Journal of Cleaner Production 206 (2019) 827–837

Contents lists available at ScienceDirect

Journal of Cleaner Production

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

A data-driven two-level clustering model for driving pattern analysis of electric vehicles and a case study

Xuefang Li [a, b], Qiang Zhang [a, b, *], Zhanglin Peng [a, b], Anning Wang [a, b], Wanying Wang [a, b]

[a] School of Management, Hefei University of Technology, Hefei 230009, China
[b] Key Laboratory of Process Optimization and Intelligent Decision-making, Ministry of Education, Hefei 230009, China

ARTICLE INFO

Article history:
Received 21 May 2018
Received in revised form
20 September 2018
Accepted 21 September 2018
Available online 24 September 2018

Keywords:
Daily driving pattern
Electric vehicles
Two-level clustering
Multifaceted driving patterns

ABSTRACT

The driving patterns of the electric vehicles describe how users use their vehicles, reflecting the users' habits. These patterns have a positive effect on vehicle energy consumption. In this paper, a two-level clustering model is proposed to determine the driving patterns of electric vehicles. Firstly, the driving pattern characteristics are extracted from the data set of the electric vehicles. Then, the driving patterns including daily driving patterns and multifaceted driving patterns are obtained by a two-level clustering model. The data of 1463 electric vehicles in China were collected from September 1, 2015, to September 1, 2016. Using the proposed model, we obtain five types of daily driving patterns and four types of multifaceted driving patterns. Then, the features of clusters are extracted, and the geographical distribution analysis of the multifaceted driving patterns is conducted. The experimental results reveal that there are many driving patterns of the electric vehicles. Moreover, the effectiveness of the clustering models is verified by the experiments. The customer segmentation based on the driving patterns of electric vehicles is of a great significance for the development of personalized and targeted marketing strategies of vehicle manufacturers and energy efficiency improvement.

1. Introduction

With the aim to promote the cleaner production, the energy and environment efficiency should be analyzed (Chafic-Thomas et al., 2016; Song and Zheng, 2016; Song and Zhou, 2015). The proportions of energy consumption and carbon dioxide emission in the transportation sector are up to 28% and 33%, respectively, whereas the light-duty vehicles comprise 57% of the total energy consumption (Kelly et al., 2012). Comparing with the internal combustion engine petrol vehicles, the electric vehicles (EVs) have numerous benefits that reduce the transportation-related emissions (Greene et al., 2010) and driving cost, improve energy efficiency (Zhang et al., 2018), and in some cases eliminate the use of the fossil fuels (Montazeri-Gh and Mahmoodi-K, 2016). However, a big flaw of the EVs is the limitation of their batteries; namely, EVs have a limited driving range and require a longer charging time compared with the conventional types of vehicles (Noorden, 2014). Considering the range limitations, recent developments of driving patterns have attracted much attention.

The driving patterns show the usage behaviors of drivers (Ericsson, 2001) and directly determine the energy consumption of the EVs. These patterns are becoming an important issue that affects the EVs energy consumption (SAE, 2010), economic costs (Neubauer et al., 2013), and environmental parameters (Elgowainy et al., 2010). Moreover, the accurate identification of driving patterns could greatly improve and optimize both component design and energy control strategy of the EVs, such as the battery capacity design (Gu and Rizzoni, 2006), electric motor size (Zhang and Xiong, 2015), driving range estimation (Yu et al., 2012), and evaluation of vehicle emission (Lin et al., 2004), as well as the energy infrastructure construction of the EVs (Kelly et al., 2012). Besides, the optimization of design and energy control strategies can improve the energy conversion rate of the EVs.

With the increasing application of the sensing and network communication technology in the EVs, a large volume of EVs driving data can be collected in real time (Borgstedt et al., 2017). The EVs driving data are typical time-series that represent the driving status including the status data of a vehicle, geographical data, extreme data of battery, etc., which reflect the driving behaviors either directly or indirectly (Lu et al., 2017). Therefore, the

* Corresponding author. School of Msanagement, Hefei University of Technology, No.193 Tunxi Road, Hefei, Anhui, China.
E-mail address: qiang_zhang@hfut.edu.cn (Q. Zhang).

https://doi.org/10.1016/j.jclepro.2018.09.184
0959-6526/

identification of driving patterns based on the real-time driving data becomes achievable. Generally, there are two key issues in the identification of driving patterns: 1) determination of proper dimensions of a driving pattern, 2) determination of a driving pattern based on the dimensions of the driving pattern obtained from the driving data.

Regarding the first issue, the existing indicators of the driving patterns can be classified into time-dimension indicators and spatial-dimension indicators. The time-dimension indicators: driving velocity (Adornato et al., 2009; Karin Brundell-Freij, EvaEricsson, 2005), driving acceleration (Gong et al., 2011; Wang et al., 2008), driving daily distance (Achour and Olabi, 2016; Brady and O'Mahony, 2016), and idle time of driving (Dembski et al., 2002; Smith et al., 2011; Wu et al., 2010) are used to describe the temporal characteristics of driving patterns. Also, the indicators such as daily driving range (Corchero et al., 2014) and trip range (Wang et al., 2015) are often extracted to measure the driving demand in the spatial dimension. Generally, the current studies mainly focus on driving range and related environmental factors to identify the driving patterns (Karabasoglu and Michalek, 2013). However, the parking behavior, including the parking location, parking frequency, and parking time, is often neglected by most researchers. Nevertheless, the parking behavior, as an important part of a driver's usage behavior, can reflect the activity scope and parking habits, and further estimate the driver's usage purpose combined with the driving range (Smith et al., 2011; Staackmann et al., 1997).

