Computers, Environment and Urban Systems 67 (2018) 21–28

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

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

TripCube: A Trip-oriented vehicle trajectory data indexing structure

Tao Xu [a,b], Xihui Zhang [c], Christophe Claramunt [d], Xiang Li [a,*]

[a] Key Laboratory of Geographical Information Science, Ministry of Education, East China Normal University, 500 Dongchuan Road, Shanghai 200241, China
[b] Key Laboratory of Analysis and Processing on Big Data of Henan Province, Henan University, 85 Minglun Road, Kaifeng, Henan 475001, China
[c] Department of Computer Science & Information Systems, College of Business University of North Alabama, Florence, Alabama, United States
[d] Naval Academy Research Institute, Brest, France

A R T I C L E I N F O

Article history:
Received 19 November 2016
Received in revised form 10 August 2017
Accepted 12 August 2017
Available online xxxx

Keywords:
Spatio-temporal data management
Indexing structure
Vehicle trip
Vehicle trajectory data

A B S T R A C T

With the dramatic development of location-based services, a large amount of vehicle trajectory data are available and applied to different areas, while there are still many research challenges left, one of them being data access issues. Most of existing tree-shape indexing schemes cannot facilitate maintenance and management of very large vehicle trajectory data. How to retrieve vehicle trajectory information efficiently requires more efforts. Accordingly, this paper presents a trip-oriented data indexing scheme, named *TripCube*, for massive vehicle trajectory data. Its principle is to represent vehicle trajectory data as trip information records and develop a three-dimensional cube-shape indexing structure to achieve trip-oriented trajectory data retrieval. In particular, the approach is implemented and applied to vehicle trajectory data in the city of Shanghai including >100 million locational records per day collected from about 13,000 taxis. *TripCube* is compared to two existing trajectory data indexing structures in our experiments, and the result exhibits that *TripCube* outperforms others.

1. Introduction

With the dramatic development of location-based services, vast amount of vehicle trajectory data can be easily gathered with GPS receivers equipped on vehicles. The abundance of trajectory data presents a valuable opportunity for scholars to discover previously unknown but potentially valuable information about vehicle movements and traffic situations, such as developing trajectory data mining methods (Dodge, Weibel, and Forootan, 2009; Izakian, Mesgari, and Abraham, 2016; Liu and Karimi, 2006; Pfoser and Theodoridis, 2003; Zhou et al., 2015), inferring residents travel characteristics and patterns (Hu, Miller, and Li, 2014; Kang, Liu, and Wu, 2015; Liu, Wang, Xiao, and Gao, 2012; Torrens et al., 2012), discovering spatio-temporal features of traffic flow (Ge et al., 2010; Liu and Ban, 2012; Wang, Wang, Song, and Raghavan, 2017; Wei, Zheng, and Peng, 2012; Zheng, Liu, Yuan, and Xie, 2011), and predicting travel time (Chen and Rakha, 2014; Jiang and Li, 2013). In the meantime, such data analysis avenues bring novel challenges, a crucial one being the development of appropriate solutions for the most efficient management of vehicle trajectory data (Jiang and Li, 2013; Kwan, 2016). Efficient data structures and algorithms need to be tailored for vehicle trajectory data (Jiang and Yao, 2006; Katal, Wazid, and Goudar, 2013).

As the moving of vehicles is usually constrained by road network, the spatial distribution of vehicle trajectory data is linear along with road segments. With predefined spatio-temporal granularity, original trajectory data can be divided into trajectory segments. Many indexing structures based on trajectory segments have been proposed. Most of them employ R-tree-based indexing structure and set some spatial attributes (e.g., longitude and latitude) as keywords to index vehicle trajectory points or segments. The principle of R-tree (Guttman, 1984) is to group nearby objects based on their spatial locations and to represent them with their Minimum Bounding Rectangles (MBRs) stored in tree nodes. However, in urban road network, long-term and massive vehicle trajectory data must generate a great number of overlapping or redundant MBRs and the corresponding indexing structure must be a bloated multilevel R-tree. Such a R-tree indexing structure is difficult to maintain, which dramatically increases operational cost and reduces query efficiency.

