Power Grid Safety Assessment Based on an Improved Generative Adversarial Network

XU Aidong¹, ZHANG Yunan² and JIANG Yixin²

¹ Electric Power Research Institute, China Southern Power Grid
² Guangdong Provincial Key Laboratory of Power System Network Security
Corresponding author’s e-mail: zhanght@seu.edu.cn

Abstract. Critical clearing time (CCT) is one of the important indices for transient stability evaluation. From the data-driven perspective, the research on CCT prediction is troubled by changing models and insufficient samples, that is, the small sample characteristics of the target system may be submerged by the large sample characteristics of the source system. In this paper, a prediction method of CCT based on improved generative adversarial network (WGAN) is put forward to solve this problem, and the network structure of WGAN is redesigned. This method is suitable for the transition phase of power grid when the operation mode changes. Through the unsupervised training of WGAN, the neural network will automatically learn the complex relationship between the small sample data of the target system, so that the generator can produce small sample data with high precision. Then the CCT prediction model was built on the expanded balanced dataset. Results validate that the method can effectively learn the distribution law of samples, improve the CCT prediction ability, and has the advantage of precision in the transition phase of power grid with insufficient samples.

1. Introduction
With the high proportion of new energy grid connected generation and the large capacity ac/dc transmission system network operation, the operation characteristics of modern power system have changed significantly[1][2]. These characteristics lead to the increasing risk of transient stability of power system[3]. Critical clearance time (CCT) is an intuitive and useful index to describe power system large-disturbance rotor angle stability limit[4]. For CCT calculation, time domain simulation is one of the most basic methods. In order to solve the problem of CCT computing speed, some researches adopt direct method. In the transient energy function method, the accurate calculation of the critical energy is the key to calculate the CCT, and this process usually takes a lot of time[5]. For this reason, sensitivity analysis method was adopted in TEF (transient energy function) method to avoid the CCT search process[6][7]. Besides TEF method, extended equal area criterion (EEAC) method for CCT calculation is also very important[8][9]. All these direct methods usually have more or less model simplifications in assumptions, which may lead to unacceptable computational errors. In addition to these methods based on analysis models, data-driven methods represented by machine learning methods also show potential in fast CCT computing. In [10], ensemble learning framework is introduced to enhance adaptability to power system operation changes. The above method has strong fitting ability and fast calculation speed, but it does not consider the influence of the time-varying characteristics of the power system on the prediction of transient stability.

From the perspective of time variability of transient image, it is assumed that the change in the topology structure of the power grid results in the corresponding change in the transient stability charac-
teristics of the power grid. Therefore, the prediction sample and the training sample do not have the same distribution independently so that the current prediction model cannot adapt to the prediction requirements after the system changes. In this paper, the grid before the change is called the source system and the target system after the change. Due to insufficient data accumulation in the target system, it is difficult to train a complete prediction model based on its own data.

The above problems can be improved from two aspects: one is using less training data requirements algorithm. However, under the constraint of the number of samples, the information content of the training set naturally determines the theoretical upper limit of the training effect of the algorithm; the other one is transferring data and features from the source system to the target system to increase the available training samples and improve the ability of feature expression. In [11], transfer learning is proposed to establish the feature set and sample set migration channel of source system and target system. However, the number of samples from the source system after feature migration is still much larger than target system’s samples, and there is a risk that the small sample features of the target system will be submerged by the large sample features of the source system.