Regarding the second issue, the traditional studies often develop the theoretical models based on the questionnaires data, such as activity-based micro-simulation (Amjad et al., 2011) and time-based parameters statistics (Wang et al., 2008; Wu et al., 2010) to identify the driving pattern. Due to a large volume of real-time driving data, by using these theoretical models, it is difficult to capture the driving patterns accurately. Hence, a series of the data mining techniques, such as clustering (Adornato et al., 2009), correlation analysis (Xydas et al., 2016), stochastic prediction method (Ashtari et al., 2012), and time-series clustering, are widely used to analyze the driving data. Zhang et al. developed a time-series clustering with variable weights to analyzes the driving cycle of hybrid-electric vehicles (Zhang et al., 2016). Zhou et al. used the *k*-means time-series clustering to recognize the driving patterns (Zhou et al., 2017). Wang et al. studied the time-series clustering to analyzes the pure-electric buses driving cycle in big cities in China (Wang et al., 2017). The existing time-series clustering methods focus on the whole clustering of driving patterns, emphasizing the characteristics of a usage behavior in a period. However, because of the multi-purpose usage of the EVs, the EVs may show several driving patterns in different time periods. In Fig. 1, the five-day driving patterns of the EV are presented, where it can be seen that the EV has different driving patterns in different parts of the day. So, if only the whole clustering is used, the daily driving patterns are neglected (Keogh and Lin, 2005). Hence, it is necessary to perform the subsequence clustering based on the whole clustering, thereby accurately mining the multifaceted driving pattern.

Based on the two above-mentioned issues, here, a novel methodological framework for identifying the driving patterns of the EVs is presented. The main contributions are two-fold:

1) The driving range and parking range are adopted to identify the driving patterns based on the real driving data, and the corresponding measure indicators, i.e., the distribution of driving mileage and the parking concentration degree, are developed.
2) We develop a two-level clustering model, which provides deeper insights into driving patterns with consideration of different time granules. In the proposed model, the EVs driving data are first processed to identify the daily driving patterns by the level-1 clustering. Then, based on the result of the level-1 clustering, the multifaceted driving patterns of the EVs are obtained by the level-2 clustering.

The remainder of this paper is organized as follows. In Section 2, the analysis of driving patterns based on the driving data is presented. A two-step cluster-based pattern designed for the driving pattern mining is introduced in Section 3. The experimental results of a two-level clustering and the geographic distribution analysis of multifaceted driving patterns are provided in Section 4. Lastly, the conclusions are given in Section 5.

2. Driving patterns based on driving data

Driving patterns characterize the usage behavior of a user (Xie et al., 2018). Driving patterns vary with users and time, mainly because of the users' driving habits, trip requirement, traffic conditions, etc. Here, the driving range and parking range are adopted to identify the driving patterns based on the real driving data.

2.1. Driving range

The driving range reflects the user's mileage requirement, driving habits, and speed preference at a different time (Ericsson, 2000). The daily driving range of a vehicle is presented in Fig. 2, where it can be seen that mileage is different in different periods. Moreover, it can be noticed that users have different mileage requirements in different periods (Zhang et al., 2017). Simultaneously, because of the different speed preferences, there are also diverse speeds during the day.

Currently, scholars usually consider either the driving speed or the driving mileage of a single trip to represent the driving range (S. Jeon et al., 2002). However, these methods ignore key information shown in Fig. 2, i.e., the mileage requirement or the speed preference. Therefore, this paper adopts the driving mileage per unit of time to measure the driving range. In that way, not only the driving mileage is measured, but the user's speed preference is also reflected.

The distributions of the driving range and the daily driving range are denoted as set S and set D, respectively, and they are defined as:

$$S = \{D_1, D_2, D_3 \ldots D_{N_s}\} \tag{1}$$

$$D_i = \{t_{1i}, t_{2i}, t_{3i} \ldots t_{ji}\} \tag{2}$$

In set S, D_i represents the daily driving range of an EV, and N_S denotes the total number of travel days. In the daily driving range D_i, t_{ji} denotes the driving mileage at the jth unit time of an EV on the ith travel day. Here, a time unit of 10 min is adopted.

2.2. Parking range

The parking behavior is a part of the usage behavior of an EV user. The distribution of the parking locations and parking frequency can indicate the range of activity of the EVs, e.g., the parking frequency and the number of different parking locations are higher for rental cars than for private cars.

Therefore, the parking frequency and distribution of parking location of an EV are the key indicators of the parking behavior, and simultaneously, the key indicators of the driving behavior. If the number of parking locations of a vehicle is relatively high, and the parking frequency of each location is uneven, then, the parking locations of an EV are relatively dispersed, so that an EV may be a