Power system transient stability assessment based on rough set and deep residual neural network

Jian Chai¹, Xihuai Wang² and Jianmei Xiao¹*

¹,²School of Logistics Engineering, Shanghai Maritime University, Shanghai,201306, China
*Corresponding author’s e-mail: jm Xiao@shmtu.edu.cn

Abstract. Machine learning algorithms have been widely used in power system transient stability evaluation. The combined application of data analysis and evaluation and neural network provides a new direction for power system transient stability analysis. After the actual power grid is running, there is obviously an imbalance between stable samples and unstable samples. The current deep learning network realizes the power system transient stability assessment method with too many redundant attributes, and the characteristics will inevitably be lost during the data transmission process. This leads to serious problems with the tendency of the training of the data-driven transient stability assessment model. The rough set theory algorithm is introduced to reduce the redundant attributes of power system transient data sets, which simplifies the difficulty of data training. At the same time, as the neural network deepens, the deep residual neural network model has a higher accuracy rate and effectively avoids the "gradient explosion" and "gradient dispersion" problems. Compared with the traditional neural network, it has better Evaluate performance.

1. Introduction
With the rapid development of the power system, the trend of smart grid and regional interconnection is becoming more and more rapid. Large-scale complex and volatile new energy sources are connected to the power system, making the power system bear greater pressure. According to research[1], power angle instability caused by large disturbances in the power system is the main cause of large-scale power outages. Traditional transient stability assessment methods are time-domain simulation and direct methods, and machine learning technology based on data-driven processing is becoming more and more advanced, and this method is not affected by the complexity of the model, and it is more and more popular by current power system transient stability analysis researcher. The shallow learning method completes the mapping of the system state parameters to the transient stability by learning the input training samples, and realizes the transient evaluation of the power system[2]. Such as using artificial neural network(ANN), support vector machine(SVM), decision tree(DT), random forest(RF), etc. However, the shallow network model has limitations for the input feature processing, and the generalization ability of complex classification problems is greatly restricted. Therefore, the shallow machine learning method cannot well realize the transient stability assessment of the power system. In recent years, deep learning research has made major breakthroughs and has been gradually applied to power system transient stability assessment. However, the power system is a very complex nonlinear system, and the topological structure of its internal electrical components is also extremely complex, and there is a lot of redundancy in the coupling information between variables[3].
Therefore, to solve the above problems, a rough set and deep residual neural network power system transient stability assessment method is proposed.

2. Introduction to Rough Sets

2.1 Rough set theory
When dealing with a large amount of fuzzy and uncertain data information, using rough set theory, after constructing the initial transient feature set, the data information is analyzed and researched, and further knowledge reduction is carried out[4]. The set $U \neq \emptyset$ constituted by the researched objects is called the universe of discourse; The set $A$ composed of a finite number of state attributes is called the attribute set; Attribute range set $V = \{V_1, \ldots, V_n\}$, where $V_i$ is the range of $A_i$; Information function: $U \times A \rightarrow V, f(x, A_i) \in V_j$. An information system is composed of the quadruple $IS = \{U, A, V, f\}$ composed of the universe $U$, the attribute set $A = C \cup D$, the attribute value range set $V$ and the information function $f$. Subsets $C$ and $D$ are condition attributes and decision attributes respectively.

2.2 Rough set attribute reduction algorithm
Every time the importance of all attributes is calculated, if the importance of this attribute for the current condition set is 0, it can be deleted. Then select the most important attributes to add to the current set of conditions, and proceed to the next cycle. Until the conditional information entropy of the current condition set is equal to the conditional information entropy of the original data[5]. Explain that the current collection can replace the original collection. This function is expressed as follows:

$$H(p) = -\sum_{i=1}^{n} p(x_i) \log(p(x_i))$$  \hspace{1cm} (1)

Let $R$ be an equivalence relation, like $\exists r \in R$, have $\text{ind}(R) = \text{ind}(R \setminus \{r\})$, Then $r$ is unnecessary, call a reduction of $R$, the intersection of all reduction sets of $R$ is called $R$'s core; The importance of the reduction is $\text{sig}(\alpha, R) = |\text{ind}(R \cup \{\alpha\})| - |\text{ind}(R)|$. The reduction of rough set is to keep the U condition of the universe unchanged, find the decisive attribute, delete the redundant attribute, and achieve the simplification of the system information.

