Article

Typical Power Grid Operation Mode Generation Based on Reinforcement Learning and Deep Belief Network

Zirui Wang [1,2,*], Bowen Zhou [1,2,*], Chen Lv [3], Hongming Yang [1,2], Quan Ma [3], Zhao Yang [3] and Yong Cui [4]

1 College of Information Science and Engineering, Northeastern University, Shenyang 110819, China; 2271037@stu.neu.edu.cn

2 Key Laboratory of Integrated Energy Optimization and Secure Operation of Liaoning Province, Northeastern University, Shenyang 110819, China

3 China Electric Power Research Institute, Beijing 100192, China; lvchen@epri.sgcc.com.cn (C.L.); xitong-maquan@epri.sgcc.com.cn (Q.M.); yangzhao@epri.sgcc.com.cn (Z.Y.)

4 State Grid Shanghai Municipal Electric Power Company, Shanghai 201507, China; cuiyong@sh.sgcc.com.cn

* Correspondence: 2271032@stu.neu.edu.cn (Z.W.); zhoubowen@ise.neu.edu.cn (B.Z.)

Abstract: With the continuous expansion of power grids and the gradual increase in operational uncertainty, it is progressively challenging to meet the capacity requirements for power grid development based on manual experience. In order to further improve the efficiency of the operation mode calculation, reduce the consumption of manpower and material resources, and consider the sustainability of energy development, this paper proposes a typical power grid operation mode generation method based on Q-learning and the deep belief network (DBN) for the first time. Firstly, the operation modes of different generator combinations located in different regions are obtained through Q-learning intelligent generation. Subsequently, the generated operation modes are clustered as different operation mode sets according to the data characteristics. Furthermore, comprehensive evaluation indexes are proposed from the perspectives of the steady state, transient state, and the economy. These multi-dimensional indexes are integrated via the analytical hierarchy process–entropy weight method (AHP-EWM) to enhance the comprehensibility of the evaluation system. Finally, DBN is introduced to construct a rapid operation mode evaluation model to realize the evaluation of operation mode sets, and typical operation mode sets are obtained accordingly. In this way, the system calculator only needs to compare the composite values to obtain the typical operation modes. The proposed method is validated by the Northeast Power Grid in China. The experimental results show that the proposed method can quickly generate typical power grid operation modes according to actual demand and greatly improve the efficiency of operation mode calculation.

Keywords: power grid operation mode; Q-learning; deep belief network; clustering; artificial intelligence

1. Introduction

In the context of large-scale access to a power grid by a high proportion of renewable energy sources, the power grid structure is becoming progressively more complex and the number of power grid operation modes that need to be taken into account and the difficulty of analysis are increasing. Traditional calculation and analysis methods based on manual experience are unable to address the difficulties caused by power grid improvements, so there is an urgent need to generate efficient and accurate typical operation modes.

The purpose of operation mode calculations is to ensure that the power production and load consumption of the whole power grid are equal; however, with the increasing scale and complexity of the power grid, the operation mode calculation workload is large and the calculations are inefficient and dependent on manual labor [1]. Lan et al. [2] proposed a sample generation framework and training method combining the generative adversarial

network (GAN) and model-based transfer learning, which could efficiently obtain a high-performance typical operation mode sample generation model. In order to ensure secure and economical operation modes, Iqbal et al. [3] proposed a novel optimized coordination strategy for frequency regulation via electric vehicles. Li et al. [4] proposed a fast generation method of power system operation modes based on optimal power flow, and the experimental results showed that this method had the characteristics of fast calculation speed and high accuracy. Xu et al. [5] introduced an improved deep-reinforcement-learning-based approach to obtain a convergent and feasible power flow state automatically, significantly saving the workload of the operation state calculation. However, the methods proposed in the above literature are verified in the examples of the IEEE model, and do not consider how to apply them in the actual power grid. Reinforcement learning has demonstrated superiority in improving learning performance. With the abilities of autonomous learning and exploration, adaptability to environmental changes, and interaction with the environment, it can effectively support the sequential decision-making process [6]. Q-learning is a widely used reinforcement learning method [7], which is characterized by not relying on an a priori model of the environment and requiring fewer parameters. It can make good decisions in the face of high-dimensional, complex environments, and has been applied to load forecasting [8] and multi-intelligence body reinforcement learning [9]. Xi and Lei [10] presented the QTLBO and OTLBO algorithms, which integrate Q-learning and metaheuristics, to proficiently address the challenges of solving the distributed two-stage hybrid flow shop scheduling problem characterized by fuzzy processing times. Kushwaha et al. [11] introduced a pioneering approach grounded in Q-learning, enabling intelligent wind speed sensor-less maximum power point tracking. This innovative methodology facilitates real-time peak power tracking even under fluctuating conditions. To solve the energy-sharing problem, Cao et al. [12] designed an energy-sharing algorithm based on the Q-learning algorithm. Q-learning can provide good decision making in the face of different situations, so this paper proposes to combine it with the actual power grid to generate typical power grid operation modes.

System operators analyze historical typical operation modes, summarize the operation characteristics, and formulate safe and stable operation regulations. To guarantee the relevance of power grid security and economic operations, along with the applicability of operation regulations to ensure stability, these personnel meticulously monitor each mode's performance under boundary circumstances. A comprehensive and reasonable index system is the key to evaluating operation modes, while a scientific and appropriate index integration method can reflect the complete and objective level of power grid operation [13]. Common index integration methods include fuzzy decision theory [14], fuzzy comprehensive evaluation [15], the entropy weight method [16], AHP-EWM [17], and so on. Zhou et al. [18] introduced an original method rooted in fuzzy comprehensive evaluation to optimize suppression techniques, thereby reducing the impact of magnetic flux leakage generated by air core reactors in static var compensators. Lo et al. [19] proposed an innovative multi-criteria decision analysis model to discern critical failure modes in products and systems, demonstrating its efficacy through a practical case study. Wang et al. [20] advocated for evaluating high-voltage direct-current protection system reliability using the AHP-EWM, which provided both qualitative and quantitative assessments of equipment performance, confirming its effectiveness and suitability. However, the above methods are limited by the computational efficiency of the evaluation indexes.

The recent remarkable advancements in deep learning have ushered in a novel cognitive approach for efficient assessments of operation modes. This entails the construction of machine learning models endowed with multiple concealed layers and an extensive corpus of training data, thereby enabling an enhanced determination of pivotal attributes [21]. Among deep learning methods, the deep belief network (DBN) [22] has received wide attention; it adopts a two-stage learning method of forward pre-training and reverse fine-tuning of parameters to train network parameters. A trained DBN model can fully utilize the feature extraction advantages of the deep architecture. DBN has been used in uncertainty

factor generation, such as customer-side load [23] and uncertainty generation, as well as fault diagnosis [24] and power grid transient analysis [25]. Su et al. [26] integrated DBN and the Non-dominated Sorting Genetic Algorithm (NSGA-III) to develop a new preventive control method for a power system. Li and Wu [27] integrated DBN and active learning based on information entropy to conduct a transient stability assessment within a power grid. Zhang et al. [28] introduced an innovative framework that seamlessly incorporates DBN alongside Adaboost algorithms, aimed at achieving precise and efficient power demand forecasting.

With the aim of enhancing index computation efficiency, in this study, we introduce a comprehensive power grid operation mode evaluation model based on DBN. This approach capitalizes on the DBN's intrinsic feature extraction ability to establish the correlation between pivotal grid variables and all-encompassing evaluation metrics.

The contributions of this paper are as follows:

1. A power grid operation mode set extraction method based on Q-learning and DB-SCAN clustering is proposed. Based on several typical operation modes used in previous years, the load level is expanded by 5% year-on-year in order to simulate the growth trend of load, and through the reinforcement learning algorithm Q-learning, the operation modes of different optimal combinations of generators located in different regions are intelligently generated. Then, based on the data characteristics of each operation mode, the DBSCAN clustering algorithm is applied to divide the operation modes into different clusters to extract operation mode sets. It should be noted that in this paper, operation modes are generated without considering power grid structure changes.

