ARTICLE IN PRESS

Computers, Environment and Urban Systems xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

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

Mining massive taxi trajectories for rapid fastest path planning in dynamic multi-level landmark network

Dongbao Zhao[a,*], Emmanuel Stefanakis[b]

[a] *School of Resources and Environment, North China University of Water Resources and Electric Power, Zhengzhou, Henan, China*
[b] *Geodesy and Geomatics Engineering, University of New Brunswick, Fredericton, NB, Canada*

ARTICLE INFO

Keywords:
Path planning
Trajectory data mining
Landmark network
Experiential trajectory

ABSTRACT

This paper presents a speed up algorithm for real-time path planning based on massive trajectory data mining, which comprises three stages: preparatory work, preprocessing stage and the online fastest path query. At the preparatory stage, this algorithm constructs multi-level landmarks and divides the original road network into multiple levels accordingly. At the preprocessing stage, this algorithm first estimates the travel time of all road segments according to the real-time traffic information, then compares all taxi trajectories to extract the experiential fastest paths, and finally makes use of the multi-level landmarks to obtain the rough fastest paths for all landmark pairs. At the online fastest path query stage, the server side first responds by returning a rough fastest path based on the preprocessing result, and then refines it by iteratively introducing the experiential fastest paths. A series of experiments are made to compare the proposed algorithm with the other three algorithms. Experiments indicate that the proposed algorithm has the ability to find more time-saving paths in response to client requests. More importantly, because this algorithm is capable of ensuring the fast completion of pre-computation on the server side, it has an evident advantage in the time cost of the online fastest path query compared with the other three algorithms, which is particularly suitable for the online optimal path query from a larger number of end users.

1. Introduction

Nowadays, using map navigation services to find efficient driving directions has become people's daily activity when traveling outdoors. As the core component, path planning algorithms play a significant role in various map navigation applications. In recent years, a large number of vehicles traversing on road networks of big cities have already been embedded with GPS sensors that enable them to report their location to a data center in a certain frequency. The massive time-stamped GPS trajectory data provide a novel strategy for the solution of the classical path planning optimization problem. Some path planning algorithms based on vehicle trajectory data analysis and mining can be found in the literatures. For instance, Gonzalez, Han, Li, Myslinska, and Sondag (2007), Tang, Chang, and Li (2010), Q Li, Zeng, Zhang, and Li (2011), and Hu, Huang, Deng, and Xie (2013) all first divide the road network into different levels based on the computation of the speed and vehicle access frequency of road segments, and then still use the traditional Label-Setting or Label-Correcting algorithms in the hierarchical road network to find the shortest path. Although these methods take into account the traffic condition during several fixed time periods (e.g. a day is split into four time periods), but the time spans of these fix time periods are too large to reflect the real-time traffic information accurately. In essence, they are designed to find the optimal path in a static road network during these fixed time periods, instead of in a time-dependent dynamic road network, where ‘time-dependent’ means that the travel time of road segments is always changing dynamically.

Compared with finding the shortest distance path, finding the fastest path in a time-dependent dynamic road network is undoubtedly more challenging though more adapted to the actual needs. With regard to finding the fastest paths based on the analysis on massive taxi trajectory data, Zheng, F, Li, and Duan (2010) point out and compare two main ideas. One is to continually compute and update the current fastest path according to the recent time interval (e.g., 5 min) of traffic conditions, and the other is to find the fastest path in a pre-defined time-dependent road network according to an analysis on long-term historical traffic condition. Yuan, Zheng, Xie, and Sun (2013) propose a representative fastest path planning algorithm under the time-dependent road network, which first makes use of massive taxi trajectory data to build a time-dependent landmark graph that satisfies First In First Out property (FIFO, proposed by Kaufman, Lee, and Smith (1993)) to model the

---

[*] Corresponding author.
E-mail address: zdongbao@outlook.com (D. Zhao).

https://doi.org/10.1016/j.compenvurbsys.2018.07.001
Received 2 January 2018; Received in revised form 30 April 2018; Accepted 2 July 2018

Please cite this article as: Dongbao, Z., Computers, Environment and Urban Systems (2018), https://doi.org/10.1016/j.compenvurbsys.2018.07.001

dynamic traffic pattern as well as the intelligence of experienced drivers, and then use the improved Label-Setting algorithm to compute the practical fastest and customized path for end users. Recently, Yang, Pan, and BoWan (2017) first present the space-time trajectory cube as a framework for dividing and organizing the trajectory space, and then generate space-time constrained graph by merging large-scale taxi drivers' experience with the road network, and finally compute the optimal paths in the space-time constrained graph.

