Transportation Research Part D 54 (2017) 114–141

Contents lists available at ScienceDirect

Transportation Research Part D

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

Statistical inference-based research on sampling time of vehicle driving cycle experiments

Zhuang Yu [a,b], Shuming Shi [a,*], Sumin Wang [c], Yu Mu [a], Wenru Li [a], Chao Xu [a], Man Zhang [a]

[a] School of Transportation, Jilin University, Changchun 130025, China
[b] Shenzhen's Key Laboratory of Traffic Information and Traffic Engineering, Shenzhen 518021, China
[c] School of Mathematics and Statistics, Northeast Normal University, Changchun 130024, China

ARTICLE INFO

Article history:

Keywords:
Vehicle driving cycle
Sampling time
Statistical inference
Data mining

ABSTRACT

A driving cycle is a speed–time profile for a vehicle driving under a specified condition. It is usually developed from vehicle driving data collected by experiments to represent the real-life diving patterns. Because of lacking corresponding sampling theory, it is difficult for engineers to determine when vehicle driving cycle experiments should be stopped. In order to obtain sufficient experimental data, engineers normally choose to prolong the time of experiments wasting time and money. How to build a synthetical sampling subset of data representing a larger one becomes a main problem of sampling experiments. This paper, based on statistical inference theory, proposed a method to solve this problem at the city zone scale. First, the information entropy of road intersections was applied to determine the reasonable zone size. Then, according to one-month driving data of Changchun taxis and one-week driving data of Beijing taxis, it was found that the traffic distribution in city zone were able to be described by Nakagami distribution. It can pass the K-S test under the 0.05 significance level. In the order to fully use driving data, the bootstrap method was employed to conduct three resampling experiments in Changchun and five in Beijing. After analyzing the confidence intervals of distribution parameters, this paper discovered that the quality of the sampling data could be indicated by the accuracy of each zone's per car per day per square kilometers travel times. The linear relationship between the expectation of zone travel times variable coefficient and the expectation of α_{ab} which was used to evaluate the similarity between sampling distribution and population distribution was discovered. This relationship was also proved in this paper theoretically. Since the expectation of variable coefficient can be computed by sampling data, engineers are able to estimate the quality of these data in real time. If the α_{ab} reaches the preset threshold, experiments can be stopped.

1. Introduction

Vehicle driving cycle (VDC) has been applied to many diverse areas including vehicle powertrain matching, fuel economy and emission analysis, etc. (Xu, 2004). It is derived from driving data collected by VDC experiments. Then, some statistical and optimization methods are applied to generate it. Recently, the Markov chain approach is the most popular method for generating representative driving cycles (Nyberg et al., 2015; Lee and Filipi, 2011; Gong and et al., 2011).

* Corresponding author.
E-mail address: shishuming@jlu.edu.cn (S. Shi).

http://dx.doi.org/10.1016/j.trd.2017.04.029

The first step of VDC design is driving data collection. Whether driving data collected by experiments can reflect the real-world driving behaviors is the key of VDC design. The widely used collection methods can be categorized into three categories, namely, (1) the representative routes method; (2) the on-board measurement method; and (3) the chase-car method.

The key of the representative routes method is route selection. The road type, traffic flow and average velocity were often considered to select representative roads. Then, the driving data of experimental vehicles that run along the representative routes are recorded. These data are used to generate VDCs. In 1956, the California 7 Driving Cycle was the first cycle that adopted this approach. Subsequently, the ECE Driving Cycle, the FTP 75 Driving Cycle, among others also employed this method. Considering the geographical differences, Tong et al., chose two representative routes in Hong Kong to develop the Hong Kong Driving Cycle (Tong et al., 1999). In Edinburgh, 6 commuter routes were picked to collect driving data. The experiment lasted one year (Esteves-Booth et al., 2001). Chiang et al. selected three main routes in Taichung to gather driving data, and 202 test runs were conducted (Chiang et al., 2014).

The on-board measurement method involves gathering data using target cars equipped with measuring devices. It is also adopted by many researchers. Andre et al. built the European Driving Database including 2200 h, 88,000 km driving data which collected from 77 cars in Germany, Greece, and Britain. ARTEMIS European driving cycles were developed from this dataset (André et al., 2006). Ashtari et al. chose 76 volunteers to gather driving data in the city of Winnipeg with On-board equipment. Based on these data, they generated the Winnipeg VDC (Ashtari et al., 2012). Han et al. employed a military vehicle to collect 11-h driving data on military routes. These data were used to develop a driving cycle of a military area (Han et al., 2012). The Amman VDC was developed according to a midsized taxi's driving information (Mallouh et al., 2014). Schwarzer et al. generated the Honolulu VDC with 10 h of driving data (489 km) from a standard compact car (Schwarzer and Ghorbani, 2013).

