A Data-driven Approach for Charging Characteristic Parameter Identification Method of Electric Vehicles

Zhenyu Gao¹, Zhonghua Chen¹, Qian Hu¹, Hua Xue², Chen Ling³ and Xiaowen Zhang²

¹ Hangzhou Electric Power Design Institute Co., Ltd., Hangzhou 310014, China
² College of Electrical Engineering, Shanghai University of Electric Power, Shanghai 200090, China
³ evdataanalysis@163.com

Abstract. In order to accurately identify new energy electric vehicles charging behaviour characteristic parameter, the theory of consistency is put forward based on the k-means clustering method. The complex coupling network including consistency control is introduced into the data clustering analysis to accurately describe the consistency characteristics of the electric vehicles charging load data in different periods, and the dissimilarity measure with the state update of the adjacent cluster data is used to quickly calculate the initial cluster center. The k-means method is used to quickly identify the expected value of EV initial charging time in typical scenes, and to accurately extract the probability distribution function of EV charging probability and charging initial time. Combined with a practical case, it is verified that the proposed method has the advantages of simple calculation, fast clustering.

1. Introduction

New energy electric vehicles have attracted increasing attention and application due to their advantages of superior handling performance, low pollution emission and high energy security [1]. However, with the electric car mass connected to the electricity grid charging load random impact load, the safe and stable operation of power grids caused adverse effects. For electric vehicle charging probability and charging time probability distribution function can comprehensively reflect the driving habits, mileage, holidays and seasonal factors on the electric car battery load model [2].

In order to achieve rapid and accurate modelling of EV charging load, the k-means method can be used to complete EV charging load data analysis and extract charging behaviour characteristic parameters [3]. However, in order to improve the accuracy of k-means clustering, the number of clustering centers needs to be accurately judged in advance. M.G. Quiles [4] et al. proposed a region detection method based on particle competition. Each particle can control as many data points as possible according to the set competition mechanism, and finally form a particle population with different characteristics to achieve the accurate calculation of clustering center. Using the idea of coupling resonance synchronization, L. Zhao [5] et al. cluster the data points with synchronization characteristics into a data group in a similar period of time. When all the data points reach a common state, the consistency control is realized [6]. Therefore, the consensus control method can make the coupled resonant network have the characteristics of consistency and synchronization. By applying the consistency control method, T. Chen [7] et al. proved that the complex coupling network could realize the expected group solution of data with different characteristics. Therefore, the key problem to be
solved in data-driven EV charging load modelling is to use the consensus control method to solve the selection problem of the initial clustering center of the k-means method, while maintaining the advantage of small computation amount of the k-means method, and realize the accurate online identification of EV charging behaviour characteristic parameters.

In order to solve the technical difficulties of the existing data-driven EV charging parameter identification methods, which are complicated in operation and difficult to achieve engineering application, a new EV charging parameter identification method based on consensus k-means clustering was proposed [8], which automatically completed the initial clustering center selection without human intervention and improved the clustering accuracy. The probability distribution function of EV charging period was quickly and accurately calculated to establish EV charging time distribution under typical scenes to improve the accuracy of the model.

2. K-means clustering method based on constraint consensus control

2.1. Consensus control method of containment

2.1.1 Initial cluster center calculation. A constraint consensus control method was designed, and each data updated its own clustering state according to the dissimilarity measure with neighboring data. When all data completed the updating process, the preliminary selection of clustering center was completed. Based on the consensus control theory, the dissimilarity measure $d_i$ between data point $i$ and data point $j$ in the charging load data network is defined as:

$$d_i = \sum_{j \neq p} a_{ij} [V_j(t_j) - V_i(t_i)]$$ (1)

Where, $V(t)$ represents the charging load of electric vehicles at time $t$. If there is a communication contact edge between data point $i$ and data point $j$, $a_{ij} \neq 0$. If there is no communication contact edge between data point $i$ and data point $j$, $a_{ij} = 0$.

Since the data network formed by EV charging load and charging start time is undirected, after the dissimilarity measurement $d_i$ calculation of all data points is completed according to Equation (1), the ideal state reached gradually is to divide the clustering boundary with the collective decision-making average $\alpha = n^{-1} \sum_{i=1}^{n} V_i(t_i)$, and the adjacent data points close to $d_i$ form the initial clustering.

However, the clustering boundary of collective decision mean $\alpha$ is not reasonable in practical application. Therefore, the consensus control method is proposed, and the dissimilarity measurement $d_i$ between data point $i$ and data point $j$ is designed to better meet the requirements of engineering application:

$$d_i = f[V_i(t_i)] - \varepsilon \sum_{j=1}^{n} [a_{ij} h[V_j(t_j) - V_i(t_i)] - u_i(t_i)]$$ (2)

Where: $u_i(t_i) = g_i [h(s(t_i) - V_i(t_i))]$, $f[V(t_i)]$ as self feedback function, data consensus control parameters $\varepsilon > 0$ said network coupling strength, $h(\cdot)$ for the interaction between adjacent data points, the internal coupling function of $u_i(t_i)$ as consensus control containment, $s(t_i)$ is expected to clustering boundary, contain control parameters $g_i$, a stronghold of traction control gain, when $i$ was elected to pin point data point, $g_i = z > 0$; When the data point $i$ is not the containment point $g_i = 0$.

