Embedded Scenario Clustering for Wind and Photovoltaic Power, and Load Based on Multi-head Self-attention

Lijun Liu, *Member*, *IEEE*, Xin Hu, Junsheng Chen, Ruixing Wu, and Feixiong Chen

***Abstract*—The source and load uncertainties arising from increased applications of renewable energy sources such as wind and photovoltaic energy in the power system have had adverse effects on optimal planning and dispatching. Models for generating typical renewable energy and load scenarios are constructed to reduce such effects and improve the applicability of a planning and optimal dispatching model of power systems with a high proportion of renewable energy. The traditional clustering-based model for representing such scenarios cannot handle high-dimensional time-series data and consequently the feature-related information obtained cannot fully reflect the characteristics of the data. Thus, a deep convolutional embedded clustering model based on multi-head self-attention is proposed. First, a variational mode decomposition model is optimized to reduce the influence of noise-related signals on the feature extraction. The deep features are then extracted from the data using an improved convolutional autoencoder, and the appropriate number of clusters is determined using the elbow method. Following this, the network parameters are optimized based on the sum of losses during reconstruction and clustering. Subsequently, typical scenarios are then generated based on the optimized network model. Finally, the proposed method is evaluated based on data visualization and evaluation metrics. It is shown that the quality of features and the accuracy of clustering can be effectively improved by the proposed scenario generation method.**

***Index Terms*—Deep embedding for clustering, extracting features of time-series data, multi-head self-attention mechanism, scenario generation, uncertainty.**

I. Introduction

With continuous advance in power systems with a large integration of renewable energy, the volatility and periodicity of wind and photovoltaic (PV) power, and load pose significant challenges to grid planning and power dispatch [1], [2]. To improve planning and dispatch, models for power systems with a high proportion of renewable energy need to cater for the uncertainties in wind, PV power, and load demand. To do this is essential for effective planning and economical dispatch [3], [4].

The uncertainty models for wind and PV power, and load demand have been extensively studied, and three main methods have been proposed: probabilistic sampling, robust optimization, and scenario clustering. An accurate description of uncertain variables is required in the traditional probabilistic sampling method, while the probability distribution of these variables is actually unknown in real projects. The use of artificial pre-set probability distribution functions falls short in accurately representing the probability distributions and data aggregation occurs during sampling, leaving out the nature of the time-series data [5]–[7]. Although the model of probability distribution is not required in robust optimization, it seeks solution to optimize grid planning and dispatching in extreme scenarios. This may lead to overly conservative results [8]. In addition, the non-linear characteristics of high-dimensional time-series data cannot be accurately described in traditional methods of clustering, such as mean [9], spectral [10], and hierarchical [11], as they can cause the generated scenarios to differ from empirical scenarios of grid scheduling [12].

To improve the clustering quality of scenarios involving the output of renewable energy and load demand amidst high-dimensional data, principal component analysis (PCA) and singular value decomposition are generally used to reduce the dimensionality of the data and extract features from them, followed by clustering based on feature-related information in low-dimensional space [13]–[15]. The accuracy of clustering is thus improved, but the processes of dimension reduction and clustering are independent of each other so that the captured feature-related information cannot accurately reflect the characteristics of the outputs of wind and PV power as well as the load demand. In recent years, deep clustering models have provided a suitable solution to the

Received: April 19, 2023
Accepted: October 9, 2023
Published Online: January 1, 2024
Lijun Liu (corresponding author) is with School of Electrical Engineering and Automation, Fuzhou University, Fuzhou 350108, China (e-mail: liulijun0120@fzu.edu.cn).
DOI: 10.23919/PCMP.2023.000296

high-dimensional data problem. So far these have primarily been applied to image recognition [16]–[18]. In [19], a deep embedded clustering (DEC) model is proposed to generate scenarios in low-dimensional embedding space involving hydroelectric and PV power, and load. However, the processes of feature extraction and clustering in the model are separated, and therefore fail to take into account the distortion of the embedding space caused by the clustering process. Consequently, this weakens the relevant information about the features and leads to degradation of the quality of the generated scenarios. In [20], a model of deep convolutional embedding clustering (DCEC) is proposed, one that guarantees the representativeness of the features of load in the embedded space. However, the use of a small convolutional kernel and variable step length in the convolutional layer to capture the feature-related information poses challenges in accurately reflecting the global characteristics of wind and PV power, and the demand of load.

To address this, a widely used attention mechanism in natural language processing tasks provides a solution to the problem by allowing the dependencies between data to be modeled regardless of their distances in the input or output sequences [21]. However, the single-head attention mechanism may not fully exploit the synergistic effect between multiple elements in long sequences [22]. Thus, in [23], a multi-head self-attention mechanism is introduced in the prediction model. This learns the distribution of feature weights by calculating the similarity of the elements in the sequences. This makes the model capable of characterizing the global features created by the multiple elements in the sequences, resulting in promising outcomes in the study. Therefore, in this paper, a multi-head self-attention mechanism is introduced in deep embedding clustering to compensate for the lack of global feature representation ability of the convolutional layer.

Compared with previous studies, this paper makes the following contributions:

1) Given the contingency of parameter selection in variational mode decomposition (VMD), a multi-strategy fusion of the improved slime mould algorithm (SMA) is proposed in this paper to optimize VMD's parameter combination. Based on the optimal parameter combination, the time-series data for wind power, PV power and load are processed to mitigate the influence of outliers on feature extraction.

2) Addressing the limitations of convolutional layers by using a multi-head self-attention mechanism and establishing a multi-head self-attention-based convolutional autoencoder to accurately capture the global feature within the time-series data.

3) A deep convolutional embedding clustering with multi-head self-attention (DCEC-MS) model is proposed to combine feature extraction with the clustering process. This ensures that the embedded features can accurately capture the temporal characteristics of wind and PV power, and load, resulting in the generation of a high-quality joint-scenario set.

The rest of the paper is structured as follows: the framework and process of scenario generation are demonstrated in Section II, while the data processing based on the VMD model optimized by the improved SMA is described in Section III. The joint-scenario generation method based on DCEC-MS is described in Section IV. Case studies are conducted and corresponding results are presented in Section V. Conclusions are drawn in Section VI.

II. Research Framework for Scenario Generation

To represent uncertainties in the output of renewable energy and the load demand in the problems of grid planning and optimal dispatch, it is unfeasible for planners and dispatchers to use all the relevant data. The proposed method of scenario generation is used to extract and reduce the number of features in given data to generate scenarios that can reflect the original output of wind, PV, and load. Thus, a typical set of scenarios can be presented for the optimal dispatching or planning model. The flowchart of the model for scenario generation is shown in Fig. 1, and it consists of the following four stages:

1) Data processing. VMD parameters are optimized based on an improved multi-strategy SMA, while anomalous historical data are identified and cleaned using the improved VMD.

2) Reducing data dimensionality. The improved convolutional autoencoder with multi-head self-attention is used to scale down the data on wind and PV power, and load for their deep feature extraction. Then the original temporal signals are reconstructed by using the convolutional decoder. The appropriate number of clusters is selected in the embedded low-dimensional feature space based on the elbow method, and the initial centers of clustering are obtained by K-means.

