International Journal of Multiphysics
Volume 18, No. 2, 2024
ISSN: 1750-9548

A Method for Extracting Generalized Typical Operating Modes of High-Proportion Renewable Energy Power Systems Driven by a Combination of Data and Knowledge

Xiaobiao Fu[1], Xu Jiang[1], Xinmeng Li[2,*], Yunpeng Li[1], Xin Liu[3], Jiakai Wu[2]

[1]Power Dispatching Control Center, State Grid Jilin Electric Power Supply Company, Changchun 130021, China

[2]School of Electrical Engineering, Northeast Electric Power University, Jilin 132012, China

[3]State Grid Changchun Electric Power Supply Company, Changchun 130021, China

*Corresponding Author.

Abstract

Typical operation modes serve as an important basis for guiding the operation of power systems and identifying potential safety vulnerabilities. In renewable energy power systems, however, the operation modes exhibit a trend of de-typification, making it difficult for typical modes selected based on manual experience to fully capture the impact of complex operating conditions of renewable energy on grid security. To address this issue, this paper proposes a data-knowledge hybrid-driven method for extracting generalized typical operation modes in high-proportion renewable energy power systems. Firstly, the key operational characteristic sequences of the system are selected, and the basic dataset of operation modes is divided using a hierarchical clustering method, forming a classification sample set of operation modes represented by various cluster centers. Then, considering differences in renewable energy penetration rates, load levels, and other factors, the operational point samples within each category with boundary operational characteristics are further subdivided to form a set of samples for verification. Finally, N-1 security checks are conducted on the operation modes in the verification set, and operation modes with similar safety issues are reduced to form a generalized typical operation mode set. Additionally, safety risk indicators are used to visually represent and evaluate the generalized typical operation modes in the risk characteristic space. The effectiveness of this method is validated using the actual topology and historical operational data of a provincial power grid, providing strong support for the operational departments in formulating grid operation rules and ensuring grid safety.

Keywords: Knowledge-data hybrid-driven, hierarchical clustering, generalized typical operation modes, safety risk, operational dataset.

1. Introduction

The operation mode of a power grid is determined by the combination of various power system elements such as unit status, grid topology, AC/DC line flows, and node load sizes [1]. In traditional power grids, the direction and magnitude of power flows are relatively fixed, and typical operation modes can be selected from historical peak, valley, and flat periods [2-4]. Power grid operators can formulate grid operation regulations by analyzing and calculating typical operation mode sets [5]. However, with the increasing proportion of wind and solar power generation and pumped storage in the system, the operational scenarios of the power grid are becoming increasingly atypical. This poses significant engineering challenges for efficiently and comprehensively

obtaining typical operation mode samples while considering the risk conditions of safe and stable grid operation and addressing the limitations of traditional methods that select typical operation modes from a single scenario [6-8].

Current methods for extracting typical operation scenarios and generating samples can be mainly divided into knowledge-driven methods based on expert experience and data-driven methods [9-10]. Knowledge-driven methods often plan thermal power installation capacity based on the maximum annual load, thereby extracting scenarios including seasonal variations in power generation modes, maximum or minimum load operation modes, and potential grid operation boundary modes [11]. However, as the grid structure becomes more complex, scenarios extracted based on manual experience are highly subjective and cannot simultaneously relate multiple factors causing extreme scenarios. To address this issue, existing studies have used data-driven methods to extract typical grid scenarios, which can be further divided into sample-driven stochastic analysis methods and machine learning and artificial intelligence methods [12-13]. Common sample generation methods include stochastic simulation and optimization models. The former obtains a large number of analyzable samples through random sampling, such as the Monte Carlo method; the latter establishes optimization models based on constraints and objectives to solve for samples [14]. However, the high dimensionality, nonlinearity, and uncertainty of power systems often lead to combinatorial explosion problems in stochastic simulation methods, making it difficult to generate typical samples in a targeted manner. Optimization models also face challenges of high computational complexity and long generation times. Therefore, the drawbacks of sample-driven stochastic analysis methods are their lack of targeting, low efficiency, and insufficient typical samples [15].

Machine learning and artificial intelligence methods can extract pattern rules from massive data, making them suitable for identifying complex operational characteristics of power systems. For instance, combining clustering methods with production simulation methods, selecting characteristic variables to represent operation features, reducing dimensionality, and extracting typical grid modes using K-means clustering algorithm can compensate for the lack of reference operational scenarios when using production simulation alone [16]. For future high-proportion renewable energy scenarios, capturing the spatial distribution characteristics between load and wind power unit output can adapt clustering to various shapes and sizes, optimizing computational efficiency [17]. Existing studies on extracting operation modes through data-driven methods mainly use mean or density clustering algorithms and their improved algorithms to obtain high-frequency operational scenarios in the grid. These methods minimize intra-class variance by iteratively optimizing the Euclidean distance between data points and cluster centers, but the physical meaning of using cluster centers to judge typical modes is unclear, and there is a lack of research on extreme scenarios at the grid safety operation boundary [18]. Literature [14] improved the K-means algorithm by detecting outliers as risk operation scenarios. Literature [19] proposed that searching for grid safety operation regions can be equated to finding the vertices of the safety operation region in the sample set, using Lawson's algorithm to find the convex hull vertices of high-dimensional discrete point sets. However, its safety domain is divided based on specific section screening results, posing potential operational risks [20].