By comparing Equations (1) and (2), it can be seen that by introducing the consensus control constraint item, only a small number of local feedback controllers can be injected into the data network to achieve reasonable setting of the clustering boundary of the data network.
By type (2) analysis shows that contain consensus control $s(t_i)$ boils down to is expect clustering boundary restrain the consistency of the data points, to simplify the operation, both contain consensus, quickening the response of the control law in the electric vehicles charging load data network will only a small number of data points is set to contain, do not break general, choose every $p$ point data contain points, set up internal coupling function:

$$h(x) = x$$

(3)

The linear self-feedback function is

$$f[V_i(t_i)] = \beta V_i(t_i)$$

(4)

Equation (2) can be written as

$$d_i = \beta V_i(t_i) - \varepsilon \sum_{j=1}^{n} I_{ij} V_j(t_j) + \varepsilon g_i[s(t_i) - V_i(t_i)]$$

(5)

After completing the calculation of the dissimilarity measure $d_i$ between data point $i$ and data point $j$ according to Equation (6), gradually achieve the ideal state of divisions of expected clustering boundary $s(t_i)$ scales, with close to $d_i$ data points to form the initial clustering fast, determine the initial clustering center number k-means algorithm for k-means clustering accurately calculate the probability of electric vehicles charging and charging time probability distribution function preparation conditions.

2.1.2 K-means clustering algorithm based on consensus control. According to the initial number of clustering centers k obtained by the consensus control method, the weekly charging load data set $V$ of EV charging stations under typical scenes is divided into $k$ groups ($k < n$), and each group represents a class. On the basis of the initial clustering completed by the constraint consistency control, the k-means algorithm is adopted to calculate the similarity between the data points and construct the adjacent matrix. Through repeated iteration, the clustering center is updated and the grouping is refined until the clustering results no longer change, which is the optimal clustering result. The computational complexity of the k-means clustering method has a linear relationship with the data set size, and the calculation is fast. The constraint consistency control method completes the calculation of the initial number of clustering centers, which ensures the accuracy of the clustering results of the k-means algorithm.

The adjacent matrix was constructed by similarity measurement, and the two closest groups were found out and represented by G1 and G2. The two closest element points between the two groups were connected, and the average dissimilarity between the vertices in G1 and G2 of each group was calculated. If its dissimilarity is less than the defined threshold, merge G1 and G2 into a larger group.

When the number of groups reaches the pre-defined group number $k$, the grouping is ended, and the clustering center of each group is calculated. This process is iterated repeatedly until every $k$ clustering center value does not change, that is, the k-means clustering result is obtained, and the mathematical expected value of the $k$ clustering center value is obtained.

Figure 1 shows the clustering results of a random data set. The proposed k-means clustering method based on constraint consistency control is applied to perform clustering analysis on the data set. Information is exchanged between each data point and at least the nearest two data points. As can be seen from the analysis of Figure 1, the proposed constraint consistency control method rapidly forms the initial clustering by measuring the dissimilarity between adjacent data points, and determines the number of initial clustering centers of the k-means algorithm, which has good convergence. The proposed k-means clustering method can quickly complete the accurate calculation of each grouping clustering center, and has a better clustering effect.

2.1.3 Calculation of probability distribution function of starting time of electric vehicles charging. The probability distribution function of EV charging start time can reflect the relationship between EV
users’ charging demand and time [9]. Taking the mathematical expectation value of the clustering center obtained by the consensus k-means clustering method as the expected value $\mu_c$ of the initial charging time of the charging pile, the probability distribution function of the initial charging time of electric vehicles in the charging station can be written as:

$$h(t) = \left(\sigma^2\pi^{0.5}\right)^{-1} e^{-\frac{(t-\mu)^2}{2\sigma^2}}$$

(6)

Where, $\sigma$ is the variance of the initial charging time of the charging pile. The flow chart of the system is shown in figure 2.

3. Case studies
Based on the charging load data of an EV charging station in a financial center in Hangzhou in 2020, clustering analysis was conducted on the charging load data of EV charging stations in each season to test the correctness and feasibility of the proposed EV charging load modelling method based on consensus k-means clustering.

3.1 Consensus k-means clustering performance test
In 2020, spring, summer, autumn and winter in financial center electric vehicles charging station a week were charging load data: January 6-January 12 (324 points), April 13-April 19th(368 points), August 10-August 16(352 points),November 9-November 15(307 points), validate the proposed consistency k-means clustering method is feasible. Figure 3(a)-(d) shows the clustering analysis results of charging load data in four natural weeks. In the figure, the abscissa is the starting time of charging, and the ordinate is the charging load of electric vehicles.

