An improved decision tree-based method for predicting overvoltage peak values integrating a model-driven scheme

Zhe Zhang[1], Boyu Qin[1], Xin Gao[1], Yixing Zhang[1], and Tao Ding[1]

[1]Xi'an Jiaotong University

April 26, 2023

Abstract

The commutation failure is the most prevalent fault in line-commutated converter based HVDC systems, which may result in transient overvoltage on the sending-side system. Overvoltage level evaluation has become a crucial task for power industries to assess the tripping risk of large-scale wind turbines and implement effective stability control measures. In this paper, decision tree (DT) model is adopted to extract the mapping relationship between transient overvoltage and massive electrical quantities of power grids. The common DT algorithm is transformed by modifying the error weight assignment, which reflects the error tolerances for different actual overvoltage regions. To compensate for potential inaccuracies in the data-driven method, a derivation of the mathematical relationship between the reactive power consumed by the rectifier and AC voltage is presented, along with an analytical expression for the peak value of transient overvoltage. On this basis, an overvoltage analysis method integrating the model-driven and data-driven techniques is proposed, and the improved DT algorithm is ap-plied to fast error correction, enhancing the interpretability of regression prediction results. Case studies were performed in the actual Northwest China local region hybrid AC/DC power grid with transient overvoltage problems, and the simulation results verified the effectiveness of the proposed method.

An improved decision tree-based method for predicting overvoltage peak values integrating a model-driven scheme

Zhe Zhang [1], Boyu Qin [1*], Xin Gao[1], Yixing Zhang [1], Tao Ding[1]

[1] State Key Laboratory of Electrical Insulation and Power Equipment, School of Electrical Engineering, Xi'an Jiaotong University, Xi'an, China

[*] *qinboyu@xjtu.edu.cn*

Abstract: The commutation failure is the most prevalent fault in line-commutated converter based HVDC systems, which may result in transient overvoltage on the sending-side system. Overvoltage level evaluation has become a crucial task for power industries to assess the tripping risk of large-scale wind turbines and implement effective stability control measures. In this paper, decision tree (DT) model is adopted to extract the mapping relationship between transient overvoltage and massive electrical quantities of power grids. The common DT algorithm is transformed by modifying the error weight assignment, which reflects the error tolerances for different actual overvoltage regions. To compensate for potential inaccuracies in the data-driven method, a derivation of the mathematical relationship between the reactive power consumed by the rectifier and AC voltage is presented, along with an analytical expression for the peak value of transient overvoltage. On this basis, an overvoltage analysis method integrating the model-driven and data-driven techniques is proposed, and the improved DT algorithm is applied to fast error correction, enhancing the interpretability of regression prediction results. Case studies were performed in the actual Northwest China local region hybrid AC/DC power grid with transient overvoltage problems, and the simulation results verified the effectiveness of the proposed method.

Introduction

To achieve the "double carbon" target and promote the construction of the new power system, renewable energy sources have ushered in leapfrog growth. Due to the advantages of large transmission capacity, low power loss, and flexible transmission power adjustment, the line-commutated converter (LCC) based high voltage direct current (HVDC) transmission technology has gained widespread adoption in China owing to its capability of long-distance and large-capacity transmission [1,2]. Currently, China has formed the typical hybrid AC/DC power grid with the largest scale and most complicated network structure in the world [3]. Commutation failure (CF) is a unique fault of LCC-HVDC transmission system [4]. During the CF, the voltage amplitude of the sending AC grid will first decrease then increase [5-7]. For the sending system with large scale renewable energy integration, transient overvoltage can cause off-grid accidents of renewable energy, jeopardizing the secure and stable operation of the hybrid AC/DC power grid [8-10]. Therefore, it has become a growing concern for the power industries to analyse the overvoltage level under typical DC faults, providing a basis for the stability analysis of hybrid AC/DC power grid and guiding the formulation of renewable energy high voltage ride through standards.

