Transportation Research Part C 71 (2016) 284–302

Contents lists available at ScienceDirect

Transportation Research Part C

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

Spatial-temporal traffic flow pattern identification and anomaly detection with dictionary-based compression theory in a large-scale urban network

Zhenhua Zhang [a], Qing He [a,b,*], Hanghang Tong [c], Jizhan Gou [d], Xiaoling Li [e]

[a] Department of Civil, Structural and Environmental Engineering, The State University of New York, Buffalo, NY 14260, United States
[b] Department of Industrial and Systems Engineering, The State University of New York, Buffalo, NY 14260, United States
[c] School of Computing, Informatics, and Decision Systems Engineering, Arizona State University, Tempe, AZ 85281, United States
[d] Serco Inc., Reston, VA 20190, United States
[e] Northern Region Operations, Virginia Department of Transportation, Fairfax, VA 22030, United States

ARTICLE INFO

Article history:
Received 20 November 2015
Received in revised form 29 April 2016
Accepted 11 August 2016

Keywords:
Traffic anomaly detection
Traffic flow pattern identification
Dictionary-based compression theory

ABSTRACT

Traffic flow pattern identification, as well as anomaly detection, is an important component for traffic operations and control. To reveal the characteristics of regional traffic flow patterns in large road networks, this paper employs dictionary-based compression theory to identify the features of both spatial and temporal patterns by analyzing the multi-dimensional traffic-related data. An anomaly index is derived to quantify the network traffic in both spatial and temporal perspectives. Both pattern identifications are conducted in three different geographic levels: detector, intersection, and sub-region. From different geographic levels, this study finds several important features of traffic flow patterns, including the geographic distribution of traffic flow patterns, pattern shifts at different times-of-day, pattern fluctuations over different days, etc. Both spatial and temporal traffic flow patterns defined in this study can jointly characterize pattern changes and provide a good performance measure of traffic operations and management. The proposed method is further implemented in a case study for the impact of a newly constructed subway line. The before-and-after study identifies the major changes of surrounding road traffic near the subway stations. It is found that new metro stations attract more commute traffic in weekdays as well as entertaining traffic during weekends.

1. Introduction

Studies on traffic flow patterns within a scale of road facilities have aroused increasing attentions in recent years. The traffic flow patterns can be taken as those characteristics of vehicle groups passing a point or a short segment during a specified span or traveling over longer sections of highway (Lan et al., 2008). The spatiotemporal features of traffic flow, occupancy, and speed over different time scales can provide insights into traffic operation and control, urban planning, incident management, etc. The time-of-day and day-of-week features of traffic flow or occupancy reveal the fluctuations

* Corresponding author at: Department of Civil, Structural and Environmental Engineering and Department of Industrial and Systems Engineering, The State University of New York, Buffalo, NY 14260, United States.

E-mail addresses: zhenhuaz@buffalo.edu (Z. Zhang), qinghe@buffalo.edu (Q. He), Htong6@asu.edu (H. Tong), Jizhan.Gou@serco-na.com (J. Gou), Ling.li@vdot.virginia.gov (X. Li).

http://dx.doi.org/10.1016/j.trc.2016.08.006
0968-090X/

of the jammed conditions and traffic operations of road links, intersections or even networks and can serve different research purposes. For instance, White (2007) focused on the impact of the daily visitor transportation on the public infrastructures and suggested approaches to improve the environmental sustainability of national parks; Ramaswami and Sivarajan (1996) studied the network traffic patterns to design the logical topology and the routing algorithm so as to minimize the network congestion; Cassidy and Bertini (1999) investigated the characteristics of freeway traffic flow patterns and their findings have practical implications for freeway traffic planning and management; Lee et al. (2014) proposed a method to identify the congestion patterns from the big traffic data and predicted when the road congestion events will dissipate. Of all these studies, despite their different emphasis and purposes, the traffic pattern identification usually performs as an initial step in the entire research process in transportation management, and there is still research gaps to explore further.