3) Optimizing features of the scenarios. The joint loss function is formed by combining the reconstruction loss and clustering loss. The Adam optimizer is then used to fine-tune the network parameters of the convolutional self-encoder and update the clustering results until the iterative error falls below a predetermined threshold.

4) Scenario generation. The clustered set of scenarios is obtained based on feature optimization, and the outputs of wind and PV power as well as the demand for load are calculated in each scenario using the mean value method, as shown in (1). The index to evaluate clustering and data visualization are used to compare and analyze the results with traditional methods of scenario generation.

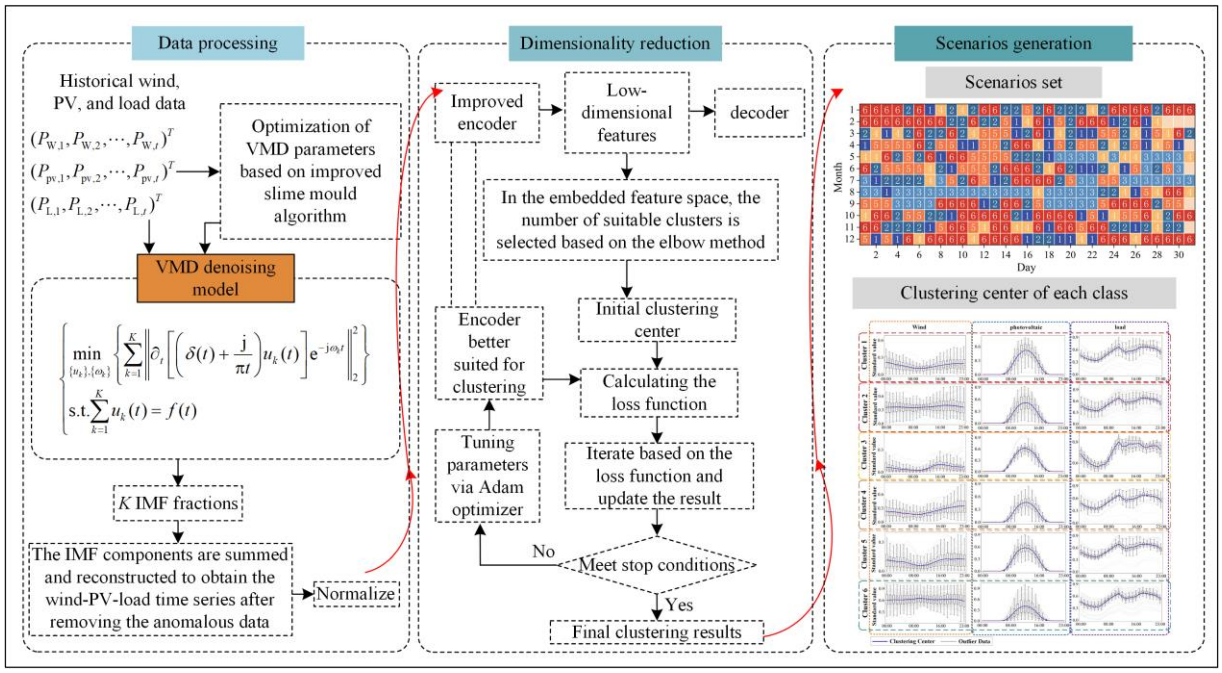

Fig. 1. Flowchart for generating a model for scenarios involving wind power, photovoltaic power, and the load demand.

$$\begin{cases} P_{\mathrm{w},t}^{j} = \sum_{i=1}^{N_{\mathrm{cluster},j}} P_{\mathrm{w},t,i}^{j} / N_{\mathrm{cluster},j} \\ P_{\mathrm{pv},t}^{j} = \sum_{i=1}^{N_{\mathrm{cluster},j}} P_{\mathrm{pv},t,i}^{j} / N_{\mathrm{cluster},j} \\ P_{\mathrm{load},t}^{j} = \sum_{i=1}^{N_{\mathrm{cluster},j}} P_{\mathrm{load},t,i}^{j} / N_{\mathrm{cluster},j} \end{cases} \tag{1}$$

where $P_{\mathrm{w},t}^{j}$, $P_{\mathrm{pv},t}^{j}$, and $P_{\mathrm{load},t}^{j}$ denote the outputs of wind power, PV power, and the load demand at time t in the scenario in category j, respectively; $P_{\mathrm{w},t,i}^{j}$, $P_{\mathrm{pv},t,i}^{j}$, and $P_{\mathrm{load},t,i}^{j}$ denote the ith outputs of wind power, PV power, and the load demand at time t, respectively; while $N_{\mathrm{cluster},j}$ is the number of curves in the scenarios in category j.

III. Data Processing Based on the VMD Model Optimized by an Improved SMA

A. Data Cleaning Based on VMD

The accuracy of the temporal data on wind and PV power, and load forms the basis for improving scenario generation accuracy. However, the presence of noise stemming to electromagnetic interference, failure of signal transmission equipment and communication can lead to abnormal values of wind and PV power data. Because traditional methods of noise reduction, such as median and mean filtering, cannot be applied to non-stationary time-series data [24], a multi-strategy fusion-based SMA is used to optimize the VMD model, while the time-series data on wind and PV power, and load $f(t)$ are decomposed into K IMF components $u_k(t)$. The component summation method is then applied to reconstruct these decomposed K IMF components, i.e., $\hat{f}(t) = u_1(t) + u_2(t) + \cdots + u_k(t)$. The specific process is shown in the data processing section of Fig. 1.

First, $f(t)$ is decomposed into K IMF eigenmodes with central frequency $u_k(t)$. Once the sum of the finite bandwidths of $u_k(t)$ has been minimized, the components of effective decomposition of the given temporal data on wind and PV power, and load are obtained, as shown in (2). These components are added together to obtain clean temporal data as described in [25].

$$\begin{cases} \min\limits_{\{u_k\},\{\omega_k\}} \left\{ \sum\limits_{k=1}^{K} \left\| \partial_t \left[\left(\delta(t) + \dfrac{\mathrm{j}}{\pi t} \right) \times u_k(t) \right] \times \mathrm{e}^{-\mathrm{j}\omega_k t} \right\|_2^2 \right\} \\ \mathrm{s.t.} \sum\limits_{k=1}^{K} u_k(t) = f(t) \end{cases} \tag{2}$$

where ω_k is the central frequency of the kth instance of $u_k(t)$; ∂_t is a partial derivative operator; and $\delta(t)$ is the impulse function.