Aiming at the problem of sample imbalance in the second thought, this paper proposes a method of using generative adversarial network to expand small sample data set. In recent years, the generative adversarial network has made a breakthrough in missing image repair and high resolution image reconstruction. For example, in [12], an image recognition method based on conditional convolutional GAN (C-CGAN) is proposed to overcome the dependence of traditional deep learning methods on the number of samples. In traditional GAN, JS divergence is used to measure the distribution deviation between the real data and the generated data. In the training process, gradient disappear is easy to occur, resulting in training difficulties. In [13], the instability samples of the power system are generated and the online transient stability assessment is realized with the enhanced original sample training classifier based on the improved CGAN. In [14], GAN is combined with CNN powerful feature extraction ability. A large number of the complete measurement data is used to extract data features in order to overcome physical modeling data representation problem. Literature [15] establishes a conditional Wasserstein generative adversarial network model for gradient punishment optimization to guide the generation process of multi-category power transformer fault samples. It overcomes the training instability problem of the original generation adversative network model. In [16], Wasserstein distance is used to replace JS divergence, and Wasserstein distance is minimized as the target to train GAN network, which effectively improves the stability of GAN training.

There are obvious similarities between the problem of missing image filling and segmentation graph recovery and the problem of sample generation in the power system. Both of them generate data in accordance with objective laws under given context constraints. At present, GAN has been applied to data generation in the power system in existing studies. In order to avoid the instability of the original GAN training, a GAN-based load scene generation method was proposed in [17], and the model was trained based on DCGAN architecture and JS divergence as the objective function. The experimental results show that with the gradual progress of training, the data quality generated by GAN is gradually improved, which can produce a load sequence that is true enough and satisfies the diversity. Most of the existing GAN network structures are designed for image data. The typical image input is a 2-d 3-channel matrix, while the measurement data of the power system is a 1-d time series. All the above methods transform one-dimensional data into two-dimensional images for sample generation, and there is a risk of feature loss in the process of data transformation. Therefore, the existing network is not suitable for the processing of power system data, and GAN’s network structure parameters must be redesigned to suit the topic of small sample generation.

Therefore, this paper integrates the advantages of Wasserstein GAN (WGAN) and designs the GAN network structure parameters with 1 dimensional (1D) convolutional layer as the main structure according to the characteristics of power system data (1 dimensional sequence data). The generation loss of data is generated through Wasserstein distance constraint, so as to generate high-precision small sample data. The data set is balanced by expanding small samples to prevent the large sample features of the source system from drowning the small sample features of the target system. The re-
sults show that the CCT prediction model based on the expanded equilibrium data set has higher accuracy.

2. An Improved Generative Adversarial Network Based On Wasserstein Distance
The problem of building small sample data can be fundamentally transformed into a generating problem of learning data distribution, that is, training a GAN generation model that can generate small sample data, and selecting the generated data with the least difference from the original sample as the expansion.

Suppose there are $i$ groups measurement in the system, and the corresponding measurement value is $x_i$. Due to the complicated distribution relationship between these measurements, we set them as $p_r(x)$, and it can be seen from the above introduction that $p_r(x)$ is difficult to be described by an explicit mathematical model. With a set of noise vectors (hidden variables) $z$, which satisfy the joint gaussian distribution $p_z(z)$, the mapping relationship between $p_z(z)$ and $p_r(x)$ can be established through the deep neural network. In this way, new data satisfying the original data distribution relationship can be generated by sampling in the known distribution as input. The mapping establishment process is realized through GAN’s training, and the composition includes generator outputs $G(z; \theta^{(G)})$ and Discriminator output $D(z; \theta^{(D)})$. Meanwhile, $\theta^{(G)}$ and $\theta^{(D)}$ respectively represents the weight of the two networks.

During training, the generator’s input is noise $z$, through the up-sampling step of the multi-layer neural network, the distribution rule of the generated data $G(z)$ will gradually fit the sample data $p_r(x)$. The discriminator and the generator are trained at the same time. The input is both the data generated by the generator and the real sample data. Through the sampling step opposite to the generator, the final output is the probability $p_r$ about whether the input data is a real sample. The loss functions of generator and discriminator are as follows:

$$L_{\theta^G} = -E_{z \sim p_z(z)}[D(G(z))]$$  \hspace{1cm} (1)

$$L_{\theta^D} = -E_{x \sim p_r(x)}[D(x)] + E_{z \sim p_z(z)}[D(G(z))]$$  \hspace{1cm} (2)

