Power Grid Fault Diagnosis Method Based on Stacked Sparse Denoising Auto-Encoder and GRU Network

Ling Zheng, Panfei Yan and Fei Chen

Department of Computer Technology, Faculty of Control and Computer Engineering, North China Electric Power University, No.2 Beinong Road, Huilongguan Town, Changping District, Beijing, China
Email: 810781500@qq.com

Abstract. This research aims to solve the problems of feature extraction, data dimension reduction, gradient dissipation and low training efficiency in power grid fault classification. Aiming at PMU measurement data with high dimensionality and strong real-time characteristics, a fault diagnosis method based on Stacked Sparse Denoising Auto-Encoder and GRU network is proposed. The Stacked Sparse Denoising Auto-Encoder is used to reduce the sequence dimension to obtain the sparse feature expression of the data, and then the time-dependent features of the data extracted by the GRU are used to obtain the fault type. Simulations and experiments show that compared with the traditional neural network algorithm, the proposed method can effectively extract high-dimensional data features, reduce data dimensions, improve the efficiency of GRU network classification, accelerate the convergence speed and reduce the training time, and has better stability and higher accuracy.

Keywords. PMU measurement data; stacked sparse noise reduction autoencoder; GRU; power grid fault diagnosis.

1. Introduction
With the rapid development of artificial intelligence technology, many new and more effective methods have been provided for power grid fault diagnosis. Artificial neural network is a non-linear, adaptive information processing system composed of a large number of interconnected processing units. It has the characteristics of fast execution speed, good robustness, and strong learning function. Scholars at home and abroad have done a lot of research in this direction [1]. Ref. [2] uses genetic algorithm to calculate the initial value of network parameters on the basis of BP neural network, find the optimal number of hidden layer nodes, and optimize the BP neural network to solve the problem of fault diagnosis. Ref. [3] proposed the application of a new Radial Basis Function (RBF) neural network to solve the problem of fault diagnosis and extended the Orthogonal least square algorithm to optimize RBF neural network parameters. Although the above methods alleviate the problems of local optimization and gradient dispersion faced by neural networks during fault diagnosis, there are still problems such as long training time, slow convergence speed, and poor stability. When applied to the multi-class fault diagnosis problem of time series PMU measurement data, the parameters are difficult to determine and the classification efficiency is low.

According to the advantages and disadvantages of these diagnostic methods, it also aims at the shortcomings of existing neural networks in power grid fault diagnosis [4]. This paper proposes a grid fault diagnosis method based on Stacked Sparse Denoising Auto-Encoder and GRU network. After experimental comparison, While the data dimension is reduced, the characteristics of time series data...
are also fully expressed. At the same time, the model has good robustness and anti-noise performance, which can effectively improve the classification efficiency, reduce the model complexity, and make the classification results more accurate, reasonable and stable. It can provide effective reference for fault diagnosis of dispatch center.

2. Power Grid Fault Diagnosis based on SSDAE and GRU Networks

2.1. Sparse Denoising Auto-Encoder Network

In traditional Auto-Encoders, in general, the number of hidden layer neurons is larger or even exceeds the input dimension, and there is only one hidden layer. The reconstruction ability of the input data is limited, and the ability to extract data features is poor [5]. The high-dimensional and sparse expression is good. Sparse Auto-Encoder (SAE) is obtained by adding some sparseness constraints on the basis of the traditional Auto-Encoder, so that most neurons in the hidden layer are in a suppressed state. Only a few are activated, the data can be effectively transferred in the neural network, so as to obtain a more effective and advanced expression of the input data [6]. Sparse Denoising Auto-Encoder (SDAE) is based on SAE, adding noise to the input data, forcing the encoder to learn to extract the most important features and learn more robust representations in the data. The SDAE network structure is shown in figure 1.

Figure 1. Structure diagram of sparse denoising auto-encoder.

The calculation form of the loss function of the DAE is equation (1):

\[ J(x, y) = \frac{1}{2n} \sum_{i=1}^{n} \| o_i - x_i \|^2 + \frac{\lambda}{2n} \sum_{j=1}^{n} w_j^2 \]  

(1)

In the equation, the first term is the autoencoder loss function, \( o_i \) is the output data, \( x_i \) is the input data, the second term is the regularization term, \( \lambda \) is the regularization parameter, and \( w_j \) is the weight matrix.