In summary, although data-driven methods have made some progress in the study of operation mode extraction, current methods face two main challenges: first, the difficulty in considering the possible combinations of a large number of renewable energy outputs, leading to the omission of typical operation modes and resulting in potential safety hazards in grid planning and operation; second, the difficulty in characterizing and physically interpreting the extracted typical operation modes, incorporating a large number of redundant operation modes into the sample set. This paper proposes a knowledge-data hybrid-driven method for extracting power system operation mode samples, using hierarchical clustering algorithms to classify the preselected operation mode sets characterized by operational feature sequences such as renewable power station output, load levels, and sectional power flow directions. It considers the differences in renewable energy penetration rates and load levels to subdivide samples with safety boundary operation characteristics within each category and conducts N-1 security checks. Operation modes with similar safety issues are merged and reduced to form a generalized typical operation mode set for the power system. The main innovations of this method are as follows:

(1) Proposing a hierarchical partition clustering method for extracting dominant factors of power system operation modes. The advantage of hierarchical clustering in classifying power system operation modes lies in processing different operational data characteristics layer by layer, analyzing the independent influence of each characteristic on system behavior, thereby gradually constructing more accurate and representative operational patterns. This enhances the controllability and interpretability of the clustering process, making each layer's clustering results easy to understand and aiding the application of traditional clustering algorithms in analyzing complex power system operational patterns.

(2) Proposing a method for expanding and reducing typical operation modes in high-proportion renewable energy power systems. When extracting typical operation modes with operational risks, the method introduces derivative characteristics related to system safety boundaries by expanding the dataset. This method enhances the ability to capture the behavior of power systems under extreme conditions, helping to identify operational states that may lead to system risks. By focusing on these boundary characteristics, it effectively predicts and evaluates system stability and safety when facing severe challenges, providing a scientific basis for formulating response measures and optimizing system design.

2. The Framework for Extracting Generalized Typical Operation Modes Integrating Safety Operation Risks in High-Proportion Renewable Energy Power Systems

This paper proposes a power grid operational mode extraction framework that integrates safety operation risks, comprising three main components as shown in Figure 1:

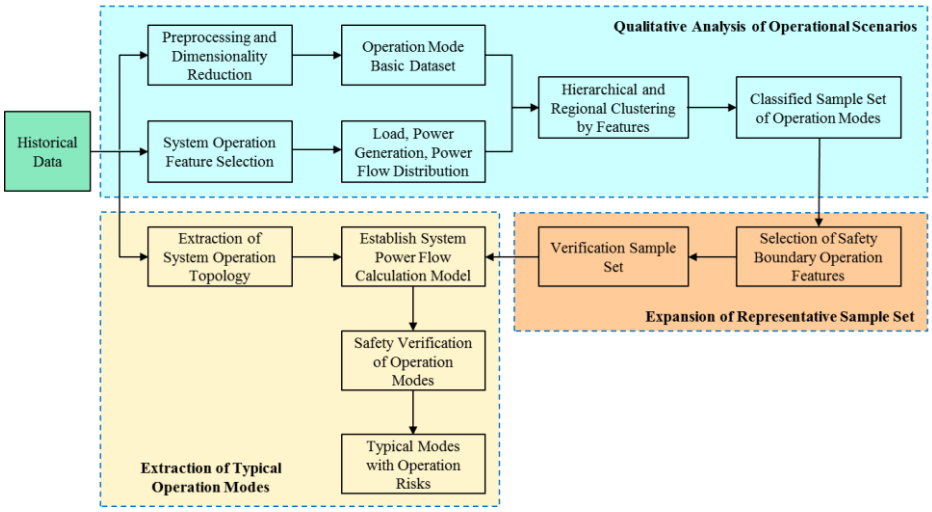


Figure 1 Framework for extracting generalized typical operation modes

(1) Qualitative clustering and analysis of operational scenarios are the core components of the entire framework. After preprocessing the raw operational data collected from the power company, the key characteristics of grid operation are selected to define the operation modes. Then, hierarchical and partitioned clustering is performed based on the preselected feature sequences, and samples from each cluster center in the final results are selected as the classification sample set of operation modes. This set is used to analyze the differentiated characteristics of various operation modes and provides a data foundation for subsequent work.

(2) The expansion of the representative sample set is the intermediate stage of the framework. After hierarchical clustering is completed, features that may contain the system's safe operational limits are added to filter out boundary samples within each cluster. These, along with the cluster center modes, establish the validation sample sets for various subsets of operation modes. These intra-cluster samples cover a range of extreme conditions from maximum load to minimum generation, forming a sample system of operational characteristics under the safe operating boundaries of the power system.

(3) Using historical operational data, an equivalent power flow calculation model of the grid is established based on the system topology information at various times. Under the framework of static security analysis of the

system, N-1 security checks are performed on the validation sample sets of operation modes to quantitatively assess the operational risks at different time points. The generalized typical operation modes are then extracted, and their distribution characteristics are identified and analyzed using visualization techniques.