Where $E$ represents the distribution of expectation; $G(z)$ means generator generates data, and $D(\cdot)$ means discriminator network output. GAN training process is a two-person zero-sum game problem. The objective function of the game process can be defined with following equation.

$$\min_{\theta^G} \max_{\theta^D} V(G, D) = E_{x \sim p_r(x)}[D(x)] - E_{z \sim p_z(z)}[D(G(z))]$$  \hspace{1cm} (3)

The above objective indicates that the generator is trying to generate data that is close to the distribution pattern of real data, thus making it impossible for the discriminator to determine whether the data is from real data. After the training, the generator will obtain the distribution pattern of real data unsupervised.

Specifically, the Wasserstein distance is used to measure the optimization objective in the description. Compared with the traditional JS distance, the Wasserstein distance can alleviate the problem of gradient disappearance in the training process and improve the training stability. Wasserstein distance is defined as follows:

$$W(p_r, p_g) = \inf_{\gamma} \left\{ \mathbb{E}_{(x,y)\sim \gamma} [\|x-y\|] \right\}$$  \hspace{1cm} (4)

Where $\Pi(p_r, p_g)$ is the collection of joint probability distribution $\gamma$; $W(p_r, p_g)$ is the infimum of $\gamma(x,y)$. Since it is difficult to calculate the Wasserstein distance between any division, the Kantorovich-Rubinstein dual form is adopted as:

$$W(p_r, p_g) = \frac{1}{K} \sup_{f \in \mathcal{F}} E_{x \sim p_r} [f(x)] - E_{x \sim p_g} [f(x)]$$  \hspace{1cm} (5)
The optimization goal under the Wasserstein distance can be achieved when $D(x)$ write as $f_{\theta_b}(x)$ and $G(z)$ write as $g_{\theta}(x)$.

3. CCT Prediction Model Using WGAN

3.1 WGAN Structure and Training

Considering that the sample data is one-dimensional time series data, 1D convolutional layer is used to reduce the dimension. The detailed network structure design is shown in figure 1.

The input of the generator is 100d random noise vector, and batch standardization operation is used to accelerate convergence and slow down over-fitting between levels, so that the gradient propagation can be deeper. In addition, tanh activation function was used in the output layer, and ReLu function was used in the other layers to generate false data with channel number of 1 and size of 10×10. The parameters of each convolutional layer are artificially designed to ensure that the final output dimension after the convolution operation is consistent with the sample dimension in the calculation example.

The discriminator network structure parameter design is basically symmetric with the generator network. The difference lies in that the activation function of the convolutional layer is replaced by LeakyReLU to improve the recognition performance. The network finally uses the full connection and the sigmoid activation function output to represent the probability that the input data belongs to the real measurement data.

In the training discriminator network, samples are first collected from the joint gaussian distribution and historical data to construct a batch of training data. Z input generator was used to generate measurement data, and then the loss value of discriminator was calculated according to the optimization objective, and the network parameters were updated with Adam optimizer. When the generator network is trained, the weight of discriminator network is fixed, the loss value of generator network is also calculated, and the network parameters are updated by Adam. Before each update of generator network parameters, perform the update of discriminator network parameters to improve the training speed.

3.2 CCT prediction model

Considering the requirements of model speed and accuracy for transient stability prediction, extreme learning machine (ELM) showed high speed and accuracy in similar prediction tasks. Therefore, ELM is used as a CCT prediction algorithm.
Extreme learning machine is a feed-forward neural network\cite{18}. The network structure is usually divided into three layers, namely the input layer, the hidden layer and the output layer. The hidden layer weight and bias are randomly given before training, and remain unchanged during training. As long as the activation function and the number of hidden layer neurons are determined, the output layer weight can be calculated, so that the unique optimal value can be obtained after a single training. For online CCT prediction evaluation, \(X_t = [x_{t1}, x_{t2}, \ldots, x_{tm}]^T \in \mathbb{R}^m\), \(t_i = [t_{i1}, t_{i2}, \ldots, t_{im}]^T \in \mathbb{R}^m\). Therefore, for the input \(N\) training samples \((x_i, t_i)\), a single hidden layer neural network with \(L\) hidden layer nodes can be expressed as:

\[
\sum_{j=1}^{L} \beta_j g(w_i, x_j + b_j) = o_j, j = 1, \ldots, N
\]