The Sparse Auto-Encoder is to add sparseness constraints to each layer of the hidden layer, so that most nodes are constrained to zero, and only a few are not. When the output of the neuron is close to 1, the neuron is considered to be activated. When approaching 0, the neuron is suppressed [7]. The calculation form of the average activation value of hidden neurons is equation (2):
In the equation, \( h_j(\tilde{x}) \) represents the activation value of hidden neuron \( j \).

Adding a penalty factor to SDAE can keep the average activation value of hidden neurons in a small range. The penalty factor is calculated as equation (3):

\[
KL(\rho||\hat{\rho}_j) = \rho log \frac{\rho}{\hat{\rho}_j} + (1 - \rho) lg \frac{1 - \rho}{1 - \hat{\rho}_j}
\]

In the equation, \( \rho \) is a sparse parameter, usually a small value close to 0 (generally \( \rho = 0.05 \)), \( KL(\rho||\hat{\rho}_j) \) is to measure the difference between \( \rho \) and \( \hat{\rho}_j \), training in the network. In the process, if \( \hat{\rho}_j \) and \( \rho \) are different, they will be punished to suppress the hidden layer neurons [8].

The calculation form of SDAE’s loss function is equation (4).

\[
L(x,y) = J(x,y) + \beta \sum_{j=1}^{k} KL(\rho||\hat{\rho}_j)
\]

In the equation, \( J(x,y) \) is the loss function of the noise reduction autoencoder, and \( \beta \) is the weight of the sparsity penalty factor.

2.2. Grid Fault Diagnosis Model based on SSDAE and GRU Network

Long short-term memory model (LSTM) belongs to a type of Recurrent Neural Network (RNN). It uses a special three-gate structure to control the flow and operation of information. It is suitable for processing time-series data and can be used to solve traditional algorithms. Learning long-term feature relationships and gradient dispersion [9]. The GRU network is a very effective variant of the LSTM network. The number of parameters is less than that of the LSTM, and the structure is simpler. On the whole, the training speed of the GRU is faster than that of the LSTM. As shown in figure 2, by combining the advantages of SSDAE and GRU, this paper first uses SSDAE to extract the depth features of each dimension of the measurement data, reducing the dimension while obtaining less and more effective data information, and then using GRU to capture the time dependence of the dimensionality reduction data Features, and finally combined with SoftMax classifier to obtain fault classification.

![Figure 2. SSDAE-GRU power grid fault diagnosis model.](image-url)
3. Implementation and Analysis of Power Grid Fault Diagnosis Algorithm

3.1. Algorithmic Data Collection and Data Preparation
The experiment is based on the 39 Bus New England System network in Power Factory to collect simulation data. The experiment selects 7 kinds of fault state data for each 3s time length as the experimental samples, and obtains the experimental sample information as shown in Table 1. Among them, each fault state data randomly selects 80% of the samples for model training and 10% of the samples for the model. For verification and adjustment, 10% of the samples are used for model testing.

| Sample failure type                              | Sample length | Number of samples | Fault label |
|--------------------------------------------------|---------------|------------------|-------------|
| 3-phase short circuit fault                      | 800           | 18000            | 0           |
| 2-phase short circuit fault                      | 800           | 18000            | 1           |
| Single ground fault                              | 800           | 18000            | 2           |
| Two ground faults                                | 800           | 18000            | 3           |
| Multiple devices fail at the same time           | 800           | 54000            | 4           |
| Cascading failure                                | 800           | 36000            | 5           |
| Circuit breaker refuses to operate               | 800           | 18000            | 6           |

3.2. Sparseness and Robustness Analysis
In order to verify the effect of different degrees of sparse constraints and noise on the accuracy of the model, the experimental sparseness parameter $\rho$ was selected 0.01, 0.05, 0.1, 0.15, 0.2, respectively, and the noise ratio was selected 0%, 10%, 20%, 30%, 40%, the average value of each type of experiment is taken as the final accuracy rate. The experimental results are shown in Figure 3.

![Figure 3](image_url)

**Figure 3.** The effect of different degrees of sparse constraints and noise ratio on the classification accuracy.

As shown in Figure 3, under the same sparse constraint, the accuracy of classification tends to be stable and then decreases with the increase of the noise ratio, which shows that adding a certain proportion of noise will make the model more robust, but a high proportion of noise ratio some information will be lost. Under the same noise, as $\rho$ increases, the accuracy of classification shows a downward trend, which indicates that as $\rho$ increases, the sparse constraints of the model gradually
decrease, and the model cannot effectively extract the deep abstract features of the data. Too high, the extracted feature expression is incomplete. Therefore, choosing appropriate sparsity parameters and adding an appropriate noise ratio can make the model more robust, more robust, and better in classification.

