

Research on Big Data Mining Technology of Electric Vehicle Charging Behaviour

Yujun Liu*, Yi Hong, Cheng Hu

*School of Electrical Engineering, Southeast University,
210096 Nanjing, China
liuyujun@seu.edu.cn*

Abstract—Thousands of electric vehicles (EV), which are large in number and flexible in their use of electricity, will be connected to the power system in the near future, which will bring more uncertainty to the power system. Therefore, it is necessary to study the general characteristics of EV charging behaviours. In the charging process, big data regarding charging behaviour of EVs are generated. This paper proposes a big data mining technique based on Random Forest and Principle Component Analysis for EV charging behaviour to identify and analyse clusters with different charging characteristics from the big data. This paper uses Dundee's January 2018 EV charging data to conduct experiments, and obtains the charging behaviour clusters of the workdays, weekends, and holidays of January. The superiority of the random forest algorithm in the EV clustering problem is reflected when compared to the Euclidean distance method. The clusters obtained by the random forest algorithm have clearer characteristics, including the user's charging method and travel behaviour. The results show that the charging behaviour of EVs has certain regularity, and the charging load has obvious peak-to-valley difference that is necessary to be regulated.

Index Terms—Electric vehicle; Charging behaviour; Big data; Random Forest; Cluster analysis.

I. INTRODUCTION

As environmental pressures and the lack of energy shortage become more serious, electric vehicles attract more and more attention because of their high energy efficiency and low emissions of pollutant gases [1]–[3]. Analysing the spatial and temporal distribution of EV charging load is the basis for studying the impact of large-scale development of electric vehicles on the power grid, capabilities of participating in grid interaction, and charge/discharge-control strategies [4]. However, the charging behaviours of EVs are normally random and diverse, which seemingly makes them complicated and hard to analyse [5].

At present, the modelling of EV charging behaviour is mainly based on the fitting of statistical data and the probability distribution function [6]. Reference [7] proposed a calculation method that comprehensively considered the charging time distribution of different types of EVs, especially private cars, buses, and taxis. Reference [8]

introduced the gravity model in traffic demand forecasting and analysed the spatial distribution of EV charging loads. However, data used in [7] and [8] was relatively vague. Reference [9] analysed the differences in EV charging behaviour between residential and commercial areas based on demographic and traffic information. In [10], the multi-agent simulation was used to establish the spatiotemporal distribution model of EV charging loads. However, the probability distribution functions used in [9] and [10] were mainly based on simulation, and hardly considered a large amount of actual charging data.

Because the results of the above-mentioned two methods may have a large deviation from the actual, this paper will create the third method of EV-charging-behaviour modelling, which is based on the measured charging information and the big data mining technology. The current trend shows that the functions of charging facilities are constantly updated, and more multiple types of charging data will be recorded to form Big Data on the charging behaviour of EVs in a specific region [11]. In the absence of human intervention, the collected EV charging data samples usually contain information, such as charging start time, end time, charging energy, and charging location, which are not marked with a clear category. Thus, the EV charging data sample is Unlabelled Data. However, the study of the regularity and cluster of EV from the perspective of measured charging information is quite lacking. Currently, the classification of electric vehicles is mainly based on the types of vehicles. Reference [12] classified the charging mode of hybrid EVs by considering the difference in the structures of hybrid EVs. It did not cover all types of EVs and do not consider the differences in the behaviour patterns of EV users. However, if the actual measured charging information, including all kinds of EVs in a certain region, can be fully used, the accuracy of the charging behaviour description will be greatly improved, which is beneficial to manage the charging behaviour of EVs.

The big data mining technology used in this paper is mainly based on Principle Component Analysis (PCA) [13] and Random Forest (RF). The Random Forest algorithm designed by L. Breiman in the early 21st century is one of the most successful methods currently available for processing Big Data [14]. The RF algorithm refers to a machine learning method that uses multiple decision trees to train and predict samples [15]. It is capable of highly parallelized processing

Manuscript received 24 April, 2019; accepted 5 October, 2019.