(6)

Where \(g(x)\) is the activation function, \(w_i = [w_{i1}, w_{i2}, \ldots, w_{in}]^T\) is the input weight, \(\beta_i\) is the output weight, \(b_i\) is the \(i\)th bias of hidden layer unit. \(w_i \cdot x_j\) is the inner product of \(w_i\) and \(x_j\). The learning process is expressed as an optimization problem as follows:

\[
\min f = \frac{1}{2} \| \beta \|^2 + \frac{M}{2} \sum_{i=1}^{N} \epsilon_i^2
\]

\[s.t \ h(x_i)\beta = t_i - \epsilon_i, i = 1, 2, \ldots, N\]

(7)

Where \(\epsilon_i\) is the \(i\)th training error, \(M\) is the weight value of the training error. During the training, \(w_i\) and \(b_i\) are randomly initialized, \(\beta\) is determined by the optimal model.

Based on the above generation algorithm and prediction algorithm, this paper proposes a CCT prediction model using WGAN. The model is divided into two levels: training and application. In model training, the original training data set includes the source system samples and the target system samples. The target system data in the training set is generated by improving generative adversarial network, and the generated data is used to expand the training set so as to make the training data set samples balanced. The ELM network was trained on the expanded training set, and the optimal CCT prediction model was obtained by optimizing the input weight of ELM and the bias of the hidden layer. In the application of the model, new data are predicted to realize CCT prediction.

4. Case Study

4.1 Fault Sample Generation

Monte carlo sampling method is applied to generate samples by simulating in Matlab PSTv3.0 toolbox. Fault type is set as three-phase short-circuit fault. The load level factor for each bus is assumed to follow uniform distribution ranging from 80% to 120% on the base. Fault locations are set by randomly choose one transmission line from the whole system. On this basis, 1500 samples for large samples (source domain) are generated. In addition, the base operation mode is also changed by cutting off the 26-29 line and removing the load on bus 26 to simulate the under operational variability, with which 135 samples are generated for small samples (target domain).

4.2 Test Parameter Setting

4.2.1 Hyper-parameter Optimization Of ELM

The only hyper-parameter of the ELM model is the number of hidden nodes. In [15], an empirical formula is defined as:

\[
N = \sqrt{a + b + c}
\]

(8)

Where \(N\) is the number of nodes in the hidden layer, \(a\) is the input number, \(b\) is the output number, and \(c\) needs to optimize the selection of positive integers.

As shown in figure 2, the test accuracy increases with the increase of the number of hidden layer nodes. To balance the training speed and test accuracy, set \(N\) to 20.

5
The number of hidden layer nodes

| Number of Hidden Layer Nodes | RMSE |
|------------------------------|------|
| 7                            | 0.008|
| 14                           | 0.023|
| 21                           | 0.017|
| 28                           | 0.014|
| 35                           | 0.008|
| 42                           | 0.020|
| 49                           | 0.026|
| 56                           | 0.011|
| 63                           | 0.025|
| 70                           | 0.013|

The number of hidden layer nodes is 50 so that the accuracy is balanced with the complexity of the model.

Fig 2 Hidden layer node number optimization

4.2.2 Iteration Times Of WGAN

During the experiment, the iteration times of 100, 300, 500, 700 and 900 were selected in this paper to conduct WGAN training sample expansion experiment. After the iteration times of 500, the accuracy tends to be stable, and the accuracy does not improve significantly with the increase of the iteration times. Considering the time efficiency and experimental results, WGAN iteration times were selected for 500 times in the subsequent experiments.

