Scholar Publishing Group
Academic Journal of Energy
https://doi.org/10.38007/RE.2021.020102
ISSN 2790-2714 Vol. 2, Issue 1: 9-16

# *Energy Optimal Scheduling Problem Based on Clustering Algorithm*

**Ruoxin Lin and Linyuan Fan***

*College of Mathematics and Data Science, Minjiang University, Fuzhou 350108, China*

*fanlinyuan@mju.edu.cn*

**corresponding author*

***Keywords:*** Clustering Algorithm, Energy Crisis, K-means Algorithm, Simulation Experiment

***Abstract:*** The increasingly serious global energy crisis has provided important opportunities for the development of new energy. The concept of sustainable energy development advocated by the state has promoted the development and use of new energy, and a large number of new energy power generation units have been connected to the power system. K-means algorithm is a major breakthrough in the field of clustering analysis. Because of its convenience, it has been widely used in fraud detection, image processing, market analysis and other fields. In this paper, K-means algorithm is used for clustering mining of power grid data. In order to obtain more specific relationships between data, association rules mining based on Apriori algorithm is carried out for clustering results, and regions and systems with unreasonable power grid energy consumption are found; Finally, aiming at the problem area, a scheduling scheme is designed by using fuzzy control method to realize the optimal scheduling of power grid energy.

## 1. Introduction

As a major energy consumer, China's oil consumption ranks second in the world, and its coal and electricity consumption ranks first in the world. China's energy demand is increasing with the rapid economic growth, which leads to the continuous rise of energy prices. However, even if the energy prices rise, some regions will even face energy shortages. Specifically, the supply-demand relationship between electricity and coal is very tight, the oil price has been maintained at a historically high level, but the inventory of oil and coal has been at a low level for a long time [1-2]. Nowadays, the waste of resources in China is also quite serious. As a high-energy consumption industry, the energy consumption of steel industry alone accounts for 15% of the total energy consumption of the national economy [3]. At the same time, according to statistics, the average energy consumption required by eight high energy consumption industries such as steel industry and manufacturing industry in the production of unit products is higher than 47% of the energy consumption level under the world's advanced technology [4]. Improving energy efficiency and saving resources has become one of the important tasks for all walks of life to improve competitiveness, increase economic benefits and achieve sustainable development.

Today, with the continuous development and innovation of power system, distributed generation technology has gradually become the focus of power system economic operation research field because of its flexibility and economy [5]. In addition, distributed generation technology and large power grid can form effective complementarity, so that the reliability of power supply can be improved. It also has obvious advantages over ordinary distribution network in terms of environmental protection. Therefore, distributed energy generation can be regarded as a key direction for the development of power system in the future [6]. However, distributed energy also has obvious defects, that is, natural energy has strong uncertainty, such as wind and solar energy, which greatly limits the efficiency of distributed generation and the utilization of power grid. Therefore, the optimal scheduling of distributed energy is an urgent problem to be solved in the current power system. With the continuous development of modern science and technology, it has had a far-reaching impact on people's production mode [7]. The application of new technologies in industrial production has significantly improved both in improving production efficiency and in strengthening safety assurance. In the process of energy conservation and emission reduction, the active use of electronic communication technology and Internet technology can effectively promote the implementation work [8]. At present, the leading technology in the world is artificial intelligence technology, which can be applied to various fields. Of course, intelligent power consumption and industrial intelligence have emerged and are in a rapid development trend. The combination of artificial intelligence and new energy power generation will be a development trend, which can carry out real-time online monitoring of distributed generation power status, load information and other switchgear, Artificial intelligence technology can be used to predict the output and load demand of distributed energy units [9].

Based on the above background, this paper studies the real-time energy monitoring and energy management optimization system of the industrial park, monitors and analyzes the operation of distributed energy generation equipment in the industrial park, and adjusts the power generation output on the premise of considering the objective environment and the power generation characteristics of each power generation unit, so as to make effective use of resources and achieve economic optimization of microgrid operation. The establishment of the energy monitoring and dispatching system is of great significance. From the perspective of the user side, it greatly reduces the operation cost of the power system and greatly improves the energy efficiency. From the perspective of the power generation side, on the one hand, it improves the energy waste, reduces the energy loss and unnecessary waste, on the other hand, it greatly improves the power generation efficiency and ensures the reliability and uninterrupted use of electricity. By establishing an energy flow monitoring system, on the one hand, it can monitor the operation status of distributed energy, on the other hand, it can put forward optimal scheduling strategies for energy flow, improve power generation efficiency, make the microgrid operate under the economically optimal conditions, and provide real and effective data support and power generation strategies for intelligent power consumption, so as to improve the intelligent management level of power application links.