2. A rational evaluation index system for the operation modes collection is established. The key indexes to meet the demand of actual work procedures are selected from three perspectives: steady state, transient state, and economy mode. Index calculation can replace the system operator calculation process. At the same time, in order to facilitate the system operator to compare each mode, and the application of the analytical hierarchy process–entropy weight method (AHP-EWM) for the fusion of the weights of the multi-dimensional indexes, the results of the operation of the multi-indicator mode are synthesized into a single composite value, which effectively reduces the workload of the system operator.

3. An evaluation model of operation mode sets is established based on DBN. This paper proposes a fast evaluation method of operation modes based on DBN, which no longer needs to calculate each index of operation modes, but constructs the correlation relationship between the feature data of operation modes and the composite values through a neural network, to quickly and accurately obtain the composite values of operation modes. In this way, the system operator can select the highest value of each operation mode set according to the comprehensive value, and then obtain the typical operation mode.

The sections of this paper are organized as follows: Section 2 presents the methods, including the reinforcement learning algorithm for intelligent operation mode generation, the operation mode clustering method based on the historical typical operation mode, and the operation mode evaluation indexes and the deep learning-based operation mode evaluation framework. Section 3 presents simulation examples and results of the Northeast Power Grid in China. Section 4 discusses the results and compares them with other research. Section 5 concludes the full paper.

2. Methods

2.1. Reinforcement-Learning-Based Power Grid Operation Mode Generation Model

2.1.1. Actual Engineering Needs

In practical engineering, after calculating the converged power flow, the system operator still needs to adjust the power of key transmission sections [29] and analyze the transient stability to determine the safe and stable operation boundary of the power grid.

The whole process is cumbersome, inefficient, overly dependent on manual labor, and unable to keep up with the increasing scale and complexity of the power grid operation mode calculations. Therefore, an operation mode generation method combining manual experience and reinforcement learning is proposed. In this paper, we consider the adjustment of generator switching without preserving the reserve capacity.

The power grid power flow convergence adjustment process can be regarded as a sequential decision-making process based on reinforcement learning, in which the system operator is regarded as the agent and the power flow calculation program is regarded as the environment. The system operator, as an agent, adjusts power grid components by performing a series of actions, and the effect of this adjustment on each action can be quantitatively described by feeding back the instantaneous reward value, which reflects the change in operation mode after each adjustment step. Through continuous interaction with the environment, the intelligent agent gradually learns which actions lead to significant improvements in the operation mode, and thus can make more optimal decisions. The increase in cumulative rewards represents the improved performance and operational efficiency of the system. In this process, the intelligent agent continuously adjusts its actions through interaction with the environment and updates its strategies according to the feedback from the environment, thus gradually improving the overall performance of the system.

2.1.2. Computational Process Modeling in Operation Modes

The interaction between the intelligent agent and the environment can be described by a Markov decision process (MDP), which consists of five variables (S, A, P, R, γ). S is the state space of the system; s_t is the state of the system at moment t; A is the action space; a_t is the action of the intelligent agent at moment t; P is the state transfer function; $P(s_{t+1} \mid s_t, a_t)$ is the probability of transferring the action a_t after taking it from state s_t to state s_{t+1}; R is the reward function; r_t is the reward value obtained after taking action a_t in state s_t; and γ is the discount factor ($0 \le \gamma \le 1$), which is used to weigh the effects of immediate and future reward values on the decision-making process.

In this paper, the Q-learning algorithm is used to calculate the operation mode generation, while the State–Action value function concept is introduced and a Q-table is created, which serves to record the reward value after performing action a_t on state s_t at moment t. The state space s_t at moment t is defined as

$$s_t = [p_1, p_2, \ldots, p_m, q_1, q_2, \ldots, q_m, v_1, v_2, \ldots, v_n] \tag{1}$$

where p_i is the active power of the i-th generator; q_i is the reactive power of the i-th generator; v_i is the value of the voltage at the i-th bus; m is the number of generators (excluding the slack machine); and n is the total number of buses. The action space A is discrete. Let the action space $A = [1, 2, \ldots, m]$ denote the corresponding number of the generator. $a_t = i$ indicates that generator i is switched on at time t. The algorithm adopts an $\varepsilon - greedy$ strategy to select the action. This entails random selection of an action with the probability of ε to explore the environment, while opting for the action with the highest Q-value in the current Q-table with the probability of $1 - \varepsilon$. After executing the chosen action, the Q-table state is updated accordingly, and the specific state update formula is as follows:

$$Q_{t+1}(S, A) = Q_t(S, A) + \alpha(R_{t+1} + \gamma \max Q_{t+1}(S, A) - Q_t(S, A)) \tag{2}$$

where $Q_{t+1}(S, A)$ is the updated value of the Q-table, $Q_t(S, A)$ is the current value of the Q-table, α is the learning rate, R_{t+1} is the reward function, γ is the discount factor, and $\max Q_{t+1}(S, A)$ is the maximum $Q_{t+1}(S, A)$ of the succeeding state S_{t+1}. The algorithm iteratively updates the Q-table values and then determines what action to take in the succeeding states based on the updated values.

From (2), it can be seen that the instantaneous reward R_{t+1} affects the estimation result $Q_{t+1}(S, A)$, and $Q_{t+1}(S, A)$ also affects the selection of the action A_t, which determines

the adjustment direction of the power flow. Thus, a reasonable reward mechanism is the key to gradually adjusting the power flow to the target interval. In this paper, we only consider two indexes in the power flow convergence adjustment process: (a) power flow convergence, which is denoted by *conda*.1, and (b) the limit of the slack machine output power, which is denoted by *conda*.2. Therefore, the reward function is defined as follows:

$$R_t = \begin{cases} 1, & conda.1\&conda.2 \\ 0, & others \end{cases} \tag{3}$$

After the execution of action A_t, if the power flow calculation converges and the output power of the slack machine does not exceed the limit, then $R_t = 1$. Otherwise, $R_t = 0$; thus, the fewer the number of adjustment steps, the larger the cumulative reward obtained under the action of this reward function.

2.2. Feature Extraction for Power Grid Operation Modes

There are many data dimensions of the operation mode and there is a nonlinear correlation between different dimensions [30], which specifically manifests in a change in the operation law of related electrical quantities such as the line power flow, generator output, and node voltage phase angle. Thus, the generated operation mode is preprocessed to initially compress redundant dimensions, thereby enhancing the operational efficiency of subsequent algorithms. The original input feature set of each operation mode p_i is defined as

$$p_i = [\mathbf{G}_m, \mathbf{V}_b, \mathbf{P}_l, \mathbf{Q}_l, P, Q] \tag{4}$$

where $\mathbf{G}_m$ is the switching status of all generation units (1 for power on, 0 for power off); $\mathbf{V}_b$ is the bus voltage magnitude vector of all generation units; $\mathbf{P}_l$ is the active power vector of all loads; $\mathbf{Q}_l$ is the reactive power vector of all loads; P is the active power vector of the slack machine; and Q is the reactive power vector of the slack machine.

In order to further explore and analyze the potential features in a series of generated operation modes, the compressed and processed feature data are subjected to a cluster analysis. Cluster analysis is an unsupervised learning method to group similar operation modes into one category and reveal the relationships and commonalities between them. Ultimately, the operation mode with the highest evaluation value among the clusters obtained by clustering division is selected as the typical operation mode. The algorithms for each segment are described in detail below.

2.2.1. Data Preprocessing

In data processing, as the individual electrical quantities have different scales, if the value of a certain feature is large, the feature will have a large weight in the overall error calculation; thus, features with smaller values will be ignored. Therefore, during processing, the first choice is z-score standardized processing of the operation mode data, that is, the data with different outlines are converted to the same measure for comparison. The calculation formula is as follows:

$$x' = \frac{x - \mu}{\sigma} \tag{5}$$

where x is the observed value of a characteristic in a certain operation mode, μ is the mean value of the characteristic data, σ is the standard deviation of the characteristic data, and x' is the normalized data value.