4.3 Results analysis

In order to test the effectiveness of WGAN model sample generation, the original training set was compared with the training set after adding the generated sample. In order to ensure the rationality of the comparison, the method in this paper generates new data samples based on the original training set, and randomly selects samples of a certain proportion of the original training set from the generated new samples. The absolute error and percentage of absolute error of the predicted results on the test sets of the two methods are shown in figure 3. Boxplot shows the median and upper and lower quartile of the prediction error. The violin figure shows the distribution of the data.

(a) Absolute error distribution of prediction results
Fig 3 The analysis of prediction error

As can be seen from figure 3, the absolute error of prediction of the original method is concentrated around the median, which is 0.027 seconds. The absolute error of the method in this paper is concentrated around the lower quartile, which is 0.019 seconds. The absolute error percentage of the initial method is distributed near the median, which is 2.9%. The absolute error percentage of the method in this paper is distributed near the lower quartile, which is 1.7%. In this paper, the number of samples with large error is reduced and the stability of prediction performance is improved. This result shows that the improved GAN can generate new samples similar to but different from the original samples by learning the distribution rules of samples, and such samples can be used in subsequent experiments to equalize the training data set.

In order to analyze the influence of the expansion size of the training set on the prediction accuracy, GAN was used to expand the small sample data of the target system. The expansion proportion was 20%, 40%, 60%, 80% and 120% of the original sample number, respectively, for 6 groups of comparative experiments. Table 1 shows the test prediction accuracy of different expansion ratios.

| Expansion ratios | MAE   | MAPE   | RMSE  |
|------------------|-------|--------|-------|
| 20%              | 0.0315| 5.98%  | 0.0410|
| 40%              | 0.0248| 4.77%  | 0.0326|
| 60%              | 0.0235| 4.49%  | 0.0308|
| 80%              | 0.0159| 2.98%  | 0.0204|
| 100%             | 0.0196| 3.58%  | 0.0245|
| 120%             | 0.0183| 3.52%  | 0.0241|

As can be seen from table 3, the prediction accuracy of CCT first improves and then decreases with the increase of the expansion ratio. The prediction result is the best when the expansion ratio is 80%. MAPE reaches 2.98%, which is 3% higher than 20%, and RMSE reaches 0.0204, which is 0.0206 higher than 20%. Therefore, the expansion proportion of the training set in the subsequent experiment was 80%.

To test whether the addition of generated samples can prevent the large sample features of the source system from drowning the small sample features of the target system, comparative experiments were conducted on different training sets. The original training set contains 1500 migrated samples from the source system and 100 samples from the target system. Based on the original training set, the samples are expanded and generated. The predicted results before and after the expansion are shown in figure 4.
As can be seen from figure 4, after adding the generated samples, the overall prediction error and error ratio of the test set are better than the prediction results based on the original training set. The average prediction error is reduced by 0.0124 seconds, and the MAE is reduced by 2.38%.

The results show that the income generated by improving GAN expansion of training sample set, the balance between classes, using both the original characteristics of the large sample source system, and keep the target system of new characteristics of medium and small sample, after the expansion of training on the training set of regression, greatly enhance the CCT prediction accuracy, can effectively solve the large sample characteristics will be the problem of small sample characteristics of submerged.

5. Conclusion
In this paper, an improved generative counter neural network is used to expand the small sample data of the target system and reduce the unbalance of the data. ELM is used on the expanded data. The results show that the new samples generated by the improved GAN model are not simple copies and splicing of the original samples, but the distribution rules of real samples are learned and new and effective samples are generated. The data set with generated sample can also effectively prevent the problem of large sample characteristics drowning small sample characteristics and improve the accuracy of CCT prediction.

Considering the current development of deep learning technology in smart grid, it is of great research value to design an algorithm that can effectively expand the number of samples and improve the number of samples.