Although the algorithms above can provide navigation services according to real-time traffic information, finding the fastest path is more complicated and time-consuming than finding the shortest path because the path planning in a stochastic time-dependent dynamic network is proved by Ahuja, Orlin, Pallottino, and Scutella (2003) to be an NP hard problem. The huge time overhead of finding the fastest path is unacceptable when applied over a large-scale dynamic road network, and the problem becomes more serious if further considering online concurrent navigation path queries from a large number of end users. In order to improve online query performance of the fastest path, one way to work around this problem is to pre-compute the fastest paths for all point pairs on the server side during preprocessing phase in exchange for high efficiency of online path planning. However, for real-time online path planning, the factors of preprocessing time cost, space storage cost and online query time cost must be synthetically considered. This is because if highly time-consuming preprocessing is needed to be performed again from scratch every several minutes, preprocessing stage will definitely drag down the efficiency of online optimal path query. In view of the existing problems, this study attempts to propose an efficient real time path planning algorithm based on massive taxi trajectory data analysis. The characteristics of this algorithm are as follows:

(1) It continually computes the current fastest path according to the most recent snapshot of traffic information in the dynamic road network.
(2) Through building multi-level landmark network, this algorithm achieves the rapid calculations of the fastest paths of all landmark pairs on the server side every five minutes, thereby greatly improving the online query efficiency of the optimal paths.
(3) It takes advantage of massive taxi experiential trajectories and introduces the experiential fastest paths into the final path planning results so as to better learn from taxi driver's experience knowledge.

2. Related work

Firstly, there are also some other path planning methods based on trajectory data mining. For instance, Wei, Zheng, and Peng (2012) and Zhu, Jiang, Liu, Liu, and Zhao (2015) all attempt to mining the top-k most popular paths for a given source, destination and time-span, while Delling, Goldberg, Goldszmidt, Krumm, Talwar and Werneck (2015) and Ceikute and Jensen (2015) are devoted to finding the personalized path based on the mining of driver's preferences for driving paths. Although these methods can all offer significant references for map navigation service, their optimization goals are not intended to find the general-purpose shortest or fastest path for any road node pairs.

Secondly, In order to improve the online shortest path query efficiency greatly, some scholars have proposed several famous speedup techniques for the shortest path planning, such as SILC method proposed by Samet, Sankaranarayanan, and Alborzi (2008), Arc-flag method proposed by Köhler, Möhring, and Schilling (2006),CH method proposed by Geisberger, Sanders, Schultes, and Delling (2008), and TNR method proposed by Bast, Funke, Matijevic, Sanders, and Schultes (2007). These methods are able to attain efficient online query at the expense of expensive preprocessing time overhead, but they are not suitable for real-time path planning because their preprocessing procedures are designed for large-scale static road network rather than large-scale dynamic road network. Although Efentakis and Pfoser (2013), Delling, Goldberg, Pajor, and Werneck (2015) all conduct important research on the pathfinding acceleration technology in the dynamic road network, these methods do not involve the research on massive trajectory data mining and analysis.

Finally, trajectory data can also be employed to estimate travel time of road segments and predicate the future change of traffic flow, which is also of vital importance to the real-time path planning. Wang, Zheng, and Xue (2014b) point out that traffic delays brought out by traffic lights or congestion will influence the accurate determination of the travel time of road segments. Jenelius and Koutsopoulos (2013) present a statistical model for urban road network travel time estimation using vehicle trajectories obtained from low frequency GPS probes as observations. Wang,Zheng, and Xue (2014b) propose a citywide and real-time model for estimating the travel time of any path by using the vehicle trajectories received in current time slots and over a period of history as well as map data sources. Isaenko, Colombaroni, and Fusco (2017) address traffic dynamics estimation by using floating car data in order to develop an integrative framework to recognize and select the suitable method for traffic forecasting. Although these methods cannot be directly used to solve the problem of real-time path planning, they play an important role in determining some important parameters of path planning.

3. Framework

In the urban road network, some higher level roads are often visited frequently, carrying a large amount of traffic flow. These important road segments or road nodes constitute landmarks. People often prefer to use a sequence of landmarks to highlight key directions to destination according to people's natural thinking habit. Therefore, this study constructs multi-level landmarks to achieve the real time path planning process. As shown in Fig.1, the idea of the proposed algorithm is to continually compute the current fastest path according to the recent traffic flow to provide real-time navigation service. The entire framework comprises three stages: preparatory work, preprocessing stage, and the online fastest path query. They are deployed offline or online respectively for balancing the computational load and increasing efficiency of online optimal path query as much as possible.

The task of the preparatory work is to generate multi-level landmarks based on analysis on long term of trajectory data and divide the original road network into a multi-level road network accordingly. This stage is performed only once on the server side.

The preprocessing stage is to use the speedup technique to precompute the fastest paths for all landmark pairs on the server side every five minutes. It comprises three aspects of work. Firstly, massive trajectory data is used to estimate the travel time of road segments within the current five minutes. Secondly, massive taxi trajectories of the latest one hour are compared to obtain a large number of experiential fastest paths. Thirdly, the Floyd algorithm is performed respectively on the different levels of the road network to connect the experiential fastest paths to obtain the rough fastest paths for all landmark pairs.

As for the online fastest path query, a large number of end users submit query requests for the fastest paths. For each query request, the server side first responds by returning a rough fastest path, and then refines the rough fastest path by introducing the experiential fastest paths into it iteratively, and finally transmits the road nodes sequence to the client side, which is converted to actual navigation path on the client side eventually.

4. Construction of multi-level road network

This section realizes the construction of multi-level road network through building multi-level landmarks.