From the analysis of figure 3, the use of check consensus control method, through charging load data set between adjacent data points not similarity measure computation, fast accurate finished the initial clustering charging load data, through the k-means clustering method of realization of electric vehicles charging feature difference between different set of classes is clear, 6-13 financial center is relatively high demand for electric vehicles charging, at around 9 in the peak load. From 0 to 6 o'clock, 13-16
o'clock and 21-24 o'clock, EV charging demand is stable. From 16 to 21 o'clock, EV charging demand enters the sub-peak stage. Figure 1 results show that the proposed method uses no similarity measure between adjacent data points can quickly determine the initial clustering center, electric vehicles load data using k-means clustering method to accurately capture the clustering center of mathematical expectation, fast convergence, calculate the accurate clustering center, prepared condition for electric vehicles charging characteristic parameter identification. According to the measured data, the k-means algorithm, the consensus k-means clustering method of two communication edges, the one-property k-means clustering method of three communication edges and the consensus k-means clustering method of four communication edges are adopted to conduct quantitative analysis on the clustering center of the surrounding charging load data. The results are shown in table 1. Known from the analysis of table 1, the proposed contain consistency control option 2, 3, 4 respectively communication number of edges, with the result of clustering center is not big, the influence of error is very small, not similarity measure can rapid convergence to near the mathematical expectation, the k-means clustering algorithm of clustering center accurate characterization of the complex coupling the consistency of the data network features in different seasons, so the communication number of edges is different, for K-means clustering is small, the influence of the clustering of high precision, good stability.
Table 1. Quantitative comparison of cluster centers with the traditional k-means clustering method and the proposed consensus k-means clustering method

|                        | k-means algorithm | improved algorithm \((m=2)\) | improved algorithm \((m=3)\) | improved algorithm \((m=4)\) |
|------------------------|-------------------|-------------------------------|-------------------------------|-------------------------------|
| the first cluster center | 3.01 ± 0.07       | 2.89 ± 0.02                   | 2.90 ± 0.02                   | 2.88 ± 0.02                   |
| the second cluster center | 8.98 ± 0.08       | 9.14 ± 0.02                   | 9.13 ± 0.01                   | 9.12 ± 0.01                   |
| the third cluster center | 19.30 ± 0.07      | 19.33 ± 0.02                  | 19.32 ± 0.02                  | 19.34 ± 0.01                  |

3.2 Parameter extraction performance test

![Figure 4](image-url) Distribution function of starting time of EV charging.

Combined with figure 3 and table 1, based on the mathematical expectation value of the clustering center obtained by the consensus k-means clustering method, the distribution function of EV charging starting time can be obtained from Equation (6), as figure 4. Among them: \( \sigma_c \) set the variance of the initial charging time of commercial charging pile as 1.68 based on the data of National Household Travel Survey (NHTS) collected by the US Department of Transportation in 2017 [10].

4. Conclusion

Complex coupling data network consistency control method is introduced to analyse the electric car charging parameters, only calculation between adjacent data points of similarity measure, for k-means clustering can quickly identify the initial clustering center, by k-means clustering method to complete the electric vehicles charging characteristic parameters to calculate accurately, the proposed k-means clustering method based on consensus check for electric vehicles charging load unconstrained data set size and shape, with a small amount of calculation, fast clustering, the characteristics of the feature parameter identification is accurate.

5. References

[1]  S A Khan, M R Islam, Y Guo and J Zhu 2019 A new isolated multi-port converter with multi-directional power flow capabilities for smart electric vehicle charging stations[J] IEEE Transactions on Applied Superconductivity vol 29 no 2 pp 1-4.
[2]  H Jiang, S Ning and Q Ge 2019 Multi-objective optimal dispatching of microgrid with large-scale electric vehicles[J] IEEE Access vol 7 pp 145880-8.
[3]  Kwon S H, Bae Y B, Liu J and Ahn H S 2020 From matrix-weighted consensus to multipartite average consensus IEEE Transactions on Control of Network Systems vol 7 no 4 pp 1609 – 20
[4]  M G Quiles, L Zhao and R L Alonso 2008 Particle competition for complex network community detection[J] Chaos vol 18 no 3 pp 033107.
[5]  L Zhao and T H Cupertino 2008 Chaotic synchronization in general network topology for scene segmentation[J] Neurocomputing vol 71 no 16 pp 3360-6.
[6] S H Kwon, Y B Bae, J Liu and H S Ahn 2020 From matrix-weighted consensus to multipartite average consensus[J] IEEE Transactions on Control of Network Systems vol 7 no 4 pp 1609-20.

[7] T Chen, X Liu and W Lu 2007 Pinning complex networks by a single controller[J] Transactions on Circuits and Systems I: Regular Papers vol 54 no 6 pp 1317–26.

[8] Cupertino T H, Huertas J and Zhao L 2013 Data clustering using controlled consensus in complex networks Neurocomputing vol 118 pp 132 – 40

[9] Liu P and Yu J 2018 Identification of charging behavior characteristic for large-scale heterogeneous electric vehicle fleet Journal of Modern Power Systems and Clean Energy vol 6 no 3 pp 561 – 81

[10] Jia J 2019 Analysis of alternative fuel vehicle (AFV) adoption utilizing different machine learning methods: a case study of 2017 NHTS IEEE Access vol 7 pp 112726 – 35