A MAXIMUM LIKELIHOOD MAP-MATCHING ALGORITHM

Nanne J. van der Zijpp
Delft University of Technology
Faculty of Civil Engineering and Geosciences
Transportation Planning and Traffic Engineering Section
zijpp@ct.tudelft.nl

1 INTRODUCTION

With the development of increasingly advanced traveller information and traffic control systems (ATIS/ATCS), the requirements for traffic surveillance systems that provide the input data for these applications become more and more difficult to meet. ITS applications require accurate and up to date estimates of traffic state variables, such as travel time, occupancies, link flows and Origin-Destination patterns.

This paper considers the problem of deriving these quantities from vehicles that report their positions regularly and therefore act as *probes*. The paper will present a technique that can be used to match a sequence of probe-positions to a network, even if the individual positions in this sequence contain large observation errors. Once vehicle positions are matched to map locations by a process known as map matching, estimates of traffic state variables can be derived.

With the techniques currently available it is already feasible to build a centralised, high quality traffic surveillance system. However, the initial investments and the costs of communication associated with such a dedicated system are considerable, while little is known about the willingness of consumers to pay for services derived from such a traffic surveillance system.

This makes it worthwhile to investigate the possibility of building a traffic surveillance system that uses data that are available without large investments. Anonymous cell phone might provide such a datasource. When a cellular phone is actively used, its location can be determined by a range of techniques. By retrieving the cell phone position periodically, a trail of estimated positions results.

The accuracy of GSM positioning techniques varies widely, but is generally less then the accuracy that is reached with GPS. However, this is compensated by the high number of observations that are potentially 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 trail of reported probe-positions to a network. The novelty of this paper lies in the size of the observation errors that can be dealt with.

Existing map-matching approaches can be categorised in point-to-point, point-to-curve and curve-to curve algorithms. Point-to-point and point-to-curve algorithms have in common that probe positions are matched individually to the network without considering earlier or later observed probe-positions. In point-to-point algorithms a finite number of network positions are considered, while in point-to-curve algorithms an infinite number of positions along a (finite) number of road sections are considered. The application of point-to-point and point-to-curve algorithms, requires that the observation errors are small because there is no possibility to avoid erroneous matches by imposing consistency requirements on subsequent matches.

This paper will therefore solely be concerned with curve-to-curve algorithms. 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 jointly the locations to which the messages are matched form a feasible path, and that the implied vehicle trajectories are in line with speed constraints that apply.

The novelty of this 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.

2 PROBLEM STATEMENT

As mentioned in the previous section, this paper is concerned with map matching. A sequence of Measurement Reports (MR's) that contain a location vector and a time stamp need to be matched with positions along a route. This is known as an example of curve-to-curve matching.

If the measurement errors are small and the positions of all network sections are accurately known, a simple strategy would suffice to construct a route from a series of MR's: simply match each MR to the nearest network section and concatenate these sections to a route.

In the more realistic case where individual errors of MR's are too large to obtain absolute certainty about the road section a vehicle is on, map matching is only possible if multiple MR's are considered simultaneously. In this case additional constraints can be imposed, such as the requirements that the matched positions form a feasible route and that the speed constraints are satisfied.

Each time a particular MR is matched to a specific point on a route, the size of the error in the position vector of the MR is implied. Provided that a probabilistic model for this error is available, one can compute the likelihood of this match. Likewise one can compute the likelihood of simultaneously matching a number of MR's to a number of points.

When a MR-sequence is matched to a sequence of points this implies that a vehicle has experienced specific travel times on the intermediate route sections. One can add additional information to the route matching problem by supplying a probability distribution of these travel times. This probabilistic model automatically incorporates the constraints on minimum and maximum speed that were mentioned above. The shape of such a probability distribution reflects the apriori knowledge that is available about travel times. If no apriori knowledge is available, or one chooses to ignore this information, a uniform probability distribution can be used.

Above information is sufficient to define the map-matching problem in a Maximum Likelihood manner. Suppose the MR-sequence that needs to be matched contains K elements and the number of routes that should be considered is R. Then all MR's should be matched to a longitudinal position along one of these R routes. In this case, the *map-matching solution vector* of the map-matching problem contains one discrete element and K continuous elements. The map-matching problem therefore has an infinite number of potential solutions, implying that the Maximum-Likelihood solution can only be determined analytically.

3 ALGORITHM

The algorithm will be described in the full paper.

4 EXPERIMENTS

Overall set-up of experiments

A series of experiments were done to demonstrate the map matching algorithm and to analyze its performance. These experiments involve generating artificial measurement reports, matching these measurement reports to a map, and expressing the quality of the map matching process in a number of performance indicators.