To investigate the impact of CF on AC system, simulation analysis [11] and discussion [12,13] have been conducted to study the mechanism of transient overvoltage caused by CF. However, these studies only briefly analysed the effect of the DC current rise and drop stage on transient overvoltage, without considering the impact of CF caused by different fault severity and duration on transient overvoltage. In terms of calculating the overvoltage peak value, model-driven techniques relying on power system mechanism models have been proposed, which are comprised of the AC equivalent method [14], reactive power short circuit ratio method [15], and single branch voltage drop method [16]. The AC equivalent method and reactive power short circuit ratio method are proposed based on the ratio of reactive surplus level and system short-circuit capacity during the transient period. Nevertheless, the derivation of above-mentioned two methods involves model simplifications that could lead to unacceptable computation errors. Considering the impact of active power fluctuation to the transient overvoltage, the single-branch voltage drop method is proposed to improve the prediction accuracy, while the computational burden could be challenging in practical power systems. Therefore, the trade-off needs to be made between computation accuracy and speed when adopting these model-driven methods for online overvoltage peak value prediction.

With the application of wide area measurement systems (WAMS) [17], artificial intelligence methods have shown application potential for data-driven overvoltage peak value prediction through data relationship mining [18]. Among traditional machine learning methods, neural networks (NN) have been widely used for transient overvoltage prediction due to the powerful non-linear mapping capabilities [19]. In [20], support vector machines (SVM) have also demonstrated good performance in transient overvoltage classification problems. In addition, core vector machines (CVM) are constructed to extract the mapping relationship between transient overvoltage and massive electrical quantities of power grids [21]. However, traditional machine learning algorithms typically require manual feature extraction from the data, which may affect the prediction effectiveness for complicated and unstructured data types. Consequently, deep learning methods such as long-short term memory network (LSTM) [22] and deep imbalanced learning framework [23] have been adopted to predict the overvoltage peak value due to the capability of automatic feature extraction. In spite of fast computation speed, the rigorous theoretical analysis of power system evolution mechanism is abandoned in data-driven methods, and improper feature selection will lead to over-fitting phenomenon when the number of training samples is insufficient, which affects the accuracy of prediction results. Moreover, higher prediction accuracy is crucial for high-risk scenarios in actual power system operations, while error tolerance considered in the traditional data-driven algorithms is treated equally.

To leverage the application potential of traditional data-driven methods for overvoltage prediction, this paper proposes an improved decision tree (DT) based method integrating a model-driven scheme. The main contributions of this paper are summarised as follows: (a) The decision-making principle of DT model are presented, and the application potential of DT model for predicting the overvoltage peak value is elaborated. To improve the prediction accuracy in high-risk scenarios, the traditional DT algorithm is modified by differentiating the error tolerances for different actual overvoltage regions. (b) A theoretical analysis method for overvoltage peak value of converter buses is studied, with an acceptable calculation accuracy and the potential for online application. On this basis, the data-driven method is integrated with the model-driven method to enhance the robustness to insufficient training sample and the interpretability of prediction results. The proposed DT method is adopted to reveal the association pattern between theoretical analysis results and true values. The advantages of the proposed approach include: (a) Compatibility of computation speed and accuracy for online application. (b) Strong interpretability of regression prediction results.

The remainder of this paper is organized as follows: In section 2, the traditional DT algorithm is modified to enhance the predicting performance in high-risk scenarios. Section 3 establishes an integrating method for predicting overvoltage peak value. Time-domain simulations are performed in section 4 as a verification. Section 5 concludes the paper and highlights future research directions.

Improved data-driven method for predicting overvoltage level

In this section, the decision-making principle of DT model are presented, and the application potential for predicting the overvoltage peak value is elaborated. In addition, to improve the prediction accuracy in high-risk scenarios, the traditional DT algorithm is modified by differentiating the error tolerances for different actual overvoltage regions.

DT model

DT is an powerful supervised machine learning tool to solve the classification and regression problems in high-dimensional data space [24]. The illustration of DT model is depicted in Fig. 1. The basic principle of DT is to recursively partition the input space into smaller subsets based on the values of the input features. The prediction process starts from the root node and ends at a terminal node, and the node with two successors in the DT model is considered as internal node. Each internal node of the tree represents a test on one of the input features, and each branch corresponds to one of the possible outcomes of the test. The leaves of the tree represent the final predictions or classifications for each input instance.

