Transportation Research Part C 93 (2018) 13–35

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

Likelihood-based offline map matching of GPS recordings using global trace information

Luk Knapen*, Tom Bellemans, Davy Janssens, Geert Wets

Hasselt University, Transportation Research Institute (IMOB), Wetenschapspark 5 bus 6, 3590 Diepenbeek, Belgium

A R T I C L E I N F O

Keywords:
GPS traces
Map matching
Transportation modeling
Big data analysis

A B S T R A C T

In batch map matching the objective is to derive from a time series of position data the sequence of road segments visited by the traveler for posterior analysis. Taking into account the limited accuracy of both the map and the measurement devices several different movements over network links may have generated the observed measurements. The set of candidate solutions can be reduced by adding assumptions about the traveller's behavior (e.g. respecting speed limits, using shortest paths, etc.). The set of feasible assumptions however, is constrained by the intended posterior analysis of the link sequences produced by map matching. This paper proposes a method that only uses the spatio-temporal information contained in the input data (GPS recordings) not reduced by any additional assumption.

The method partitions the trace of GPS recordings so that all recordings in a part are chronologically consecutive and match the same set of road segments. Each such trace part leads to a collection of partial routes that can be qualified by their likelihood to have generated the trace part. Since the trace parts are chronologically ordered, an acyclic directed graph can be used to find the best chain of partial routes. It is used to enumerate candidate solutions to the map matching problem.

Qualification based on behavioral assumptions is added in a separate later stage. Separating the stages helps to make the underlying assumptions explicit and adaptable to the purpose of the map matched results. The proposed technique is a multi-hypothesis technique (MHT) that does not discard any hypothesized path until the second stage.

A road network extracted from OpenStreetMap (OSM) is used. In order to validate the method, synthetic realistic GPS traces were generated from randomly generated routes for different combinations of device accuracy and recording period. Comparing the base truth to the map matched link sequences shows that the proposed technique achieves a state of the art accuracy level.

1. Introduction

Map matching combines a road transport network description consisting of nodes and directed links with a time series of coordinate tuples that describes the movement of a traveler. A *trace* is a chronologically ordered sequence of all GPS recordings associated with a movement. The purpose is to reconstruct the sequence of links crossed by the traveler in chronological order. In this section a short overview of existing map matching techniques and their respective fields of application is given in order to sketch the state of the art. Two main classes of map matchers are distinguished. The technique proposed in this paper is aimed at offline (batch) map matching of GPS traces.

* Corresponding author.
E-mail address: luk.knapen@uhasselt.be (L. Knapen).

https://doi.org/10.1016/j.trc.2018.05.014
Received 29 August 2015; Received in revised form 23 March 2018; Accepted 18 May 2018

1.1. Online map matching

Online near real time map matching processes coordinate pairs as soon as they come available and aim to determine the network link that is actually being traveled. Map matchers in this class are deployed in navigation aids. Their software operates on dedicated microprocessors and typically data sampling is in the order of 1–100 Hz. In many cases data from several sensors (odometer, gyroscope, accelerometer, etc.) are available for data fusing along with GPS coordinates. Quddus et al. (2007) provide a comprehensive overview of online map matchers. Greenfeld (2002), Ochieng et al. (2010), Li et al. (2013), and Abdallah et al. (2011) discuss the data fusion techniques and inference methods. The aim of online map matching is to determine the link on which the vehicle is moving and to calculate the position of the vehicle on the link as accurately as possible (e.g. for traffic signal influencing by buses (Quddus et al., 2007)). The latter is essential in ITS (Intelligent Transportation Systems) applications and in Advanced Driver Assistance Systems. Nowadays map matchers are based on the multi-hypothesis technique (MHT) about the position of the vehicle. Such methods are called Multi-Hypothesis Map Matching (MHMM) in Bonnifait et al. (2009). In many cases, MHT and sensor data fusing feed maximum likelihood Bayesian inference engines and often Kalman filtering is used. Recent online map matchers, starting with Greenfeld (2002), usually incorporate topology constraints.