This paper was completed under the support of Electric Power Research Institute, State Grid Tianjin Electric Power Company (KJ17-1-02). The Dundee City Council gave data support.

and can meet the clustering requirements of high-dimensional and large-sample data sets in the era of Big Data. At the same time, the RF algorithm does not produce over-fitting problems and can also assess the importance of variables [15], [16]. Besides, it has successfully addressed a variety of practical areas. Due to its characteristics and advantages, Random Forest is very suitable for cluster analysis of EV behaviour based on Big Data. Since the EV charging data sample is Unlabelled Data, the category corresponding to the samples cannot be trained. Thus, Random Forest can only be trained from the original sample set. This problem belongs to Unsupervised Learning.

The purpose of this paper is to establish a big data mining technology for charging behaviour of electric vehicles. PCA will be used to reduce the dimensions of the data. Cluster analysis will be based on the Random Forest algorithm and on the EV charging data of Dundee in the UK. The application results of the big data mining model will be given.

II. THE MODELLING OF BIG DATA MINING TECHNOLOGY

A. Model Framework

As shown in Fig. 1, the big data mining model can be divided into 4 parts.

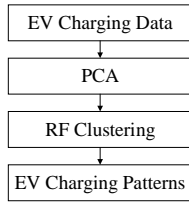


Fig. 1. Model framework of big data mining technology.

Firstly, the original data are cleaned and standardized. Secondly, PCA is applied to reduce the dimensionality of the big-data-mining problem without losing much EV charging information. Thirdly, the big data should be clustered according to characteristics of EV charging behaviour. Random Forest is used to complete the cluster function. Finally, the clusters of EV charging behaviour are achieved after the correlation analysis of the data samples is done by RF.

B. Principle Component Analysis Algorithm Flow

Principal Component Analysis is a kind of Deep Learning method. It is used to extract important sample features to reduce the number of features, but still retain most of the information of the original samples.

In order to obtain the principal components of the EV charging data set, the following calculation steps are required, including:

Step (1): Standardize the raw EV charging data and delete the error data;

Step (2): Establish a coefficient matrix \mathbf{R} between variables

$$\mathbf{R} = (r_{ij})_{m \times m}, \quad (1)$$

where r_{ij} is the element of \mathbf{R} and m is the number of features owned by each sample.

$$r_{ij} = \frac{\sum_{k=1}^n (x_{ki} - \bar{x}_i) \times (x_{kj} - \bar{x}_j)}{\sqrt{\sum_{k=1}^n (x_{ki} - \bar{x}_i)^2 \times (x_{kj} - \bar{x}_j)^2}}, \quad (2)$$

where x_{ki} is the value of k th charging event. \bar{x}_i is the average value of the i th charging characteristic.

Step (3): Find the eigenvalue of \mathbf{R} . The order of the feature roots is $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m \geq 0$. At the same time, the feature vectors \mathbf{a}_m corresponding to each feature root are obtained as follows

$$\mathbf{a}_l = [a_{1l} \quad a_{2l} \quad \dots \quad a_{ml}]^T, \quad (3)$$

where $l = 1, 2, 3, \dots, m$.

The principal component F_i can be expressed as

$$F_i = a_{1i} \times x_1 + a_{2i} \times x_2 + \dots + a_{mi} \times x_m, \quad (4)$$

where $i = 1, 2, 3, \dots, m$.

The cumulative contribution rate of the first N principal components C_N is recorded as follows

$$C_N = \frac{\sum_{k=1}^N \lambda_k}{\sum_{k=1}^m \lambda_k}. \quad (5)$$

It can be commonly considered that only the first N principal components can represent most of the information of the original feature when the cumulative contribution rate of the current N principal components is as high as from 85 % to 95 %.

C. Random Forest Algorithm Flow

Random Forest belongs to an integrated algorithm, which is based on the Bagging algorithm, and the Classification and Regression Tree (CART) algorithm [17]. The Bagging algorithm is short for Bootstrap Aggregating, and it is based on Bootstrap sampling [18]. The core idea is to use the results of Bootstrap sampling to construct a number of independent classifiers.

The implementation process of Random Forest is as follows.

Step (1): Assume that the training data set regarding EV charging S have M features. It is repeatedly sampled by Bootstrap sampling to obtain randomly generated training data sets S_1, S_2, \dots, S_n . These training data sets have m ($m \leq M$) features. The probability that this new data set does not contain the sample is about 36.8 %. This part of the data is called Out of Bag (OOB).