Then, Principal Component Analysis (PCA) [31] is used to compress the redundant dimensions of the high-dimensional operation mode data, the basic principle of which is to linearly transform the data and thus reduce the dimensionality under the condition of maintaining features with the largest variance in the sample points. The processing flow is as follows:

1. Center all operation modes vectors:

$$P = (p_1 - \overline{p}_1 \ p_2 - \overline{p}_2 \ \cdots \ p_M - \overline{p}_M) \tag{6}$$

where P is a matrix consisting of M N-dimensional operation mode vectors, p_i is the i-th mode vector, and $\overline{p}_i$ is the average of the i-th mode vectors;

2. Calculate the covariance matrix $Cov = PP^T/M$ to obtain the eigenvalue $\lambda_1, \lambda_2, \ldots, \lambda_N$ and the corresponding eigenvectors $h_1, h_2, \ldots, h_N$, where $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N$;

3. Take the first K eigenvectors to form a new matrix $H = (h_1, h_2, \ldots, h_K)$:

$$K = \text{argmin}_K \left[\left(\sum_{i=1}^{K} \lambda_i / \sum_{i=1}^{N} \lambda_i \right) \geq (1 - \theta) \right] \tag{7}$$

where θ is the compression factor and $\theta = 0$ means uncompressed;

4. Perform a linear transformation to obtain operation mode data $P' = (p'_1 p'_2 \cdots p'_M)$:

$$P' = H^T P \tag{8}$$

2.2.2. Operation Mode Clustering

In this paper, the density-based DBSCAN clustering algorithm is used; this algorithm [32] avoids the necessity to predefine the quantity of clusters and possesses the capability to discern any multitude and configuration of clusters within a dataset, even one containing perturbing data. Only the radius of the neighborhood *Epsilon* and the minimum number of neighboring points, *MinPts*, required for the core point need to be determined. However, in a scenario based on operation modes generated by reinforcement learning, it is difficult to rely on experience to select the appropriate *Epsilon*; therefore, in this paper, a k-distance map (Figure 1) is drawn to determine the selection parameters by judging the inflection point. This is executed as follows:

1. Calculate the distance from each data point $x_i(i = 1, 2, \ldots, M)$ to the k-th nearest neighbor, denoted as $d(x_i, k)$;

2. Incrementally sort and display the resulting k-distance sequence $\{d(x_i, k) \mid i = 1, 2, \ldots, M\}$, with data points on the horizontal axis x_i and the incremental k-distance sequence on the vertical axis;

3. Determine the location of the inflection point in the graph; the y-value of the inflection point is *Epsilon*.
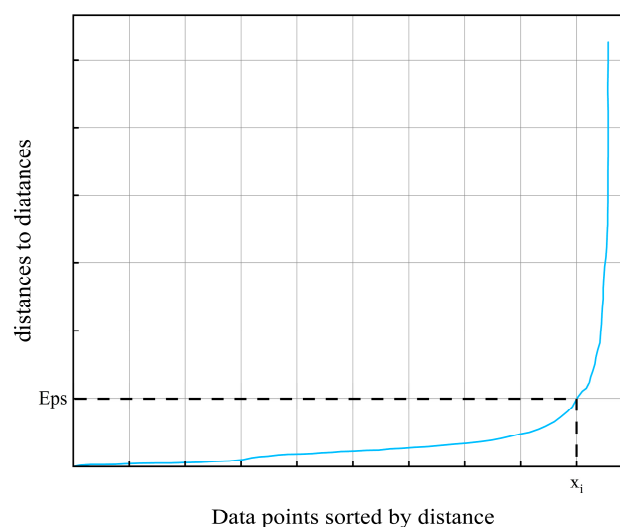

Figure 1. k-distance map.

After determining the neighborhood radius *Epsilon* and the minimum number of neighborhood points, *MinPts*, required for the core point, the DBSCAN algorithm is used to extract features from the preprocessed operation mode data P'. The specific steps are as follows:

1. Arbitrarily select an operation mode p' as the current point from the operation mode data P' and create a new cluster C for p'. The cluster count is initialized to 1.
2. Find all the operation modes in the neighborhood of the current operation mode p'. If the number of all operation modes in the neighborhood is less than *MinPts*, mark the current operation mode as noise; otherwise, mark the current operation mode as the core point of cluster C.
3. Traverse each operation mode (new current point) in the neighborhood and repeat step 2 until no new operation mode that can be marked as belonging to the current cluster C is found.
4. Choose the subsequent unlabeled operation mode of the operation mode data p' as the current point and increase the cluster count by 1.
5. Repeat steps 2–4 until all the operation modes in p' are labeled and the clustering results obtained are the different operation modes.

2.3. Operation Mode Evaluation System and Index Calculation

The adjustment of the traditional operation mode mainly relies on the experience of experts, and under the prerequisite of ensuring power flow convergence, stability calculations have to be constantly carried out to ensure the safe operation boundary of the power grid, while taking into account evaluation index optimization. However, with the increase in the diversity of power grid evaluation indexes, not only does the computational cost increase and the adjustment efficiency decrease, but it is also difficult to choose the optimal one quickly when facing interrelated and mutually constrained indexes. Therefore, in this section, an evaluation system and index calculation for the operation mode are proposed, which integrate multi-dimensional indexes through the AHP-EWM to construct a comprehensive operation mode evaluation model.

2.3.1. Indexes for the Comprehensive Evaluation of Operation Mode

The evaluation index system established in this paper includes steady security, transient security, and economy evaluation indexes, and each perspective includes 1–3 indexes (Figure 2). The comprehensive operation mode evaluation indexes can be obtained through the comprehensive calculation of the evaluation indexes from these three perspectives.
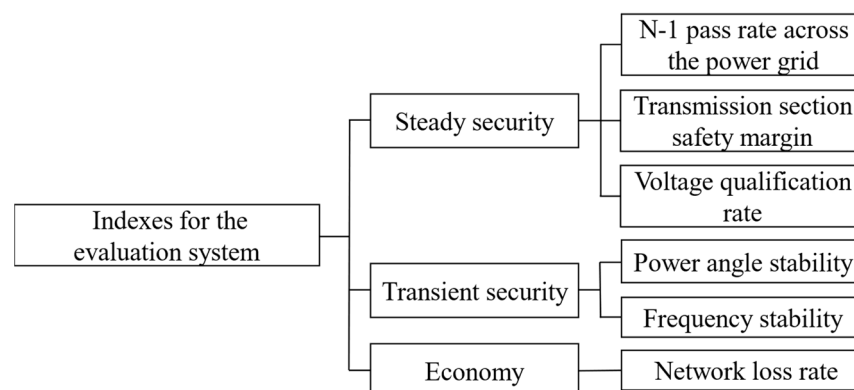

Figure 2. Operation mode evaluation system.