Different from tree-shape indexing structure, the size of cube-shape indexing structure is controllable by predefining spatio-temporal dimensions, whereas the depth and breadth of tree-shape indexing structures extend continuously. Most of cube-shape indexing approaches divide trajectory data into segments by fixed spatial granularity or fixed distance, and aggregate trajectory segments with the same spatio-temporal features into cells of cube. Trajectory segment retrieval can be made from the cube with given query conditions. However, such trajectory segmentation with fixed granularity is only beneficial to trajectory retrieval under given granularity. Moreover, it splits the semantic integrity of vehicle trajectory that brings obstacle to trajectory retrieval by given origin and destination.

In this paper, we attempt to explore a flexible indexing scheme for vehicle trajectory data with cube structure. Generally, a vehicle trajectory indicates vehicle driving process and can be divided by meaningful origin-destination pairs (e.g., the pick-up and drop-off points of taxicab, the entrances and exits of express road system, etc.) to represent vehicle mobility. Therefore, instead of fixed spatial granularity, we define vehicle

* Corresponding author.
E-mail addresses: txucn@hotmail.com (T. Xu), xli@geo.ecnu.edu.cn (X. Li).

http://dx.doi.org/10.1016/j.compenvurbsys.2017.08.005

trip to represent a travel case with an origin-destination pair and to make flexible trajectory segmentation. Moreover, we design a trip-oriented cube structure with three dimensions, namely origin, destination, and departure time, to index vehicle trajectory data and achieve efficient trip information retrieval.

Accordingly, this research introduces a vehicle trajectory data indexing scheme for trip information retrieval. A trip-oriented vehicle trajectory data indexing structure, named *TripCube*, is developed and implemented. *TripCube* consists of a three dimensional index cube and a set of trip information record. It is designed to evaluate queries on trip information between any pair of origin and destination locations at any time. The approach is applied to a large taxi trajectory data set in the city of Shanghai, and performance tables and figures as well as comparison to two existing methods are presented. Our contributions can be summarised as follows:

- *TripCube* is developed to address the challenge of organizing and indexing massive vehicle trajectory data.
- Different from most existing tree-shape indexing structures, *TripCube* is based on a cube-shape indexing structure and allocates indexing storage space before inserting new entries in order to facilitate maintaining complicated indexing structure when data volume dramatically increases.
- *TripCube* is especially applicable to long-term and massive vehicle trajectory data confined to a given road network.
- Compared with two existing trajectory-segment-based indexing structures, *TripCube* exhibits more excellent query efficiency.

The rest of this paper is organized as follows. The next section gives a detailed review of related works. *TripCube* is developed in Section 3 and is validated in Section 4 through a series of experiments. The last section concludes the paper.

2. Related works

In past few years, many research works for vehicle trajectory data indexing have been made. R-tree (Guttman, 1984) is the most common one. R-Tree and its variants, e.g., R + tree (Sellis, Roussopoulos, and Faloutsos, 1987), R*-tree (Beckmann, 1990), X-tree (Berchtold, Keim, and Kriegel, 1996), RT-tree (Xu, Han, and Lu, 1990), HR-tree (Nanopoulos, Theodoridis, and Manolopoulos, 2006), B dual-Tree (Yiu, Tao, and Mamoulis, 2008),R k-d Tree (Anandhakumar, Priyadarshini, Monisha, Sugirtha, and Raghavan, 2010),Vor-Tree (Sharifzadeh and Shahabi, 2010),HTPR*-Tree (Fang, Cao, Wang, Peng, and Song, 2012),and HBSTR-Tree (Ke et al., 2014), are often retained as the indexing structure of vehicle trajectory data. They use the Minimum Bounding Rectangle (MBR) to cluster spatial objects and create the height-balanced tree to index spatial objects. The spatial range retrieval based on MBR can be made with logarithmic query performance. However, maintaining such indexing structure is considerably complex, and insertion, update, and deletion operations, which might change the correlation and the volume of tree nodes, are time-consuming compared to data retrieval queries (Guttman, 1984). Although vehicle trajectory data are usually managed as a sort of "append-only record set" with only insertion operations but no update or deletion operations, the rapid growth of data volume still leads to a significant challenge for tree index maintenance, which not only expands the index space, but also lengthens the index query execution time.