Step (2): Each decision tree is constructed by the binary recursive method in the CART algorithm based on the corresponding training data set S_1, S_2, \dots, S_n . When processing each node of the decision tree, all possible splits are calculated based on m features of the training data set.

Choose the best split mode (e.g. get the largest Gini metric) to split each node. Repeat the above splitting process until a certain condition is met. Each tree will not be pruned, when classifiers C_1, C_2, \dots, C_n are trained corresponding to each training data set.

Step (3): Use the original EV data set X that has been separated, and use the decision trees have not been pruned obtained in Step (2) to discriminate. Then, obtain the sample distribution at the end of each decision tree $C_1(X), C_2(X), \dots, C_n(X)$.

Step (4): Count the samples at the end of each decision tree. If two samples appear at the same end, the correlation between the two will increase.

D. Principle Component Analysis Algorithm Flow

The correlation between any two samples in the EV charging big data is defined as the ratio of the numbers of these two samples appearing on the last node of the same decision tree. Assuming that the total number of EV charging data samples is n , a similar matrix of $n \times n$ dimensions can be constructed. Each element in this matrix belongs to $[0, 1]$ and represents the similarity of two corresponding data samples in each tree of the Random Forest. The probability of aggregation of similar data samples at the end of the decision tree is greater than the probability of the aggregation of dissimilar data samples [16]. The correlation of RF data samples can be used for cluster analysis.

1) Operation Principles of RF Clustering

The similarity of data samples can be used as input for the traditional clustering operation in the clustering problem of EV charging data sets. However, not all multi-structured data sets can show the form of clustering [19]. The RF algorithm can detect EV charging data sets through unsupervised machine learning and may not require prior assumptions about the clustering characteristics of the data set [20].

The main steps of unsupervised machine learning are as follows. Firstly, set the initial EV charging big data as Data Set 1. Secondly, the values of each parameter in the same data sample are independently replaced to generate Data Set 2. The methods of independent permutation are diverse. In this paper, the method of random adoption is adopted according to the empirical boundary distribution of the parameters. Data Set 1 and Data Set 2 form a mixed data set. The Data Set 2 has independent random parameter distributions, but all parameters in Data Set 2 have the same univariate parameter distribution characteristics as that of the corresponding parameters in Data Set 1. Thus, Data Set 2 destroys the non-independent structure of Data Set 1. Random Forest needs to extract Data Set 1 from the mixed data set when trees are being trained. This double-type problem can be simulated by the RF algorithm. The biggest benefit of describing it as a double-type problem is the increased feasibility of clustering.

2) Multidimensional-Scale-based Clustering Results Display

Multidimensional scaling analysis is one of the ways to help Random Forest to analyse the characteristics of a data set. This method can be used to represent the degree of correlation between EV charging data samples. The EV charging data set can be represented in coordinates in a

low-dimensional coordinate system by appropriate dimensionality reduction. The distance between any two points in the coordinate system reflects the correlation between two corresponding EV charging samples, which help to explore the factors that affect the correlation of the EV charging samples.

The RF algorithm does not need to specify the distribution characteristics of the parameters before the operation and can estimate the ability of each parameter to influence the correlation prediction results. The cross-validation within the RF algorithm can be used to evaluate the error rate of the correlation prediction. This kind of evaluation has a high accuracy.

III. APPLICATION OF BIG DATA MINING MODEL IN ELECTRIC VEHICLE CHARGING BEHAVIOUR ANALYSIS/CLUSTERING

A. Data Source

The electric vehicle charging data used herein were collected from 29 charging stations or charging piles in Dundee. The charging locations in Dundee are shown in Fig. 2.

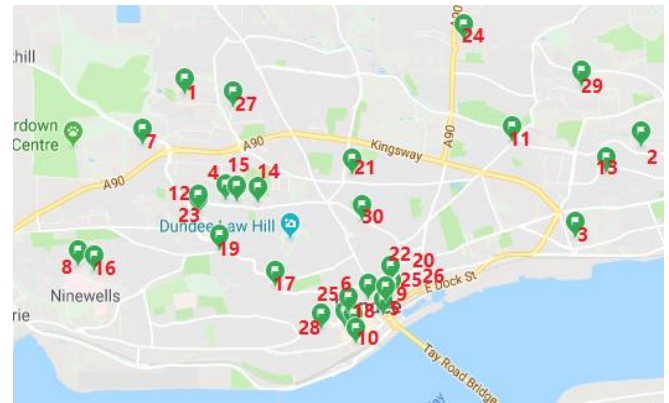


Fig. 2. The map of charging locations in Dundee.

