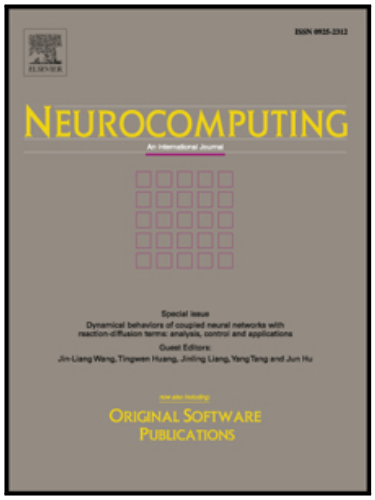
Communicated by Weiming Xiang

Accepted Manuscript



Efficient Traffic Congestion Estimation using Multiple Spatio-temporal Properties

Yongjian Yang, Yuanbo Xu, Jiayu Han, En Wang, Weitong Chen, Lin Yue

PII: S0925-2312(17)31072-X
DOI: 10.1016/j.neucom.2017.06.017
Reference: NEUCOM 18573

To appear in: *Neurocomputing*

Received date: 16 November 2016
Revised date: 3 June 2017
Accepted date: 8 June 2017

Please cite this article as: Yongjian Yang, Yuanbo Xu, Jiayu Han, En Wang, Weitong Chen, Lin Yue, Efficient Traffic Congestion Estimation using Multiple Spatio-temporal Properties, *Neurocomputing* (2017), doi: 10.1016/j.neucom.2017.06.017

Efficient Traffic Congestion Estimation using Multiple Spatio-temporal Properties

Yongjian Yang[1], Yuanbo Xu[1], Jiayu Han[1], En Wang[1,*], Weitong Chen[2], Lin Yue[3]

[1]School of Computer Science and Technology, Jilin University, Changchun 130012, China
[2]School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane, Australia
[3]School of Computer Science and Information Technology, Northeast Normal University, Changchun 130117, China

***Abstract*—Traffic estimation is an important issue to analyze the traffic congestion in large-scale urban traffic situations. Recently, many researchers have used GPS data to estimate traffic congestion. However, how to fuse the multiple data reasonably and guarantee the accuracy and efficiency of these methods are still challenging problems. In this paper, we propose a novel method Multiple Data Estimation (MDE) to estimate the congestion status in urban environment with GPS trajectory data efficiently, where we estimate the congestion status of the area through utilizing multiple properties, including density, velocity, inflow and previous status. Among them, traffic inflow and previous status (combination of time and space factors) are not both used in other existing methods. In order to ensure the accuracy and efficiency, we apply dynamic weights of data and parameters in MDE method. To evaluate our methods, we apply it on large-scale taxi GPS data of Beijing and Shanghai. Extensive experiments on these two real-world datasets demonstrate the significant improvements of our method over several state-of-the-art methods.**

Keywords:Traffic congestion estimation, Large-scale road networks, Multiple spatio-temporal properties, Dynamic weight calculation, GPS data.

I. Introduction

Currently, advances in terms of Wireless Sensor Networks (WSN) enrich the variety of human mobility information [1, 2]. However, how to use various kinds of massive data (e.g. taxicab GPS trajectory data, road segments, etc.) to solve the real problems is still challenging. Urban traffic congestion is a kind of the real problems, and it has become a critical problem because it not only affects the inhabitants' daily life, but also damages the social and economic development of the city [1]. Nevertheless, urban traffic situation is complex and constantly changing with time and space. The information of traffic usually includes multiple spatio-temporal data, so it is quite difficult for the inhabitants to obtain the current and future traffic condition at a certain road section in time. Estimating the status of urban traffic congestion effectively is the first step to solve the urban traffic congestion problem.

There are three major challenging problems in urban traffic **congestion estimation (CE)**. Firstly, which kinds of data in massive and various information should be analyzed in large-scale urban sections for CE? The sensors equipped in taxicabs, which can collect the Global Position System (GPS) data, are effective facilities [3]. They can offer traffic researchers with massive and detailed real-time traffic data in large-scale urban sections. Ubiquitous mobile sensors can probe a city's rhythm and pulse [4]. GPS data can provide detailed taxicab's information of traffic flow such as longitude, latitude, directions, speed, etc. Using GPS data of taxicabs, we can deduce the whole city's traffic situations because 1) the data can be collected easily and 2) the taxicab has the ability to spread over all road networks in the city than other public traffic [1][5] (shown in Fig.1) (note that every red point is a taxt GPS point). Secondly, because sensors collect the information of traffic periodically and frequently, it is a common view that we have an information explosion. For example, the data collected by 1000 taxicabs in Beijing one day is nearly 10GB text-file. It contains huge information about the congestion in the historical trajectory data. However, tackling all historical data is expensive and meaningless for congestion estimation. Therefore, we should choose the most useful and significant properties in traffic flow data to estimate congestion. Finally, how to improve the accuracy and efficiency of traffic congestion estimation? The existing methods for CE have not addressed the three problems systematically.

This paper proposes a **Multiple Data Estimation (MDE)** method, which estimates congestions in urban sections using traffic flow data (GPS data collected by taxicabs) and road segments. The existing methods for CE usually use one or two properties, while MDE uses traffic density, average velocity, inflow, and previous status as spatio-temporal properties to evaluate a section's congestion situation. These properties are multiple for traffic congestion and difficult to calculate directly. In addition, the weights of properties in MDE are varying in consonance with time and space according to the traffic networks. We use dynamical determination for weights of properties, which can ensure the accuracy and efficiency in

*Corresponding author