3.3. Performance Comparison and Analysis
It can be seen from the data in table 2 that when processing time series data, LSTM and GRU have unique advantages compared with traditional BP neural networks and SVM, but there is a problem of the length of iteration. Compared with other networks, SSDAE-Softmax has a fast iteration speed but its accuracy has not been improved. Compared with LSTM and GRU, SSDAE-LSTM and SSDAE-GRU not only greatly speed up the iterative speed of the model but also improve the classification accuracy, which shows that SSDAE can not only make the abstract features of the depth of extraction more effective expression it can also effectively reduce the dimension to obtain less and more effective information. Although the difference in classification effect between SSDAE-GRU and SSDAE-LSTM is not very large, due to the relatively large training overhead of the LSTM network and the simplified structure of GRU, the network parameters are less effective and the iteration speed is improved. Therefore, the SSDAE-GRU model is superior to other networks in terms of accuracy and iteration speed when processing PMU measurement data, and has stronger robustness and stability.

|                          | Training accuracy/% | Test accuracy/% | Time required for one iteration/s |
|--------------------------|---------------------|----------------|----------------------------------|
| **BP**                   | 95.37               | 92.56          | 140.21                           |
| **SVM**                  | 93.80               | 91.08          | 121.54                           |
| **LSTM**                 | 95.74               | 95.51          | 208.50                           |
| **GRU**                  | 95.84               | 95.13          | 177.14                           |
| **SSDAE-Softmax**        | 94.97               | 92.35          | 33.69                            |
| **SSDAE-LSTM**           | 96.87               | 96.34          | 126.8                            |
| **SSDAE-GRU**            | 97.38               | 96.12          | 65.43                            |

4. Conclusion
This paper proposes a grid fault diagnosis model that combines a stack-type noise reduction autoencoder and a GRU network. This model adds a certain proportion of noise and sparse constraints on the basis of the autoencoder to extract more representative deep layers of PMU measurement data. Abstract features reduce the dimensionality of the measurement data sequence and provide GRU with initial weights and hidden layer offsets. Experiments and analysis show that the algorithm proposed in this paper can effectively extract the deep abstract features of PMU measurement data, reduce the complexity of GRU classification, and have stronger robustness and generalization ability. On high-dimensional data sets, fault classification higher efficiency. Compared with the traditional BP neural network and SVM, it overcomes the problems of traditional network feature extraction, data dimensionality reduction, gradient dissipation, and low training efficiency. At the same time, it can provide an effective reference for the fault diagnosis of the dispatch center.

References
[1] Zhang X, Wei J, Zhao D, Zhang D and Liu Y 2013 Research history and prospect of power grid fault diagnosis Grid Technology 37 2745-53.
[2] Zhang X, Guo X, Yu Y, Chen J, Wang J and Peng B 2019 A power grid fault diagnosis method based on genetic algorithm optimized BP neural network Science and Technology Innovation and Application 29 15-17.
[3] Bi T, Ni Y, Wu F and Yang Q 2002 Power network fault diagnosis method based on new neural network Chinese Journal of Electrical Engineering (02) 74-79.
[4] Guo Y, Ran C, Ji X and Ma J 2013 Analog circuit fault diagnosis based on combinatorial optimization BP neural network Journal of Northwestern Polytechnical University 31 (01) 44-48.

[5] Vincent P, Latouchelle and Lajoie I, et al. 2010 Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion The Journal of Machine Learning Research 11 (12) 3371-3408.

[6] Qiu Y 2018 Epilepsy EEG Classification Based on Sparse Denoising Autoencoder (Jinan: Shandong University).

[7] Chen R, Huang X and Yang L, et al. 2017 Bearings life state recognition using deep sparse auto-encoder neural network with noise adding sample expansion Journal of Vibration Engineering 30 (5) 874-882.

[8] Zhang X and Qi L 2019 Study on classification of distribution network station area based on sparse noise reduction self-coding and clustering algorithm Power Information and Communication Technology 17 (12) 15-23.

[9] Cheng Q, Chen Z, Zhang Y, Jiang W, Liu X and Liang S 2020 Research on fault diagnosis of solar photovoltaic modules based on CNN-LSTM Application of Electronic Technology 46 (04) 66-70.