The EV charging data is provided by Dundee City Council (<https://data.dundee.gov.uk>). The time period for collecting data is from January 1st to January 31st, 2018. The variables used for observation include the charging start time, the charging end time, the charging duration, the charging energy, and the charging location. After removing the inefficient charging behaviour with incomplete parameters or obvious parameter errors, the total of the remaining effective charging behaviour is 5654 times, including 4220 for workdays, 1180 for weekends, and 254 for holidays.

B. Results of Principal Component Analysis

Based on PCA, the original five observed variables are converted resulting in five new principal components, whose contribution degrees to express the information of entire charging data set are shown in Fig. 3, respectively. The first four principal components can represent 93 % of the information and characteristics of the original data. Therefore, the fifth component is not needed in the clustering analysis. Practice results show that this method reduces the computational time of the RF algorithm and the occupied computer memory.

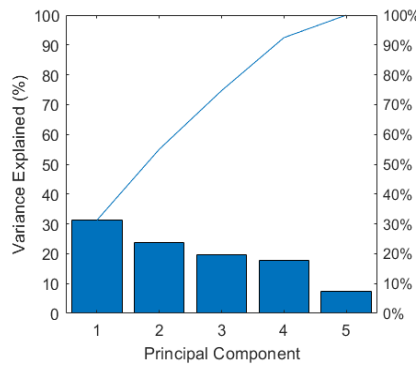


Fig. 3. Cumulative contribution of principal components to the entire charging data set.

C. Clustering Implementation Process

In this paper, *R* programming language is used to implement the clustering function of the RF algorithm. The coordinates of the classical multidimensional scale (CMS) are used to characterize the relationship between the charging behaviours. The specific process is as follows.

Step (1): Set the number of forests and the number of decision trees in each forest. This paper is set to 3 and 500, respectively.

Step (2): Set the number of parameters used by each node on the decision tree to fork, making it equal to an integer. This integer is close to one-half of the total number of parameters and its value is set to 2 in this paper.

Step (3): Enter the charging behaviour data into the random forest algorithm, which is implemented using the *RFdist* function in *RStudio* software. So, the importance degree of

each parameter (Gini index) is obtained.

Step (4): Using the *cmdscale* function in *RStudio* software to generate two-dimensional coordinates of each charging behaviour in the classical multidimensional scales.

Step (5): The classical multidimensional scales of all charging behaviours are drawn based on the two-dimensional coordinates of the charging behaviour obtained in Step (4). The distance between the two points represents the dissimilarity of the two charging behaviours. Each point is represented by (x, y) .

Step (6): According to the shape feature of the classical multidimensional scale obtained in Step (5), the image is divided, and the dense points are classified into the same class. Thereby, the clustering results of the charging behaviour of the EV users are obtained.

In this paper, the charging behaviours of workdays, weekends, and holidays are clustered separately and the corresponding classical multidimensional scales are obtained, respectively, as shown in Fig. 4.

D. Clustering Results

According to the shape features of the CMSs of the three periods, the clusters are performed separately, so that the clustering results of the three periods can be obtained. Figure 6 shows the distribution characteristics of different clusters of parameters. The abscissas of Fig. 5 represent time or charging energy and the ordinates represent the number of times that the corresponding value of the abscissas appear in the big data.

Table I provides a detailed explanation of Fig. 5.