3. Introduction to Deep Residual Neural Networks
The deep residual neural network is an improved convolutional neural network proposed by He Kaiming’s team in 2015[6]. It skips the middle network layer and passes the previous activation value directly to the later network to avoid gradient explosion and the gradient disappears and a deeper network is trained. The deep structure can link neural network layers that are "closer" and "earlier" to the input layer to achieve identity mapping, so that gradient parameters are effectively transmitted, and the difficulty of training network parameters is greatly reduced, so that higher accuracy can be trained Deep network model.

3.1 Deep residual network structure
Available from Document, the overall network structure includes input layer, convolutional layer, multiple deep residual modules, batch normalization layer, Activation function, average pooling layer and output fully connected layer. From the related introduction in literature [7] The overall network structure is shown in the figure 1:
The input layer, receive the incoming transient feature set that has been processed by rough set, and the input parameter is the 32-dimensional data of the combined feature variable. In this paper, the initial processing of the data set input to the model is carried out by normalization to make the neural network converge faster. The conversion function is:

\[ x^* = \frac{x - \text{min}}{\text{max} - \text{min}} \]  

Convolutional layer, contains multiple convolution kernels, extracts input features, and updates and trains weights and biases. The deep residual neural network greatly reduces the amount of training parameters. The operation process of the convolutional layer is:

\[ y_j = \sum_{i \in M} X_i * K_{ij} + b_j \]  

Batch standardization layer, as much as possible to reduce the difference between the data sample and the characteristics of each dimension. After traditional input data is processed with a series of weights, paranoia and activation operations, it cannot guarantee the standardization of implicit feature quantities. Preprocessing the data before the hidden layer through the batch normalization layer accelerates the optimization of the neural network [8]. In the average pooling layer, the average value is selected as the final feature in the pooling area, and the average pooling is performed in this area to reduce the dimensionality. Improve the anti-interference ability of the model. The calculation process is:

\[ H_i = p(H_{i-1}) \]  

The output fully connected layer, through the activation function, converts the incoming data of the previous layer into probability. \( y_{\text{pred}} \) defines the two-class threshold value \( \delta \), the default is 0.5, and the model output probability is:

\[ y_{\text{pred}}(i) = \begin{cases} 0, & \overline{Y}(i) < \delta \\ 1, & \overline{Y}(i) \geq \delta \end{cases} \]  

Among them, 1 means the system is stable, and 0 means the system is unstable.

3.2 Residual block unit

The residual network is a sub-network, which can form a deep network through stacking. Check the research theory of He Kaiming’s team, refer to the literature [9], The residual neural network layer is shown in the figure 2:
Through jump connection, copy $a^{[l]}$ directly to the deep layer of the neural network. No longer passing along the main path, the transfer formula is

$$a^{[l+2]} = g(a^{[l+2]} + a^{[l]})$$

In the back propagation, the gradient consists of two parts, where the direct mapping gradient to $x$ is 1, and the other part is the gradient of the multi-layer ordinary neural network mapping; assuming the error is $\varepsilon$, the partial derivative of it is:

$$\frac{\partial \varepsilon}{\partial x} = \prod_{l=1}^{L-1} \lambda_i \sum_{i=1}^{L-1} F(x_i, w_i)$$

Assuming that the loss function is loss, the backpropagation formula is:

$$\frac{\partial \text{loss}}{\partial x} = \frac{\partial \text{loss}}{\partial x} \frac{\partial F(x_i)}{\partial x} = \frac{1}{\prod_{i=1}^{L-1} \lambda_i} \sum_{i=1}^{L-1} F(x_i, w_i)$$

The In-depth residual infrastructure diagram is shown in the figure 3.

Through the "shortcut connections" method, the input $x$ is directly passed to the output as the initial result, and the output result is:

$$H(x) = F(x) + x$$

When $F(x) = 0$, then $H(x) = x$, and the identity mapping is realized.