## 2. Research on Energy Optimal Scheduling Theory Based on Clustering Algorithm

### 2.1. K-means Algorithm

The algorithm minimizes the similarity between points (data) in each cluster and maximizes the similarity between clusters. After setting the K value, first randomly generate the initial K centroid positions, select appropriate distance measurement methods, such as Euclidean distance, Manhattan distance, Minkowski distance, etc., traverse and calculate the distance from each point to each centroid, and divide each point into the centroid closest to it according to the principle of minimum distance, so that each point is assigned to a cluster [10].

At this time, the centroid may not be the optimal location, so it is necessary to constantly update the centroid of each cluster to achieve the purpose of optimization. The method of updating the centroid position is to calculate the mean value of each point in the cluster, and the position of the mean value is the new cluster centroid position. After constantly seeking the mean value to constantly update the position of the cluster, until the centroid of the cluster does not change, and the range of change reaches the optimization standard, the algorithm ends, and the final clustering result is formed [11].

As shown in Table 1, the distance measurement in K-means clustering algorithm can choose many forms, such as Euclidean distance, Manhattan distance, Minkowski distance, etc. different distance measurement methods are adopted, but the calculation methods are different in the step of calculating the distance, and the other algorithm steps are the same. The distance mode used in practical application should be selected according to the characteristics of practical application [12-13].

*Table 1. Input and output of K-means clustering algorithm*

| Type | Content |
|---|---|
| Input | Data set A, number of clusters K |
| Output | K clustering centers |

### 2.2. Advantages and disadvantages of K-means clustering algorithm

As one of the classical clustering algorithms, K-means clustering algorithm has prominent advantages. First of all, its algorithm idea is relatively clear, the algorithm is simple to implement, and the running speed is fast. It is widely used in many fields [14]. However, K-means clustering algorithm also has some shortcomings. It has good clustering effect on spherical clusters and poor clustering performance on other forms of clusters; The k-value of K-means algorithm is artificially set, and the clustering effect of different K-values is significantly different; The initial value of K value is randomly generated. If the initial value happens to take a poor initial centroid, the program will terminate when the centroid may be locally optimal during update. The K-means measures the similarity of points in the cluster according to the distance. In many applications, other parameters such as time are required as clustering parameters. At this time, the distance measurement needs to be appropriately improved [15-16].

## 3. Data Mining of Energy Dispatching Based on Clustering Algorithm

### 3.1. Power data Clustering

Cluster analysis is a very active research topic in the research of data mining. It is different from the traditional classification. The classification model is sample data, and its class label has been known. Therefore, the purpose of classifying sample data is to find classification rules in samples, so as to classify other objects that do not have clear class labels. Clustering is to cluster all data without knowing any information about the classification of the target data [17]. Therefore, clustering analysis belongs to an unsupervised learning, which is a process of dividing the data set into multiple subsets, and finally making the objects in each subset similar and the objects in different subsets different. The goal of clustering mining is to find the nature of "class" of all data in multidimensional space. Therefore, the domain knowledge of data sources and the specific meaning of each data need not be considered in clustering, and variables can be regarded as one dimension in multidimensional space for data analysis.

Power data clustering analysis refers to the clustering analysis of power consumption data by analyzing the time series characteristics of the data itself, without considering the source and significance of the data, after preprocessing the collected power energy data to obtain the complete initial data to be analyzed. The clustering process in this chapter is to divide the daily power data into multidimensional data according to the time series characteristics of power consumption data, so as to characterize the characteristics of power consumption data in each region. Because Euclidean distance can intuitively show the similarity between multidimensional data, Euclidean distance is used as a measure to select the initial center point. Finally, the power data is classified after several iterations based on the Euclidean distance.

