Transportation Research Part C 57 (2015) 55–67

Contents lists available at ScienceDirect

Transportation Research Part C

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

Real-time identification of probe vehicle trajectories in the mixed traffic corridor

Yu Mei [1], Keshuang Tang *, Keping Li [1]

Department of Comprehensive Transportation Information and Control Engineering, Tongji University, No. 4800, Cao'an Road, Shanghai 201804, China

ARTICLE INFO

Article history:
Received 22 December 2013
Received in revised form 10 February 2015
Accepted 9 June 2015
Available online 20 June 2015

Keywords:
Semi-supervised learning technique
Probe vehicle trajectory
Mixed corridor
Urban expressway

ABSTRACT

This paper proposes three enhanced semi-supervised clustering algorithms, namely the Constrained-K-Means (CKM), the Seeded-K-Means (SKM), and the Semi-Supervised Fuzzy c-Means (SFCM), to identify probe vehicle trajectories in the mixed traffic corridor. The proposed algorithms are able to take advantage of the strengthens of topological relation judgment and the semi-supervised learning technique by optimizing the selection of pre-labeling samples and initial clustering centers of the original semi-supervised learning technique based on horizontal Global Positioning System data. The proposed algorithms were validated and evaluated based on the probe vehicle data collected at two mixed corridors on Shanghai's urban expressways. Results indicate that the enhanced SFCM algorithm could achieve the best performance in terms of clustering purity and Normalized Mutual Information, followed by the CKM algorithm and the SKM algorithm. It may reach a nearly 100% clustering purity for the uncongested conditions and a clustering purity greater than 80% for the congested conditions. Meanwhile, it could improve clustering purity averagely by 21% and 14% for the congested conditions and 6.5% and 6% for the uncongested conditions, as compared with the traditional K-Means algorithm and the basic SFCM. The proposed algorithms can be applied for both on-line and off-line purposes, without the need of historical data. Clustering accuracies under different traffic conditions and possible improvements with the use of historical data are also discussed.

1. Introduction

Probe vehicle systems have been widely implemented in many developed countries as a promising technology. It has also gained increasing attentions in China in recent years. So far, more than 10 mega cities in China including Beijing and Shanghai have already launched the application of probe vehicle systems since 2002, most of which are based on taxi GPS (Global Positioning System) data. Those systems usually upload taxi's GPS data with a time interval of 5–60 s. Probe vehicle systems have many applications, e.g., dynamic traffic state estimation, OD (Origin Destination) estimation, route travel time estimation. In Shanghai, the numbers of GPS-equipped buses and taxies have reached 15,390 and 48,714 respectively at the end of 2011 according to Shanghai comprehensive transportation annual report (Shanghai City Comprehensive Transportation Planning Institute, 2012). More than 40,000 GPS-equipped taxies provide GPS data to Shanghai Traffic Information Center with a time interval of 10–20 s.

* Corresponding author. Tel.: +86 21 69583656; fax: +86 21 33626318.
E-mail addresses: 90mei@tongji.edu.cn (Y. Mei), tang@tongji.edu.cn (K. Tang), keping_li@vip.163.com (K. Li).
[1] Tel.: +86 21 69583656; fax: +86 21 33626318.

http://dx.doi.org/10.1016/j.trc.2015.06.008
0968-090X/

A high proportion of urban roads are elevated expressways that often exist in parallel with and close to ground arterials in many mega cities in China. It creates the so-called the mixed traffic corridor. Again, taking Shanghai as an example, the total mileage of elevated expressways has reached 308 km at the end of 2011 and most of which are parallel with and close to the arterials (Shanghai City Comprehensive Transportation Planning Institute, 2012). Usually, accurate altitude data together with a three-dimensional (3D) digital map can effectively identify vehicles running on the elevated expressways and the ground arterials. However, the accuracy of altitude data and the availability of 3D digital map often restrict such an application. It is firstly due to that it is difficult to accurately obtain altitude data in an environment full of skyscrapers and elevated expressways because a GPS device needs to receive signals from at least four satellites in order for the estimation of altitude of probe vehicles (Velaga et al., 2009). Secondly, many commercial probe vehicle systems in China only provide the post-processed horizontal GPS positions, i.e., easting and northing fixes (He et al., 2013). Hence, how to identify probe vehicle trajectories in the mixed corridor using horizontal GPS data has become a technical problem in the application of probe vehicle systems in China's mega cities.