B. VMD Model Optimized by an Improved SMA

The process of analyzing the VMD model shows the importance of determining the pre-set values of K and α. If K is too large, over decomposition occurs, resulting in overlapping modes, whereas if α is too large, the center frequency is lost. Most current related studies use the observational method to select the preset values of the parameters, but the method is subject to uncertainty. Therefore, Kullback–Leibler (KL) divergence is used to measure the similarity between the components of the eigenmode $u_k(t)$ and the original data $f(t)$ in

this paper. The minimization of the KL divergence is chosen as the objective function, and the optimal combination of parameters is selected based on the solution to the multi-strategy fusion-based SMA. SMA is an intelligent optimization algorithm proposed in 2020. It is optimised by simulating the behaviour of slime molds during foraging. The process of the standard SMA model has been detailed in [26]. However, its mechanism for position update remains deficient in terms of the speed of convergence in the early stages, while it can also easily fall into local optima in later iterations. Thus, SMA is improved by fusing the adaptive adjustable feedback factor with the adaptive backward-learning mechanism, and the process is shown as follows:

1) Adaptive Adjustable Feedback Factor

It is known from the principle of the standard SMA that a linear reduction in the feedback factor v_c cannot provide accurate and timely feedback on the concentration and quality of food, resulting in a low rate of convergence in the early stages. The adaptive adjustable v_c is thus introduced to expedite the decline in v_c in the early stages to enhance the ability of the algorithm to search for the global optimum and avoid falling into a local optimum. The improved mathematical expression of v_c is given as:

$$v_c = \left(\mathrm{e}^{(T_{\max}-T)/(T_{\max}-1)} \times \frac{1}{\mathrm{e}-1} \right)^m \tag{3}$$

where m is the adaptive adjustment factor; T is the number of iterations; and $T_{\max}$ is the upper limit of the number of iterations.

2) Adaptive Backward-Learning Mechanism

The adaptive backward-learning mechanism introduces a vector X_i^o to the area of slime exploration, which is an opposite vector to the position X_i^f of each slime individual. The mechanism compares their fitness values to avoid falling into the local optimum. The position X_i^c of the individual mucilage i in the Tth iteration is given by:

$$X_i^o(T) = \min(X_i^f(T)) + \max(X_i^f(T)) - X_i^f(T) \tag{4}$$

$$X_i^c(T) = \begin{cases} X_i^o(T) & S(X_i^o(T)) < S(X_i^f(T)) \\ X_i^f(T) & S(X_i^o(T)) \geqslant S(X_i^f(T)) \end{cases} \tag{5}$$

The ith mucilage uses adaptive decision-making to explore food by comparing the current adaptation value $S(X_i^f(T))$ with the previous optimal adaptation value $S(X_i^f(T))$ to determine whether to use the backward-learning mechanism for additional exploration. It updates its position for the next iteration as follows:

$$X_i(T+1) = \begin{cases} X_i^f(T) & S(X_i^f(T)) \leqslant S(X_i(T)) \\ X_i^c(T) & S(X_i^f(T)) > S(X_i(T)) \end{cases} \tag{6}$$

C. Evaluation Metrics for Data Cleaning

The mean absolute error (MAE) is selected to assess data cleaning, i.e.:

$$MAE = \sum_{t=1}^{N} \left| f(t) - \hat{f}(t) \right| / N \tag{7}$$

where N is the total number of data samples on wind and PV power, and load demand.

MAE measures the ability of the data processing algorithm to retain the original features of the data while removing outliers. In this context, a smaller *MAE* indicates more effective data processing.

IV. A Joint-Scenario Generation Method Based on DCEC-MS

The structure of the DCEC-MS-based scenario generation is shown in Fig. 2. The model contains an encoder and a decoder. An encoder based on a small convolutional kernel with three convolutional layers of variable step length is used to enhance the ability of the network to extract the initial features of time-series data after data processing. This is followed by three multi-head self-attention layers, to explore feature-related information and improve the accuracy of clustering. Finally, the embedding layer integrates the timing feature information captured from wind power, PV power, and load to create a low-dimensional space representing their characteristics. The decoder part reduces features in the low-dimensional space through a fully connected layer and a reshaping layer, and the reconstruction of the time-series data is completed by processing them through three deconvolution layers. The features embedded in the low-dimensional space are used for clustering and calculating the clustering loss. Then, the DCEC-MS model fine-tunes the network parameters based on the joint loss function to produce the best possible joint-scenario for wind and PV power, and load.

A. Multi-Head Self-Attention Mechanism for Enhancing Model Feature Extraction Capability

To address the limitations that the feature-related information captured by the small convolution kernel and the convolution layer with a variable step length cannot adequately reflect the characteristics needed [27], the approach aims to reduce the dependence of the model on external information. This is achieved by using a multi-head self-attention mechanism to capture the correlation within these data, enabling their features in the low-dimensional space to accurately reflect the characteristics of the original data. In this method, multi-head self-attention uses multiple queries to capture the feature-related information from the data in multiple groups of subspaces in parallel, and combines them together by weights. The process is shown below.

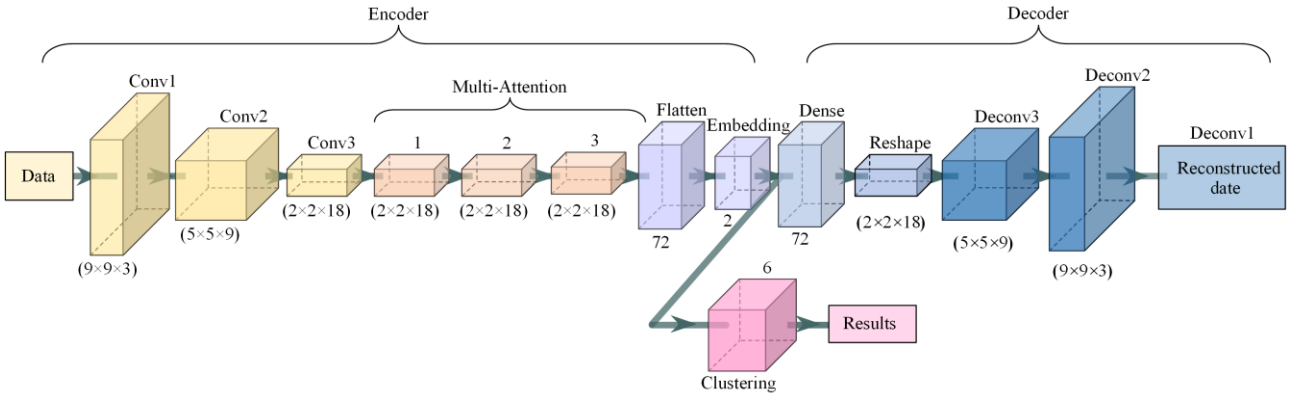

Fig. 2. Structure of the DCEC-MS network.

1) A linear transformation is used to transform the input data Y of the multi-head self-attention layer into a query matrix $\boldsymbol{Q}_a$, a key matrix $\boldsymbol{K}_a$, and a value matrix $\boldsymbol{V}_a$ as:

$$\begin{cases} \boldsymbol{Q}_a = \boldsymbol{W}_Q Y \\ \boldsymbol{K}_a = \boldsymbol{W}_K Y \\ \boldsymbol{V}_a = \boldsymbol{W}_V Y \end{cases} \tag{8}$$

where $\boldsymbol{W}_Q$, $\boldsymbol{W}_K$, and $\boldsymbol{W}_V$ are the transformation matrices.