4. Case analysis

4.1 Initial feature construction

Time-domain simulation was performed on a 10-machine 39-node system in New England. The test system contained 39 buses, 46 transmission lines, and the system frequency was 60 Hz. Consider the topological structure and N-1 fault operation mode in the simulation process. With 5% as the step size, set a total of 10 different load levels from 80% to 125%, and at the same time change the generator output accordingly. After the fault is cleared, the lines overlap and the system topology remains unchanged. The simulation duration is set to 5s. Transient stability assessment is a binary classification problem.
4.2 Transient stability assessment process
In [10], The confusion matrix can describe the classification performance of the model for test data with known labels. In the confusion matrix, $T_S$, $T_{US}$, $F_S$ and $F_{US}$ represent the number of stable samples correctly classified, the number of unstable samples correctly classified, and the number of misclassifications for stable samples and unstable samples. The number of misclassifications for predicting stable samples as unstable samples.

| Predicted label | Actual label | Stabilize | Instability |
|-----------------|--------------|-----------|-------------|
| Stabilize       | $T_S$        | $F_S$     |
| Instability     | $T_{US}$     | $F_{US}$  |

4.3 Model performance comparison
After inputting the transient feature set into the model training, the final classification result is shown in the figure 4 below.

![Schematic diagram of result classification](image)

Table 2. Evaluation and comparison of each model.

| Model | Accuracy/% | Recall rate/% | precisions/% | $F1$ value |
|-------|------------|---------------|--------------|------------|
| DNN   | 96.55      | 90.26         | 96.98        | 95.62      |
| DT    | 96.10      | 81.64         | 95.89        | 95.23      |
| SVM   | 95.63      | 87.88         | 96.16        | 98.09      |
| CNN   | 96.78      | 98.11         | 96.23        | 97.20      |
| DRNN  | 98.98      | 98.33         | 98.65        | 98.89      |

5. Conclusion
1) The proposed rough set attribute reduction algorithm can not only remove redundant information data between the topological structure of the power system, and reduce the impact of redundant information on the model; at the same time, it can also strengthen and re-characterize the input features to make the provided features more sufficient.
2) Compared with deep neural networks such as CNN, DRNN can improve the model's practice feature extraction ability, and it still has a higher accuracy rate in the case of insufficient and excessive samples. The generalization ability of the model is strong.
References

[1] Liu, Y. (2018) Analysis and Enlightenment of Disconnection and Outage of Brazil’s "9.13" Far Northwest Power Grid. J. Proceedings of the Chinese Society for Electrical Engineering, 38(11): 3204-3213.

[2] Yu, Y., Y. Yang, Q., W. (2007 (2018 reprint)) Power System Analysis. China Electric Power Press Publishing. Beijing:

[3] Odun Ayo, T. Crow, M, L. (2012) Structure-Preserved Power System Transient Stability Using Stochastic Energy Functions, J. IEEE Transactions on Power Systems27(3):1450-1458.

[4] Su, F. Yang, S., H. Wang, H., Y. (2017) Research on Fast Termination Algorithm for Time Domain Simulation of Power System Transient Stability. J. Proceedings of the CSEE.37(015):4372-4378.

[5] Wang, T. Liu, J., L. Zhu, S., X. Wang, Z., P. (2020) Power system transient stability assessment and emergency control strategy based on random forest. J. Power System Technology, 44(12): 4694-4701.

[6] Song, L. (2020) An Integrated Scheme for Online Dynamic Security Assessment Based on Partial Mutual Information and Iterated Random Forest. J. IEEE Transactions on Smart Grid11(4):3606-3619.

[7] Zhang, K. (2017) Overview of rough set theory and its applications. J. Intelligent Processing and Applications.6:93-98

[8] Li, B., Y. Xiao, J., M. Wang, X., H. (2020) Power System Transient Stability Assessment Integrating Neighborhood Rough Reduction and Deep Forest. J. Transactions of China Electrotechnical Society35(15): 3245-3257.

[9] Zhou, Y., Z., Zha, X., Lan, J. (2020) Power system transient stability prediction based on data augmentation and deep residual network. J. China Electric Power (001):22-31.

[10] He K M, Zhang X Y, Ren S Q, et al. (2016) Deep residual learning for image recognition. J. IEEE Computer Society:770-778.