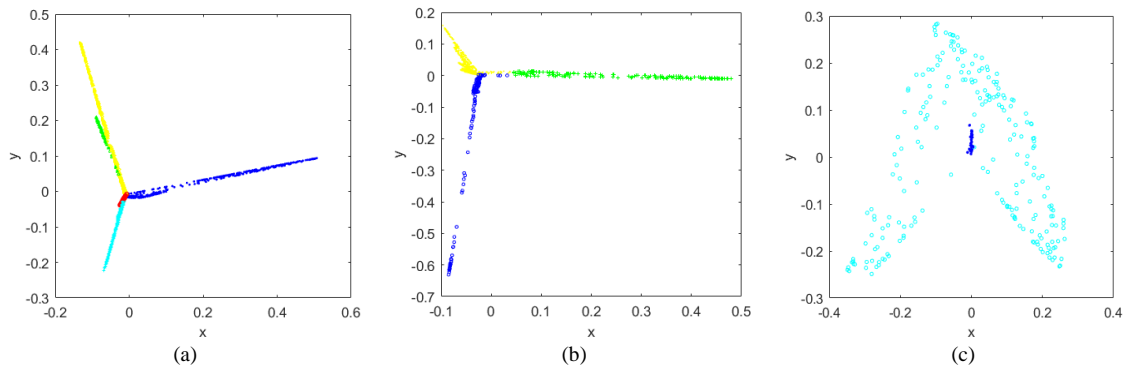


Fig. 4. Classic Multidimensional Scale of EVs Charging Behaviour: (a) is the CMS on workdays; (b) is the CMS for weekends; (c) is the CMS for the holidays.

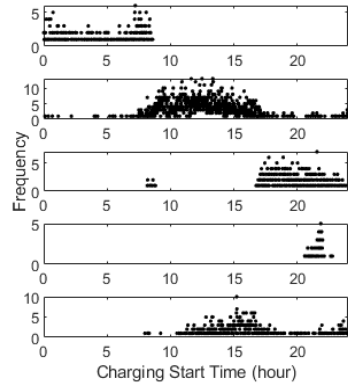
TABLE I. THE CHARGING BEHAVIOUR CHARACTERISTICS OF EACH CLUSTER.

Period	Cluster	Proportion (%)	Start time range	End time range	Average charging time (hours)	Average charging capacity (kWh)	Main charging location
Workdays	1	9.9052	00: 03 ~ 08: 35	00: 04 ~ 09: 49	0.7536	8.6963	7, 10, 21, 22, 24
	2	56.7299	08: 21 ~ 18: 51	08: 53 ~ 23: 59	3.4944	7.4130	6, 9, 10, 20, 21, 22, 29
	3	14.9763	16: 45 ~ 23: 50	17: 14 ~ 23: 59	0.6001	9.2825	7, 10, 21, 22
	4	2.2512	20: 35 ~ 22: 46	06: 00 ~ 08: 38	9.6960	10.5245	2, 3, 13
	5	16.2085	10: 30 ~ 23: 54	00: 02 ~ 20: 50	26.2296	9.3603	6, 21
Weekends	1	14.4915	00: 00 ~ 05: 33	00: 01 ~ 05: 48	0.4127	10.8698	7, 10, 21, 22
	2	13.6441	00: 07 ~ 23: 58	00: 06 ~ 23: 51	18.9864	13.0891	2, 3, 9, 12, 13, 21, 22
	3	71.8644	03: 18 ~ 23: 51	05: 35 ~ 23: 58	0.9612	8.5612	7, 10, 21, 22
Holidays	1	13.3858	06: 30 ~ 23: 56	00: 18 ~ 20: 18	20.6003	13.2809	3, 9, 13, 16, 21
	2	86.6142	00: 00 ~ 23: 18	00: 07 ~ 23: 49	0.6901	8.5522	7, 10, 21, 22

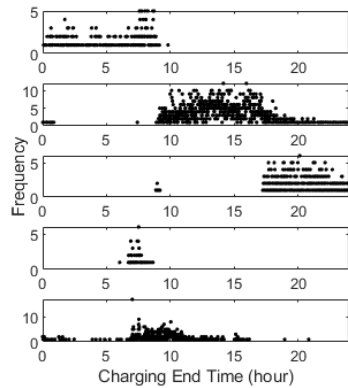
IV. DISCUSSION OF RESULTS

E. The Superiority of Random Forest Clustering

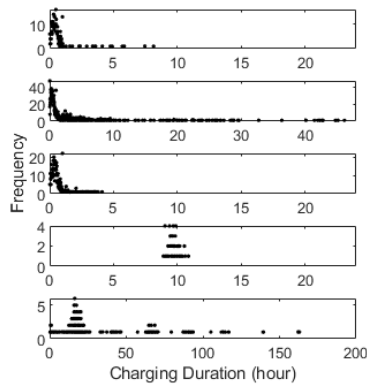
Firstly, the effectiveness and superiority of the Random Forest algorithm for the clustering of EV charging behaviour will be analysed in this section. The Euclidean distance method is used to express the correlation between all charging behaviours in the workdays of January 2018. The result is shown in Fig. 6.



