence of a link between node (a,r,k) (node a on route r in period k) and node $(b,r,k+1)$ (node b on route r in period $k+1$) indicates that the movement of a vehicle between node a and b during the period t_{k+1}-t_k is feasible in terms of constraints (4) and (5).”

Traversing a link will be given the following interpretation:

“Traversing the link that connects node (a,r,k) with node $(b,r,k+1)$ represents the action of matching message k with node a on route r AND matching message $k+1$ with node b on route r”

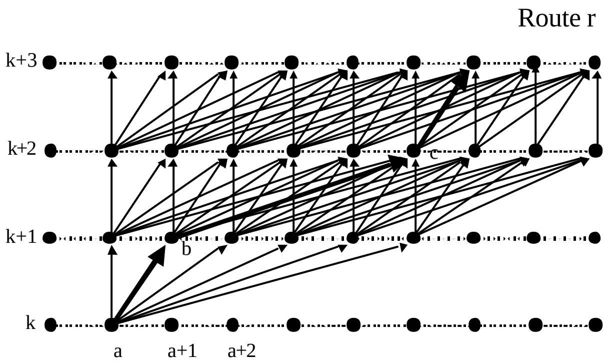


Figure 2: *Hypernetwork in which each link represents a feasible transition from start-node to end-node.*

This implies that travelling over a route through the nodes $\{(a,r,k), (b,r,k+1), (c,r,k+3)\ldots\}$ is equivalent to matching message k with node a, message $k+1$ with node b, etc. Hence each route matching map p that is feasible in terms of constraints (3), (4), (5), (6) corresponds to a path that connects the layer 1 of the hypernetwork to the layer that represents period K, and vice versa.

The likelihood of this matching map is given by equation (7). Instead of maximising the likelihood of equation (7), we might as well minimise the negative logarithmic value of this expression:

$$\log L(p) = \sum_{r=1}^{R}\sum_{a=1}^{A}\sum_{k=1}^{K} p_{ka}^{r} \,.-\log(\ \mathrm{p}^{k}[x_k - x_{ar}, y_k - y_{ar}]\,) \tag{8}$$

If we associate with each link in the hypernetwork a *cost* of:

$$-\log(\ \mathrm{p}^{k}[x_k - x_{ar}, y_k - y_{ar}] \tag{9}$$

where (a,r,k) are the node-number, route number and period number of the *end-node*, it follows that $\log L(p)$ corresponds to the *cost* of the path that is defined by p.

Finding the route matching map p that maximises the loglikelihood is hence equivalent to finding the shortest path that connects the nodes $(a,r,1)$ and (b,r,K) over all combinations of initial match node a, end match node b and route r. By extending the earlier defined hypernetwork with an origin-centroid *A* that is connected to all potential initial match nodes $(a,r,1)$, and a destination-centroid B that is connected to all potential end match nodes (b,r,K) (see Figure 3) the problem is simplified to finding the minimum cost path between A en B.

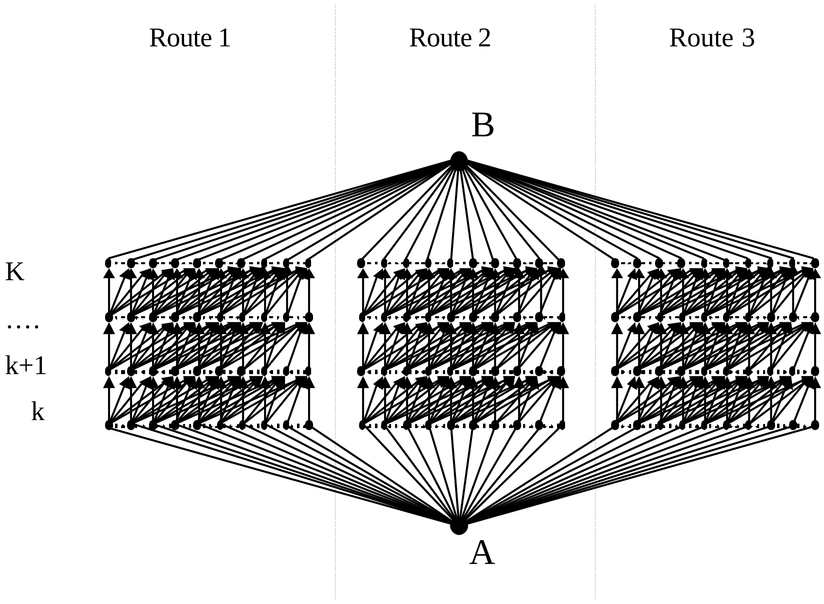


Figure 3: *Hyper network, extended with origin-connector A and destination connector B. The connector links to A have costs equal to* (9) *and links to B have zero costs.*

To find this minimum cost path standard shortest path algorithms can be used, once the network is defined. These algorithms are characterised by a high efficiency.

Experiments

The algorithm has been implemented and tested for a few example networks using a Monte Carlo approach. A typical map-matching result is shown in Figure 4. The sensitivity of the algorithm for parameters like average call duration, accuracy of measurements and frequency of measurements has been assessed.

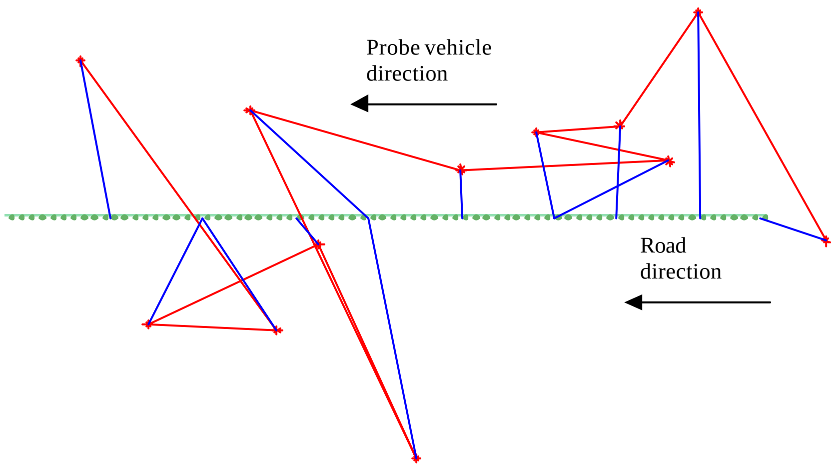


Figure 4: *Result of matching messages of run 1 (of 1000) to route 2 of scenario 1*

Acknowledgement

The research that has led to this publication was initiated and sponsored by Syntrack, a joint venture of Libertel-Vodaphone and Ericsson