### 3.2. K-means Clustering Algorithm

The K-means clustering algorithm uses sum of squared errors (SSE) to optimize the centroid position. For the specified K clusters, the smaller the distance between the points in the cluster, the smaller the corresponding SSE, and the better the clustering effect.

$$SSE_1 = \sum_{m=1}^{j} \sum_{a_{m1 \in Km_1}}^{j} \|a_{m1} - Km_1\|^2 \quad (1)$$

Where sse1 represents the first SSE, M represents the variable representing the number of centroids during programming, and its integer value is from 1 to J. AM1 represents all data belonging to the m-th centroid after the first centroid determination, KM1 represents the m-th centroid determined for the first time, and AM1 ∈ KM1 represents that AM1 is divided into clusters centered on KM1.

Calculate the mean value of the points in each cluster as the new centroid K2:

$$K_2 = \{K1_2, K2_2, \ldots Kj_2\} \quad (2)$$

$$Kj_2 = \mathrm{mean}(a_{m1}) \quad (3)$$

In equation (3), mean () represents the function of finding the mean value. Through this function, the second centroid K2 can be determined, and the data in a is redistributed to the new cluster according to the principle of the nearest distance, so as to calculate the second SSE2:

$$SSE_2 = \sum_{m=1}^{j} \sum_{a_{m2 \in Km_2}} \|a_{m2} - Km_2\|^2 \quad (4)$$

Calculate the gap between SSE2 and sse1 Δ SSE2

$$\Delta SSE_2 = |SSE_1 - SSE_2| \tag{5}$$

Judgement Δ Whether SSE2 meets the error requirements, set δ Is the set positive number, which is the threshold of error judgment, which is determined according to the actual situation. if Δ SSE2 less than δ, Namely:

$$\Delta SSE_2 < \delta \tag{6}$$

It is considered that the clustering error meets the accuracy requirements, and the algorithm stops. If the requirements are not met, calculate the mean value of the points in each cluster with K2 as the centroid to form a new centroid K3, and redistribute the data in a to the new cluster according to the principle of the shortest distance. If sse3 is not calculated for the third time, calculate Δ Sse3 and determine whether to terminate or cycle again until the end of the algorithm [18].