Fig.1 *Illustration of DT model*

The objective of constructing a DT is to determine the optimal sequence that minimizes the impurity measure at each split. The impurity reduction is calculated based on the difference between the impurity of the parent node and the weighted impurity of the child nodes after the split. A maximal tree is initially trained by recursively splitting a node into two purer successors, where all the available splitting rules are traversed until further splitting cannot improve overall accuracy. Eventually, splitting process partition all the samples in a multidimensional space into different subregions with homogeneous samples [25,26], and the samples in each subregion should have the same or similar prediction objective. Based on the well-trained DT model, the complicated classification or regression problem can be converted to a series of "if-then" questions based on the thresholds of partial input features or their linear combinations [27].

For the regression problem of overvoltage peak value prediction, the electrical quantities related to voltage responds of power grids as selected as input features during the off-line training process. The key factors of overvoltage peak value can then be extracted by determining the splitting rules. When applying online, DT model can achieve the overvoltage prediction according to the operation characteristics of power grids.

Improved DT algorithm

As for traditional regression DTs, the splitting rule is defined as follows. Firstly, t is an internal node in the regression DT, and the purity of node t can be obtained by .

Where N is the number of samples in the internal node t ;y_i is the label of i_{th}sample; is the average value of N samples in the node t . Next, the purity loss between the internal node s and two successors split by s is adopted to determine the splitting rules, and the branching quality index ΔR is defined to quantitatively assess the purity loss.

Where $R\ (t_R\ )$ and$R\ (t_L\ )$ are the purity of the right and left subtrees split by s , respectively; N_R andN_L are the number of samples in the right and left subtrees, respectively. Therefore, to make each subregion more homogeneous, the splitting attribute X_j and standard $K\ _2$ should be selected to maximize the purity loss of node s .

Similarly, all internal nodes in the DT model are split according to the above-mentioned process until relevant constraints such as node purity meet requirements, and the terminal nodes can be obtained. In the terminal node, the ultimate prediction value is determined as the average of all samples.

According to the actual operation requirement, higher overvoltage level poses a greater threat to the secure and stable operation of power systems. Effective control measures should be implemented to eliminate the risk of severe faults. However, it is obvious that the node purity in treats training errors of different samples equally. To address this limitation and better reflect the actual operation requirement, larger weights are assigned to prediction errors in high-risk scenarios, and the modification of the purity and branching quality indices as follows.

According to and , the improved DT algorithm limits the prediction error of high-risk samples and introduces the knowledge of risk differences in overvoltage problems.

Specific DT construction scheme

The specific construction of DT model involves two steps. In step 1, the sample set is generated offline for overvoltage peak value prediction. The composition of sample set is depicted in Fig. 2, where the key electrical quantities of power systems are selected as input features and the corresponding overvoltage peak values obtained by the PSASP software are adopted as the output label.

Fig.2 *Composition of sample set*

In step 2, the improved DT algorithm is applied to determining the splitting rule for each node in the DT model. Specifically, pseudo code for DT model construction is presented below.

Algorithm for DT Model Construction
Initialize: Set appropriate values for the minimum purity of samples ($P_{\min}$) and the minimum number of samples ($S_{\min}$) i

Overvoltage prediction method integrating the model-driven and data-driven techniques

The improved DT-based method proposed in Section 2 realizes reliable prediction accuracy and strong adaptability to high-risk scenarios. However, for data-driven methods, the lack of theoretical analysis and the potential over-fitting phenomenon due to insufficient training samples should be taken into account. Furthermore, the interpretability of prediction results are crucial for decision-making in power system operation. To further address the above issues, this section proposes an overvoltage analysis method integrating the model-driven and data-driven techniques, and the integration framework of two methods is presented.

Model-driven overvoltage analysis method

The transient overvoltage mechanism and an analytical expression on transient overvoltage peak value are studied, and key influencing factors leading to overvoltage problems are extracted. The proposed model-driven method has the potential for online application, which provides the effective support for the integration with data-driven methods.

(1) Transient overvoltage mechanism

The equivalent circuit of a typical wind-thermal-bundled HVDC transmission system is shown in Fig. 3.

Fig.3 *Typical two-terminal AC/DC hybrid system model*

When the AC/DC hybrid system is operating normally, . The active and reactive powers should be balanced, which can be expressed as:

Where the subscript N indicates the normal operating condition, Q_{drN} is the reactive power consumption of the converter station, Q_{CrN} , $Q_{ac\ 1N}$, Q_{wN} denote the reactive output of reactive power compensation device, sending AC system, and wind farm, respectively. A short-circuit fault which occurs in the receiving AC system may lead to CF. For the sending AC system, the voltage is directly related to the reactive power, and the transient overvoltage of converter bus can be expressed as follows.