3. Knowledge-data hybrid-driven Characterization of Power Grid Operation Modes and Hierarchical K-Means Clustering Method

3.1 Operation mode data characterization method

3.1.1 Selection of operation mode characteristic variables

The actual operating data of the power grid includes details such as plant and station information, unit output data, AC/DC line topology information and power flow data, transformer tap positions, and the open/close status of substation switches and disconnectors. If all these details are used to characterize the features of the grid operation modes, the abundance of redundant information may obscure the main features. Therefore, the selection of operational features should be based on the operating logic of the power system and the correlations between the data. Important features that significantly affect the safe and stable operation of the grid include the output of thermal power units, wind and solar output from the renewable energy side, load, and line power flow distribution, which can be used as evaluation indicators for the safe operation of the power system.

In this paper, the power flow data obtained from historical operational data (or production simulation) is used to generate operational features according to the following rules, achieving initial dimensionality reduction of the operational data in a new feature space:

$$P_t = (P_{N_{load},t}, P_{N_{line},t}, P_{N_{gen},t}, P_{N_{re},t}) \quad (1)$$

$P_{N_{load},t}$, $P_{N_{line},t}$, $P_{N_{gen},t}$, $P_{N_{re},t}$ are the feature sequences representing the system load, line power flow direction, thermal power unit output, and renewable energy unit active power output at time t, respectively. These sequences represent the characteristics of the power system's operation on the source side, grid side, and load side.

3.1.2 Operation mode data partitioning and dimensionality reduction

This paper selects the main attribute variables that can be differentially described from the original dataset based on different system operational characteristics at each time point. However, the data dimensions and scale are still large, necessitating further dimensionality reduction.

(1) Power direction coding of interconnection lines

Based on the network topology characteristics, the power system is divided into regions, forming power flow sections. The active power direction of the tie lines between regions represents the power exchange characteristics between the cut sets of grid areas, achieving a spatial distribution representation of the power system operation modes. The encoding rules are set as follows: the initial direction of active power flow in the tie lines under the base state operation mode is designated as the positive direction and is encoded as 1, the opposite direction is encoded as -1, and no power flow is encoded as 0.

(2) Equivalent output and load within regions

According to the power system topology partitioning results, features with high similarity are merged to achieve dimensionality reduction. Considering that load variations in the power system have a certain predictability and the load characteristics of plant stations within a region have small differences, the load data $P_{N_{load},t}$ of plant stations within each partition are linearly summed to replace the independent plant station load characteristics with the overall regional load characteristics. The load data after further dimensionality reduction is represented as p'_t

$$P'_t = (P'_{N_{load},t}, P'_{N_{line},t}, P_{N_{gen},t}, P_{N_{re},t}) \quad (2)$$

Through partitioning the power system based on network topology characteristics and performing dimensionality reduction on the data, the structure of the dataset is effectively simplified while retaining key features.

3.1.3 Evaluation indicators for differences between power system operation mode samples

To better assess the differences in the spatial distribution of power flows between operation modes and quantitatively describe the system's operational characteristics to further guide power system operation planning, this paper proposes a consistency indicator for describing the spatial distribution of power flows. Specifically, the Power Flow Direction Consistency Rate (PFDCR) is defined for the power flow direction characteristics of the power system. Compared with general clustering evaluation indicators (such as the silhouette coefficient), this indicator can provide more direct feedback on the clustering effectiveness.

$$PFDCR\% = \frac{N_C}{N_{Total}} \times 100\% \tag{3}$$

In the formula, N_C represents the number of power flow direction samples that are completely consistent with the cluster center, and N_{Total} represents the total number of samples within the cluster.

3.2 Introduction to the hierarchical k-means clustering method

The operating mode data of the power grid system is characterized by large data scale, non-linear correlations between data, and data sparsity in high-dimensional data spaces. These characteristics lead to a decline in the quality of decision boundaries when using Euclidean distance measures in traditional mean clustering methods. This paper proposes a hierarchical mean clustering method for scenario sample clustering analysis. Representative scenarios are selected from the clustering results as typical scenarios, optimizing the traditional mean clustering scheme.

Hierarchical mean clustering extends traditional mean clustering by systematically refining clusters layer-by-layer, reducing intra-class variability and enhancing interpretability. The process starts by assigning high weight to the first focus label and low weights to others, completing the first clustering layer. The dataset is divided into k subsets, each corresponding to a cluster center. For each subset, the previous focus label's weight is lowered, and the next focus label's weight is increased, proceeding to the next clustering layer. This continues until all labels have been sequentially focused. The final layer's results form a hierarchical clustering structure, providing deep insights into data distribution and relationships.

