Transportation Research Part C 82 (2017) 337–357

Contents lists available at ScienceDirect

Transportation Research Part C

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

Understanding the distribution characteristics of bus speed based on geocoded data

Yuchuan Du [a,*], Fuwen Deng [a], Feixiong Liao [b], Yuxiong Ji [a]

[a] Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, Shanghai 201804, China
[b] Urban Planning Group, Eindhoven University of Technology, Eindhoven 513, 5600 MB, The Netherlands

A R T I C L E   I N F O

Article history:
Received 19 September 2016
Received in revised form 8 July 2017
Accepted 11 July 2017

Keywords:
Geocoded data
Bus operation
Speed distribution
Finite mixture model
Cluster analysis

A B S T R A C T

Data-driven traffic management and control has attracted much attention recently. This paper conducts a series of coherent analyses based on geocoded data to understand the distribution characteristics of bus operational speed and to explore the potential applications of speed distributions. First, an original bipartite model is adopted for capturing instantaneous speed where the suspended and moving states are considered separately and a two-component mixed Weibull distribution is used to model the speed distribution in moving states. The mixed Gaussian distribution with variable components is found to be capable of expressing the speed distribution patterns of different road sections. Second, elaborate analyses on the basis of speed distribution modelling are conducted: (i) regression analyses are conducted to explore the correlations between parameters of instantaneous speed distributions and traffic related factors; (ii) a powerful clustering method using Kullback-Leibler divergence as the distance measure is proposed to grade the road sections of a bus route. These results can be utilized in fields such as bus operations management, bus priority signal control and infrastructure transformation aiming to improve the efficiency of bus operations systems.

1. Introduction

Vehicle speed is one of the most important measurements of traffic performance (May, 1990). Speed distribution is required for many traffic engineering applications. For example, an appropriate speed distribution model is the fundamental input for vehicle generation in traffic microsimulation systems (e.g., Park and Schneeberger, 2003; Llorca et al., 2015) and activity-travel scheduling simulation applications (e.g., Liao et al., 2013; Liao, 2016). Also, speed distribution can be utilized in theoretical analyses of traffic flow characteristics and to devise appropriate traffic operational measures (e.g., Yu and Abdel-Aty, 2014). Specifically, understanding the characteristics of bus speed distribution is essential for public transit operations management. There are two key problems in the investigation of speed distribution: (1) to develop an appropriate model accurately representing the speed distribution; and (2) to conduct analysis on the distribution characteristics, i.e., explaining the distribution patterns or parameters.

To address the first problem, numerous studies have been conducted to model the vehicle speed distributions by applying different statistic distributions, such as normal, log-normal, Student's *t*-distribution and diverse mixture distributions. Leong (1968) and McLean (1978) analyzed free speeds measured on two-lane two-way rural highways and concluded that car

* Corresponding author.
E-mail address: ycdu@tongji.edu.cn (Y. Du).

http://dx.doi.org/10.1016/j.trc.2017.07.004

speed distributions could be represented by a standardized normal distribution. However, more studies have subsequently shown that normal distribution is only applicable under homogenous traffic flows. For example, Lindner (1965) pointed out that normal distribution could only be applied in limited cases. Instead, a method was introduced in order to represent the non-normal distribution, allowing the approximation of the distribution using an expansion series with successive derivations of the normal distribution. Gerlough and Huber (1975) proposed utilizing log-normal distribution to reflect the skewness of speed distribution. Dey et al. (2006) found that speed distribution on highway could be unimodal or bimodal depending upon the speed variations of different vehicle categories. Consequently, they defined a spread ratio and found out that the speed data followed a unimodal curve only when the spread ratio was in the range of 0.69–1.35. Park et al. (2010) explored the applicability of finite mixtures of normal distributions to capture the heterogeneity in vehicle speed data, and adopted Bayesian estimation via Markov Chain Monte Carlo sampling for parameter estimation. Jun (2010) evaluated the traffic congestion patterns during the Thanksgiving holiday period using a Gaussian mixture speed distribution. Zou and Zhang (2011) investigated the suitability of mixture models using skew-normal and skew-*t* distributions which can fundamentally accommodate skewness and excess kurtosis. Expectation Maximization type algorithm was adopted to estimate distribution parameters. Rossi et al. (2014) compared the performances of different methods, including four-component normal, two-component skew-normal and two-component skew-*t* mixture models. Wang et al. (2012) proposed truncated normal and lognormal distributions to describe vehicle speed distribution because any traffic quantity of interest makes sense only within a limited value range.