In most of the existing map matching algorithms, common solutions to identify probe vehicle trajectories in the mixed corridor are either applying judgment rules based on topological relations of the mixed corridor, or utilizing travel speed threshold values based on an assumption that vehicles travel faster on the elevated expressways. The former is more often used in the case of offline application, in which a complete trajectory of every probe vehicle during a long time interval is already available. However, it can usually identify a small proportion of sample vehicles due to the limited information available from the topological relations and has rarely been used in the case of online application due to partial trajectories. The latter can be used for both online and offline application purposes as its calculation process is fairly simple and fast. However, proper travel speed threshold values are often difficult to determine particularly for the congested traffic conditions, which need a lot of historical data. In addition, its accuracy would decrease to an unacceptable level for the congested conditions due to small travel speed differences between the elevated expressways and the ground arterials, although it can obtain good performance for the uncongested conditions.

To overcome those drawbacks of the existing methods, this paper proposed three enhanced semi-supervised clustering algorithms, namely the Constrained-K-Means (CKM), the Seeded-K-Means (SKM), and the Semi-Supervised Fuzzy c-Means (SFCM). The proposed algorithms are able to take advantage of the strengthens of topological relation judgment and the semi-supervised learning technique by optimizing the selection of pre-labeling samples and initial clustering centers of the original semi-supervised learning technique based on horizontal GPS data. They were validated and evaluated based on the probe vehicle data collected at two mixed corridors on Shanghai's urban expressways. Clustering accuracies under different traffic conditions and possible improvements with the use of historical data were also discussed in the paper.

2. Literature review

To date, probe vehicle systems based on the GPS data from private cars, buses or taxies have already applied widely in the Intelligent Transportation Systems (ITS). Traffic state estimation and prediction was one of the basic applications (De Fabritiis et al., 2008; Kong et al., 2013; Lund and Pack, 2010; Pu et al., 2009). Another application of probe vehicle systems was updating digital road maps, which had the advantages of low cost, short updating circle and high accuracy (Li et al., 2012; Vartziotis et al., 2012). Some studies have also applied probe vehicle data to estimate dynamic OD for traffic demand forecasting purpose (Baek et al., 2010; Friedrich et al., 2010). Recently, the extended floating car systems (xFCD) that were based on the floating cars equipped with extra visual cameras and on-board Electronic Control Units have appeared to be a more efficient way for traffic state estimation, the level-of-service estimation, and real-time road safety assessment, e.g., Kyamakya et al. (2011), Messelodi et al. (2009), Pell et al. (2012) and Diaz et al. (2012).

Map-matching (MM) process is vitally important in the application of probe vehicle systems as its performance has a significant effect on obtaining real-time traffic information. MM algorithms usually integrate GPS data with a spatial road map in order to identify the road segment on which a user (or a vehicle) is traveling and the location on that segment. Recent research on MM algorithms has been based on either a conventional topological analysis or a probabilistic approach. Among them, topological MM algorithms are relatively simple, easy and quick, enabling them to be implemented in real-time. Therefore, a topological MM algorithm is used in many navigation devices manufactured by industry. Although the performance of some of these algorithms is good in relatively sparse road networks, they are not always reliable for complex roundabouts, merging or diverging sections of motorways, and complex urban road networks (Velaga et al., 2009). Hence, Quddus et al. (2006) proposed a map matching algorithm based on fuzzy logic theory and tested it on different road networks of varying complexity. It was found that the fuzzy logic-based map matching algorithm could provide a significant improvement over existing map matching algorithms both in terms of identifying correct links and estimating the vehicle position on the links. Velaga et al. (2009) developed an enhanced weight-based topological MM algorithm in which the weights are determined from real-world field data using an optimization technique to improve the performance of matching. Smaili et al. (2014) proposed an approach based on hybrid dynamic Bayesian networks enabling to implement in a unified framework two of the most successful families of probabilistic model commonly used for localization: linear Kalman filters and Hidden Markov Models. The combination of these two models enabled to manage and manipulate multi-hypotheses and multi-modality of observations characterizing MM problems and improve integrity approach.