As such, some existing solutions about trajectory segmentation have already tried to deal with the data volume challenges of vehicle trajectory data to some extent. Brakatsoulas, Pfoser, and Tryfona (2004) proposes a trajectory data management scheme to model, store, and mine moving object database. In this schema, vehicle trajectory is divided into segments corresponding to the edge of road network. Leonardi et al., (2014) deigns a spatio-temporal hierarchies framework, and decomposes a complete trajectory into trajectory segments with respect to given distance and time. In addition, trajectory segment can be delineated according to spatial region and time (Leonardi, Marketos, Frentzos, and Giatrakos, 2010; Masciari, 2012; Masciari, 2015; Pelekis and Theodoridis, 2014; Surya Prakash, 2014). With predefined spatio-temporal granularities, these approaches split trajectory data as trajectory segments with fixed distance or fixed range and index trajectory segments with tree-shape indexing structure (B-tree, or R-tree). This can reduce the complexity of tree indexing structure to some extent and facilitate trajectory segment retrieval under the defined spatio-temporal granularity. However, the fixed trajectory segmentation may break the semantic integrity of vehicle trajectory and cause data errors and losses inevitably (Masciari, 2015). The challenges caused by the large vehicle trajectory data have not been solved yet.

As the description of a travel case, vehicle trip from an origin to a destination contains complete semantic information. Using vehicle trip information from vehicle trajectory data, many research works have been made to predict travel time (Jiang and Li, 2013; Xu, Li, and Claramunt, 2017; Xu, Xu, Hu, and Li, 2017), analyze driving behavior (Ren, Tao, and Xiang, 2014), and reveal traffic patterns (Dai, Yang, Guo, and Ding, 2015; Izakian et al., 2016; Liu and Ban, 2012), etc. How to retrieval vehicle trip information efficiently has become a critical issue. Therefore, it is very valuable to trajectory segmentation based on vehicle trip for the management and application of vehicle trajectory data.

Since vehicle trip can be easily identified by origin, destination, and departure time, a controllable indexing space can be designed with spatial features (origin and destination points) and temporal features (departure time). A cube structure can directly map the relationship of spatio-temporal data, and some research results (Cao et al., 2015; Lins, Klosowski, and Scheidegger, 2013; Pelekis and Theodoridis, 2014; Surya Prakash, 2014) about cube structures for the management of spatio-temporal data have been presented. Based on them, we attempt to develop a novel trip-oriented trajectory data indexing scheme, named *TripCube*, to support the maintenance and retrieval of vehicle trajectory data. The *TripCube* consists of a three-dimensional indexing cube and a set of vehicle trip information records. Vehicle trajectory data are organized as vehicle trip information records and indexed by the cube structure on its attributes (i.e., origin, destination, and departure time).

3. Methodology

3.1. Principles

TripCube is a two-level indexing structure, i.e., a trip level and a cube level. At the trip level, we define trip information record as *vts* to represent trip information from raw vehicle trajectory data, where *vts* is a composite structure that includes vehicle ID, origin, destination, departure time, travel route, travel time, and vehicle trajectory, etc. At the cube level, a three dimensional index cube, consisting of one origin dimension, one destination dimension, and one time dimension, is generated to manage *vts* with specific origin, destination, and departure time. The origin and destination of a trip represent the travel starting and ending positions, which can be determined by application purposes, e.g., the pick-up/drop-off points of taxicab, the entrances and exits of express road system, and commuter destinations, etc. *TripCube* converts vehicle trajectory data to *vts* and indexes it by origin, destination, and departure time. By this means, *TripCube* is not a typical spatial indexing structure. Instead of performing a spatial search for vehicle trajectory points, *TripCube* retrieves the set of vehicle trips with explicitly encoded spatial and temporal query conditions (e.g., origin, destination, and departure time). The structure of *TripCube*, its initialization procedure, and retrieval algorithms are presented in the remaining part of this section.

3.2. Vehicle trip structure

To extract trip information from raw vehicle trajectory data, we define *vehicle trip structure* to represent trip information and corresponding locational sample points.