

PAPER • OPEN ACCESS

A Virtual Power Plant Load Curve Clustering Method Based on Improved K-means Algorithm and Its Application

To cite this article: Hui Li *et al* 2020 *IOP Conf. Ser.: Earth Environ. Sci.* **619** 012055

View the [article online](#) for updates and enhancements.

You may also like

- [Design of Virtual Power Plant System Model under Master-Slave Multi-chain](#)
Qingshui Xue, Zhen Xue, Haifeng Ma et al.
- [Research on declaration strategy of virtual power plant spot market considering transaction price risk](#)
Jie He, Xi Luo and Wen Zhao
- [Research on the development trend of virtual power plant in the spot market](#)
Yicheng Mu



ECS
The
Electrochemical
Society
Advancing solid state &
electrochemical science & technology

DISCOVER
how sustainability
intersects with
electrochemistry & solid
state science research

A Virtual Power Plant Load Curve Clustering Method Based on Improved K-means Algorithm and Its Application

Hui Li¹, Lang Zhao^{1,a}, Dong Peng¹, Zhidong Wang¹, Xueying Wang¹ and Xin Ai²

¹State Grid Economic and Technological Research Institute Co., Ltd., Beijing 102209, China;

² School of Electrical and Electronic Engineering, North China Electric Power University, Beijing 102206, China;

^azhaolang@chinasperi.sgcc.com.cn

Abstract. In view of how virtual power plants can effectively participate in power grid operation, a method of load curve clustering of virtual power plants based on principal component analysis reduction and aggregation level clustering and k-means clustering is proposed, and the application of clustering results is studied. Firstly, combined with the data obtained from the information physical network, the principal component analysis method is adopted to analyze the characteristics of different loads participating in the virtual power plant aggregation, so as to standardize the data and reduce the dimension. Then, the algorithm combining aggregation hierarchical clustering and k-means clustering is used to cluster all load output curves participating in the aggregation, to obtain load curve clusters of the same class and find out the clustering center. Finally, the clustering results are analyzed, and the corresponding evaluation system is established. Through comprehensive evaluation, appropriate load combinations are selected to participate in the virtual power plant aggregation.

1. Introduction

In the context of the construction of the Energy Internet, small and micro smart sensors will be widely distributed in all corners of the power grid. The data obtained by the power terminal is huge, which will promote the transformation of all power and related businesses to intensive, intelligent and automated [1]. Virtual Power Plant (VPP) as the specific form and basic unit of the energy Internet, has various aggregation methods. This paper analyzes the output curve characteristics of each load, uses artificial intelligence technology to aggregate loads with different characteristics, uses clustering centers to describe their output characteristics, incorporates these characteristics into the aggregation indicators of the virtual power plant, and uses comprehensive evaluation to determine the final aggregation program [2].

There are many ways to aggregate virtual power plants. First, the dimension of the resource data participating in the aggregation of virtual power plants is reduced, and then the clustering is performed. At present, the application of cluster analysis in power systems is mainly focused on load forecasting. With the construction of the ubiquitous power Internet of Things, cluster analysis will be applied to many levels of the power grid. Literature [3] uses the k-means clustering method to model through a multi-scenario method to describe the uncertainty of the load. After obtaining the cluster centers of various clusters participating in the virtual power plant aggregation, this paper establishes a matching index system to determine the optimal aggregation method [4]. The wind power and photovoltaic



power plants which participate in the aggregation of virtual power plants are taken as examples. In order to highlight the seasonality of wind power and photovoltaic, this paper, based on the data of one year, aggregates the curves by quarter and finds their clustering centers. The complex curve is divided into several specific clusters, the curve characteristics of each cluster are analyzed, and different combination methods are comprehensively evaluated to determine the optimal load combination.

2. Principal component analysis method

Massive data brings great value to the prediction, aggregation and optimal scheduling of virtual power plants, and also increases the computational burden. Data dimensionality reduction can effectively reduce computational pressure.