Where [?] Q_r is the reactive surplus of converter station, S_{cr} is the short circuit capacity of converter station. During the CF, the reactive power consumed by the converter is dynamic. Therefore, the [?] Q_r in can be expressed as follows.

Combining with , the transient overvoltage level under different DC faults can be obtained.

(2) Expression of transient overvoltage level

According to the dynamic characteristic of reactive power compensation device, can be derived from .

Combining and , the transient overvoltage of converter bus can be derived.

According to [28], the Q_{dr} can be calculated as follows.

Where U_{dr} and $U_{dr\ 0}$ are DC voltage and no-load DC voltage of the rectifier bus, respectively; I_{dr} is the DC current. It is obvious that the I_{dr} is always greater than 0 during the CF. According to , the Q_{dr} is greater than 0, which indicates that the rectifier and inverter only absorb reactive power. Furthermore, the derivative of Q_{dr} is shown as follows.

When a CF occurs, the DC voltage of the inverter will directly plummets to 0, and the DC current will rapidly increase. Then, the DC current will reach $I_{\ \min}$ ($I_{\ \min}$ is depending on VDCOL), and the rectifier constant current controller will escalate the firing angle to reduce the DC current. Consequently, the DC current will increase first and then decrease to $I_{\ \min}$, and the U_d will also decrease to $U_{dr\ \min}$. Make $\mathrm{d}Q_{dr}$ $/\mathrm{d}t$ =0, then

when $U_{dr} = U_{d\ \min}$ and $I_{dr} = I_{d\ \mathrm{rmin}}$, is satisfied, and the rectifier will absorb the minimum reactive power. The $U_{dr\ \min}$ can be calculated as follows.

Substituting into , the Q_{dr} can be expressed as:

Therefore, the transient overvoltage of converter bus can be calculated as follows.

It can be seen from that the crucial factors which contribute to overvoltage issues are the short-circuit capacity, the reactive output of the AC system during normal operation, and the reactive power consumption of the converter station during the fault. Therefore, alternative transient overvoltage control measures can be summarized as follows: reducing reactive-voltage sensitivity [29] and suppressing reactive surplus source [30,31].

*Framework of the integrated method*The proposed theoretical analysis method for calculating the overvoltage peak value of converter buses achieves the compatibility of computation speed and accuracy for online application, providing effective support for the integration with data-driven method. Specifically, as shown in , the theoretical overvoltage values can be obtained efficiently by utilizing the equivalent parameters of AC system and the operation parameters of DC system. To avoid the over-fitting phenomenon caused by improper feature selection and enhance the interpretability of regression prediction results, the theoretical analysis results are regarded as additional input features to the original training samples for the data-driven method. The detailed integration mode between model-driven and data-driven overvoltage analysis methods is depicted in Fig. 4. The objective of improved DT model is transformed from massive data relationship mining to association pattern revealing between the theoretical evaluation values and the true values. Therefore, for typical fault scenarios, the key electrical quantities and corresponding theoretical overvoltage values are selected as input features, and the overvoltage peak values obtained by the time domain simulation method are taken as the output. The DT model is trained by the improved samples to achieve fast error correction.

Fig.4 *Integrated prediction network structure*

*Specific procedure of overvoltage level prediction*The specific procedure of overvoltage peak value prediction is depicted in Fig. 5, which is comprised of offline training and online prediction.

Fig.5 *Specific procedure of overvoltage level prediction*

In the stage of offline training, typical fault scenarios are simulated in PSASP software, taking into account factors such as load levels, renewable energy penetration rates, and active power transmitted by the DC link. From the simulation results, characteristic quantities related to overvoltage levels are extracted to form the sample set. The input features of the sample set consist of key electrical quantities and corresponding overvoltage peak values calculated by theoretical analysis method, while the output labels are the overvoltage peak values obtained by time domain simulation method. The tree growing and pruning procedures are then carried out based on the sample sets, and the splitting rules of each node are determined according to the prediction performance to obtain the optimal overvoltage level prediction model. In the stage of online application, when a fault occurs in the actual power grid, key electrical quantities are collected by WAMS, and the corresponding theoretical overvoltage peak value is obtained through . The combined input features are then fed into the well-trained improved DT model, which accurately predicts the overvoltage level and guides the secure and stable operation of power systems.