With regard to second problem, relatively less research has focused on explaining the practical meanings of vehicle speed distributions or applying these results in the specific contexts. Ko and Guensler (2005) proposed that a mixed speed distribution over a given time period includes congested and uncongested cases. Based on that assumption, they showed it was possible to identify congestion characteristics by exploring the distribution parameters. Park et al. (2010) found that speed separation mainly are resulted from mixed vehicle composition for the flow level, rather than different time periods of a day. Maurya et al. (2015) compared the applicability of several distributions for different classes of vehicles and concluded that the log-normal distribution fitted well with motorized two-wheeled vehicle speed whereas the beta-distribution seemed to be a better fit for speeds of cars, buses and light commercial vehicles. Kyriakopoulou et al. (2016) formulated a methodology to obtain quantitative congestion measures and roadway performance on the basis of the speed distribution parameters.

Most of these aforementioned studies take the mixture speeds of various vehicle types on highways as the research focus, apart from a few cases considering urban roads. However, few studies analyze the operation speed distribution of city buses specifically. In fact, there are significant operational differences between buses and other types of vehicles, e.g., on operational stops and vehicle driving performances, which lead to significant speed distribution differences. Urban ground public transit is a vital infrastructure for cities and the lifeline of urban transportation in modern metropolises. "Transit metropolis" is a development goal for an increasing number of cities, since public transit is conducive to easing traffic congestion and reducing exhaust emissions. Therefore, the enhancement of public transit service, and particularly improving bus speeds in congested areas is an important consideration. The foundation of such work offers sound understanding of bus speed distribution characteristics.

In this study, bus instantaneous and section speed were calculated and analyzed based on GPS-data, which were recorded by GPS devices on buses in the city of Shanghai (China). An original bipartite model was proposed for the instantaneous speed distribution to solve the zero-inflation problem caused by operational stops and congestion delays in the downtown area. In this model, the zero and nonzero components were independently considered. For the nonzero component, a comparison between multiple distribution types was performed, and a two-component Weibull mixture distribution is proposed. A mixed Gaussian distribution model was proposed for the section speed distribution. The Expectation-Maximization algorithm was used for parameter estimation of the proposed models. In addition, simple regression analysis was used to effectively capture the correlation between model parameters and road traffic attributes, considering road section length, green ratio, red light duration in one period, and daily stops. This paper also proposed a powerful and flexible clustering technique using Kullback-Leibler divergence as the measurement of different statistical distributions, which has potential applications in rating bus operating status for different road sections.

The remainder of this paper is organized as follows. Section 2 presents data processing methods and exploratory data analysis results. Section 3 describes the methodology for modelling instantaneous and section speed distributions. Section 4 drills into the analyses of speed distribution parameters and morphology. Section 5 is a further discussion of the research on public transit. Section 6 provides a summary of conclusions and planned future work.

2. Data description and preprocessing

2.1. Data sources

Field data were collected by GPS devices installed on buses in Shanghai from 15th February 2016 to 10th April 2016. The frequency of data recording is 10 s, i.e., a GPS device records positioning information in every 10 s. The data format is shown in the right-hand side of Fig. 1. The study area was limited to 21 sections of Xizang Road located in the central urban area of Shanghai. In this context, a road section is defined as the roadway between two neighboring signalized intersections. The