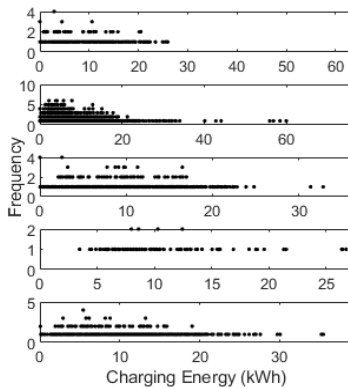
(a)



(b)



(c)



(d)

Fig. 5. Characteristic distributions of five clusters of EVs charging behaviour in workdays: (a) is the charging start time distribution; (b) is the distribution of charging end time; (c) is the charge duration distribution; (d) is the charging energy distribution. Each subgraph contains five distributions, which in turn represent the charging behaviour of the first to fifth clusters from top to bottom.

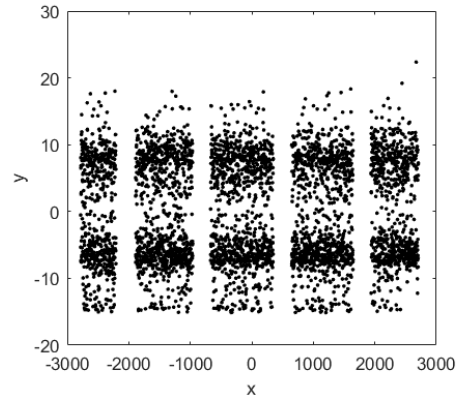


Fig. 6. CMS of EV Charging Behaviour on Workdays by Euclidean Distance.

It can be seen that there are more than 10 clusters in Fig. 6. So, the clustering results obtained by the Euclidean distance method are too complicated. Moreover, according to the evaluation of the importance of the feature parameters by the *R* programming, the scatter distribution obtained by the Euclidean distance method depends only on one feature parameter and is independent of the remaining feature parameters. In addition, as can be seen from the values of the axes, the distance between each point is much larger than the result obtained by Random Forest. Therefore, the clusters obtained by the Euclidean distance method are diverse, but not suitable for the general characteristics of the reaction behaviour. However, the shape of the scatter plot obtained by RF clustering is regular and dense. The images are striped in strips, which makes it easy to segment and discover categories of images. Besides, the difference between the abscissa and the ordinate of each point is small. This is consistent with the characteristics of the RF algorithm, i.e., the correlation value of the data samples generated by the random forest algorithm should be between 0 and 1. This also makes the CMS of RF more versatile. In summary, the random forest algorithm has obvious advantages in the cluster analysis of EV charging behaviour.

F. Analysis of EV Cluster General Characteristics

According to the results of Table I, the information of EV charging data is further mined in this section. Table II summarizes the links between EV clusters and social behaviours.

As can be seen from Table II, the characteristics of each cluster are clear and closely related to different social behaviour. This also shows that the Random Forest algorithm effectively implements the function of clustering the charging behaviour. However, there are fewer clusters for holidays and weekends and their characteristics are not as obvious as in the

clusters of workdays. Because of less data samples for holidays and weekends, there are less clusters that exist objectively. Also, the behaviour patterns of EV users on holidays and weekends do differ from the working days. In addition, due to the small amount of data on holidays, the

images are scattered and irregular, which is not conducive to clustering based on CMS. However, the number of holidays is much smaller than the number of workdays during the whole year. So, even if there is inaccuracy in the holiday clustering, it will not have a huge impact on the overall results.

TABLE II. THE GENERAL CHARACTERISTICS OF VARIOUS CLUSTERS OF EVs CHARGING BEHAVIOUR.