Case study

In this section, the Northwest China local region hybrid AC/DC power grid depicted in Fig. 6 is adopted as the test system to verify the accuracy and effectiveness of the proposed method. The test system is constructed based on 750kV grid structure. Qingyu DC transmission projects are adopted to achieve the transmission of renewable energy electricity in Northwest China region, forming a typical hybrid AC/DC power grid.

Fig.6 *Northwest China local region hybrid AC/DC power grid*

Performance of improved DT model

(1) Evaluation indices

Performance evaluation indices, containing the mean absolute error (MAE), mean absolute percentage error (MAPE), root-mean squared error (RMSE) and coefficient of determination (R^2), are adopted in this paper to evaluate the overvoltage prediction effect of different models, and calculation formulas are shown as:

where m is the number of testing samples,$y\,'_i$ is the predicting value,y_i is the actual value, and y is the mean value of y_i .

(2) Improved DT model training

Taking the operating mode and fault type of power systems into account, the simulation software PSASP is adopted to establish a dataset of overvoltage under large disturbances. For the operating mode, the output of traditional power stations and renewable energy stations are adjusted under the load levels of 90%, 100% and 110%. In addition, the DC transmission is adjusted in increments of 10% within the range of 60% to 100%. As for the fault type, CF is set at the rectifier station, including single and double pole faults. The number of times that CF occurs is set as 1, 2 and 3, respectively. In addition, the duration of CF ranges from 0.15s to 0.25s. The simulation time is 20s, and the rated frequency of the test system is 50Hz. For each DT, the total number of samples is 4480, where 60% of the samples is selected as the training data set, and the remaining 40% are used for the testing dataset.

The key electrical characteristics are selected as the input features, and the overvoltage peak value of Qingnan 750kV bus is collected as the output. As a common approach for modelling and verifying model parameters, the '10-fold cross-validation' [32] is adopted to determine the depth of DT model. With the increasing of DT depth, the prediction effect of testing set are depicted in Fig. 7. Considering the coordination between calculation efficiency and prediction accuracy, the optimal DT depth is determined as 11.

Fig.7 *Prediction effect under different DT depth*

During the training process, as the number of samples increases, the RMSE and R^2 of training set and testing set are depicted in Fig. 8 and 9, respectively. The consistency of prediction accuracy between training set and testing set is demonstrated, verifying the effectiveness and generalization capability of the integrated method.

Fig.8 *RMSE of training and testing set*

Fig.9 $R^{\,2}$ *of training and testing set*

(3) Prediction effect and visualization of improved DT model

Fig. 10 and 11 depict the MAPE index of traditional DT model and improved DT model under different ranges of actual overvoltage peak values. It can be concluded from the prediction effect of the traditional integrated method that the MAPE for testing samples with high actual peak values is considerably higher than those with low actual peak values. Hence, modifying the DT algorithm is expected to reduce the MAPE for cases with high actual peak values. After modifying the common DT algorithm, the overall prediction effect in testing samples is improved, especially in the cases with high actual peak values.

In detail, to evaluate the efficacy of the improved DT algorithm, four regions have been delineated based on overvoltage levels, namely region 1 with less than 1.34 p.u., region 2 ranging from 1.34 p.u. to 1.38 p.u., region 3 ranging from 1.38 p.u. to 1.42 p.u., and region 4 with more than 1.42 p.u.. The corresponding MAE and MAPE indices for each of these regions are presented in Tables 1 and 2. It can be seen that the maximum improvement in both MAE and MAPE indices, amounting to 6.9% and 34.3%, respectively, occurs in region 4. These findings indicate that the proposed approach has more significant enhancement in high-risk scenarios.