Algorithm: Hierarchical Mean Clustering

Inputs: Dataset X with m samples, each having n features
Number of clusters K
Number of labels n

Output: Cluster centers C

```
//Initialize weight matrix W for n labels in the dataset
for j = 1 to n do
    W[j] = 0
end for
W[1] = 1  // Set high weight for the first focus label
//Initialize cluster centers C randomly
for t = 1 to n do
    // Assign samples to clusters based on current weights
    for each sample i in X do
        Assign x_i to cluster c_k where k = argmin_k (sum(W * (x_i - c_k)^2))
    end for
    // Update cluster centers based on assignments
    for each cluster k from 1 to K do
        c_k = mean of all x_i assigned to cluster k
    end for
    // Prepare weights for the next label
    if t < n then
        W[t] = 0  // Reset previous focus label weight
        W[t + 1] = 1  // Set high weight for the next focus label
    end if
```

21: **end for**
22: **return** cluster centers C

4. Extraction of Generalized Typical Operation Modes Considering Static Safety Operation Risks

Although the hierarchical clustering method enhances the controllability and interpretability of features along the clustering vertical path, it also reduces the comparability of samples between clusters. When dealing with requirements such as extreme operation modes that involve multiple cross-layer associated features such as plant station output, line power flow, and transformer load, merely identifying extreme operation modes through the final extracted cluster center samples may lead to the omission of risky samples within the clusters. Additionally, the clustering results on a single clustering path may fail to fully capture the complex characteristics of extreme operation modes. Therefore, it is necessary to moderately expand the basic dataset of typical operation modes formed by the cluster centers, to improve the hit rate of extreme operation mode samples with potential operation risks in the safety verification process without excessively increasing the workload of safety verification.

4.1 Subdivision method for intra-cluster operation points with operational risks

Operation modes that may reach the static safety operational limits of the system under specific conditions can be considered as being at the boundary of the power system's safe operation. Special operational characteristic conditions such as heavy load power flows and limit violations are strongly correlated with specific overall operational modes of the system, such as substation load and regional total load, thermal power plant load and regional total thermal power generation, and changes in power flow characteristics caused by the interplay between renewable energy and thermal power. These conditions can be addressed by introducing derived features related to the operational boundary to expand the representative dataset of operation modes.

By adding features that may contain the system's safe operational limits to filter the intra-cluster boundary samples after clustering, the following characteristics are used: maximum load $L_{\max}$, minimum load $L_{\min}$, mid-level load L_{mid}, maximum thermal power generation Gmax, and minimum thermal power generation $G_{\min}$. After hierarchical clustering, the operation modes corresponding to these features within each cluster are combined with their cluster center mode $S_{cent,i}$to establish an expanded operation mode sample set D. The subset of samples for each class

S_i can be represented as:

$$S_i = (S_{cent,i}, S_{L_{max,i}}, S_{L_{min,i}}, S_{L_{miD,i}}, S_{G_{max,i}}, S_{G_{min,i}}) \tag{4}$$

The hypothetical set of operation modes on the safety operation boundary, D', encompasses all operation modes within the six-dimensional boundary characteristic system. This system is used to describe the operational modes in terms of safety boundary characteristics. Figure 2 provides a schematic diagram showing the samples belonging to two different subsets, with the final values taken along each axis according to the magnitude of the safety risk.

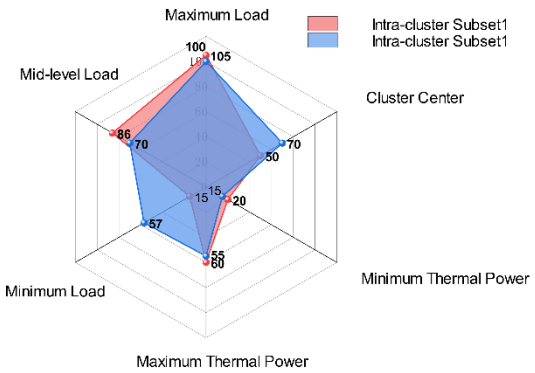


Figure 2 Schematic diagram of the safety boundary operation characteristic system

The set of operation modes on the safety operation boundary, D', can be mathematically expressed as the set of samples that meet the following conditions:

$$D' = \{S_i \in D: \text{Satisfies extreme operational conditions}\} \tag{5}$$

The operation modes $S_i \in D$ are identified as generalized typical operation modes with operational risks in this paper. The extreme operational conditions are defined as those where, after N−1 static security verification, the line load rate β_{L} or the transformer load rate β_{T} exceeds the thermal stability limits of the lines and main transformers. The load rate indicator calculation formula is given by equation 8:

$$\beta_i = \left|\frac{P_i}{P_{imax}}\right| i = 1,2,\cdots,N \tag{6}$$

In the formula, P_i represents the actual power flow or power of the line or main transformer, P_{imax} is its rated power, and N is the number of lines or main transformers.