By mapping $\boldsymbol{Q}_a$, $\boldsymbol{K}_a$, and $\boldsymbol{V}_a$ to the θ feature subspace, we obtain the query matrix $\boldsymbol{Q}_{a\theta}$, the key matrix $\boldsymbol{K}_{a\theta}$, and the value matrix $\boldsymbol{V}_{a\theta}$ in the θ subspace as:

$$\begin{cases} [\boldsymbol{Q}_{a1}, \boldsymbol{Q}_{a2}, \cdots, \boldsymbol{Q}_{a\theta}] = [\boldsymbol{Q}_a \boldsymbol{W}_{Q1}, \boldsymbol{Q}_a \boldsymbol{W}_{Q2}, \cdots, \boldsymbol{Q}_a \boldsymbol{W}_{Q\theta}] \\ [\boldsymbol{K}_{a1}, \boldsymbol{K}_{a2}, \cdots, \boldsymbol{K}_{a\theta}] = [\boldsymbol{K}_a \boldsymbol{W}_{K1}, \boldsymbol{K}_a \boldsymbol{W}_{K2}, \cdots, \boldsymbol{K}_a \boldsymbol{W}_{K\theta}] \\ [\boldsymbol{V}_{a1}, \boldsymbol{V}_{a2}, \cdots, \boldsymbol{V}_{a\theta}] = [\boldsymbol{V}_a \boldsymbol{W}_{V1}, \boldsymbol{V}_a \boldsymbol{W}_{V2}, \cdots, \boldsymbol{V}_a \boldsymbol{W}_{V\theta}] \end{cases} \tag{9}$$

where $\boldsymbol{W}_{Q\theta}$, $\boldsymbol{W}_{K\theta}$, and $\boldsymbol{W}_{V\theta}$ are the transformation matrices of the θ feature subspace.

2) The scaled dot product and the softmax function are used to calculate the self-attention values in the θ feature subspace as:

$$head_\theta = \text{softmax}(\boldsymbol{Q}_{a\theta}, \boldsymbol{K}_{a\theta}^{\mathrm{T}} / \sqrt{d}) \boldsymbol{V}_{a\theta} \tag{10}$$

where d is the scaling factor and $head_\theta$ is the self-attention value in the θ feature subspace.

3) The process of fusing self-attention into the θ feature subspace is as follows:

$$M_{\text{multi-head}} = C_{\text{Concat}}(head_1, head_2, \cdots, head_\theta) \boldsymbol{W}^o \tag{11}$$

where $M_{\text{multi-head}}$ is the fused value of self-attention; C_{Concat} is the matrix concatenation operation; and $\boldsymbol{W}^o$ is the parameter matrix.

B. Convolutional Self-Encoder Based on Multi-Head Self-Attention

The convolutional self-encoder has a good ability to extract non-linear features, and is thus suitable for capturing dynamic features from curves of the data. The encoder constructed in this paper is an improvement of the convolutional autoencoder, and the multi-head self-attention mechanism is used to enhance the quality of the features that are extracted by the convolutional network, to generate better scenarios for grid planning and dispatching.

The autoencoder contains two parts: an encoder and a decoder. Backpropagation between them is used to iteratively train and fine-tune the parameters of the network while minimizing the loss of reconstruction to obtain the optimal features in low-dimensional space.

The convolutional layer feeds the feature-related information captured from the temporal data on wind and PV power, and load into the multi-head self-attention layer to further extract features, and its coding process is as follows:

$$h_{\text{conv}} = \sigma(X_{\text{conv}} \omega_{\text{conv}} + b_{\text{conv}}) \tag{12}$$

where h_{conv} is the output feature-related information of the convolution layer; σ is the rectified linear unit activation function; X_{conv} represents the temporal data after data processing; while ω_{conv} and b_{conv} are the numbers of convolution kernels and the bias of the convolution layer, respectively.

The process of decoding is as follows:

$$h_{\text{deconv}} = \sigma(h_{\text{conv}} \omega_{\text{deconv}} + b_{\text{deconv}}) \tag{13}$$

where h_{deconv} is the output feature-related information of the deconvolution layer; while ω_{deconv} and b_{deconv} are the numbers of convolution kernels and the bias of the deconvolution layer, respectively.

The proposed model minimizes the mean-squared error as the function of reconstruction loss to continuously optimize the parameters of the encoder and the decoder using the following expressions:

$$L_r = \sum_{j=1}^{N_d} \left\| X_{\text{conv}} - h_{\text{deconv}} \right\|_2^2 / N_d \tag{14}$$

where L_r is the function of reconstruction loss; and N_d is the number of days for wind power, PV power, and load data.

C. Joint Optimization of Features and Scenarios

Based on the low-dimensional features obtained in the autoencoder, the K-means algorithm is applied to initialize the clustering centers. Then, based on the joint loss function, the Adam optimizer is chosen to adjust the network's parameters to guarantee the representativeness of the embedded features, and to obtain the optimal clustering results. This joint loss function is given by:

$$L = L_r + \gamma L_c \tag{15}$$

where L is the joint loss function; γ is the coefficient used to suppress the degree of deformation in low-dimensional space, generally set to 0.1; and L_c is the loss function of clustering described by the KL divergence.

The clustering layer uses the center of the cluster μ_j as its connection weight to the low-dimensional feature Z_i, and maps each feature Z_i to a soft label. To improve the accuracy of fitness of Z_i and μ_j, the Gaussian distribution is used as the ideal target distribution. The clustering loss thus describes the KL divergence between the distribution on the soft labels and the Gaussian distribution, and is used to measure the similarity between them. The procedure is as follows:

$$q_{ij} = \left(1 + \left\|Z_i - \mu_j\right\|^2\right)^{-1} / \sum\nolimits_j \left(1 + \left\|Z_i - \mu_j\right\|^2\right)^{-1} \tag{16}$$

$$L_c = \mathrm{KL}(B\|Q) = \sum\sum b_{ij} \log(b_{ij}/q_{ij}) \tag{17}$$

$$b_{ij} = (q_{ij}^2 / \sum\nolimits_i q_{ij}) / \sum\nolimits_j (q_{ij}^2 / \sum\nolimits_i q_{ij}) \tag{18}$$

where q_{ij} is the probability that the low-dimensional feature Z_i belongs to the cluster center μ_j; and b_{ij} is the auxiliary function of the target distribution.

D. Optimization of DCEC-MS Network Based on Adam

The Adam optimizer integrates the advantages of first-order momentum based on the stochastic gradient descent and second-order momentum based on mean-squared propagation. The means of momenta of both are used to exploit the performance of the sparse gradient while maintaining the learning rate of each parameter so that the algorithm is robust against non-stationary problems. The computational procedure is as follows:

$$g_{tt} = \nabla_\varphi F_{tt}(\varphi_{tt} - 1) \tag{19}$$

$$M_{tt} = \beta_1 \cdot M_{tt-1} + (1 - \beta_1) \cdot g_{tt} \tag{20}$$

$$\eta_{tt} = \beta_2 \cdot \eta_{tt-1} + (1 - \beta_2) \cdot g_{tt}^2 \tag{21}$$

$$\hat{M}_{tt} = M_{tt} / (1 - \beta_1) \tag{22}$$

$$\hat{\eta}_{tt} = \eta_{tt} / (1 - \beta_2) \tag{23}$$

$$\varphi_{tt} = \varphi_{tt-1} - \psi \cdot \hat{M}_{tt} / (\sqrt{\hat{\eta}_{tt}} + \varepsilon) \tag{24}$$

where *tt* is the time interval; g_{tt} is the gradient; M_{tt} is the first-order moment estimate of g_{tt}; φ_{tt} is a model parameter; η_{tt} is the second-order moment estimate of g_{tt}; while $\hat{M}_{tt}$ and $\hat{\eta}_{tt}$ are their corresponding network outputs; ψ is the step value of the network; β_1 is the rate of exponential decay of M_{tt}; generally set to 0.9; β_2 is the rate of exponential decay of η_{tt}, generally set to 0.999; and ε is a constant used to ensure the robustness of the algorithm.