The detailed definitions of the indexes are as follows:

(a) In the $N - 1$ verification inspection of the whole power grid, the percentage of times that the power flow converges after removing any line and transformer in the system,

where the voltage and frequency are not out of the limits, is defined as the grid-wide $N-1$ pass rate:

$$I_{N-1} = \frac{A_{N-1,\text{pass}}}{A_{N-1}} \tag{9}$$

where A_{N-1} is the number of calibration tests and $A_{N-1,\text{pass}}$ is the number of calculation passes;

(b)   The transmission section safety margin is a key parameter for inter-regional power supply, where A_{TC} is the transmission power of the selected section in the current operation mode and A_{TTC} is the power limit of the selected section:

$$I_{TC} = 1 - \frac{A_{TC}}{A_{TTC}} \tag{10}$$

(c)   The voltage pass rate can reflect the voltage pass level of the nodes and show the voltage quality of the current operation mode. The ratio of the number of nodes N_q with a qualified voltage to the total number of nodes N in the whole grid is defined as the voltage qualification rate:

$$I_{VQ} = \frac{N_q}{N} \tag{11}$$

(d)   Power angle stability refers to the power angle swing of the generator caused by an expected accident and its severity. It is calculated by taking the maximum power angle offset $|\delta_{\max}|$ between generators:

$$I_{TAS} = \frac{360^{\circ} - |\delta_{\max}|}{360^{\circ} + |\delta_{\max}|} \tag{12}$$

(e)   Frequency stability refers to the degree of generator frequency shift caused by large disturbances due to expected accidents. It is calculated by taking the maximum frequency shift in the line $f_{\max}$:

$$I_{TFS} = \frac{f_{\max}}{f} \tag{13}$$

(f)   The total power generation of the power grid P_{gen} minus the total load of the power grid P_{load} is defined as the total network loss of the power grid. The ratio of total network loss to total power generation defines the network loss rate:

$$I_{loss} = \frac{P_{gen} - P_{load}}{P_{gen}} \tag{14}$$

2.3.2. Model for Calculating Index Weights

To assign and calculate weights, the subjective weighting method is mainly based on the analytical judgment of decision makers, while the objective weighting method is mainly based on the original index. However, the latter method is too dependent on the original data; therefore, the AHP-EWM was chosen to integrate multi-dimensional operation mode evaluation indexes. In the process of calculating the subjective and objective weights and in the evaluation of the two evaluation models individually, the specific steps are as follows:

Step 1.   Normalize the indexes using the ideal point approximation method to convert the values of the indexes into the (0, 100) interval. The formula is as follows:

$$I' = \frac{I - I^-}{I^+ - I^-} \times 100 \tag{15}$$

where I is an index; I^+ is the positive ideal solution for the index; I^- is the negative ideal solution for the index; and I' is the converted value of the index.

Step 2. Construct a judgment matrix A. In this paper, we use a nine-level numerical scale layer to illustrate the relationship between indexes:

$$A = \left[a_{ij}\right]_{n \times n} = \begin{bmatrix} a_{11} \cdots a_{1n} \\ \vdots \vdots \\ a_{n1} \cdots a_{nn} \end{bmatrix} \tag{16}$$

where $a_{ii} = 1$, $a_{ji} = 1/a_{ij}$, n is the number of indexes, and the scale a_{ij} indicates the importance of the i-th index compared to the j-th index in the same operation mode.

Step 3. Compute the concatenated product of the elements of each row, take the nth root $\overline{W}_i$ and the eigenvectors of the matrix A, normalize them, and check for consistency.

$$\begin{cases} \overline{W}_i = \sqrt[n]{\prod\limits_{j=1}^{n} a_{ij}} \ i = 1, 2, \cdots, n \\ w'_i = \overline{W}_i / \sum\limits_{i=1}^{n} \overline{W}_i \\ W' = (w'_1, w'_2, \cdots, w'_n) \end{cases} \tag{17}$$

Step 4. Calculate the objective weights and combine with the method used in [33] to obtain the following formula:

$$\begin{cases} H(i) = \dfrac{-\sum\limits_{k=1}^{m} I'_{ik} \ln I'_{ik}}{\ln m} \\ w''_i = \dfrac{1 - H(i)}{\sum\limits_{i=1}^{n} (1 - H(i))} \\ W'' = \left(w''_1, w''_2, \cdots, w''_n \right) \end{cases} \tag{18}$$

where I'_{ik} is the normalized data of the i-th index of the sample k, $H(i)$ is the entropy value of the i-th index, w''_i is the weight of the i-th index, W'' is the weight vector of evaluation index based on the entropy weight method, and m is the number of data samples.

Step 5. Combining the strengths of subjective prior knowledge and objective data analysis, calculate the combined weights using the following formula:

$$\begin{cases} w_i = \dfrac{w'_i w''_i}{\sum\limits_{i=1}^{n} w'_i w''} \\ W = (w_1, \cdots, w_i, \cdots w_n) \end{cases} \tag{19}$$

In the overall evaluation process, the subjective and objective weights are first calculated layer by layer from the sub-index layer to the top layer and checked for consistency, the comprehensive weights are then calculated, and finally a comprehensive operation mode evaluation model is obtained.

2.4. Rapid Operation Mode Evaluation Model

DBN is a form of neural network characterized by its inclusion of multiple concealed layers. It is composed of a series of restricted Boltzmann machines (RBMs) and backpropagation (BP) neural network layers superimposed on top of one another. The network fine-tunes the inter-layer connection weights via a layer-by-layer greedy learning algorithm. This approach optimizes the connection weights between individual layers, endowing the network with a robust feature extraction capability and enabling successive layer-wise feature extraction and representation.

The learning and training processes of DBN can be divided into two stages: unsupervised pre-training and supervised parameter fine-tuning. During pre-training, adjacent

layer pairs establish an RBM, and the output from the subsequent RBM layer serves as input for the preceding RBM layer, facilitating incremental, layer-by-layer training that employs the greedy unsupervised algorithm to extract higher-level data features. The second stage is the fine-tuning stage, where the parameters of the entire network are supervised and fine-tuned using a BP neural network. This method of unsupervised pre-training followed by supervised fine-tuning effectively reduces the parameter optimization search space, overcomes the problem of neural networks being prone to fall into local optima, and shortens the supervised training time.

The flow of the rapid operation mode evaluation model based on DBN proposed in this paper is shown in Figure 3.
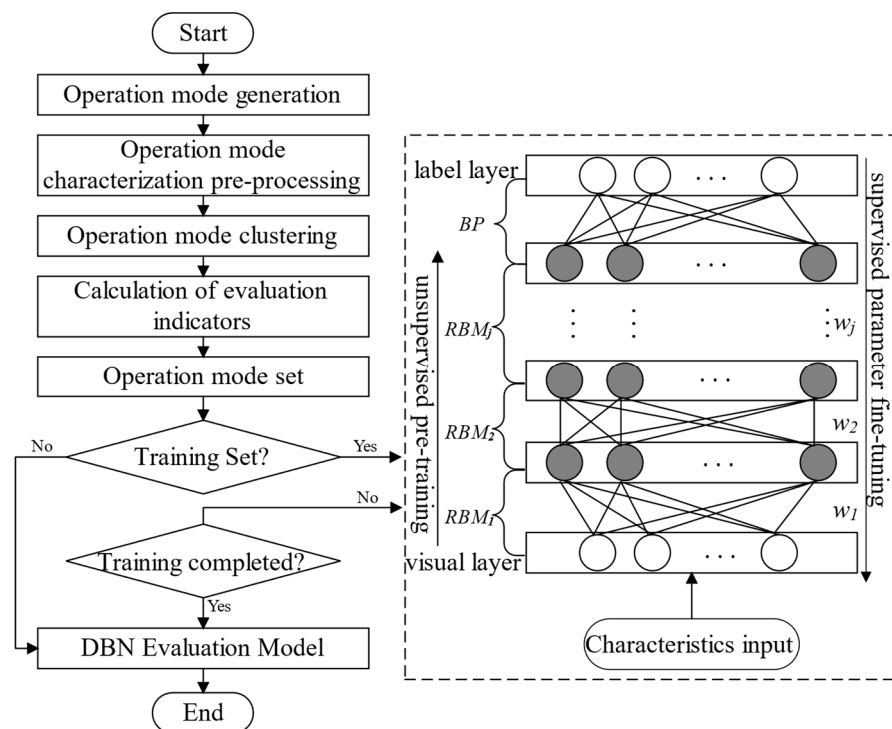
Figure 3. Flowchart of rapid operation mode evaluation model.

The steps are as follows:

Step 1. Generation of operation modes. According to the historical operation data of the power grid, operation mode generation is carried out according to the reinforcement learning algorithm proposed in this paper. Through data preprocessing methods such as standardized processing, feature extraction, and feature dimension reduction, the operation mode feature dataset for clustering and DBN training is obtained.