## 4. Analysis of Energy Scheduling Based on Clustering Algorithm

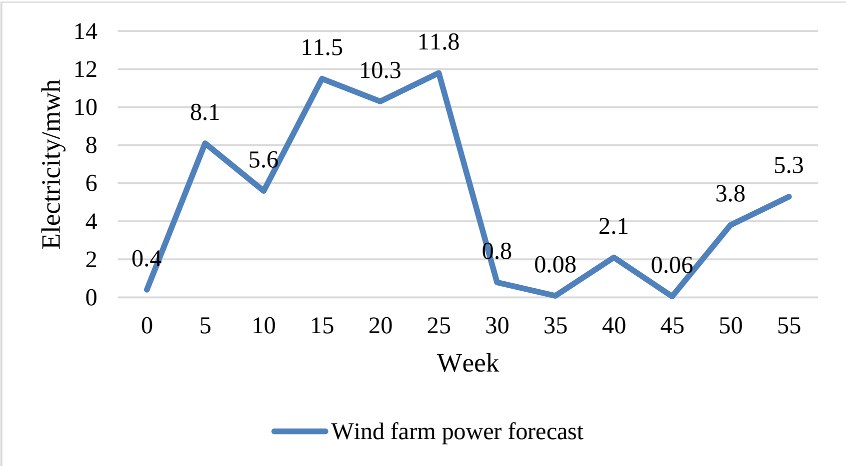


*Figure 1. Wind farm power prediction curve*

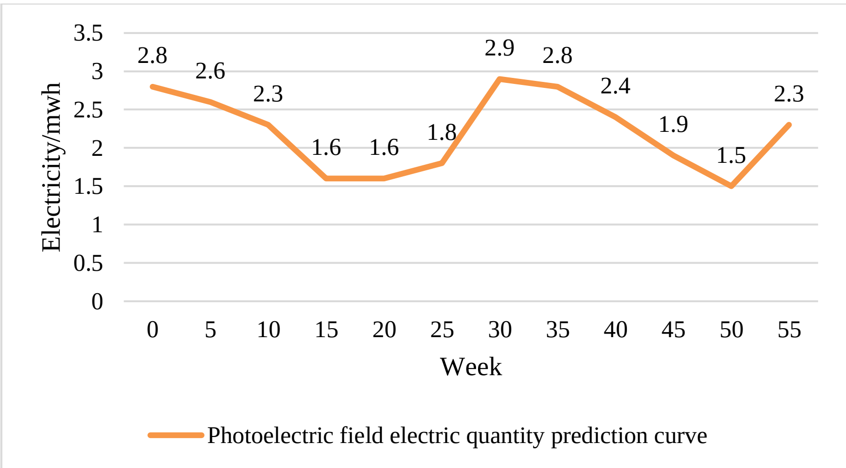


*Figure 2. Photoelectric field electric quantity prediction curve*

As shown in Figure 1 and figure 2, according to the 55 week simulation test, the energy optimal scheduling based on clustering method can more accurately reflect the actual energy use and scheduling, with less error and higher accuracy.

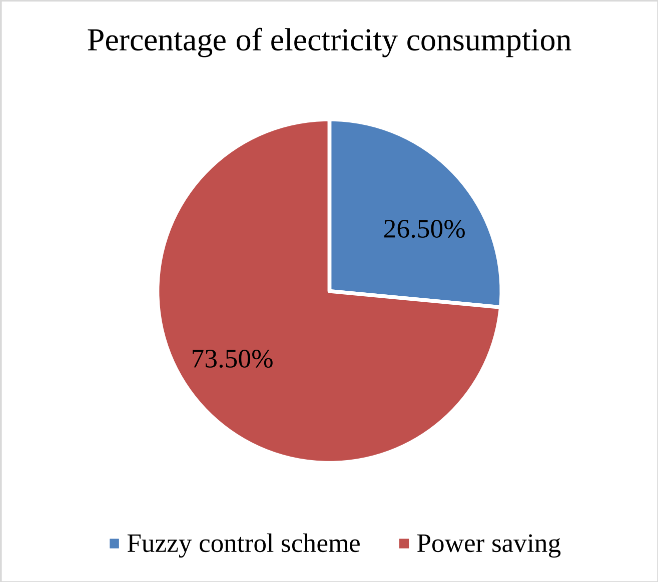


*Figure 3. Power consumption analysis diagram*

It can be seen from Figure 3 that in the power consumption test, the power consumption of fuzzy control scheme is 73.5%, and the power saving is 26.5%, which can well meet the daily industrial power demand, and also save energy expenditure to a certain extent.

*Table 2. The results of electricity decomposition index*

| Scene | Maintenance cost | Abandoned wind power | Waste electricity | Mean square deviation of progress coefficient |
|---|---|---|---|---|
| 1 | 4.3253*10^6 | 1.1333*10^6 | 1.0551*10^6 | 1.052 |
| 2 | 4.0530*10^6 | 1.0691*10^6 | 0.9532*10^6 | 1.259 |
| 3 | 3.9913*10^6 | 0.8388*10^6 | 0.1368*10^6 | 1.022 |
| 4 | None | 1.2038*10^6 | 1.2684*10^6 | 0.932 |

In order to illustrate the effectiveness of the model, this paper designs the following four scenarios.

(1) Scenario 1: decompose the contracted electricity and jointly optimize the maintenance plan.

(2) Scenario 2: take the maintenance cost as the optimization goal, and carry out the joint optimization of contract power decomposition and maintenance plan.

(3) Scenario 3: contract electricity decomposition model of multi-energy power system, that is, the model proposed in this paper. Decompose the contracted electricity and jointly optimize the maintenance plan by discarding the wind, electricity and maintenance costs.

Scenario 4: using the model in the literature, only the contracted electricity is decomposed into weeks, without considering the unit maintenance plan.

## 5. Conclusion

In the study of energy optimal scheduling based on clustering algorithm, four scenarios are designed to verify the effectiveness of the annual optimal scheduling model of multi-energy power system. Through the analysis of the example results, it shows the significance of the joint optimization of electricity decomposition and maintenance plan, and the superiority of the multi-objective model in maintenance economy and new energy consumption. The comprehensive analysis shows that the K-means online algorithm is more stable in simulation, and the simulation shows that the K-means online algorithm converges faster, which makes the charging system more stable in practical application.

## Funding

This article is not supported by any foundation.

## Data Availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

## Conflict of Interest

The author states that this article has no conflict of interest.