V. Case Study

The TensorFlow 2.6.0 runtime environment is used for case studies. The time-series data of wind power, PV power, and load (8760×3 points) from a regional grid at a sampling interval of 1 hour in 2018 is chosen to verify the effectiveness of the proposed model.

A. Data Processing

MAE is chosen to quantitatively analyze the accuracy of the data processing models. Three sets of time-series data on wind and PV power, and load are applied to assess the performances of four data processing models: the mean filter, median filter, SMA+VMD, and the improved SMA+VMD. The results are shown in Table I.

TABLE I
Comparison of Data Processing Methods

Method	Wind	PV	Load
Improved SMA+VMD	1.36	0.86	15.73
SMA+VMD	5.01	1.53	34.65
Mean value filtering	28.22	24.41	490.56
Median filtering	8.15	2.28	100.65

From Table I, it can be seen that compared with the two traditional methods of denoising of mean and median filtering, the index values of the improved SMA+VMD and SMA+VMD models are significantly smaller on different sets of time-series data, with a year-on-year rate of reduction more than 32.89% and a maximum year-on-year rate of reduction of 96.79%. It shows that the VMD model retains the original characteristics of the data while removing outliers, and verifies its superiority in terms of data processing. The improved SMA+VMD has lower values on the evaluation indices than SMA+VMD on all three datasets of the time-series, with year-on-year reductions of 72.85%, 43.79%, and 54.60%, respectively. This indicates that the improved SMA strengthens the model's capability of performing global search, prevents it from falling into a local optimum, and enables it to search for the global optimum, i.e., the best global combination of the parameters of VMD.

B. Generating Clustering Scenarios

1) Parameters of DCEC-MS Model

The normalized temporal data on wind power, PV power, and load are resized from 8760×3 to 24×3×365 as the inputs to the DCEC-MS model. Model training is divided into two parts. The first part involves generating

the initial centers of clustering by pre-training the model with the Adam optimizer over 200 iterations at a learning rate of 0.001, while the second part involves optimizing the features of the scenarios by fine-tuning the network using the Adam optimizer with a maximum of 20 000 iterations and an error of 0.001. The parameters of the DCEC-MS model are shown in Table II.

TABLE II
PARAMETERS OF THE DCEC-MS MODEL

Type	Input Size	Number of convolution kernels	Step length	Output size
Input	9×9×1			9×9×1
Conv1	9×9×1	3	1	9×9×3
Conv2	9×9×3	9	2	5×5×9
Conv3	5×5×9	18	2	2×2×18
Multi_attention1	2×2×18			2×2×18
Multi_attention2	2×2×18			2×2×18
Multi_attention3	2×2×18			2×2×18
Flatten	2×2×18			72
Embedding	72			2
Clustering	2			6
Dense	2			72
Reshape	72			2×2×18
Deconv3	2×2×18	9	2	5×5×9
Deconv2	5×5×9	3	1	9×9×3
Deconv1	9×9×3	1	1	9×9×1

2) Generating the Optimal Scenario

The elbow method is applied on the feature-related information embedded into the low-dimensional space [28] to observe the elbow values of classes with different numbers of clusters, as shown in Fig. 3. As can be seen, the curve exhibits a prominent inflection point when the number of clusters is six. Then the optimal number of clusters is set accordingly.

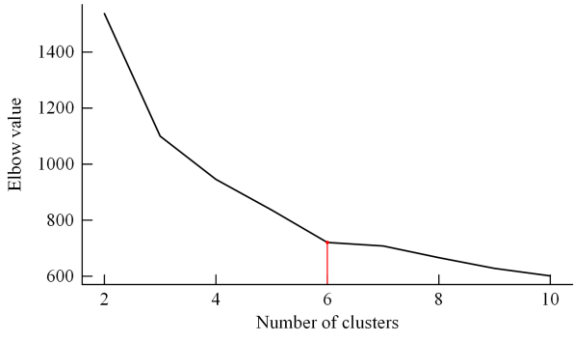

Fig. 3. Results of optimal clustering of the scenarios

The results of the optimal clustering of the scenarios during the year are shown in Fig. 4, while the data distribution and centers of clustering of each scenario are shown in Fig. 5.

Figures 4 and 5 show that scenario I is distributed over all months of the year, and reflects cloudy and breezy weather. The output of PV power is moderate in this scenario because solar radiation is blocked by clouds. The overall output of wind power is low, and its curve has a concave shape. Scenario II is mainly distributed over spring and autumn. The PV output is moderate, with a maximum of 0.52, while the output of wind power is stable. Scenario III is distributed in the summer, and features high temperatures, weak winds, and sunshine. This results in a reduction in atmospheric density, wind speed, and consequently, the wind power output. The maximum output of wind power is only 0.1. Because of the significant increase in solar radiation in summer in the region, the impact of high temperature on the output of PV power lessens as the output increases. The maximum PV output is 0.62. As a large number of cooling devices, such as air conditioners, are used in summer, the daytime load is high until the temperature drops at night. Scenarios IV and V have similar probabilities of occurrence, with moderate outputs of PV and wind power, both exhibiting smooth rises in the night and during the day. Scenario VI is mainly distributed in autumn and winter. The solar radiation in this scenario decreases, so the PV power only has a maximum output of 0.4. Wind power is more stable, and fluctuates around 0.62. As the region is located in the south, there is little need for heating equipment in the winter, so the load does not increase significantly. In summary, the differences in power generation and load among the scenarios illustrate their adequacies.

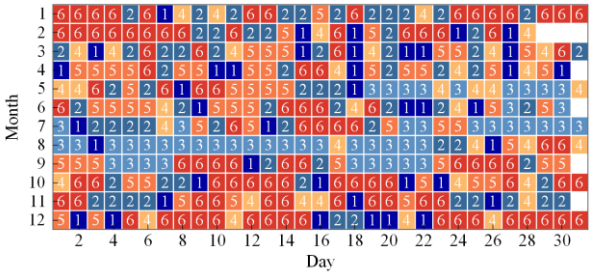

Fig. 4. Reconstruction loss of the autoencoder.

C. Assessing the Quality of the Generated Scenarios

The high-quality scenarios generated above are required to have the following characteristics: 1) the temporal characteristics of the scenarios involving wind power, PV power, and load should represent their historical characteristics; 2) the scenarios should have a high intra-cluster similarity and a low inter-cluster similarity.