Step 2. For the operation mode characterization data and DBSCAN clustering algorithm, the operation modes are divided into different operation mode sets.

Step 3. According to the steady-state indexes, transient state indexes, and economy indexes, operation mode evaluation is carried out, and then the respective weights from these three perspectives are calculated using the AHP-EWM. The multi-dimensional index values and weights are then integrated into a comprehensive value.

Step 4. All the operation modes are randomly divided into the training set and test set, where the ratio of training set to test set is 8:2, and the model is trained according to the forward unsupervised pre-training and reverse supervised parameter fine-tuning method.

Step 5. After the training is completed, the test set feature data are input into the DBN model. The evaluation value of the operation mode is then obtained to obtain the rapid operation mode evaluation model based on DBN.

3. Results

In order to substantiate the viability of the proposed intelligent generation of typical operation modes, Power System Analysis Software Package (PSASP 7.3) from the China Electric Power Research Institute (CEPRI) is used. Simulation examples are based on the actual historical typical operation modes of the Northeast China Power Grid. The Northeast China Power Grid comprises four provincial-level regions with 2128 nodes, 531 generators, and 816 load buses. The number of 500 kV buses is 221, and the number of 220 kV buses is 732. Sections 3.1 and 3.2 present results for three different load levels (low valley, flat waist, peak scenario). It should be noticed that the load levels are increased by 5% on a year-over-year basis to simulate future load growth trends, and the power grid structure changes are not considered.

In Section 3.1, the figures illustrate the selected typical operation modes for each operation mode set under the three load levels. In Section 3.2, the figures demonstrate the clustering analysis of the generated operation modes based on their data characteristics. Given the substantial number of generated operation modes and the relatively close comprehensive evaluation values, Section 3.3 focuses on explaining operation modes with significant differences in evaluation values. In Section 3.4, a total of 8695 operation modes are generated as the data sample set based on historical operation modes.

3.1. Intelligent Generation of Operation Modes

During the operation mode generation process, the algorithm gradually adjusts the power flow by controlling the start and stop of the generator to ensure power flow convergence and constrain the slack machine power within a given interval. The slack machine's given output interval is 120–600 MW, and the hyper-parameters are set to the following: the greedy coefficient is 0.8, the learning rate is 0.15, and the discount factor is 0.9. In this paper, grid training is performed using the Q-learning and State Action Reward State Action (SARSA) algorithms, with the resultant average cumulative reward value curve illustrated in Figure 4. It is evident from Figure 4 that SARSA converges at 37 episodes with an average cumulative reward value of 38, while Q-learning converges at 42 episodes with an average cumulative reward value of 45. Q-learning is able to obtain higher reward values for more similar episodes. The Q-learning algorithm outperforms the SARSA algorithm in terms of overall effect.
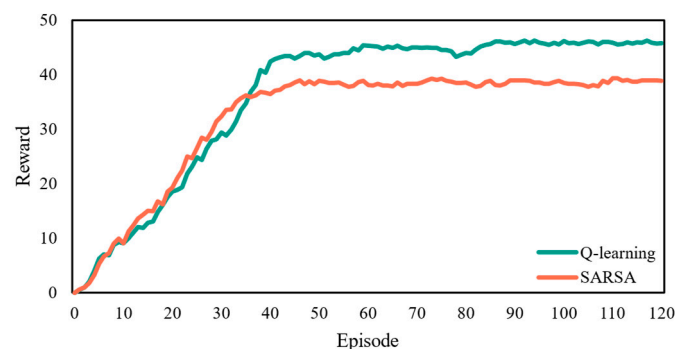

Figure 4. Comparison of average cumulative reward values.

As shown in Figure 5, among the three groups of operation modes with different load levels, each group includes one historical typical operation mode and multiple newly generated typical operation modes, and it can be seen that the generator output of each region in each mode increases. There is no situation where the new generator output of a certain region is zero; only the on-load capacity of each region is different; for example,

in the typical operation mode 1 with the flat waist level, the on-load capacity of region 1 is a little bit lower or close to the same level compared to the other two typical operation modes, while the starting capacity of the other regions is lower than or close to the other two typical operation modes. The figure illustrates that each newly generated typical operation mode exhibits varying regions where the primary actions of generator units occur. This variation arises because the algorithm interacts with the environment to identify the optimal combination of generator actions specific to the current region. This approach enables the coordinated operation of generators across multiple regions, resulting in typical operation modes that can be applied in practical power grid scenarios.
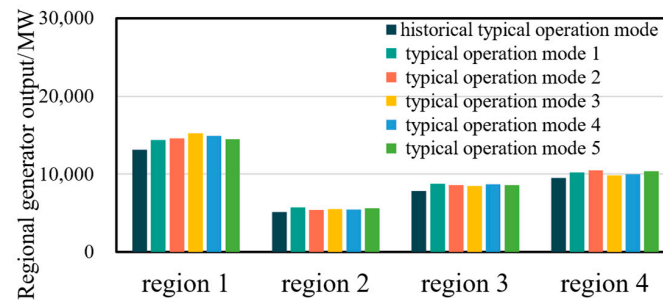
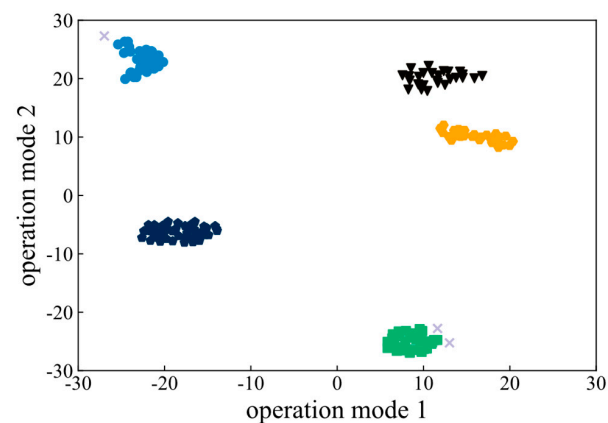(**a**) low valley
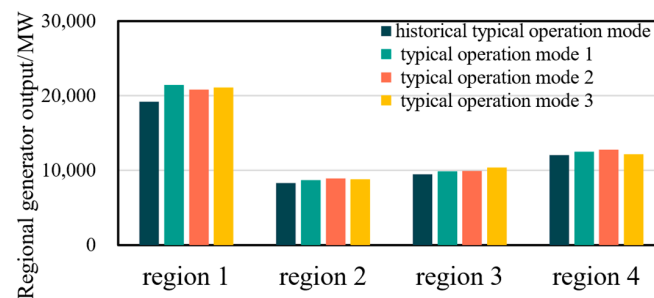
(**b**) flat waist
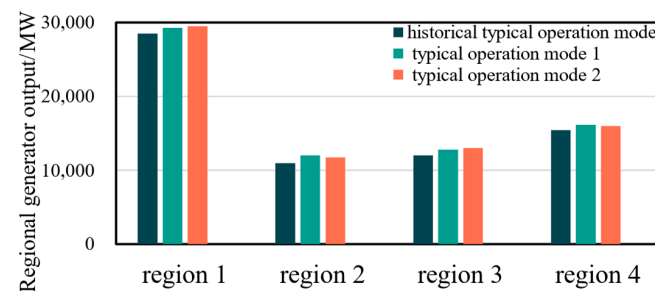
(**c**) peak scenario

**Figure 5.** Typical operation mode generation at different load levels.

3.2. Distribution of Operation Mode Sets

In this section, the high-dimensional operation mode data intelligently generated based on equivalent historical typical operation modes are first preprocessed, and then the operation mode feature data obtained after preprocessing are clustered and analyzed to obtain the state space distribution of the operation modes under the three load levels (as shown in Figure 6). Each point represents an operation mode, and different colors represent different sets of operation modes obtained by clustering. From the state space distribution graph, it can be seen that the number of sets of operation modes decreases as the load level increases from five operation mode sets at the low valley level and three operation mode sets at the flat waist level to finally two sets at the peak scenario level. This

indicates that at lower load levels, generator turn-on is low and the number of generators that can be switched on and off while performing mode generation is high. As the load side demand rises, generator turn-ons increase and the number of generators that can be actuated decreases significantly. Even though the generator output values are the same in each region for multiple operation modes, the number of generators that can be actuated is not necessarily the same, which explains why the number of operation mode sets at the low valley level is more than those of the other two load levels.