## References

*[1] Ma H . An Energy-Optimal Clustering Model Based on Event Region 2018(3):172-176.*
*[2] Li G S , Lin Q Y , Wu J H , et al. Dynamic Computation Offloading Based on Graph Partitioning in Mobile Edge Computing. IEEE Access, 2019(6), 7:1-1.*
*[3] Cheng K C , Aeron S , Hughes M C , et al. Optimal Transport Based Change Point Detection and Time Series Segment Clustering. 2019(3):13-14.*
*[4] Dharwadkar N V , Poojara S R , Kadam P M . Fault Tolerant and Optimal Task Clustering for Scientific Workflow in Cloud. International Journal of Cloud Applications & Computing, 2018, 8(3):1-19. https://doi.org/10.4018/IJCAC.2018070101*
*[5] Khafaf N A , Jalili M , Sokolowski P . A Novel Clustering Index to Find Optimal Cluster Size with Application to Segmentation of Energy Consumers. IEEE Transactions on Industrial Informatics, 2020, PP(99):1-1.*
*[6] Joo U G . Makespan Minimization Scheduling Problem with Energy-Efficient Turning On/Off Mechanism. Journal of the Korean Institute of Industrial Engineers, 2018, 44(1):1-8. https://doi.org/10.7232/JKIIE.2018.44.1.001*
*[7] Yadav P , Yadav V K , Yadav S . Distributed Energy Efficient Clustering Algorithm to Optimal Cluster Head by Using Biogeography Based Optimization. Materials Today: Proceedings, 2018, 5(1):1545-1551.*
*[8] Sharma V , Bala M . A Credits Based Scheduling Algorithm with K-means Clustering// 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC). 2018(1):5-6.*
*[9] Agbehadji I E , Millham R C , Abayomi A , et al. Clustering algorithm based on nature-inspired*

*approach for energy optimization in heterogeneous wireless sensor network. Applied Soft Computing, 2021, 104(2016):107-171. https://doi.org/10.1016/j.asoc.2021.107171*

*[10] Allahvirdizadeh Y , Galvani S , Shayanfar H . Data clustering based probabilistic optimal scheduling of an energy hub considering risk-averse. International Journal of Electrical Power & Energy Systems, 2021(4):106-107.*

*[11] Sundararaj V , Muthukumar S , Kumar R S . An Optimal Cluster Formation Based Energy Efficient Dynamic Scheduling Hybrid MAC Protocol for Heavy Traffic Load in Wireless Sensor Networks. Computers & Security, 2018, 77(AUG.):277-288.*

*[12] Moghaddam M M , Marzband M , Parhizi N . Optimal energy scheduling for a grid connected Microgrid based on Multi-period Imperialist competition algorithm// IEEE International Conference on Power & Energy. IEEE, 2017(3):113-114.*

*[13] Nazari-Heris M , Mohammadi-Ivatloo B , Haghrah A . Optimal short-term generation scheduling of hydrothermal systems by implementation of real-coded genetic algorithm based on improved Mühlenbein mutation. Energy, 2017, 128(jun.1):77-85.*

*[14] Yuan R , Dinh N T , Pipada Y , et al. Optimal activity and battery scheduling algorithm using load and solar generation forecast. 2021(8):56-58.*

*[15] Kim T H , Shin H , Kwag K , et al. A parallel multi-period optimal scheduling algorithm in microgrids with energy storage systems using decomposed inter-temporal constraints. Energy, 2020(2):117-169.*

*[16] Rehman A U , Wadud Z , Elavarasan R M , et al. An Optimal Power Usage Scheduling in Smart Grid integrated with Renewable Energy Sources for Energy Management. IEEE Access, 2021, PP(99):1-1.*

*[17] Samriya J K , Kumar N . An optimal SLA based task scheduling aid of hybrid fuzzy TOPSIS-PSO algorithm in cloud environment. Materials Today: Proceedings, 2020(12):96-97.*

*[18] Bornapour M , Hooshmand R A , Parastegari M . An Efficient Scenario-Based Stochastic Programming method for Optimal Scheduling of CHP-PEMFC, WT, PV and Hydrogen Storage Units in Micro Grids. Renewable Energy, 2018, 130(JAN.):1049-1066.*