Assuming that there are n output curves of wind power, photovoltaic, and energy storage participating in virtual power plant scheduling, each curve is a p -dimensional variable, and the sample matrix \mathbf{X} composed of these curves is:

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} \quad (1)$$

where x_{ij} is the i -th variable of the j -th sample. Record the original variable index as $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p$ set the new variable, that is, the comprehensive index after dimensionality reduction processing, as $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_m$ ($m \leq p$), then

$$\begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \\ \vdots \\ \mathbf{z}_m \end{bmatrix} = \begin{bmatrix} l_{11} & l_{12} & \dots & l_{1p} \\ l_{21} & l_{22} & \dots & l_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ l_{m1} & l_{m2} & \dots & l_{mp} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_p \end{bmatrix} \quad (2)$$

The new variables $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_m$ are called as the original variable, the first, second, ..., m -th principal component of $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p$, and the coefficient l_{ij} is called the principal component loading. First, the matrix \mathbf{X} is standardized, and then the correlation coefficient matrix \mathbf{R} of the standardized matrix is calculated:

$$\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1p} \\ r_{21} & r_{22} & \dots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \dots & r_{pp} \end{bmatrix} \quad (3)$$

Use Jacobi method to solve the characteristic equation, find the characteristic value, and find the corresponding unit characteristic vector. Among them, r_{ij} can be expressed as:

$$r_{ij} = \frac{\sum_{k=1}^n (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)}{\sqrt{\sum_{k=1}^n (x_{ki} - \bar{x}_i)^2 (x_{kj} - \bar{x}_j)^2}}, i, j = 1, 2, L, p \quad (4)$$

The score of each principal component can be obtained:

$$\mathbf{Z} = \begin{bmatrix} z_{11} & z_{12} & L & z_{1m} \\ z_{21} & z_{22} & L & z_{2m} \\ M & M & O & M \\ z_{n1} & z_{n2} & L & z_{nm} \end{bmatrix} \quad (5)$$

At this point, the dimensionality reduction of various types of load curves of the virtual power plant has been achieved, and in the subsequent calculations, these m principal components are used to replace the original complex data for calculation, so as to simplify calculations and improve efficiency.

3. K-means clustering method based on agglomerative hierarchical clustering

Hierarchical clustering methods are divided into two types: cohesive hierarchical clustering and split hierarchical clustering. This article uses the more commonly used agglomerated hierarchical clustering

[5]. First, consider n objects as n clusters, calculate the Euclidean distance d_{ij} between clusters:

$$d_{ij} = \sqrt{(z_{i1} - z_{j1})^2 + (z_{i2} - z_{j2})^2 + L + (z_{im} - z_{jm})^2} \quad (6)$$

In the formula, d_{ij} represents the distance between the i -th curve and the j -th curve; $z_{i1}, z_{j1}, z_{i2}, z_{j2}, K, z_{im}, z_{jm}$ is the corresponding principal component score in the \mathbf{Z} matrix; n clusters have a total of $n(n-1)/2$ distances. Arrange the distances from small to large, and the array contains marks for n objects.

Use agglomerative hierarchical clustering to determine the initial cluster center to perform k-means clustering, the specific steps are as follows:

- (1) Determine the number of clusters k ;
- (2) Input the \mathbf{Z} matrix and perform agglomerated hierarchical clustering analysis on the data set;
- (3) Calculate the average value of the obtained data of k clusters to obtain the initial cluster center;
- (4) Using k-means clustering, calculate the distance from each data to all cluster centers, and classify the data into the cluster where the nearest cluster center is located;
- (5) After the classification is completed, recalculate the cluster center of each cluster;
- (6) Repeat the fourth and fifth steps until the cluster center no longer changes or no data is reassigned to other clusters.

4. Virtual power plant load combination scheme based on comprehensive evaluation

This article mainly considers economy and environmental protection. The economics considers the virtual power plant's wind power, photovoltaic power generation benefits, energy curtailment costs and other factors. Used to measure the clean energy consumption level of virtual power plants. The higher the economy and environmental protection, the better the effect of participating in virtual power plant aggregation on behalf of wind and light. Curve characteristics refer to several curve clusters obtained according to improved k-means clustering, analyze the curve characteristics of different clusters, and use the curve characteristics and scale of each cluster to measure the characteristics of renewable energy itself, mainly including output characteristics and reliability [6],

where the output characteristics consider factors such as the daily load rate of energy, the daily peak-valley rate, and the daily load fluctuation rate, and the reliability considers the percentage of time that the curve output of this type of cluster meets the threshold every day and the type of cluster within a specific time. The abnormal rate of the curve, better the output characteristics and higher the reliability. This paper selects the eight indicators shown in Table 1 as the comprehensive evaluation indicators for guiding load combination, and gives the quantitative calculation methods of each indicator.