(**a**) low valley
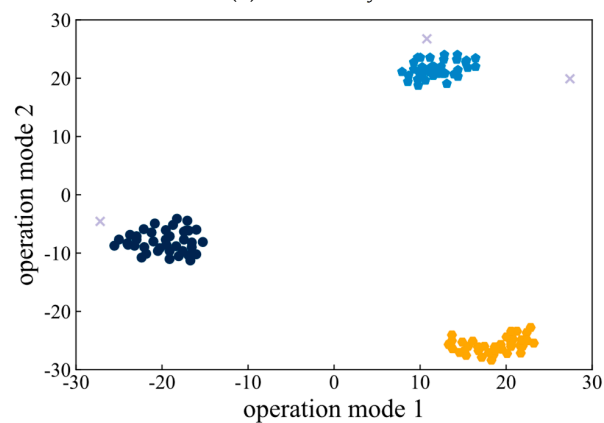
(**b**) flat waist
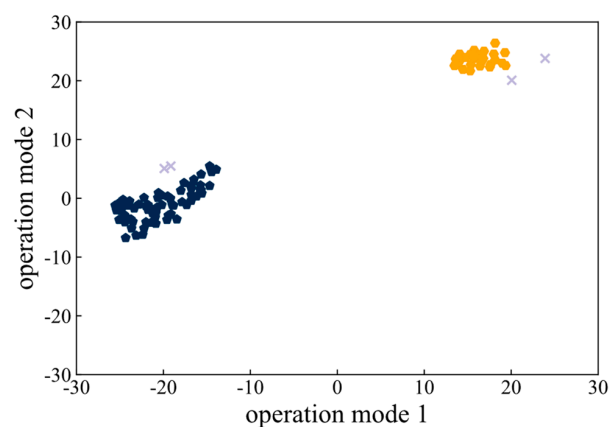
(**c**) peak scenario

Figure 6. Clustering results of operation modes at different load levels. Each point represents an operation mode, and different color blocks represent different sets of operation modes obtained by clustering.

3.3. Indexes and Analysis of Comprehensive Operation Mode Evaluation

In terms of weight calculation, subjective weights are first calculated to evaluate the indexes based on scheduling experience, and then objective weights of steady-state security indexes, transient state security indexes, and economy indexes are calculated based on the operation mode data. Finally, the coupling calculation of the subjective and objective weights according to (19) yields the comprehensive weights of the three perspectives, with the results shown in Table 1.

Table 1. Results of weight calculation.

Weight	Steady-State Security Index	Transient State Security Index	Economy Index
Subjective weight	0.2973	0.5389	0.1638
Objective weight	0.4234	0.3219	0.2547
Comprehensive weight	0.3691	0.5086	0.1223

According to the index system constructed in this paper, simulation-generated operation modes are selected for the calibration test, and the comprehensive evaluation values of the operation modes are further calculated by multiplying the index values obtained through calibration with the corresponding weights. Observing the data in Table 2, it can be seen that the transient state security index of each operation mode accounts for the largest proportion, and the economy index accounts for the smallest proportion, which corresponds to the comprehensive weighting value. The differences between the operation modes are mainly reflected in the transient state security index, and the rest of the indexes are closer in value.

Table 2. Calculation results of operation mode indexes.

| Mode Number	Steady-State Security Index	Transient State Security Index	Economy Index	Total Value
Basic mode	32.8604	46.5159	11.6732	91.0495
Mode 1	30.9461	45.9558	10.9218	87.8237
Mode 2	30.8144	43.7940	11.5482	86.1566
Mode 3	31.7483	46.6853	10.6337	89.0673
Mode 4	30.1596	44.6487	10.9485	85.7568

In order to verify the key role played by the transient state security index in the evaluation of operation modes, the following will evaluate the above modes by considering only the steady-state security index and economy index. The combined weights of the steady-state security index and economy index are 0.7678 and 0.2322, respectively. As shown in Table 3, the evaluation values of the modes decrease when not considering the transient state security index compared with mode 1 and mode 3. Mode 2 has a lower comprehensive evaluation value when considering the transient state security index, but its comprehensive evaluation value is higher when it is not considered. The main reason for this is that the transient index reflects the operation mode status after line faults occur, while the steady-state index reflects the current operation mode status. The steady-state index in most of the generated operation modes is relatively good, and the real-world grid is also more related to the operation modes after a fault occurs. Thus, the transient state security index proposed in this paper is more capable of checking the comprehensive operation mode performance.

Table 3. Comparison of calculation results of operation mode indexes.

Mode Number	Steady-State Security Index	Economy Index	Total Value
Basic mode	68.3560	22.1629	90.5189
Mode 1	64.3739	20.7362	85.1101
Mode 2	64.1000	21.9255	86.0255
Mode 3	66.0426	20.1893	86.2319
Mode 4	62.7379	20.7869	83.5248

3.4. Rapid Operation Mode Evaluation Model

The pre-processed operation mode feature data and the comprehensive evaluation index values corresponding to each group of operation modes are input into the DBN network, in which the number of implicit layers is set to 4; the number of nodes in each implicit layer is 50, 30, 20, and 10, respectively; the learning rate is 0.1; the momentum parameter is 0.5; and the number of iterations is 200. Due to the space limitation, in this paper, we only present part of the results for the test set. As shown in Figure 7, it can be seen that the DBN model has the ability to recognize the correlation relationships between key feature data in the operation mode and result indexes and shows strong assessment and prediction abilities.
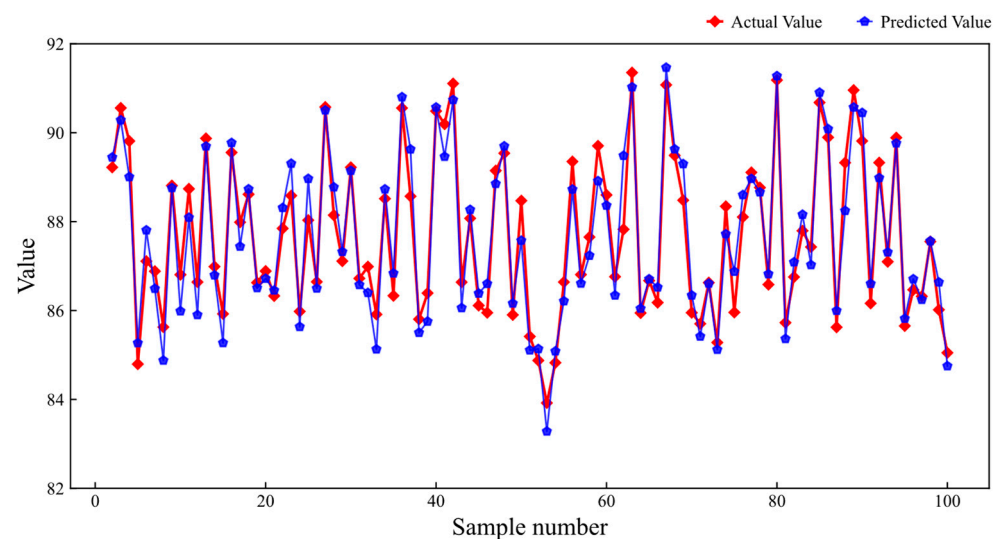
Figure 7. Comparison of actual values with DBN predictions.