1) Comparison of the Reconstruction Loss of the Encoder

Reconstruction loss is used to analyze the ability of the temporal features extracted by the DCEC-MS model to represent the historical characteristics of wind power, PV power, and load. The Adam optimizer is applied to train the DEC and DCEC-MS, while the autoencoders of both models perform feature extraction and reconstruction on the time-series data. A comparison of their reconstruction losses is shown in Fig. 6.

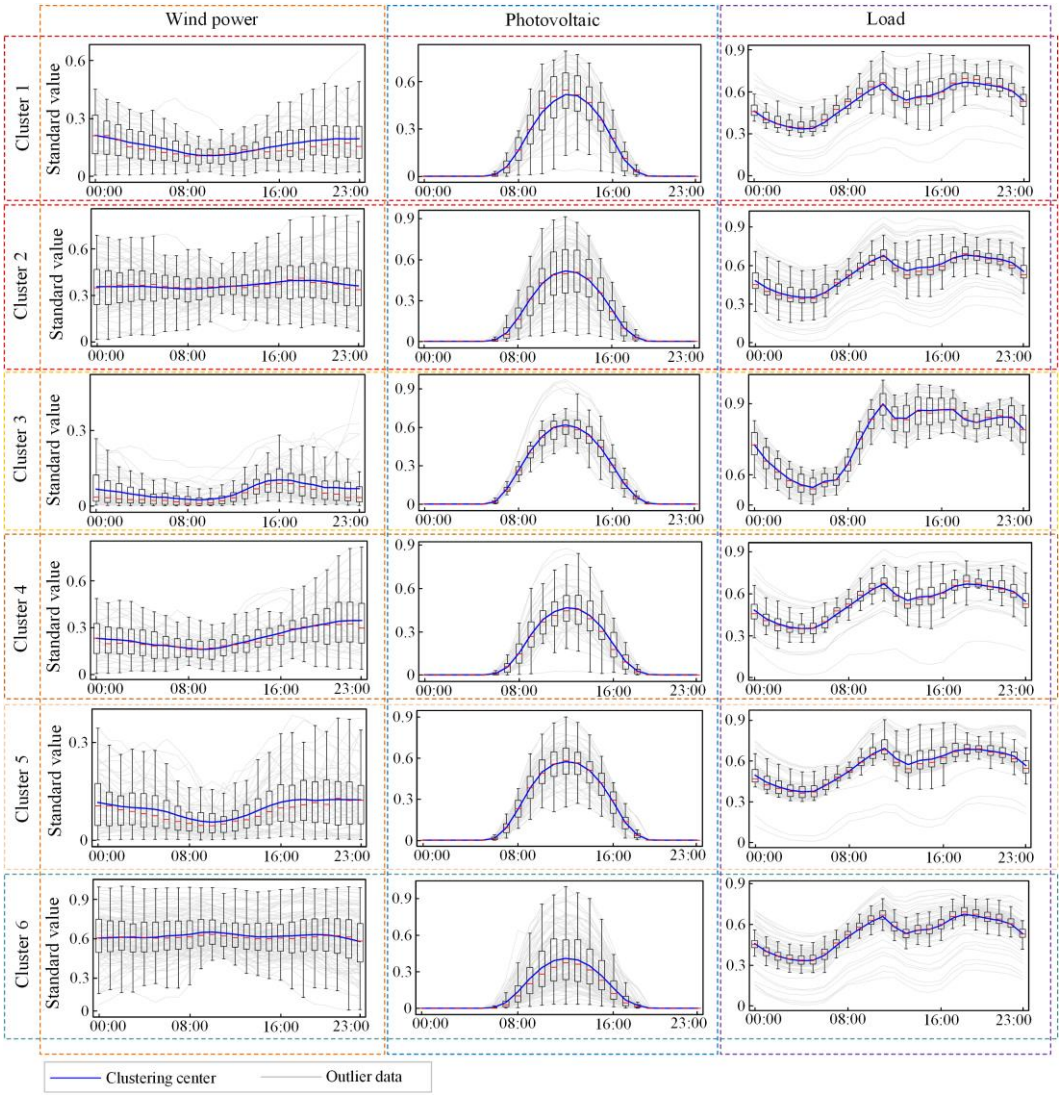

Fig. 5. Data distribution and clustering center of each scenario.

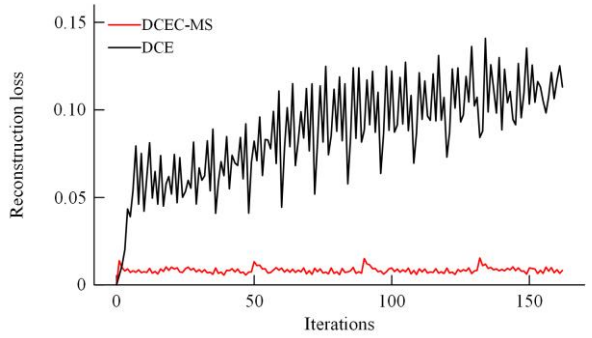

Fig. 6. Reconstruction loss of the autoencoder.

Because the traditional DEC model considers only the loss of clustering as the loss function of the overall network for its iterative fine-tuning, this distorts the low-dimensional embedding space and affects the reconstruction of data by the autoencoder. With an increasing number of iterations, the reconstruction loss experiences a slight rise, followed by stabilization, and eventually fluctuates around 0.11. The DCEC-MS suppresses spatial distortion by considering the losses of reconstruction and clustering to form a joint loss function. Consequently, its reconstruction loss stabilizes at around 0.0069 as the number of iterations increases, with a reduction of 93.7% compared with that of DEC. This shows that the ability of the DCEC-MS model for data reconstruction is superior to the DEC model, and its generated scenarios can better represent the characteristics of wind power, PV power, and load.

2) Comparative Analysis of the Validity of Clustering

Three metrics are chosen to assess the quality of clustering, i.e., the Calinski–Harabasz index (CHI), the silhouette coefficient (SC), and the Davies–Bouldin index (DBI), and to quantitatively analyze the inter-cluster and intra-cluster similarities of the five clustering models of K-means, PCA+K-means, DEC, DCEC, and DCEC-MS. The results are shown in Table III. CHI is obtained from the ratio of inter-cluster separation to intra-cluster tightness, and SC is used to measure the similarity between each data point and the cluster to

which it belongs. In this context, the larger the two values, the better is the effect of clustering. DBI, which is the ratio of the sum of intra-cluster distance to inter-cluster distance, is used to measure the mean value of the maximum similarity between scenarios in each class. A smaller DBI value indicates a higher accuracy of clustering.