First, the measures for traffic flow patterns are not limited to traffic flow measurement. The metrics to describe the patterns usually vary from traffic flow (Cassidy and Bertini, 1999; Shen and Zhang, 2009; Zhang et al., 2001), density (Cassidy and Mauch, 2001; Treiber and Kesting, 2012), speed (Banaei-Kashani et al., 2011), etc. This is mainly due the feature of each metric being unique and the researchers usually have different purposes that their problems intentionally shape the definition of the traffic flow patterns. Besides traffic-related metrics, other information such as time and link locations may further enrich and clarify the connotations of pattern identification. For example, studies show that given the combination of direction, connectivity and locality of a road segment (and without having access to the actual traffic flow in the segment), one can distinctively determine the corresponding traffic signature (pattern) of a road segment with high probability (Banaei-Kashani et al., 2011). The features of the traffic flow patterns should be further explored combining both traffic, time and location information.

Second, traditional studies focused mainly on traffic flow patterns within intersections (Teodorovic et al., 2001) or corridors (Lan et al., 2008; Schoenhof and Helbing, 2007). These may not be sufficient to explain the pattern changes in a large-scale road network with hundreds of intersections. In recent years, the size of traffic data booms both in time and space, and the large-scale floating and fixed sensors are extensively utilized to collect the traffic data. The studies of pattern identification can expand from links to a district or even to a county. This expansion brings more difficulties and has aroused increasing attentions. There have been efforts in reducing the dimensionality of the traffic flow pattern problems, such as functional data analysis (Guardiola et al., 2014), principle component analysis (Jin et al., 2008). How to archive and summarize massive historical data effectively and extract meaningful traffic flow patterns from accumulated data to support decision making has become a significant challenge, considering the huge size of the dataset (Xu et al., 2013).

Third, traditional studies of traffic patterns drew their conclusions from different spatial and temporal perspectives including the day-to-day features (Hu and Mahmassani, 1997) or time-of-day features (Anbaroglu et al., 2014). The temporal traffic patterns obtained from traffic time series studies can reveal the pattern features in a different perspective (Li et al., 2015). Of all these studies, there is a pressing need to develop a systematic approach for traffic flow pattern identification and anomaly detection that consider both different spatial and temporal perspectives. The temporal features of the traffic flow pattern, e.g. at an intersection, should be unveiled by comparing with its surrounding traffic as well as its historical ones. Instead of combining "time" as a traffic-related variable, we proceed with a temporal perspective to study the changes of traffic flow patterns over time.

To address the above research gaps, we propose a method based on the compression theory in regional traffic flow pattern identification. Compression-based approaches have been successfully implemented in pattern recognition and anomaly detection in different domains, such as image processing (Akoglu et al., 2012) and system query processing (Binnig et al., 2009). Rather than trying to compress the set of frequent items, compression-based approaches compress the database and search for the subset of all frequent item sets that compresses the database best (Siebes et al., 2006). The method can capture the best regularities of the data with as little redundancy as possible and thus can capture the most important patterns in the datasets (Tatti and Vreeken, 2008). Previous research proves that this approach provides accuracy on par with the state-of-the-art in anomaly detection (Smets and Vreeken, 2011). Some even argue that this approach is competitive or superior to many of the state-of-the-art approaches in anomaly and interestingness detection, classification, and clustering with empirical tests on time series, DNA, text, XML, and video datasets (Keogh et al., 2007). Similar to the data fusion which can maximize the utility of the available (traffic) information (Treiber and Kesting, 2013), the compression method is capable of recognizing the frequent traffic flow patterns through effective interpretations of multi-dimensional data. Based on the compression theory and corresponding algorithm, we will be able to quantify the traffic performance in a given location and detect the abnormal traffic flow patterns from our studied network.

Through our study, we unveil the important features of traffic flow patterns separately from spatial and temporal perspectives. From the spatial perspective, the problem is how to measure and evaluate the anomaly degree of a road link or intersection approach as compared to others. This relates to the identification of recurrent traffic flow patterns and provides insights in how to constantly identify problematic locations within a certain scale of the road network. From the temporal perspective, the problem is how to measure and evaluate the abnormal time-of-day traffic flow patterns as compared to historical records. This relates to the non-recurrent traffic flow patterns and helps traffic operators learn why traffic goes to the extreme in one day whereas those in other days are relatively normal. What's more, both spatial and temporal patterns can be interpreted in different geographic and time scales. We include three different geographic scales: sub-region level, intersection level, and detector (or lane) level, as illustrated in Fig. 1. The traffic flow patterns in different levels of geography are further studied to explore the hourly and daily features of the traffic flow patterns.