To determine the factors that influence the quality of the map-matching process, the sensitivity of this process with respect different parameters is investigated. This is done by repeating the experiments for different scenarios, where each scenario is characterized by a set of parameters. For each scenario a set of performance indicators is computed by averaging the results of a large number of Monte Carlo experiments. The different scenarios share a single set of random numbers, so that differences in performance indicators between scenarios reflect as much as possible the impact of the changed parameter values rather than random fluctuations.

Definition of scenarios

A scenario is characterized by the parameters that are used to generate the Measurement Reports and the parameters that are used as input to the Map Matching algorithm. Parameters that do not vary between scenario's are referred to as the *fixed* parameters. In the present analysis the following parameters are fixed: the parameters that define the network, parameters that define the routes that are considered, and parameters that define the speed that vehicles maintain on the network.

Parameters that do vary between scenario's are referred to as the *variable* parameters. In the present analysis, the following parameters are varied:

- The poll-rate. This is the time in between two subsequent messages;
- The accuracy of the measurement reports. This parameter influences the size of the randomly generated observation error (see below);
- The speed margin. Any vehicle trajectory that implies that a vehicle has exceeded the maximum speed will be rejected. The maximum speed is obtained by adding the speed margin to the true speed;
- The spatial resolution. In order to make the proposed map matching algorithm feasible, the longitudinal position of a vehicle along a route is discretized. The spatial resolution refers to the maximum distance between two subsequent position

Definition of performance criteria

The following performance criteria where used

- Percentage of correctly matched routes
- Percentage of correctly matched links
- Reduction in positioning error
- Reduction in bias
- Likelihood

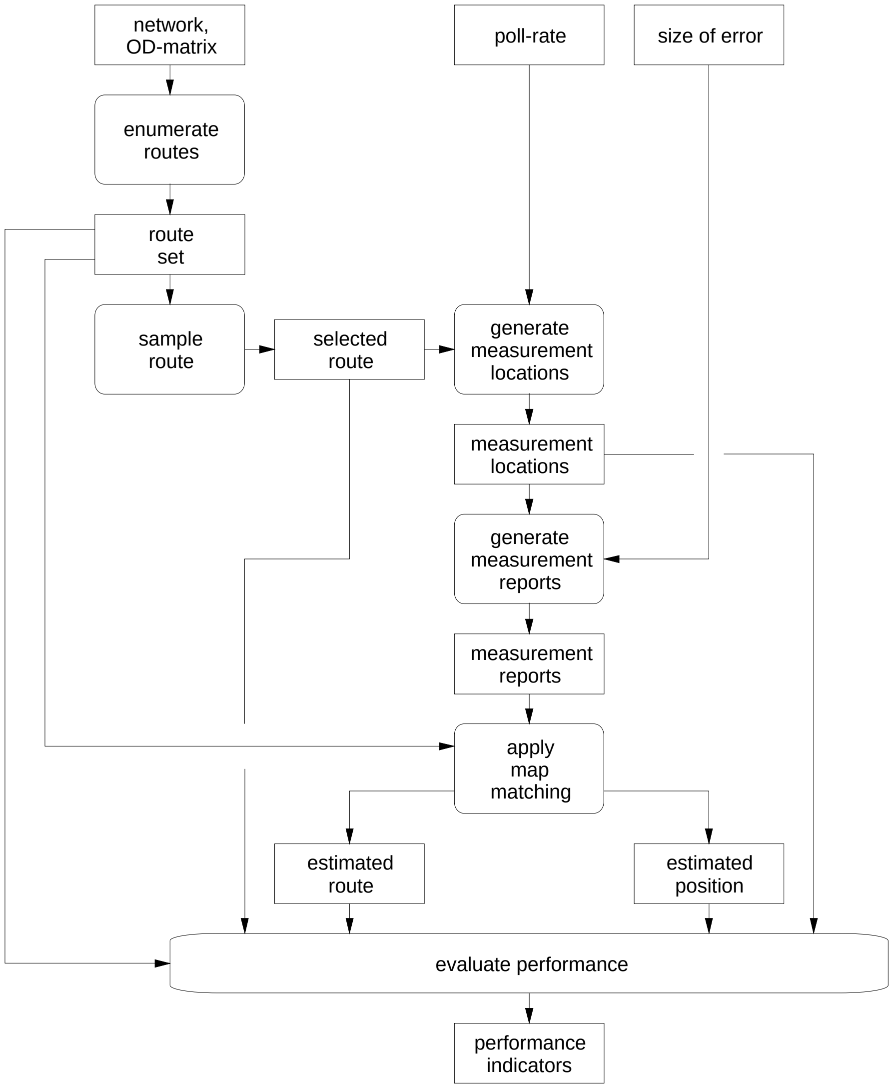

Figure 1: *Experimental set-up*

5 RESULTS

The results will be presented in the form of tables and plots. Figure 2 illustrates a Maximum Likelihood map-matching result. The sensitivity experiments where done using Monte Carlo techniques and involved solving 100 map-matching problems per scenario.