Table 1. Comprehensive evaluation index system

Selected indicators Object of evaluation	Level I indicators	Level II indicators
Characteristics of VPP	Economy	Wind and light generation benefits
		Cost of abandonment
	Environmental protection	Clean energy uptake rate
Curves of characteristics	Force characteristics	Daily load rate
		Peak-to-valley differential
		Daily load fluctuation
	Reliability	Proportion of time meeting threshold
		Number of failures

In order to determine the optimal load combination and ensure energy efficiency, this paper evaluates all load combinations in the number of combination clusters. After obtaining all the load combinations that meet the conditions, the entropy weight method is used to determine the index weight, and the final score of each combination is weighted according to the results. For a given number of aggregated clusters, compare the evaluation scores under different combinations, select the best load aggregation method for virtual power plants under the indicators considered in this paper.

5. Case analysis

This paper is based on the actual power of the 7 photovoltaic sites, 8 onshore wind power and 4 offshore wind power sites in Europe in 2017, taking into account the seasonality of the scenery, and clustering on a quarterly basis. Take the data of the first quarter as an example, take 15 minutes as an interval, draw a total of 1710 curves of 19 load 90 days for clustering, and study the law of output. The normalized output distribution of all loads is shown in Figure 1. Each curve is the normalized output distribution of a certain load in one day.

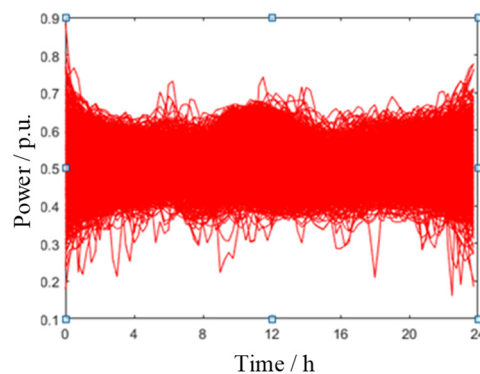


Figure 1. Standardized data output distribution

First, perform principal component analysis, take the cumulative contribution rate of 85%, and obtain the eigenvalues that meet the conditions and their contributions as shown in Table 2.

Table 2. Eigenvalue and contribution rate

Eigenvalues	39.97	27.14	6.31	5.56
Contribution rate	45.35%	30.79%	7.16%	6.31%

Then proceed to agglomerative hierarchical clustering. After the validity test, the clustering effect is best when the number of clusters is 8. After obtaining 8 initial cluster centers, perform k-means clustering. The clustering results are shown in Figure 4, where the red curve is the final cluster center of the cluster.

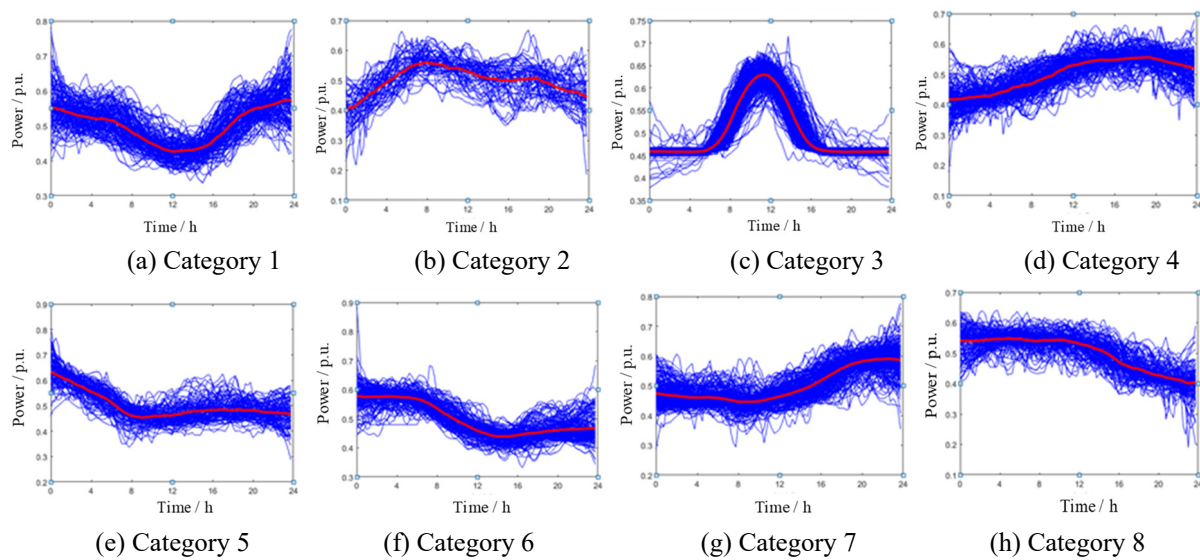


Figure 2. Clustering results

The numbers of 1-8 types of curves are: 155, 65, 631, 190, 113, 170, 246, 140, the characteristics of each cluster curve are more obvious. Secondly, the performance of the two algorithms is quantitatively compared through the effectiveness index. The improved k-means algorithm has shorter clustering time, and X is smaller than the traditional method, and the clustering effect is better.