4.2 Operational risk indicators and risk characteristic space representation

If the number of occurrences of line power flow and main transformer load limit violations at a certain time t after N-1 verification is summed and associated with the operational risk level of that mode, it is recorded as the risk indicator α_t. At the same time, the total outputs of the synchronous units $P_{N_{gen},t}$ and renewable energy data $P_{N_{re},t}$ at that moment are summed, and the ratio of the total output of renewable energy to the total load is calculated to obtain the renewable energy penetration rate η_t at that moment.

$$\eta_t = \frac{\sum P_{N_{re},t}}{\sum P'_{N_{load},t}} \tag{7}$$

Thus, each sample under a boundary mode characteristic category can be uniquely represented in a three-dimensional coordinate system composed of the total load, penetration rate, and risk indicator α_t. All risk-containing operational samples collectively form the risk characteristic space of the system. The coordinate representation method is shown in equation (8):

$$Coord(t) = \left(\sum P'_{N_{load},t}, \eta_t, \alpha_t\right) \tag{8}$$

5. Case Study Analysis

5.1 Basic conditions for the provincial grid case study

The basic information of the historical data collected by the D5000 grid intelligent dispatch system for the Jilin Province power grid is as follows: Selected time period from March to June 2023, during which wind power output was significant, with a sampling interval of 30 minutes. The dataset includes over 300 substations at 220kV and above, and approximately 520 AC lines, resulting in a total of 3640 sets of active power data. Figure 3 shows the partitioning results of the power system into cut set sections formed by selecting 17 tie lines, divided into a total of 11 regions.

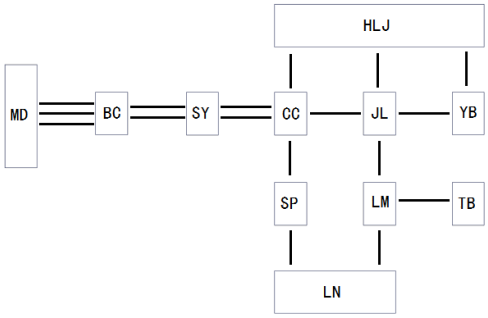


Figure 3 Partition map of jilin province power grid

After partitioning and dimensionality reduction of the data, the basic operation mode dataset is formed based on $P'_t = (P'_{N_{load},t}, P'_{N_{line},t}, P_{N_{gen},t}, P_{N_{re},t})$. The dataset is subjected to hierarchical k-means clustering in the order of load-side, grid-side, and source-side characteristics. Upon completing the final layer of k-means clustering, the cluster centers for all classes are extracted to form a representative sample set of operation modes, resulting in a total of 50 cluster centers.

The load and unit active power outputs are divided into three operational segments: small (below 30%), medium (30%-70%), and large (above 70%). The active power direction of interconnection lines is set as 1 for west-to-east and north-to-south directions in the Jilin Province grid partition map, -1 for the opposite direction, and 0 when there is no power flow. The results of the representative sample set of operation modes formed after hierarchical clustering are shown in Table 1:

Table 1 Hierarchical clustering results for typical operation modes in Jilin province in 2023 (Top 35% of samples out of 50 categories)

ID	Load (by scale)	Wind Power, Photovoltaic Power, Thermal Power, Hydropower(by scale)	Power Flow Direction	Time	Percentage
1	M	S, S, M, S	10111110010111000	202306231050	8.10%
2	M	L, S, S, S	00010110011111011	202304030400	7.65%
3	S	M, S, S, S	10010110011111000	202306190120	7.46%
4	M	L, M, S, S	00010110011111011	202304070800	5.74%
5	M	S, S, M, S	10010000010111000	202306222320	5.19%

5.2 Analysis of hierarchical clustering effectiveness

(1) Diversity of operation modes

From the table, it is evident that the clustering results encompass typical operation modes of the Jilin Province power system under different load levels (small, medium, large), various energy combinations (wind power, photovoltaic, thermal power, hydropower), and different power flow direction configurations. The results demonstrate the effectiveness of the hierarchical clustering method in distinguishing diverse operation modes in a high-proportion renewable energy power system.

(2) Representativeness of categories

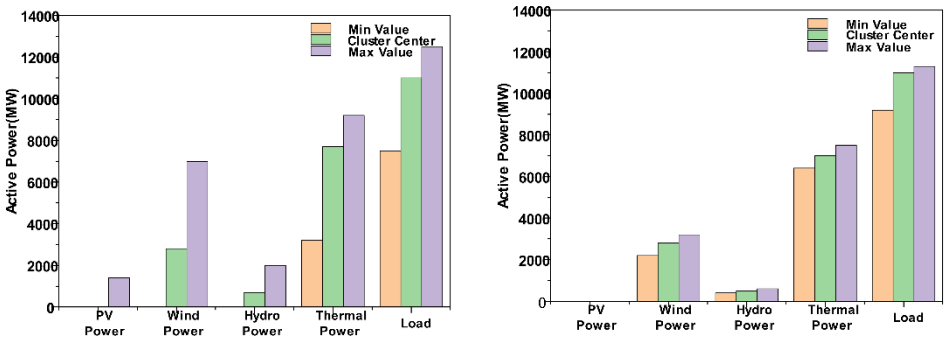
The "category proportion" indicator for each cluster shows the representativeness of the cluster center relative to the total sample. A higher percentage indicates that these cluster centers can better represent a significant proportion of the operation modes. For example, the category with serial number 1 has a category proportion of 8.10%, indicating that operation modes with this source-load characteristic account for a large proportion of the total samples and represent common operation modes.