Table 1 MAE (p.u.) of four overvoltage regions

Region	Traditional DT model	Improved DT model
1	0.030046	0.019885
2	0.024337	0.019224
3	0.029752	0.021256
4	0.030084	0.028010

Table 2 MAPE (%) of four overvoltage regions

Region	Traditional DT model	Improved DT model
1	2.2051	2.0424
2	1.8262	1.4297
3	2.2067	1.7691
4	2.2894	1.5037

Fig.10 *Mean absolute error rate and actual values of testing samples ranging from 1.30 p.u. to 1.36 p.u.*

Fig.11 *Mean absolute error rate and actual values of testing samples ranging from 1.38 p.u. to 1.44 p.u.*

In addition, the 2-D nomogram of improved DT model is depicted in Fig. 12, where the splitting rules of each internal node during the DT generation process is indicated. In addition, the predicted value, sample size, and node purity of each terminal node can be obtained, achieving the visualization of the decision-making process.

Fig.12 *2-D nomogram of improved DT model*

Prediction effect of integrated method

(1) Theoretical analysis results

The simulation parameters of test system are as follows: U_{dN}=±400kV; R_{d}=6.04Ω· U_{N}=800kV; Q_{CN}=4000MVar; I_{dmin}=0.5kA; the rated DC current is 5kA; the rated active power is 1000MW. Considering a typical operation scenario, Q_{ac}=105MVar; the output of synchronous generators and renewable energy generation system is 5000MW and 1500MW, respectively. When the S_{C} is 32000MVA, assuming that the single CF occurs in Qingyu DC, the transient overvoltage can be calculated as 1.328 according to . The simulation results are depicted in Fig. 13, which validates the effectiveness of the theoretical analysis method.

Fig.13 *Voltage response curves of Qingnan bus*

(2) Prediction effect of integrated method

Based on the overvoltage theoretical analysis method in section 3.1.1, the theoretical overvoltage peak values of 4480 samples generated offline are calculated and adopted as partial input features of the training samples. Along with the key electrical inputs and output labels obtained by PSASP in section 3.1.2, the sample set is formed to train the DT model, revealing the association pattern between theoretical analysis values and true values. Based on the well-trained integrated DT model, the regression prediction results are compared with traditional DT model. Results of the overvoltage peak value at Qingnan 750kV bus under different prediction models are shown in Fig. 14. It can be concluded that the integrated model proposed in this paper has a better regression prediction effect on the testing data set, and the predicted values are closer to the true values.

Fig.14 *Prediction results of data-driven method and integrated method*

Furthermore, performances of the traditional DT model, improved DT model and integrated method are compared in Table 3.

Table 3 Prediction effect under different evaluation indices

Indices	Improved DT	Integrated method
MAE/p.u.	0.0291	0.0235
MAPE/%	2.1628	1.7318
RMES/p.u.	0.0348	0.0293

Indices	Improved DT	Integrated method
$R^2/1$	0.1125	0.2123

Based on the comparison presented in Table 1, it can be concluded that the integration of analytical method and traditional DT model enhances the performance of both methods. Specifically, the proposed method has achieved a 24.7% improvement in the accuracy of the traditional DT method and a 20.1% improvement in the accuracy of the improved DT method, as demonstrated by the MAE index. On the one hand, overvoltage peak values calculated by theoretical analysis method are adopted as the input features for the DT-based error correction part. On the other hand, considering the advantages of DT model in mapping relationship revealing, the error of the analytical method caused by model simplification can be reduced by the data-driven error correction part. The integration of model-driven method helps enhance the robustness of the integrated method to insufficient training sample and the inappropriate input feature selection.

Conclusion

This paper proposes an improved DT-based overvoltage level prediction method integrating the model-driven scheme for hybrid AC/DC power grids. A main factor that directly affects the performance of overvoltage analysis method is the adaptability to operational scenarios. To address the above key issue, an improved DT algorithm is proposed to predict the overvoltage level considering the advantages in mapping relationship revealing, and the prediction accuracy in high-risk scenarios is enhanced by modifying the splitting rules in the DT training process. In addition, a theoretical analysis method for evaluating the overvoltage peak value of converter buses is proposed with an acceptable calculation accuracy and the potential for online application, and the mathematical relationship between the reactive power consumed by the rectifier and AC voltage is derived. On this basis, an overvoltage analysis method integrating the model-driven and data-driven techniques is proposed to enhance the robustness to insufficient training sample and inappropriate input feature selection, and the DT algorithm is adopted to reveal the association pattern between theoretical analysis results and true values, improving the interpretability of regression prediction results. Simulations on a simplified hybrid AC/DC actual power grid have been performed to verify the effectiveness of improved integrated method.

