Information Systems 54 (2015) 169–188

Contents lists available at ScienceDirect

Information Systems

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

# Real-time traffic incident detection using a probabilistic topic model

Akira Kinoshita [a,b,*], Atsuhiro Takasu [b], Jun Adachi [b]

[a] The University of Tokyo, 7-3-1 Hongo, Bunkyo, Tokyo, Japan
[b] National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda, Tokyo, Japan

ARTICLE INFO

*Article history:*
Received 15 December 2014
Received in revised form
18 May 2015
Accepted 1 July 2015
Available online 10 July 2015

*Keywords:*
Anomaly detection
Automatic incident detection
Probabilistic topic model
Probe-car data
Real-time processing
Traffic state estimation

ABSTRACT

Traffic congestion occurs frequently in urban settings, and is not always caused by traffic incidents. In this paper, we propose a simple method for detecting traffic incidents from probe-car data by identifying unusual events that distinguish incidents from spontaneous congestion. First, we introduce a traffic state model based on a probabilistic topic model to describe the traffic states for a variety of roads. Formulas for estimating the model parameters are derived, so that the model of usual traffic can be learned using an expectation–maximization algorithm. Next, we propose several divergence functions to evaluate differences between the current and usual traffic states and streaming algorithms that detect high-divergence segments in real time. We conducted an experiment with data collected for the entire Shuto Expressway system in Tokyo during 2010 and 2011. The results showed that our method discriminates successfully between anomalous car trajectories and the more usual, slowly moving traffic patterns.


## 1. Introduction

Automatic incident detection (AID) is a crucial technology in intelligent transport systems, particularly in terms of reducing congestion on freeways [1]. Traffic incidents often cause traffic congestion, causing great inconvenience and economic loss to society. A technology that can detect traffic incidents in real time and alert people accordingly would therefore be a desirable way of reducing these ill effects.

Against this background, there have been many studies on AID, e.g., [2,3]. Most of the approaches exploit data sent from stationary sensors and cameras installed on roads. However, the installation and maintenance of such sensors is expensive, with only the main routes likely to have them [4]. On the other hand, the use of probe-car data (PCD), on which we focus in this paper, is becoming increasingly important, as the number of probe cars and the size of the associated data archives increase. PCD includes time-stamps and vehicle locations, and may contain additional data such as the speed and direction of the probe cars. Although a PCD system cannot monitor all cars, it enables traffic administrators to watch a large area at a lower cost than by using stationary sensors. In addition, a PCD system can follow the sequence of movements for a probe car in detail, which is hard to achieve via stationary sensors, and trajectory mining can be applied to the collected data.

Using PCD for freeways, it is easy to detect any reduction in speed, which sometimes implies congestion, by analyzing the speeds of the probe cars. However, this method is less applicable to local streets, where the many crossings and traffic lights can cause cars to stop frequently under normal circumstances. Moreover, speed reduction is not always abnormal, even on freeways, and is not always caused by incidents such as accidents, which we would regard as sudden and unusual traffic events in this paper.

There are two types of congestion: spontaneous and abnormal [2]. Detecting spontaneous congestion is less

* Corresponding author.
*E-mail addresses:* kinoshita@nii.ac.jp (A. Kinoshita), takasu@nii.ac.jp (A. Takasu), adachi@nii.ac.jp (J. Adachi).

http://dx.doi.org/10.1016/j.is.2015.07.002
0306-4379/

important, as it originates in road design and urban planning. Any road may have potential bottlenecks such as upslopes, curves, junctions, tollgates, and narrow sections. Vehicles are likely to slow down at the bottlenecks, with vehicular gaps shortening and drivers in the following cars having to brake. Congestion will then occur even in the absence of a specific traffic incident [5]. Spontaneous congestion also occurs when the traffic demand exceeds the traffic capacity at such bottlenecks, and it is not resolved until the demand drops below the capacity [6]. Drivers may be familiar with the locations of such potential bottlenecks, and they can avoid them. On the other hand, abnormal congestion originates in traffic incidents, which need to be detected in real time to prevent or resolve any sudden heavy congestion.