After the clustering is completed, the effect of different cluster loads participating in VPP aggregation is evaluated according to the clustering results, and the optimal solution for aggregation is determined. Different types of resource output characteristics are different, so this article uses clusters as the aggregation unit, assuming that the number of clusters that can participate in the aggregation is greater than or equal to 4, and a total of $C_8^4 + C_8^5 + C_8^6 + C_8^7 + C_8^8 = 163$ combination methods are used to comprehensively evaluate and score different aggregation methods.

As the number of combination clusters increases, the economy and environmental protection of this method will increase. In order to reduce variables, this article takes the number of cluster clusters equal to 7 as an example to compare the advantages and disadvantages of the 8 combination methods, as shown in Table 3.

Table 3. Score and combination of each scheme

Rank	Score	Combination situation
1	0.575 648	Removal of Class 7 clusters
2	0.574 083	Removal of Class 5 clusters
3	0.573 546	Removal of Class 6 clusters

4	0.568 468	Removal of Class 1 clusters
5	0.558 332	Removal of Class 4 clusters
6	0.557 230	Removal of Class 2 clusters
7	0.554 821	Removal of Class 8 clusters
8	0.522 117	Removal of Class 3 clusters

Considering the index system selected in this paper, the virtual power plant scores the highest when the 7th cluster is not included, and the aggregation effect is the best. The virtual power plant scores the lowest when the third cluster is not included, and the aggregation effect is the worst. It can be seen from Figure 2 that the third cluster is mainly the output curve of photovoltaic power plants. When photovoltaic output is excluded, both the power generation revenue and the clean energy consumption rate are significantly reduced, resulting in the lowest comprehensive evaluation score, which is consistent with the actual situation.

The method proposed in this paper can be used not only for the aggregation and evaluation of virtual power plant load curves, but also for dimensionality reduction and clustering of power curves in other scenarios. At the same time, it provides ideas for the evaluation of virtual power plant aggregation results. In the selection of indicators, the contribution of wind power and photovoltaics is mainly considered, which has certain limitations. Full consideration of the contribution of various types of loads of virtual power plants will make the evaluation results more accurate.

6. Conclusion

Considering the complex components of the virtual power plant load curve and the special operating conditions, this paper proposes a virtual power plant load curve clustering method based on PCA and improved k-means algorithm, and discusses its application. Clustering analysis of the load curves participating in the aggregation of virtual power plants through the principle component analysis method of dimensionality reduction and the clustering method combining agglomerated hierarchical clustering and k-means algorithm can solve the practical problem of virtual power plants participating in the operation of the power grid. The calculation example shows that the method can be applied to the clustering of power load curves, and the clustering effect is obvious. In the application of clustering results, a comprehensive evaluation index system for virtual power plants has been established, which can accurately reflect the effect of load participation in virtual power plant aggregation. It evaluates the possible load combination set, and verifies the rationality of the evaluation through calculation examples, which provides ideas for other related applications of clustering results.

Acknowledgment

This paper was supported by the science and technology project of State Grid Corporation of China 5102-201956310A-0-0-00, "Research on Intelligent Diagnosis Analysis and Comprehensive Decision Technology of Power Grid Development Based on Data Driven".

References

- [1] Shafie-Khah M and Catalão J P S. 2015 A stochastic multi-layer agent-based model to study electricity market participants behavior *IEEE Transactions on Power System* **30** 867-881
- [2] Peng X G, Lai J W, Chen Y. 2014 Application of clustering analysis in typical power consumption profile analysis *Power System Protection and Control* **19** 68-73
- [3] Ferraz B P, Júnior J D A, Lemos F, et al. 2019 Analysis of distribution system considering load uncertainties: a K-means clustering approach *ISGT Latin America* Gramado Brazil 1-6
- [4] Wand X D, Liu J Y, Liu Y B, et al. 2019 Typical load curve shape clustering algorithm based on adaptive piece wise aggregation approximation *Automation of Electric Power Systems*, **43** 110-121

- [5] Cao M M, Shi L J, Wang Z Z, et al. 2019 Node Net Load Forecasting of Distributed Photovoltaic Considering Different User Characteristics ISGT Asia Chengdu China 1766-1771
- [6] Nisha and Kaur P J 2015 Cluster quality-based performance evaluation of hierarchical clustering method NGCT Dehradun Nisha 649-653.