TABLE III
COMPARISON OF METHODS ON INDICES TO EVALUATE THE RESULTS OF CLUSTERING

Methods	CHI	SC	DBI
K-means	174.46	0.21	1.46
PCA+K-means	488.42	0.38	0.89
DEC	2656.51	0.53	0.53
DCEC	2740.45	0.65	0.48
DCEC-MS	4948.37	0.83	0.24

Table III shows that compared to traditional K-means clustering, the model that initially uses PCA data for dimension reduction and then for clustering has a CHI value 313.96 higher, a SC value 0.17 higher, and a DBI value 0.57 lower. This verifies the feasibility of dimension reduction to improve the accuracy of clustering. A comparison of PCA+K-means with DEC shows that the CHI of the latter is 2168.09 higher, while its SC increases by 0.15 and DBI decreases by 0.36. This shows that embedding the clustering algorithm into dimension reduction and feature extraction helps improve the accuracy of clustering. Compared with DEC, the CHI of DCEC is 83.94 higher, while its SC increases by 0.12 and DBI decreases by 0.05. This shows that the joint loss function of DCEC guarantees the representativeness of the embedded features, and the convolutional encoder is able to better explore the latent feature-related information in the data on wind and PV power, and load than the stacked encoder. Compared with DCEC, the CHI of DCEC-MS is 2207.92 higher, while its SC is 0.18 higher and DBI decreases by 0.24. This shows that the multi-head self-attention mechanism is able to compensate for the shortcomings of small convolutional kernels and convolutional layers with variable step lengths to capture a large amount of one-sided information on the features of wind power, PV power, and load to improve the accuracy of clustering.

T-distributed stochastic neighbor embedding is introduced to visualize the performance of the clustering models in terms of inter-cluster and intra-cluster similarities. Larger inter-cluster distance and smaller intra-cluster distance lead to better clustering performance, as shown in Fig. 7.

As seen in Fig. 7, the boundaries of various scenarios in the K-means model are directly fed data on wind power, PV power, and load overlapped with one another, and are not adequately classified. The model based on PCA data reduces dimension and clusters, so the boundaries of various types of scenarios overlap each other and the distance between data within the clusters is larger, indicating a lower accuracy of clustering. The boundaries of various scenarios in the DEC- and DCEC-based models are clear, and the compactness of the data within the clusters of the former is slightly inferior to that of the latter. This shows the clustering process after embedded data dimensionality reduction and feature extraction outperforms the clustering process where data dimensionality reduction and feature extraction are independent of each other. However, the classification results of each scenario in DCEC are still not close enough and the compactness is lower than DCEC-MS. In the model based on DCEC-MS, the embedded features after feature optimization are visualized in the data, and it is clear that the scenario data are divided into 6 classes and the intra-cluster spacing is close.

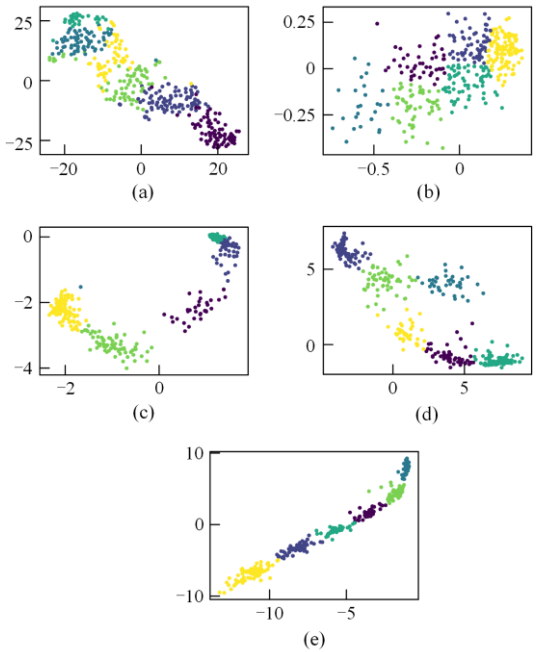

Fig. 7. Visualization of the results of clustering. (a) K-means. (b) PCA+K-means. (c) DEC. (d) DCEC. (e) DCEC-MS.

In summary, the proposed method captures important information on the features of scenarios involving wind power, PV power, and load, and contains fewer outlier data. The results of clustering of the scenarios are thus superior to those of the other models considered.

VI. CONCLUSION

This paper has proposed the DCEC-MS model to cluster data to mitigate the adverse effects of uncertainties in the power supply and demand for optimal dispatching and planning of power grids. The proposed model solves the problems of extracting and processing features from high-dimensional time-series data on wind power, PV power, and load. First, a multi-head

self-attention mechanism is used to compensate for the shortcomings of small convolutional kernels and convolutional layers with variable step lengths to capture deep feature-related information from time-series data. Consequently, the features in low-dimensional embedded space can better represent the actual characteristics of wind and PV power, and load. Second, a joint loss function is used to combine feature extraction and clustering to suppress the distortion of low-dimensional space. This guarantees the representativeness of the embedded features and improves the accuracy of clustering. The datasets on wind power, PV power, and load are applied to verify the performance of the proposed model by using indices to assess its clustering, method of data visualization, and data distribution within the clusters.

ACKNOWLEDGMENT

We thank the editors and the reviewers for their useful feedback that improved this paper.

AUTHORS' CONTRIBUTIONS

Lijun Liu: methodology, validation, formal analysis, and writing the original draft. Xin Hu: software, visualization, and reviewing and editing. Junsheng Chen: resources, review, and editing. Ruixing Wu: editing and formal analysis. Feixiong Chen: data support and some financial support. All authors read and approved the final manuscript.

FUNDING

This work is supported by the Special Grant from the Department of Finance, Fujian Province (No. 83022005) and the National Natural Science Foundation of China (No. 52107080).

AVAILABILITY OF DATA AND MATERIALS

Please contact the corresponding author for data material request.

DECLARATIONS

Competing interests: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

AUTHORS' INFORMATION

Lijun Liu was born in Fujian, China, in 1982. She received the B.S., M.S., and Ph.D. degrees from the School of Electrical Engineering and Automation, Fuzhou University, Fuzhou, China, in 2003, 2006 and 2021, respectively. Since 2016, she has been an associate professor with the School of Electrical Engineering and Automation, Fuzhou University. From 2019 to 2020, she was a Visiting Scholar in the University of Michigan, Michigan, the United States. To date, she has authored or coauthored 15 journal articles and conference papers. Her current research interests include power system planning and operation with new energy generation.

Xin Hu was born in Fujian, China, in 1999. He received the B.S. degree from the School of Information Engineering in Nanchang University, Nanchang, China, in 2021. He is currently working toward the M.S. degree with the School of Electrical Engineering and Automation, Fuzhou University, Fuzhou, China. His current interests include power system operation and control, and renewable energy application.

Junsheng Chen was born in Fujian, China, in 1998. He received the B.S. and M.S. degrees from the School of Electrical Engineering and Automation in Fuzhou University, Fuzhou, China, in 2020 and 2023, respectively. He is currently with the State Grid Fuzhou Power Supply Company. His research interests include Power system planning and extreme scenarios.

Ruixing Wu was born in Fujian, China, in 2000. He received the B.S. degree from the School of Electrical Engineering and Automation, Fuzhou University, Fuzhou, China, in 2022. He is currently working toward the M.S. degree with the School of Electrical Engineering and Automation, Fuzhou University, Fuzhou, China. His research interests include the distribution network optimization and scheduling.