(3) Comparison with traditional direct k-means clustering method

When comparing the typicality between traditional direct k-means clustering and hierarchical k-means clustering, consider setting the same number of clusters to 50. Two clusters with similar characteristics are selected from both methods, and the distribution range of active power values for source-side output and load within each cluster is compared. Traditional methods, which only consider numerical similarity (Euclidean distance), tend to overlook the physical scenarios of operation modes, resulting in clusters that still contain complex operation patterns. Specifically, the intra-cluster includes thermal power output ranging from 3258MW to 9207MW, with the cluster center sample value at 7500MW, and load active power ranging from 7300MW to 12379MW. The typicality of thermal power output and load within the cluster is not apparent, as shown in Figure 4(a). After using the hierarchical clustering method to partition the intra-class operation mode samples based on the feature sequence, the active power output of thermal power units within the class ranges from 6500 MW to 7500 MW, and the active power load ranges from 9222 MW to 11161 MW. The intra-class variability is significantly reduced, as shown in Figure 4(b).

(4) Calculation of *PFDCR* indicator

The calculation results of the Power Flow Direction Consistency Rate (PFDCR) indicator show that if the line power flow characteristics are represented only by direction, the extracted typical operation modes have a PFDCR% range of 72% to 100%. When represented by active power values, the extracted typical operation modes have a PFDCR% range of 56% to 82%, as shown in Figure 5.From the data, it is evident that using direction to represent interconnection line characteristics is more effective for extracting operation modes. However, it should be noted that the same power flow direction on a line can correspond to significantly different power values. Therefore, both the direction and value representation of interconnection line characteristics have an important impact on the extraction of typical operation modes.



(a) Direct mean clustering (b) Hierarchical mean clustering

Figure 4 Comparison of active power characteristics of operation modes in clustering results

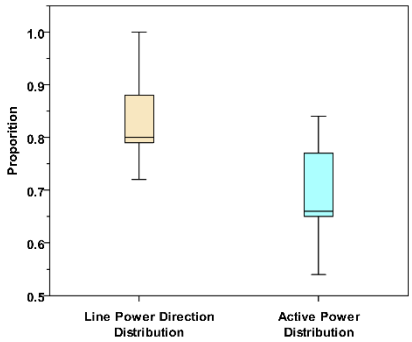


Figure 5 Differences in PFDCR indicators for clustering using different line power flow characteristics

5.3 Construction of operation mode expanded sample set and static security verification

Based on the sample set of 50 cluster center representatives selected through the hierarchical clustering method applied to provincial grid data, samples with maximum load L_{max}, minimum load L_{min}, mid-level load L_{mid}, maximum thermal power generation G_{max} and minimum thermal power generation G_{min} within each category are selected. This forms an expanded sample set of 300 operation modes. The expanded sample set undergoes N-1 security verification for the entire grid. Components with line power flow and main transformer load violations or near-violations are identified and marked by their occurrence positions and frequencies. The frequency of violations serves as the probabilistic operational risk characteristic indicator α for these modes. Figure 6 presents the statistical results after removing 58 samples without power flow violations. The resulting dataset consists of 35 dimensions and 242 columns. In the figure, cyan-colored sections represent elements without violations, while yellow to red indicates increasing frequencies of element violations.

This approach ensures that the expanded sample set captures a comprehensive range of operational scenarios, including those with potential security risks. By identifying and quantifying the frequency of violations, the method provides a robust framework for assessing and managing the operational risks of the power grid.

It is evident that by adding features that may contain the system's safe operating limits and constructing an expanded sample set of operation modes, more operation modes with overload risks, beyond those extracted by the hierarchical clustering method, are uncovered. The statistics of heavy load power flows and overloaded lines in each sample after N-1 verification further compress the feature dimensions of the samples. By counting the positions and frequencies of overloaded lines in the verified samples, part of the power flow spatial distribution characteristics is retained. The risk indicator α composed of overload frequency information and new sample feature vectors can associate safety operation risks with operational scenarios.

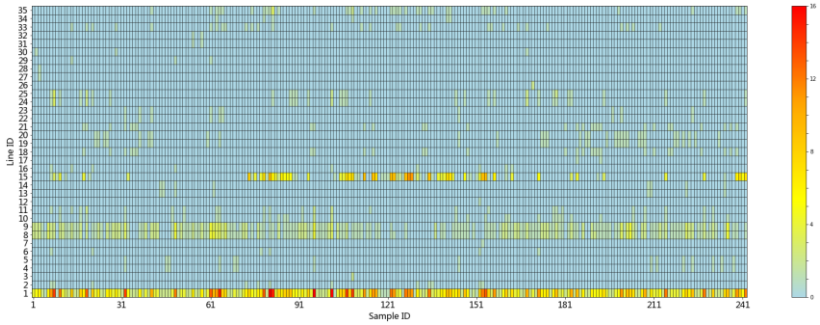
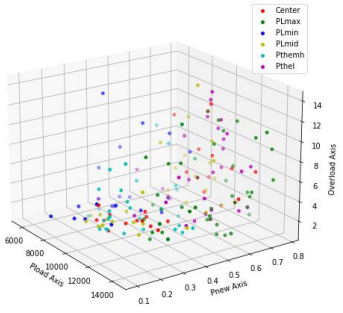


Figure 6 Statistics of overload lines after n-1 verification in the expanded sample set