References
[1] Dong Xinzhou, Tang Yong, Bu Guangquan.: Confronting problem and challenge of large scale AC-DC hybrid power grid operation[J]. Proceedings of the CSEE, 2019, 39(11): 3107-3119.
[2] Zhou Xiaoxin, Chen Shuyong, Lu Zongxiang.: Technology features of the new generation power system in China[J]. Proceedings of the CSEE, 2018, 38(07): 1893-1904+2205.
[3] Zhu Shu, Liu Kaipei, Qin Liang.: Analysis of transient stability of power electronics dominated power system: an overview[J]. Proceedings of the CSEE, 2017, 37(14): 3948-3962+4273.
[4] P. Kundur, N. J. Balu, and M. G. Lauby, Power system stability and control[M]. McGraw-hill New York, 1994, vol. 7.
[5] H.-D. Chiang.: A theory-based controlling uep method for direct analysis of power system transient stability, in IEEE International Symposium on Circuits and Systems,. IEEE, 1989, pp. 1980–1983.
[6] L. G. Roberts, A. R. Champneys, K. R. Bell, and M. di Bernardo.: Analytical approximations of critical clearing time for parametric analysis of power system transient stability[J]. IEEE
Journal on Emerging and Selected Topics in Circuits and Systems, vol. 5, no. 3, pp. 465–476, 2015.

[7] S. Sharma, S. Pushpak, V. Chinde.: Sensitivity of transient stability critical clearing time [J]. IEEE Transactions on Power Systems, vol. 33, no. 6, pp. 6476–6486, 2018.

[8] Y. Xue, T. Van Custem, M. Ribbens-Pavella.: Extended equal area criterion justifications, generalizations, applications [J]. IEEE Transactions on Power Systems, vol. 4, no. 1, pp. 44–52, 1989.

[9] M. Yin, C. Chung, K. Wong, Y. Xue, Y. Zou.: An improved iterative method for assessment of multi-swing transient stability limit [J]. IEEE Transactions on Power Systems, vol. 26, no. 4, pp. 2023–2030, 2011.

[10] Z. Chen, X. Han, C. Fan, H. Zhang, C. Liu.: Prediction of critical clearing time for transient stability based on ensemble extreme learning machine regression model [C] // IEEE Innovative Smart Grid Technologies-Asia (ISGT Asia). pp. 3601–3606 (2019).

[11] Wang Q, Zhang C, Lv Y.: Data inheritance based updating method and its application in transient frequency prediction for a power system [J]. International Transactions on Electrical Energy Systems, 2019, 29(6): e12022.

[12] Tang Xianlun, Du Yiming, Liu Yuwei.: Image recognition with conditional deep convolutional generative adversarial network[J]. Acta Automatica Sinica, 2018, 44(5): 855-864.

[13] Tan Bendong, Yang Jun, Lai Qiupin.: Data augment method for power system transient stability assessment based on improved conditional generative adversarial network[J]. Automation of Electric Power Systems, 019, 43(01): 149-160.

[14] Yang Yulian, Qi Linhai, Wang Hong, et al. Missing Reconstruction of Measurement Data Based on Generative Adversarial Network and Double Semantic Perception in Distribution Network[J/OL]. Automation of Electric Power Systems, http://doi.org/10.7500/AEPS20190605007.

[15] Liu Yunpeng, Xu Ziqiang, He Jiahui, et al. Data Augmentation Method for Power Transformer Fault Diagnosis Based on Conditional Wasserstein Generative Adversarial Network [J]. Power System Technology, 2020, 44(04): 1505-1513.

[16] Arjovsky M, Chintala S, Bottou L.: Wasserstein GAN [J]. arXiv preprint arXiv: 1701. 07875, 2017.

[17] Yeh R A, Chen C, Lim T Y.: Semantic image inpainting with deep generative models [C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017.

[18] Huang Y, Lai D.: Hidden node optimization for extreme learning machine [J]. Aasri Procedia, 2012, 3: 375-380.