The chase-car method refers to an instrumented car collecting data as it follows randomly selected vehicles, which has already been widely used around the world. Lin et al., adopted this method to obtain 102 test runs which includes 100,109 s of data (Lin and Niemeier, 2002), similarly for the case of the Singapore VDC (Wu et al., 2014), the Taipei VDC (Tzeng and Chen, 1998) and the China VDC (Wang et al., 2008) which also implemented this method. But it also remains some problems. Tong et al. noted that the chase car would easily be confused about which vehicle should be chosen according to the target vehicle random selection procedure, especially when the chase car enters a new roadway or at high volume traffic conditions (Tong and Hung, 2009).

In addition, to ensure the accuracy of the VDCs, more parameters are cosindered, such as vehicle types, grade and rational difference. With the rapid development of electric vehicles (EVs) and hybrid electric vehicles (HEVs), considering the differences between EVs and Internal Combustion Engine vehicles (ICEVs), driving cycles of EVs has been a topic under research over the last thirty years. Berzi et al. got nine-month data of EVs in Florence (Italy) including data for the use of both private and business. By grouping and selecting original driving data, they obtain DCs of EVs (Berzi et al., 2016). Zhang et al. proposed an optimization method which has the capability of providing a set of tradeoff optimal solutions among the fuel economy and various emissions and tested it with a combined UDDS/HWFET DC (Zhang et al., 2009). However, methods of cycle construction are still the same as ICEVs, the on-board measurement method and the chase-car method. And for developing VDCs, it is essential to collect enough and reliable data.

Furthermore, Road gradient has a crucial impact on vehicle fuel economy and vehicle emission, especially heavy-duty vehicles and hybrid electric vehicles. Souffran (Souffran et al., 2012) proposed a methodology for modeling real-world vehicles based on Markov chain model with the state space defined by velocity, acceleration and road grade (VAG Markov Chain Model). They designed a driving cycle with their simulation algorithm for the choice of powertrain sizing of HEV. Yue et al. (2015) analyzed the Markov property of driving cycles from phase space and correlation coefficient on VAG Markov Chain Model. They constructed the VAG Markov Chain Model with collected heavy duty truck driving data in motorway and designed a driving cycle. However, most of VDCs did not take the road slopes into account since most of passenger cars are drove in the cities. While for a VDC of heavy-duty vehicles, road slopes is a big issue.

Unlike others VDCs which were designed for one city or nation, Worldwide Harmonized Light Vehicles Test Procedure (WLTP) was developed at the United Nations level through UNECE in recent years. It is developed as a kind of transient driving cycle to fixing the growing gap between official laboratory and real-world on-road emission values, which negatively affects consumers, governments, vehicle manufacturers and society as a whole (Mock et al., 2014).

This paper aims at the sampling time of a city VDC experiment, so the driving patterns of city passenger cars are in focus. The VDC of a city can be regarded as the combination of all roads' VDCs. To obtain the VDC of each road is not difficult. The main problem is how to obtain road weights – the number of road travel times of experimental vehicles. Only if the road weights are consistent with the road traffic flow proportions of the city, can the combined VDC be similar with the real-word driving pattern. The representative routes method employs a tricky way to solve this problem. First, it divides the city roads into different categories (urban and suburban; trunk roads and branch roads). Roads in the same categories are considered similar. Therefore, researchers only need to obtain the traffic flow of each category instead of all roads', which decrease the complexity of experiments. However, it is subjective to divide city roads and it is hard to guarantee that all the roads in the same category have similar driving patterns. The on-board measurement method and chase-car method obtain the road weights in totally different ways. They directly or indirectly extract the experience of local drivers. However, given a lack of real road traffic flow, it is hard to tell the accuracy of their estimated road weights. Thus, the key of VDC experiments is to guarantee road travel times distribution of experimental cars is consistent with the road traffic flow distribution of the city. In this paper, the number of times that a vehicle uses a road is referred to as travel times in this paper.