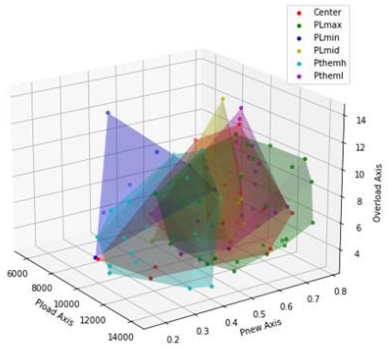
Statistics show that out of the 242 scenarios with overloads after safety verification, 38 scenarios have identical overload frequencies and component position distributions, indicating similar safety operation risks. Additionally, 89 samples have different overload distributions and frequencies, with highly dispersed risk characteristic distributions. Among the samples with overload risks, the 220kV lines have a higher overload frequency than the 500kV lines. The lines with the highest historical overload risks are the JQ Line B, JH1 Line, and HX Line A/B.

5.4 Visualization and analysis of operation characteristics with overload risks

According to the coordinate representation method of the risk characteristic space in equation (8), the extracted typical operation modes with overload risks are represented in the risk characteristic space by the three coordinate axes composed of total load $\sum P'_{N_{load}}$, renewable energy penetration rate η, and risk indicator α_t, as shown in Figure 7(a). For each category, the boundary operation samples corresponding to the maximum load, minimum load, mid-level load, maximum thermal power, minimum thermal power, and cluster centers are used to construct convex hull regions. The vertex positions of these regions describe the system risk areas under the corresponding boundary operation modes, as shown in Figure 7(b).



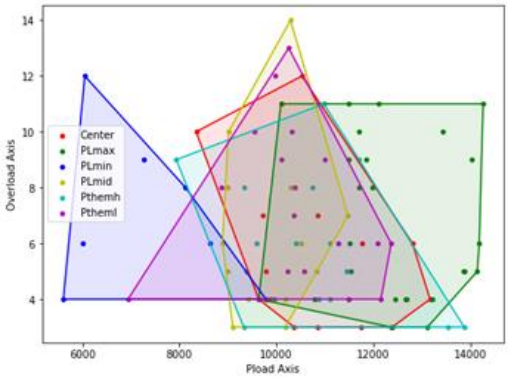
(a) Scatter plot of operation modes with operational risks



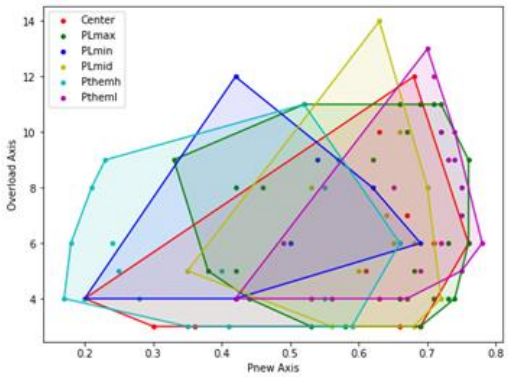
(b) Convex hull regions constructed by various boundary operation modes

Figure 7 Distribution diagram of operating points and class-partitioned operating areas

Projections on the plane formed by the load axis and the risk indicator axis, as well as the penetration rate axis and the risk indicator axis, can be obtained for further analysis. Figure 8(a) shows that the convex hull projection area of the minimum load operation mode has the lowest density of risk samples, while the maximum load operation mode contains the highest density of samples. From the plane formed by the load axis and the risk indicator axis, it is observed that system risk samples are concentrated in the 9000MW-13000MW load range, with high overlap in the convex hull projection areas of the clustering center, mid-level load, and maximum thermal power scenarios. Figure 8(b) shows that from the plane formed by the penetration rate axis and the risk indicator axis, risk samples are concentrated in the 55%-75% and 15%-30% penetration rate ranges. In particular, risk samples of the clustering center, maximum load, mid-level load, and minimum thermal power scenarios are more concentrated in the region with penetration rates greater than 60%. This indicates that by setting boundary operation samples with safety operation risk limits to expand the clustering sample set, it is possible to establish a connection between generalized typical operation modes with static safety operation risks and system operation patterns. This further allows for the identification of potential risks based on the operational state of the power system.



(a) Load Axis and Operational Risk Axis



(b) Penetration Rate Axis and Operational Risk Axis

Figure 8: Projections of convex hull regions

6. Conclusion

This paper presents a knowledge-data hybrid-driven method for extracting and analyzing generalized typical operation modes in high-proportion renewable energy power systems using hierarchical clustering algorithms. By selecting key operational features and systematically clustering the operation modes characterized by these features, we consider the differences in renewable energy penetration rates and load levels to further refine the sample operation scenarios. The method's effectiveness is validated using historical data from a provincial power grid. The main conclusions are as follows:

(1) The results indicate that the hierarchical clustering method is superior to traditional mean clustering methods in capturing the diversity and representativeness of operation modes. By using the Power Flow Direction Consistency Rate (PFDCR) as an evaluation metric, the clustering quality is directly and meaningfully assessed.

(2) The proposed method also enhances the understanding of risk operation modes by extending the classified sample set to include potential risk scenarios. This makes the risk assessment and visualization results more accurate, aiding in the identification of system weaknesses and the formulation of effective operational strategies.