Period	Cluster	General characteristics
Workdays	1	These users often work at night, charging mainly at midnight and dawn. Their charging duration is very short, generally no more than an hour, but the charging energy is relatively large.
	2	These users are basically active during daytime, mainly charging after dawn. They charge longer, but charge less.
	3	This type of user mainly starts charging in the afternoon. Their charging time is very short, but the charging energy is relatively large.
	4	This type of user mainly starts charging in the evening. The end of charging duration is the next morning. Therefore, the charging time is longer. However, the amount of charge is relatively small.
	5	This type of user mainly starts charging in the morning. However, the charging end time is irregular. The charging duration is very long, which can even contain a few days, but the charging energy is relatively small.
Weekends	1	This type of user charges mainly at midnight and dawn. Their charging duration is very short, but the charging energy is relatively large.
	2	This type of user starts charging all day long and takes a long time to charge. However, the amount of charge is relatively small.
	3	These users start charging mainly after dawn and their charging duration is very short. However, the amount of charge is relatively large.
Holidays	1	This type of user mainly starts charging in the morning. They charge for a long time, even for a few days. However, the amount of charge is relatively small.
	2	The distribution of charging start times for such users is more dispersed. The charging duration is very short, but the charge is relatively large.

G. Analysis of EV Charging Loads

It can be seen from Table I that Workday Cluster 2, of which the main charging period is morning and afternoon, has the highest proportion and consumes the most amount of electricity. This will create the first peak of charging during the day. The charging behaviour of Workday Cluster 3 and Cluster 4 is concentrated in the middle of the night and is mainly based on the fast charging. These two clusters will create the second charging peak during the day. In addition, the trends of charging behaviour on weekends and holidays are roughly similar to that of workdays. However, the number of times of charging and the charging energy of a single day on weekends and holidays are less than that of workdays, which indicates that the charging behaviours of EVs in the workday are more frequent. Therefore, it is necessary to distinguish the charging laws of the three periods.

Furthermore, the charging load of electric vehicles has a distinct peak-to-valley difference. If the charging load of the EVs can be controlled to effectively reduce the peak load, the power and grid investment can be reduced and the operating cost of the grid will possibly be cut also.

V. CONCLUSIONS

This paper proposed a big data mining technique based on Random Forest and Principle Component Analysis for electric vehicle charging behaviour to identify and analyse different types of charging behaviour characteristics. Experiments were carried out using Dundee's January 2018 charging data to obtain the charging behaviour clusters for the workdays, weekends, and holidays. It was found that each cluster has relatively clear characteristics and the users' charging methods can be inferred by analysing the clustering results. The conclusion is that the charging behaviour of EVs has certain regularity. It is necessary to control the EV charging behaviour because it can help to reduce the reconstruction and operating costs of the power grid.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

- [1] F. Wen, F. Wu, Y. Ni, "Generation capacity adequacy in the deregulated electricity market environment", *Automation of Electric Power Systems*, vol. 26, no. 19, pp. 16–22, Oct. 2002.
- [2] A. Shekhar, V. Prasanth, P. Bauer, and M. Bolech, "Economic viability study of an on-road wireless charging system with a generic driving range estimation method", *Energies*, vol. 9, no. 2, pp. 941–948, Jan. 2016. DOI: 10.3390/en9020076.
- [3] M. Brenna, F. Foiadelli, M. Longo, and D. Zaninelli, "e-Mobility forecast for the transnational e-corridor planning", *IEEE Trans. Intelligent Transportation Systems*, vol. 17, no. 3, pp. 680–689, Mar. 2016. DOI: 10.1109/TITS.2015.2480963.
- [4] K. Clement-Nyns, E. Haesen, and J. Driesen, "The impact of charging plug-in hybrid electric vehicles on a residential distribution grid", *IEEE Trans. Power Syst.*, vol. 25, no. 1, pp. 371–380, Feb. 2010. DOI: 10.1109/TPWRS.2009.2036481.
- [5] F. Rassaei, W. Soh, and K. Chua, "Demand response for residential electric vehicles with random usage patterns in smart grids", *IEEE Trans. Sustainable Energy*, vol. 6, no. 4, pp. 1367–1376, Oct. 2015. DOI: 10.1109/TSTE.2015.2438037.
- [6] P. O. Rosell, R. V. Robles, A. Sumper, and J. B. Jane, "Probabilistic agent-based model of electric vehicle charging demand to analyse the impact on distribution networks", *Energies*, vol. 8, no. 5, pp. 4160–4187, May 2015. DOI: 10.3390/en8054160.
- [7] Z. Luo, Z. Hu, Y. Song *et al.*, "Study on plug-in electric vehicles charging load calculating", *Automation of Electric Power Systems*, vol. 35, no. 14, pp. 36–42, Feb. 2011.
- [8] J. Zhao, F. Wen, A. Yang *et al.*, "Impacts of electric vehicles on power systems as well as the associated dispatching and control problem", *Automation of Electric Power Systems*, vol. 35, no. 14, pp. 2–10, Feb. 2011.
- [9] D. Steen, L. Tuan, O. Carlson, L. Bertling *et al.*, "Assessment of electric vehicle charging scenarios based on demo graphical data", *IEEE Trans. Smart Grid*, vol. 3, no. 3, pp. 1457–1468, Feb. 2012. DOI: 10.1109/TSG.2012.2195687.
- [10] M. D. Galus, R. A. Waraich, F. Noembrini *et al.*, "Integrating power systems, transport systems and vehicle technology for electric mobility impact assessment and efficient control", *IEEE Trans. Smart Grid*, vol. 3, no. 2, pp. 934–949, Sep. 2012. DOI: 10.1109/TSG.2012.2190628.
- [11] T. Yang, X. Xu, Q. Guo *et al.*, "EV charging behaviour analysis and modelling based on mobile crowdsensing data", *IET Generation, Transmission & Distribution*, vol. 11, no. 7, pp. 1683–1691, Jun.