At the same time, KNN, SVR, and XGBoost algorithms were selected to carry out a regression prediction of the comprehensive operation mode indexes in turn, and the prediction results of each model are shown in Figure 8. It is observed that traditional machine learning algorithms are unable to accurately find the correlation between the characteristics of the operation mode and the evaluation index value, while the neural network structure of deep learning can better capture the complex data patterns (Table 4).

Table 4. Error results of the rapid evaluation model for the multi-algorithm operation mode.

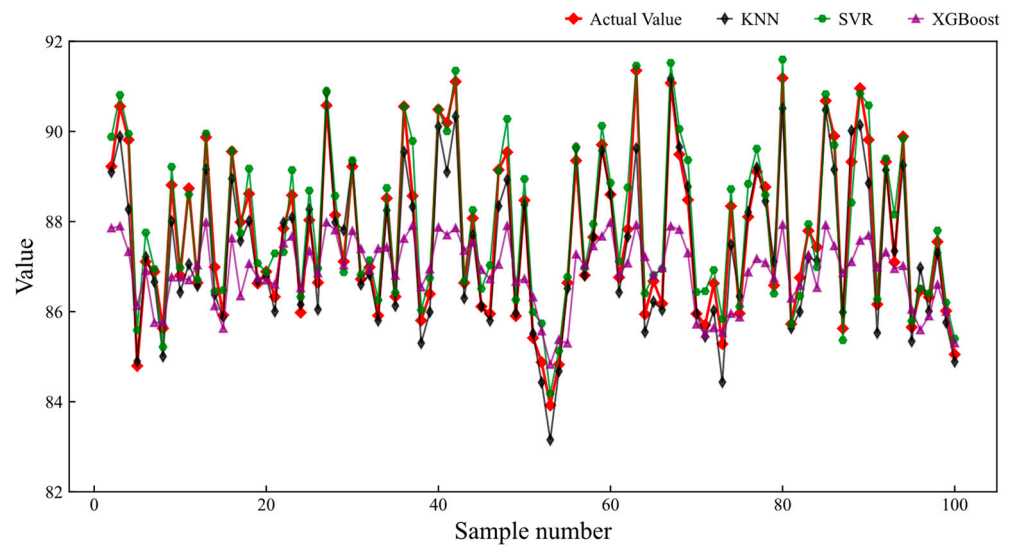| Algorithm	RMSE	MAPE
DBN	0.506	0.458
KNN	0.684	0.619
SVR	0.824	0.726
XGBoost	1.354	1.068

Figure 8. Comparison of actual values with predictions of other algorithms.

In this paper, the mean absolute percentage error (MAPE) and root-mean-squared error (RMSE) are used as quantitative indexes to evaluate the model with the following formulas:

$$
\begin{cases}
I_{RMSE} = \sqrt{\dfrac{1}{n}\sum\limits_{i=1}^{n}\left(y_i - \widetilde{y}_i\right)^2} \\
I_{MAPE} = \dfrac{1}{n}\sum\limits_{i=1}^{n}\dfrac{|y_i - \widetilde{y}_i|}{y_i}
\end{cases}
\tag{20}
$$

4. Discussion

The Q-learning algorithm proposed in this paper highlights a significant step forward in power grid operation mode generation. Compared to the SARSA algorithm [34], which first executes the action through the policy and then updates the value function according to the executed action, the Q-learning algorithm first assumes that the next step is to pick the action with the maximum reward, update the value function, and then select the action through the strategy.

For the algorithm selection for the rapid operation mode evaluation model, KNN [35], SVR [36], and XGBoost [37] algorithms are unable to exploit the correlation between the operation mode characteristics and the comprehensive evaluation index values; the DBN-based rapid operation mode evaluation model exhibits superior performance, and can effectively predict the results of the comprehensive evaluation of the operation modes and assist the system operator in preparing a typical operation mode based on the evaluation results.

Although the typical power grid operation mode generation based on reinforcement learning and DBN is innovative and effective, it should be noted that when the higher-value operation modes are concentrated in the same set, only the highest values of this set need to be paid attention to; if the higher-value operation modes are distributed across different sets, it is necessary for the system operator to make judgments based on the corresponding values and the actual situation. Future work involves the analysis of the correlation constraints among the operation mode with the highest value in the different operation mode sets to identify typical operation modes.

5. Conclusions

In order to realize efficient and accurate generation of typical power grid operation modes, this paper proposes a method of rapid typical operation mode generation based on reinforcement learning and DBN. In response to the intricate and complex process of calculating operation modes in practical power grids, especially in cases with numerous

combinations of generators, a proposal is put forth to intelligently and efficiently generate power grid operation modes based on historical typical operation modes. Furthermore, to address the challenge of a high number of generated operation modes, the DBSCAN clustering method is employed to categorize these operation modes into different operation mode sets based on their distinctive characteristics. In order to construct a rapid operation mode evaluation model, key indexes are proposed from the three perspectives of steady state, transient state, and the economy, and the weights of the multi-dimensional indexes are integrated by the AHP-EWM. Finally, DBN is introduced to establish the relationship between operation mode feature data and comprehensive values, enabling the rapid evaluation of operation modes and the identification of typical operation modes within different operation mode sets. The results demonstrate that the proposed method efficiently supports the operation modes calculation, thereby enhancing the efficiency of operation mode calculations and reducing the consumption of human and physical resources.

Author Contributions: Conceptualization, Z.W. and B.Z.; methodology, Z.W.; software, Z.W.; validation, H.Y., Q.M. and Z.Y.; formal analysis, B.Z.; investigation, C.L. and Y.C.; data curation, B.Z.; writing—original draft preparation, Z.W.; writing—review and editing, Z.W.; visualization, H.Y.; supervision, C.L. and Y.C.; project administration, B.Z.; funding acquisition, B.Z. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Iqbal, S.; Habib, S.; Khan, N.H.; Ali, M.; Aurangzeb, M.; Ahmed, E.M. Electric Vehicles Aggregation for Frequency Control of Microgrid under Various Operation Conditions Using an Optimal Coordinated Strategy. *Sustainability* **2022**, *14*, 3108. [CrossRef]
2. Lan, J.; Guo, Q.; Zhou, Y.; Sun, H. Generation of Power System Typical Operation Mode Samples: A Generation Adversarial Network and Model-based Transfer Learning Approach. *Proc. CSEE* **2022**, *42*, 2889–2900.
3. Iqbal, S.; Xin, A.; Jan, M.U.; Abdelbaky, M.A.; Rehman, H.U.; Salman, S.; Aurangzeb, M. Aggregation of EVs for primary frequency control of an industrial microgrid by implementing grid regulation & charger controller. *IEEE Access* **2020**, *8*, 141977–141989.
4. Li, C.; Wu, Z.; Dong, L.; Liu, M.; Xiao, Y. Fast Generation of Power System Operation Modes Based on Optimal Power Flow. In Proceedings of the 2021 11th International Conference on Power and Energy Systems (ICPES), Shanghai, China, 18–20 December 2021; pp. 767–771.
5. Xu, H.; Yu, Z.; Zheng, Q.; Hou, J.; Wei, Y. Improved deep reinforcement learning based convergence adjustment method for power flow calculation. In Proceedings of the 16th IET International Conference on AC and DC Power Transmission (ACDC 2020), Online Conference, 2–3 July 2020; pp. 1898–1903.
6. Li, R.Y.; Peng, H.M.; Li, R.G.; Zhao, K. Overview on Algorithms and Applications for Reinforcement Learning. *Comput. Syst. Appl.* **2020**, *29*, 13–25.
7. Jang, B.; Kim, M.; Harerimana, G.; Kim, J.W. Q-learning algorithms: A comprehensive classification and applications. *IEEE Access* **2019**, *7*, 133653–133667. [CrossRef]
8. Feng, C.; Sun, M.; Zhang, J. Reinforced deterministic and probabilistic load forecasting via Q-learning dynamic model selection. *IEEE Trans. Smart Grid* **2019**, *11*, 1377–1386. [CrossRef]
9. Shin, K.S.; Choi, H.H.; Lee, H. Knowledge Transfer-Based Multiagent Q-Learning for Medium Access in Dense Cellular Networks. *IEEE Wirel. Commun. Lett.* **2022**, *11*, 2542–2545. [CrossRef]
10. Xi, B.; Lei, D. Q-learning-based teaching-learning optimization for distributed two-stage hybrid flow shop scheduling with fuzzy processing time. *Complex Syst. Model. Simul.* **2022**, *2*, 113–129. [CrossRef]
11. Kushwaha, A.; Gopal, M.; Singh, B. Q-learning based maximum power extraction for wind energy conversion system with variable wind speed. *IEEE Trans. Energy Convers.* **2020**, *35*, 1160–1170. [CrossRef]
12. Cao, Y.; Zhao, C.; Li, D. Carbon Management for Intelligent Community with Combined Heat and Power Systems. *Sustainability* **2023**, *15*, 13257. [CrossRef]
13. Xiang, W.; Zhao, J.J.; Gu, Z.G.; Zhang, Q.R.; Fang, J.L.; Li, S.Y. A Dynamic Model Based Method for Evaluating the Operation State of Provincial Power Grid Company. In Proceedings of the 2020 8th International Conference on Power Electronics Systems and Applications (PESA), Hong Kong, China, 7–10 December 2020; pp. 1–5.