Ongoing research is focused on the extraction of input features for the DT-based integrated method. As described in section 3, the input features of DT model are comprised of the power system operation information and voltage dynamic characteristic information. However, as power grid scale expands and operation complexity increases, the original input features may contain redundant information, which can affect the model training time and evaluation accuracy. Consequently, feature dimensionality reduction technology, such as feature selection or transformation, should be adopted to obtain more expressive new features, thereby improving the model training speed and the evaluation accuracy.

Acknowledgments

This work was supported by the National Key R&D Program of China (No. 2021YFB2400800). The authors are with State Key Laboratory of Electrical Insulation and Power Equipment, and school of Electrical Engineering, Xi'an Jiaotong University, Xi'an 710049, Shaanxi Province, China.

References

1. Yamashita, K., Tsukamoto, G., Nishikata, S.: 'Steady-state characteristics of a line-commutated converter-based high-voltage direct current transmission system for series-connected wind power plants'. IEEE Trans. Ind. Appl., 2020, 56, (4), pp. 3932–3939.

2. Xue, Y., Zhang, X., Yang. C.: 'Commutation failure elimination of LCC HVDC systems using thyristor-based controllable capacitors'. IEEE Trans. Power Del., 2018, 33, (3), pp. 1448–1458.
3. Gu, M., Meegahapola, L., Wong, K. L.: 'Coordinated voltage and frequency control in hybrid AC/MT-HVDC power grids for stability improvement'. IEEE Trans. Power Syst., 2021, 36, (1), pp. 635-647.
4. Dong, X., Guan, E., Jing, L., et al.: 'Simulation and analysis of cascading faults in hybrid AC/DC power grids'. Int. J. Elect. Power Energy Syst., 2020, 115, pp. 105492.
5. Zhu, Y., Zhang, S., Liu, D., et al.: 'Prevention and mitigation of high-voltage direct current commutation failures: a review and future directions'. IET Gener. Transm. Distrib., 2019, 13: 5449-5456.
6. Xie, Q., Zheng, Z., Wang, Y., et al.: 'Analysis of transient voltage disturbances in LCC-HVDC sending systems caused by commutation failures'. IEEE Trans. Power Del., 2022, 37, (5), pp. 4370-4381.
7. Zheng, Z., Ren, J., Xiao, X., et al.: 'Response mechanism of DFIG to transient voltage disturbance under commutation failure of LCC-HVDC system'. IEEE Trans. Power Del., 2020, 35, (6), pp. 2972-2979.
8. Zhang, T., Yao, J., Sun, P., et al.: 'Improved continuous fault ride through control strategy of DFIG-based wind turbine during commutation failure in the LCC-HVDC transmission system'. IEEE Trans. Power Electron., 2021, 36, (1), pp. 459-473.
9. Mirsaeidi, S., Dong, X.: 'An integrated control and protection scheme to inhibit blackouts caused by cascading fault in large-scale hybrid AC/DC power grids'. IEEE Trans. Power Electron., 2019, 34, (8), pp. 7278-7291.
10. Yin, C., Li, F.: 'Reactive power control strategy for inhibiting transient overvoltage caused by commutation failure'. IEEE Trans. Power Syst., 2021, 36, (5), pp. 4764-4777.
11. Zhang, T., Yao, J., Sun P., et al.: 'Improved continuous fault ride through control strategy of DFIG-based wind turbine during commutation failure in the LCCHVDC transmission system'. IEEE Trans. Power Electr ., 2021, 36, (1), pp. 459-473.
12. Guo, C., Zhang, Y., Gole, A. M., et al.: 'Analysis of dual-infeed HVDC with LCC–HVDC and VSC–HVDC'. IEEE Trans. Power Del., 2012, 27, (3), pp. 1529-1537.
13. Zheng, Z., Ren, J., Xiao, X., et al.: 'Response mechanism of DFIG to transient voltage disturbance under commutation failure of LCC-HVDC system'. IEEE Trans. Power Del., 2020, 35, (6), pp. 2972-2979.
14. Wang, F., Liu, T., Ding Y., et al.: 'Calculation method and influencing factors of transient overvoltage caused by HVDC block'. Power Syst. Technol., 2016, 40, (10), pp. 3059-3065.
15. He, J., Tang, Y., Zhang J., et al.: 'Fast calculation of power oscillation peak value on AC tie-line after HVDC commutation failure'. IEEE Trans. Power Syst., 2015, 30, (4), pp. 2194-2195.
16. Li, H., Qin, B., Jiang Y., et al.: 'Data-driven optimal scheduling for underground space based integrated hydrogen energy system'. IET Renew. Power Gener., 2022, 16, (12), pp. 2521-2531.
17. Lagos, D. T., Hatziargyriou N. D.: 'Data-driven frequency dynamic unit commitment for island systems with high res penetration'. IEEE Trans. Power Syst., 2021, 36, (5), pp. 4699-4711.
18. Cremer, J. L., Konstantelos, I., Tindemans, S. H., et al.: 'Data-driven power system operation: Exploring the balance between cost and risk'. IEEE Trans. Power Syst., 2019, 34, (1), pp. 791-801.
19. Chen, H., Assala, P. D. S., Cai, Y., et al.: 'Intelligent Transient Overvoltages Location in Distribution Systems Using Wavelet Packet Decomposition and General Regression Neural Networks'. IEEE Trans. Ind. Informat., 2016, 12, (5), pp. 1726-1735.
20. Tang, W., Gu, Y., Xin, Y., et al.: 'Classification for Transient Overvoltages in Offshore Wind Farms Based on Sparse Decomposition'. IEEE Trans. Power Del., 2022, 37, (3), pp. 1974-1985.
21. Wang, B., Fang, B., Wang, Y., et al.: 'Power system transient stability assessment based on big data and the core vector machine'. IEEE Trans. Smart Grid, 2016, 7, (5), pp. 2561-2570.
22. Li, B., Wu, J., Hao, L., et al. 'Anti-Jitter and refined power system transient stability assessment based on long-short term memory network'. IEEE Access, 2020, 8, pp. 35231-35244.
23. Tan, B., Yang, J., Tang, Y., et al.: 'A deep imbalanced learning framework for transient stability assessment of power system'. IEEE Access, 2019, 7, pp. 81759-81769.
24. Zhang, J., Jia, H. Zhang, N. 'Alternate support vector machine decision trees for power systems rule