2017. DOI: 10.1049/iet-gtd.2016.1200.
- [12] F. R. Salmasi, "Control strategies for hybrid electric vehicles: evolution, classification, comparison, and future trends", *IEEE Trans. Vehicular Technology*, vol. 56, no. 5, pp. 2393–2404, Sep. 2007. DOI: 10.1109/tvt.2007.899933.
- [13] A. Zare, A. Ozdemir, M. A. Iwen, and S. Aviyente, "Extension of PCA to higher order data structures: An introduction to tensors, tensor decompositions, and tensor PCA", in *Proc. of the IEEE*, vol. 106, no. 8, pp. 1341–1358, Jul. 2018. DOI: 10.1109/JPROC.2018.2848209.
- [14] Y. Wang, S. Xia, Q. Tang *et al.*, "A novel consistent random forest framework: Bernoulli Random Forests", *IEEE Trans. Neural Networks and Learning Systems*, vol. 29, no. 8, pp. 3510–3523, Aug. 2017. DOI: 10.1109/TNNLS.2017.2729778.
- [15] Y. Dong, B. Du, and L. Zhang, "Target detection based on Random Forest Metric Learning", *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 4, pp. 1830–1838, Apr. 2015. DOI: 10.1109/JSTARS.2015.2416255.
- [16] N. Soni, R. Malekian, D. Andriukaitis, D. Navikas, "Internet of vehicles based approach for road safety applications using sensor technologies", *Wireless personal communications*, vol. 105, iss. 4, pp. 1257–1284, 2019. DOI: 10.1007/s11277-019-06144-0.
- [17] Y. Li, Z. Guo, J. Yang, H. Fang, and Y. Hu, "Prediction of ship collision risk based on CART", *IET Intelligent Transport Systems*, vol. 12, no. 10, pp. 1345–1350, Nov. 2018. DOI: 10.1049/iet-its.2018.5281.
- [18] J. R. Webber and Y. P. Gupta, "A sieve bootstrap method for correlation analysis", *IEEE Trans. Automatic Control*, vol. 52, no. 6, pp. 1079–1081, Jun. 2007. DOI: 10.1109/TAC.2007.899079.
- [19] Z. Yu, H. Chen, J. You *et al.*, "Adaptive fuzzy consensus clustering framework for clustering analysis of cancer data", *IEEE/ACM Trans. Computational Biology and Bioinformatics*, vol. 12, no. 14, pp. 887–901, Nov. 2015. DOI: 10.1109/TCBB.2014.2359433.
- [20] M. B. Pouyan and M. Nourani, "Clustering single-cell expression data using random forest graphs", *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 4, pp. 1172–1181, May 2017. DOI: 10.1109/JBHI.2016.2565561.