Velaga et al. (2009) present a topological map matching (tMM) method and describe the technique used to learn the value for the required weight coefficients. Selecting the initial link using the first fix (GPS point) is based on weight functions for *proximity* (distance) and *heading* (compared to link direction). Two additional weight functions are used for link selection at a junction: *turn restrictions* and *link connectivity*. All weight functions can have negative values. The total weight score (TWS) is a linear combination of the applicable weight function values. The respective weight coefficients are determined using traces for which the link sequence is known. For all fixes near junctions, weight coefficients are generated randomly until a tuple is found that correctly identifies the chosen link. Those tuples are applied to all fixes and the portion of wrong fixes is determined for each case. Finally the tuple leading to the minimal error is determined by regression. Weight coefficient tuples for *rural*, *suburban* an *urban* environments are presented.

1.2. Offline map matching

Offline or batch map matchers aim to process previously recorded sequences of coordinate pairs in order to extract travel behavior information either for a single moving object (either person or vehicle) over a long period or for a large set of moving objects. GPS recordings are either *vehicle traces* produced by dedicated devices mounted in a vehicle or *person traces* recorded by smartphones carried by individuals. The aim is to determine the sequence of links used by the moving object. Schüssler and Axhausen (2009) state that map matching of person traces requires high resolution network information. Available data consist of time series of GPS recordings and in some cases from other sources (Bluetooth, Wifi and mobile phone related events). Large datasets are available and need to be processed efficiently. In Quddus et al. (2007) and Schüssler and Axhausen (2009), map matching techniques (both online and offline) are classified as (i) pure geometry based methods, (ii) topological methods, (iii) probabilistic methods and (iv) advanced procedures. Pure geometric methods are further classified by Quddus et al. (2007) as point to point matching (finding the nearest node or shape point), point to curve matching (finding the polyline to which the distance is minimal) and curve to curve matching (matching the vehicle trajectory against known roads). Those methods can deliver link sequences that represent non-connected walks in the network.

The technique proposed by Marchal et al. (2005) solves the problem of non-connected walks by adding topological constraints. It starts by determining which links are identified by the first few GPS recordings. Each of those constitutes the first link in a candidate path. When the next GPS coordinate pair is processed, for each route candidate being built, only the last link in the sequence and the links that can be reached from that link (*forward star*) are investigated when looking for links matched by the new coordinate pair. Since each route candidate shall consist of a linear sequence of links, route candidates are cloned and each clone is extended by exactly one member of the forward star. The route candidates then are assigned a score and in order to avoid huge sets of route candidates, only the N route candidates having the best scores are kept ($N = 30$). Scoring is done as follows. Each GPS point can match at most one link in each route candidate and each link in the route needs at least one GPS fix. The distance between the point and the link is a measure of quality of the selection (the lower the better). For each GPS fix, the distance between the recorded position and the matched link is computed. The sum of those values constitutes the score for a candidate link sequence. If there are too many candidates, the ones having the highest scores are discarded. The computational effort and memory requirements grow with N. Making N too small, can cause promising candidates to be removed prematurely and hence can decrease the average quality of the final candidates. Schüssler and Axhausen (2009) evaluate this technique by comparing the quality (score) of the best solutions found and the corresponding computational effort for several values for the maximal candidate set size N. The paper concludes that the value reported in Marchal et al. (2005) is a valid one; the average score per GPS point does not significantly decrease with the candidate set size for $N > 30$. It also reports that the processing time per point is between 10 [ms] (for $N = 20$) and 75 [ms] (for $N = 100$).

Zhou and Golledge (2006) use a similar procedure implemented in ArcGIS. GPS recordings are processed sequentially and a pool of candidate solutions is kept. In a preprocessing stage, they first replace clusters of GPS points by their centroid (*cluster reduction*) but also add interpolated GPS points when the distance between two consecutive points is larger than half of the minimum length for the links in the buffer defined by the two GPS points. Then a 2-norm (distance) and a rotation measure are used to determine the weight for each point in the preprocessed dataset. A set of candidate partial paths is kept and extended so that a connected walk results from the method. In the link selection phase, a Dempster belief function is used to determine the plausibility of the selected link. However, the authors do not explain what criteria were used.