Feixiong Chen was born in Fujian, China, in 1990. He received the B.S. degree in electrical engineering from the Chongqing University of Technology, in 2012, and the Ph.D. degree in electrical engineering from Chongqing University, Chongqing, China, in 2017. From 2016 to 2017, he was a visiting Ph.D. student with the Department of Electrical and Computer Engineering, Wayne State University, Detroit, MI, USA. From 2017 to 2018, he was a research associate with the Department of Electrical Engineering, The Hong Kong Polytechnic University, Hong Kong, China. Since 2019, he has been with the Fujian Smart Electrical Engineering Technology Research Center, College of Electrical Engineering and Automation, Fuzhou University, Fuzhou, China, where he is currently an associate professor. His current research interests include distributed control and optimization methods of the energy internet, and energy trading in distribution networks.

REFERENCES

[1] H. Huang, M. Zhou, and S. Zhang *et al*., "Exploiting the operational flexibility of wind integrated hybrid AC/DC power systems," *IEEE Transactions on Power Systems*,

vol. 36, no. 1, pp. 818-826, Jan. 2021.
[2] S. Liu, T. Zhang, and Z. Lin *et al*., "Controlled islanding strategy considering uncertainty of renewable energy sources based on chance-constrained model," *Journal of Modern Power Systems and Clean Energy*, vol. 10, no. 2, pp. 471-481, Mar. 2021.
[3] N. Huang, W. Wang, and G. Cai *et al*., "The joint scenario generation of multi source-load by modular denoising variational autoencoder considering the complex coupling characteristics of meteorology," *Proceedings of the CSEE*, vol. 39, no. 10, pp. 2924-2934, May 2019. (in Chinese)
[4] X. Fu, X. Wu, and C. Zhang *et al*., "Planning of distributed renewable energy systems under uncertainty based on statistical machine learning," *Protection and Control of Modern Power Systems*, vol. 7, no. 4, pp. 619-645, Dec. 2022.
[5] B. Zhang, P. Dehghanian, and M. Kezunovic, "Optimal allocation of PV generation and battery storage for enhanced resilience," *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 535-545, Jan. 2019.
[6] J. GOU, F. LIU, and J. LIU *et al*., "A transmission network planning method considering high-speed railway load and wind and solar uncertainty," *Power System Protection and Contro*l, vol. 51, no. 9, pp. 156-164, May 2023. (in Chinese)
[7] X. Fu, "Statistical machine learning model for capacitor planning considering uncertainties in photovoltaic power," *Protection and Control of Modern Power Systems*, vol. 7, no. 1, pp. 51-63, Jan. 2022.
[8] P. Li, L. Song, and J. Qu *et al*., "A two-stage distributionally robust optimization model for wind farms and storage units jointly operated power systems," *IEEE Access*, vol. 9, pp. 111132-111142, Jul. 2021.
[9] L. Liu, T. Wu, X. Chen *et al*., "Multi-objective optimal allocation of DG and EV charging station based on space-time characteristics and demand response," *Electric Power Automation Equipment*, vol. 41, no. 11, pp. 48-56, Nov. 2021. (in Chinese)
[10] J. Wu, X. Chen, and S. Badakhshan *et al*., "Spectral graph clustering for intentional islanding operations in resilient hybrid energy systems," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 4, pp. 5956-5964, Apr. 2022.
[11] Y. Li, Y. Liu, and J. Liu *et al*. "Self-adaptive and robust clustering algorithm for daily load profiles based on morphological distance," *Proceedings of the CSEE*, vol. 39, no. 12, pp. 3409-3420, Jun. 2019. (in Chinese)
[12] J. Gu, S. Liu, and Y. Hu *et al*., "Optimal allocation of intermittent distributed generation based on deep convolutions generative adversarial network in scenario generation," *Power System Technology*, vol. 45, no. 5, pp. 1742-1751, May 2021. (in Chinese)
[13] O. Motlagh, A. Berry, and L. O'Neil, "Clustering of residential electricity customers using load time series," *Applied Energy*, vol. 237, pp. 11-24, Mar. 2019.
[14] J. Wang, K. Wang, R. Jia *et al*., "Research on load clustering based on singular value decomposition and K-means clustering algorithm," in *IEEE Asia Energy and Electrical Engineering Symposium* (AEEES), Chengdu, China, May 2020, pp. 831-835.
[15] S. Liu, S. You, and Z. Lin *et al*., "Data-driven event identification in the US power systems based on 2D-OLPP and RUSBoosted trees," *IEEE Transactions on Power System*s, vol. 37, no.1, pp. 94-105, Jan. 2021.
[16] J.Y. Xie , R.B. Girshick, and A. Farhadi *et al*., "Unsupervised deep embedding for clustering analysis," in *Proceedings of the 33rd JMLR International Conference on Machine Learning*, New York, USA, Jul. 2016, pp. 478-487.
[17] X. Guo, X. Gao, and X. Liu *et al*., "Improved deep embedded clustering with local structure preservation," in *IJCAI*, Melbourne, Australia, Feb. 2017, pp. 1753-1759.
[18] X. Guo, X. Liu, and E. Zhu *et al*., "Deep clustering with convolutional autoencoders," in *International Conference on Neural Information Processing*, Guangzhou, China, Aug. 2017, pp. 373-382.
[19] J. Yang, J. Liu, X. Han *et al*., "An uncertain hydro/PV/load typical scenarios generation method based on deep embedding for clustering," *Proceedings of the CSEE*, vol. 40, no. 22, pp. 7296-7306, Nov. 2020. (in chinese)
[20] Y. Bai, Y. Zhou, and J. Liu, "Clustering analysis of daily load curve based on deep convolution embedded clustering," *Power System Technology*, vol. 46, no. 6, pp. 2104-2113, Jun. 2022. (in Chinese)
[21] A. Vaswani, N. Shazeer, and N. Parmar *et al*., "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 30, Sep. 2017.
[22] Y. Jin, C. Tang, Q. Liu *et al*., "Multi-head self-attention-based deep clustering for single-channel speech separation," *IEEE Access*, vol. 8, pp. 100013-100021, Jul. 2020.
[23] D. Wang, Z. Zhang, and Y. Jiang *et al*., "DM3Loc: multi-label mRNA subcellular localization prediction and analysis based on multi-head self-attention mechanism," *Nucleic Acids Research*, vol. 49, no. 8, Aug. 2021.
[24] Z. Chen, K. Li, and Y. Li *et al*., "Novel method based on variational mode decomposition and a random discriminative projection extreme learning machine for multiple power quality disturbance recognition," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 5, pp. 2915-2926, May 2019.
[25] P. Jia, H. Zhang, and X. Liu *et al*., "Short-term photovoltaic power forecasting based on VMD and ISSA-GRU," *IEEE Access*, vol. 9, pp. 105939-105950, Jul. 2021.
[26] S. Li, H. Chen, and M. Wang *et al*., "Slime mould algorithm: a new method for stochastic optimization," *Future Generation Computer Systems*, vol. 111, pp. 300-323, Apr. 2020.
[27] L. Xiao, X. Hu, and Y. Chen *et al*., "Multi-head self-attention based gated graph convolutional networks for aspect-based sentiment classification," *Multimedia Tools and Applications*, vol. 81, no. 14, pp. 19051-19070, 2022.
[28] A. A. Munshi, "Clustering of wind power patterns based on partitional and swarm algorithms," *IEEE Access*, vol. 8, pp. 111913-111930, Jul. 2020.