(3) The generalized typical operation modes extracted in this study show significant distribution characteristics in terms of renewable energy penetration rates and load levels, reflecting the detypicalization of operation modes in renewable energy power systems. They also highlight the clustering characteristics of typical operation modes in evaluating specific system weaknesses.

In conclusion, the proposed knowledge-data hybrid-driven method provides a powerful tool for power system operators and planners to address the challenges posed by high renewable energy penetration. Future work should continue to optimize the clustering algorithm and explore its application in different grid environments to enhance its generalizability and robustness.

Acknowledgements

This research was supported by funded by the science and technology project of State Grid Corporation of China, contract number SGL0000DKJS2300267.

References

[1] Guo Q L, Lan J, Zhou Y Z, et al. Architecture and Key Technologies of Hybrid-Intelligence-Based DecisionMaking of Operation Modes for New Type Power Systems. Electric Power, 2023, 56 (09): 1-13.

[2] Luo, Y. J., Li, Y. H., Wang, P., Li, Z. X., Gao, F. Q., & Xu, F. (2016). DC voltage adaptive droop control of multi-terminal HVDC systems. Proc CSEE, 36(10), 2588-2599.

[3] Zimmerman, R., Murillo-Sandchez, C., & Thomas, R. (2011). MATPOWER: steady-state operations, planning, and analysis tools for power systems research and education. IEEE Trans Power Syst, 26(1), 12-19.

[4] Ren C, Niu S B, Ke X B, et al. Clustering Analysis of Transmission Section Modes for Power Grid with Renewable Energy. Automation of Electric Power Systems, 2022, 46 (01): 69-75.
[5] Bouffard, F., & Galiana, F. D. (2004). An electricity market with a probabilistic spinning reserve criterion. IEEE Trans Power Syst, 19(1), 300-307.
[6] Palmintier, B. (2022). Challenges of renewable energy penetration on power system flexibility. Renewable Energy, 90(1), 50-60.
[7] Goel, A., & Goel, A. (2014). Forecasting of Electricity Demand and Renewable Energy Generation for Grid Stability. Journal of Clean Energy Technologies, 2(4), 305-309.
[8] Dagoumas, A. S., & Koltsaklis, N. E. (2019). Large-scale wind power grid integration challenges and their solution: a detailed review. Environmental Science and Pollution Research, 26(22), 22244-22257.
[9] Xu Y P, Bai J, Shi H B, et al. Extreme Operation Mode Extraction Method based on Convex Hull Algorithm. Electric Power, 1-9[2024-05-14].
[10] Fernandez-Blanco, R., Dvorkin, Y., & Ortega-Vazquez, M. A. (2017). Probabilistic security-constrained unit commitment with generation and transmission contingencies. IEEE Trans Power System, 32(1), 228-239.
[11] Smith, J., & Brown, L. (2023). Data-driven next-generation smart grid towards sustainable energy evolution: techniques and technology review. Protection and Control of Modern Power Systems, 8(1), 25-38.
[12] Krause, T., Andersson, G., Donalek, P., & Rehtanz, C. (2006). Integration of Renewable Energy into Power Systems: Challenges and Solutions. Electric Power Systems Research, 77(3-4), 225-234.
[13] Mitra, P., & Murthy, C. A. (2002). Feature Selection and Extraction Methods for Power Systems Transient Stability Assessment Employing Computational Intelligence Techniques. Neural Processing Letters, 15(2), 193-203.
[14] Hou, Q., Du, E., Tian, X., Liu, F., Zhang, N., & Kang, C. (2021, January). Data-driven power system operation mode analysis. In Proceedings of the CSEE (Vol. 41, No. 1, p. 12).
[15] Wang, Y., Liu, J., & Zhang, W. (2023). A Fast Method for Uncertainty Analysis of Power System Dynamic Simulation. Processes, 11(7), 1886. doi:10.3390/pr11071886.
[16] Focken, U., Lange, M., Mönnich, K., Waldl, H., Beyer, H. G., & Luig, A. (2002). Short-term prediction of the aggregated power output of wind farms—a statistical analysis of the reduction of the prediction error by spatial smoothing effects. J Wind Eng Ind Aerodynam, 90(3), 231-246.
[17] Stott, B., Jardim, J., & Alsac, O. (2009). DC power flow revisited. IEEE Trans Power System, 24(3), 1290-1300.
[18] Yu, H., & Rosehart, W. D. (2012). An optimal power flow algorithm to achieve robust operation considering load and renewable generation uncertainties. IEEE Trans Power System, 27(4), 1808-1817.
[19] Liu J L, Yan J F, Shi D Y, et al. Automatically Generating Method of Power System Operation Mode Rules Based on Lawson Algorithm. Proceedings of the CSEE, 2023, 43 (21): 8171-8182. DOI:10.13334/j.0258-8013.pcsee.221493.
[20] Bucher, M. A., & Ortega-Vazquez, M. A. (2012). An optimal power flow algorithm to achieve robust operation considering load and renewable generation uncertainties. IEEE Trans Power System, 27(4), 1808-1817.