14. Li, H.; Wen, C.; Gu, R.; Zhou, Y. Research on intelligent decision evaluation method for substation based on fuzzy analysis and decision theory. In Proceedings of the 2016 International Conference on Fuzzy Theory and Its Applications (iFuzzy), Taichung, Taiwan, 9–11 November 2016; pp. 1–6.

15. Liu, Y.; Ouyang, Z.; Yi, H.; Qin, H. Study of the Multilevel Fuzzy Comprehensive Evaluation of Rock Burst Risk. *Sustainability* **2023**, *15*, 13176. [CrossRef]

16. Wang, J.; Hao, Y.; Yang, C. A comprehensive prediction model for VHF radio wave propagation by integrating entropy weight theory and machine learning methods. *IEEE Trans. Antennas Propag.* **2023**, *71*, 6249–6254. [CrossRef]

17. Zhuo, Y.; Cao, Y.; Chen, L.; Nie, J.; Huang, Y.; Liang, Y. Research on evaluation model based on analytic hierarchy process and entropy weight method for smart grid. In Proceedings of the 2022 5th International Conference on Energy, Electrical and Power Engineering (CEEPE), Chongqing, China, 22–24 April 2022; pp. 729–734.

18. Zhou, K.; Li, Z.; Gong, W.; Zhao, S.; Wen, C.; Song, Y. Influence of magnetic field generated by air core reactors in SVC-based substation and an optimal suppression method based on fuzzy comprehensive evaluation. *IEEE Trans. Electromagn. Compat.* **2019**, *62*, 1961–1970. [CrossRef]

19. Lo, H.W.; Hsu, C.C.; Huang, C.N.; Liou, J.J. An ITARA-TOPSIS based integrated assessment model to identify potential product and system risks. *Mathematics* **2021**, *9*, 239. [CrossRef]

20. Wang, T.; Du, Z.A.; Zhang, K.; Chen, K.; Xiao, F.; Ye, P. Reliability evaluation of high voltage direct current transmission protection system based on interval analytic hierarchy process and interval entropy method mixed weighting. *Energy Rep.* **2021**, *7*, 90–99. [CrossRef]

21. Iqbal, S.; Habib, S.; Ali, M.; Shafiq, A.; ur Rehman, A.; Ahmed, E.M.; Khurshaid, T.; Kamel, S. The Impact of V2G Charging/Discharging Strategy on the Microgrid Environment Considering Stochastic Methods. *Sustainability* **2022**, *14*, 13211. [CrossRef]

22. Tran, S.N.; Garcez, A.S.D.A. Deep logic networks: Inserting and extracting knowledge from deep belief networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2016**, *29*, 246–258. [CrossRef] [PubMed]

23. Kong, X.; Li, C.; Zheng, F.; Wang, C. Improved deep belief network for short-term load forecasting considering demand-side management. *IEEE Trans. Power Syst.* **2019**, *35*, 1531–1538. [CrossRef]

24. Tao, C.; Wang, X.; Gao, F.; Wang, M. Fault diagnosis of photovoltaic array based on deep belief network optimized by genetic algorithm. *Chin. J. Electr. Eng.* **2020**, *6*, 106–114. [CrossRef]

25. Zheng, L.; Hu, W.; Zhou, Y.; Min, Y.; Xu, X.; Wang, C.; Yu, R. Deep belief network based nonlinear representation learning for transient stability assessment. In Proceedings of the 2017 IEEE Power & Energy Society General Meeting, Chicago, IL, USA, 16–20 July 2017; pp. 1–5.

26. Su, T.; Liu, Y.; Zhao, J.; Liu, J. Deep belief network enabled surrogate modeling for fast preventive control of power system transient stability. *IEEE Trans. Ind. Inform.* **2021**, *18*, 315–326. [CrossRef]

27. Li, B.; Wu, J. Adaptive assessment of power system transient stability based on active transfer learning with deep belief network. *IEEE Trans. Autom. Sci. Eng.* **2022**, *20*, 1047–1058. [CrossRef]

28. Zhang, B.; Xu, X.; Xing, H.; Li, Y. A deep learning based framework for power demand forecasting with deep belief networks. In Proceedings of the 2017 18th International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT), Taipei, Taiwan, 18–20 December 2017; pp. 191–195.

29. Wu, W.; Yang, D.; Yu, Z.; Zhou, B.; Lv, C.; Gu, J. Determination of key transmission section and strong correlation section based on matrix aggregation algorithm. *Int. J. Electr. Power Energy Syst.* **2023**, *153*, 109387. [CrossRef]

30. Wu, S.; Hu, W.; Lu, Z.; Gu, Y.; Tian, B.; Li, H. Power system flow adjustment and sample generation based on deep reinforcement learning. *J. Mod. Power Syst. Clean Energy* **2020**, *8*, 1115–1127. [CrossRef]

31. Chen, B.; Tao, C.; Tao, J.; Jiang, Y.; Li, P. Bearing Fault Diagnosis Using ACWGAN-GP Enhanced by Principal Component Analysis. *Sustainability* **2023**, *15*, 7836. [CrossRef]

32. Lin, Y.; Giacoumidis, E.; O'Duill, S.; Barry, L.P. DBSCAN-based clustering for nonlinearity induced penalty reduction in wavelength conversion systems. *IEEE Photonics Technol. Lett.* **2019**, *31*, 1709–1712. [CrossRef]

33. Xue, T.; Chang, Z.; Zhang, N. Comprehensive evaluation of slope stability using entropy-weight fuzzy hierarchy analysis. *J. Heilongjiang Inst. Sci. Technol.* **2011**, *21*, 454–457.

34. Aslam, N.; Xia, K.; Hadi, M.U. Optimal wireless charging inclusive of intellectual routing based on SARSA learning in renewable wireless sensor networks. *IEEE Sens. J.* **2019**, *19*, 8340–8351. [CrossRef]

35. Bajpai, D.; He, L. Evaluating KNN performance on WESAD dataset. In Proceedings of the 2020 12th International Conference on Computational Intelligence and Communication Networks (CICN), Bhimtal, India, 25–26 September 2020; pp. 60–62.

36. Hanshuo, M.; Xiaodong, Z.; Xuan, T.; Fei, Q. Research on fault prediction method of electronic equipment based on improved SVR algorithm. In Proceedings of the 2020 Chinese Automation Congress (CAC), Shanghai, China, 6–8 November 2020; pp. 3092–3096.

37. Sheng, C.; Yu, H. An optimized prediction algorithm based on XGBoost. In Proceedings of the 2022 International Conference on Networking and Network Applications (NaNA), Urumqi, China, 3–5 December 2022; pp. 1–6.