Figure 3 presents the outcome of the sensitivity analysis with respect to the first performance indicator.

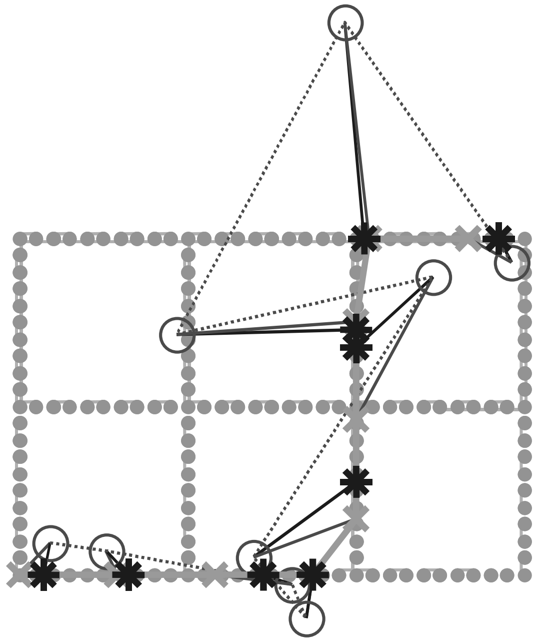

Figure 2: *Typical Map-Match results. The open circles represent Measurement Reports (MR), The '*' symbols represent the position to which a MR is matched*

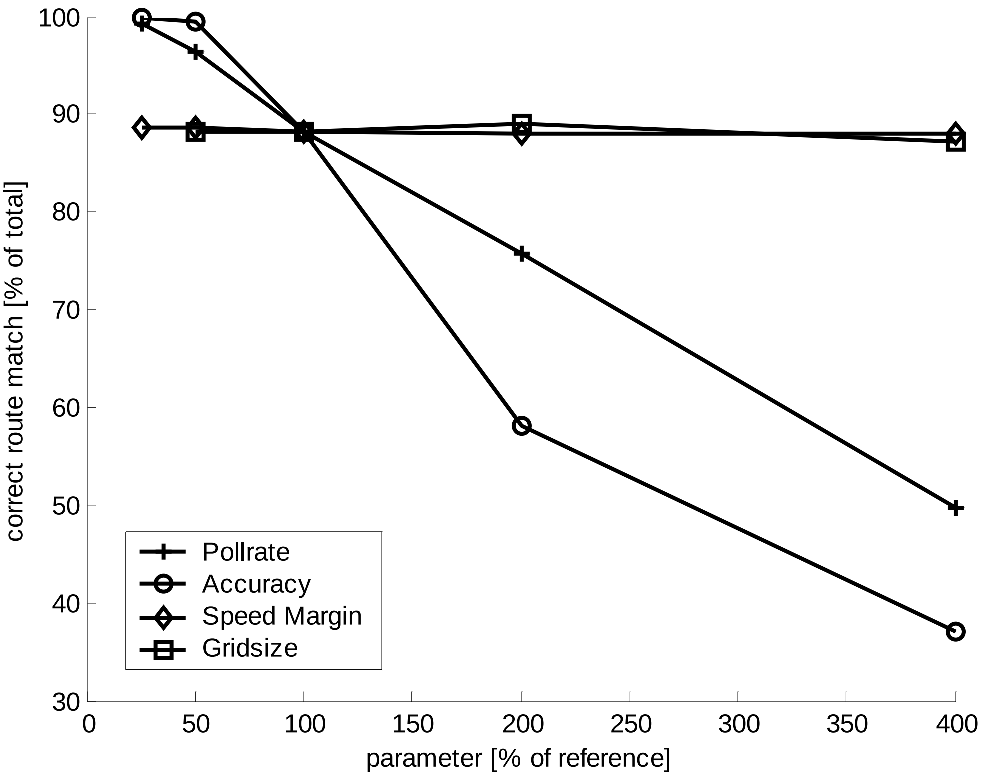

Figure 3: *Sensitivity analysis results: The percentage of successful matches is plotted as a function of the relative value of each parameter with respect to the its reference value.*

Acknowledgement

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