In this paper, we propose an AID method for detecting traffic incidents in real time by identifying abnormal car movements and distinguishing such movements from those occurring in spontaneous congestion. Our method measures differences between current traffic states (CTS) and usual traffic states (UTS), and has two aspects, namely, traffic state estimation and anomaly detection. First, we employ a probabilistic topic model [7] to model the generation of the PCD, which is influenced by hidden traffic situations such as "smooth" and "congested." The model introduces a single set of several hidden component states that are associated with probabilistic distributions over the PCD values, and each road segment during a certain time period has its own set of mixing coefficients. Using archived PCD, maximum-likelihood parameters of the model are estimated by an expectation–maximization (EM) algorithm. The estimated model reflects the usual state over the whole observation period. Our incident detection method simply follows the intuitive meaning of "anomaly." To detect incidents, the proposed method estimates the hidden state behind an observed PCD value and compares this current state with the usual state. If the current state is significantly different from the usual state, it is recognized as an anomaly.

We conducted an experiment using PCD observed for the entire Shuto Expressway system in Tokyo during 2010 and 2011. The total length of the Shuto Expressway system is approximately 300 km, and the daily traffic is about 1,000,000 vehicles per day [8]. Although the Shuto Expressway system forms the main artery system for the Tokyo area, there are many bottlenecks, and the speed limit is 60 km/h or less over most of the system [9]. Our experiment showed that the proposed method can identify trajectories involved in an incident better than existing methods.

The main contributions of this paper are as follows.

- We propose a new method for estimating traffic states by applying a probabilistic topic model to PCD, whereby road segments are characterized in terms of their expected performance hourly.
- We propose several methods for quantitative evaluation of the divergence of the CTS from the UTS using the traffic state model. We also propose several streaming algorithms that detect traffic incidents according to this divergence, whereby the detection is conducted adaptively in terms of the road segments and time periods.
- Our experiment showed that the traffic state model could be estimated using the observed PCD to reveal bottleneck sections on routes. It also showed that our AID method performed better than existing methods at identifying anomalous behavior by vehicles encountering incidents.

This paper is an extended version of the work published in the Proceedings of the Workshops of the EDBT/ICDT 2014 Joint Conference [10]. Here, we extend our previous work by introducing new divergence functions, developing a new algorithm, and conducting a new experiment using a larger-scale dataset.

The remainder of the paper is organized as follows. In the next section, we present related work. In Section 3, we introduce the traffic state model and describe our incident detection method. We conducted an experiment to evaluate our proposed method using a real PCD, and Section 4 describes the procedure and results of this experiment. We discuss the experimental results, issues and future work in Section 5. Finally, we conclude the paper in Section 6.

## 2. Related work

Anomaly detection [11] has attracted increasing research interest not only for communication networks [12,13] and social networks [14], but also for urban data. Using car-parking data, for example, useful trends as well as unusual behavior can be automatically extracted by an anomaly-detection technique [15]. AID can be considered as an application of anomaly or outlier detection to vehicular traffic data. Several AID methods have been proposed that exploit temporal data for vehicular speed [16] or flow data [17], which can be extracted from roadside surveillance cameras [18]. From the viewpoint of machine learning, AID can be regarded as a classification problem. Abdulhai and Ritchie [19] used neural networks, and Yuan and Cheu [20] used support vector machines to classify the observed vectors from stationary sensors as being incident based or otherwise. AID can also be regarded as an application of the change-point detection problem in time-series analysis, with Wang et al. [3] developing a hybrid method using time-series analysis and machine learning.

PCD, on which we are focusing, are different from the data on which existing work has been based. PCD are time-ordered sequences of points in spatio-temporal spaces, or trajectories. Piciarelli et al. [21] proposed an anomaly-detection method for trajectory data. This work used feature values extracted from the entire trajectory, implying that detection is not attempted during the ongoing movement of an object. A number of other studies have been carried out on anomaly detection from trajectory data that have been extracted from surveillance-video material [22,23]. This kind of trajectory data is different from PCD in that the sphere of movement is limited. Animal-movement data are an example of trajectory data in which the objects can move around a wide area. Lee et al. [24] proposed a method to find trajectory outliers using the example of animal-movement data. Although this method