extractions’. IEEE Trans. Power Syst., 2023, 38, (1), pp. 980-983.
25. Hou, Q., Zhang, N., Kirschen, D. S., et al.: ’Sparse oblique decision tree for power system security rules extraction and embedding’. IEEE Trans. Power Syst., 2021, 36, (2), pp. 1605-1615.
26. Wang, T., Bi, T., Wang, H., et al.: ’Decision tree based online stability assessment scheme for power systems with renewable generations’. CSEE J. Power Energy Syst., 2015, 1, (2), pp. 53-61.
27. Jia, H., Hou, Q., Liu, Y., et al.: ’Extraction of static voltage stability rule based on oblique regression tree and its ensemble algorithm’. Automat. Electric Power Syst., 2022, 46, (1), pp. 51-59.
28. Yin, C., Li, F.: ’Analytical expression on transient overvoltage peak value of converter bus caused by DC faults’. IEEE Trans. Power Syst., 2021, 36, (3), pp. 2741-2744.
29. Jin, X., Nian, H.: ’Overvoltage suppression strategy for sending AC grid with high penetration of wind power in the LCC-HVDC system under commutation failure’. IEEE Trans. Power Electron., 2021, 36, (9), pp. 10265-10277.
30. Wang, T., Pei, L., Wang, J., et al.: ’Overvoltage suppression under commutation failure based on improved voltage-dependent current order limiter control strategy’. IEEE Trans. Ind. Appl., 2022, 58, (4), pp.4914-4922.
31. Xiao, J., Heng, N., Chen Z., et al.: ’Optimal power coordinated control strategy for dfig-based wind farm to increase transmission capacity of the LCC-HVDC system considering commutation failure’. IEEE J. Emerg. Sel. Top. Power Electron., 2022, 10, (3), pp. 3129-3139.
32. Li, F., Wang, Q., Tang, Y., et al.: ’An integrated method for critical clearing time prediction based on a model-driven and ensemble cost-sensitive data-driven scheme’. Int. J. Electr. Power Energy Syst., 2021, 125, pp, 106513.