Transformer Fault Diagnosis Based on Optimized CPSO-BP Neural Network

Jiyan Zou 1, a, Quan Liang 1, Xiaoming Xu 2, Qiang Wang 3

1 School of Electrical Engineering, Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian, Liaoning, China
2 School of Software Engineering, Faculty of Software Engineering, Dalian University of Technology, Dalian, Liaoning, China
3 Shenyang State Grid Corporation of Liaoning Province, China

a jyzou@dlut.edu.cn

Abstract. Oil chromatographic analysis (DGA) is an important way to transformer fault diagnosis, combining research topics based on I-K-means clustering, t SNE visual clustering in data mining, fault classification number. In order to improve the convergence speed of neural network, an improved back-propagation BP neural network using ADAM gradient optimization algorithm instead of traditional stochastic gradient descent optimization to update the weight of neural network is proposed. Fuzzy C-means clustering and particle swarm optimization are proposed to optimize the initial parameters of neural network. By using 3500 data samples of transformers from a power plant in a city of Liaoning Province to carry out simulation experiments, and comparing the traditional BP network algorithm, CPSO-BP network algorithm and the CPSO-BP network algorithm optimized by Adam, it is proved that the CPSO-BP network optimized by Adam has fast training convergence, strong generalization ability and high accuracy. At the same time, the accuracy, precision, recall and F-1 values were used to evaluate the CPSO-BP network algorithm optimized by ADAM to verify the effectiveness and stability of the algorithm in transformer fault diagnosis.

1. Introduction

With the development of artificial intelligence, scholars have proposed a variety of transformer fault diagnosis algorithms based on oil chromatography, such as support vector machine method, fuzzy strategy method, grey correlation method, expert system and neural network. To a certain extent, the traditional oil chromatography analysis of transformer fault diagnosis method of low accuracy, poor generalization and other shortcomings, directly or indirectly improve the accuracy of transformer fault diagnosis. Among them, BP neural network is widely used in transformer fault diagnosis with the ability of self-adaptation, self-organization, self-learning and nonlinear approximation [1], however, the selection of initial weights and threshold parameters in BP neural network is random, this will slow down the convergence speed of BP neural network and lengthen the training time. To solve these problems, scholars have proposed a variety of algorithms to optimize BP neural network. Chaopeng Song [2] puts forward a transformer fault diagnosis algorithm based on improved particle swarm optimization BP neural network, and the accuracy of fault diagnosis is improved obviously. Shan Luo
[3] calibrates the fitness value of the genetic algorithm, improves the adaptive crossover probability and mutation probability, and improves the optimization ability of the genetic algorithm, so as to optimize the initial weight threshold of the BP neural network. Zeming Lun [4] used the improved chaotic genetic algorithm to optimize the neural network model. The improved algorithm model has higher accuracy in predicting short-term traffic flow. XuMeng Dou [5] proposed an APSO-BP neural network fault diagnosis model, with an accuracy rate of 95%. Jiatang Cheng [6] used the quantum particle algorithm to obtain the weight and threshold optimization parameters of BP neural network, and the fault diagnosis accuracy reached 91.67%. Although these optimization algorithms have achieved good results, there are still some shortcomings, such as the increase of computation, easy premature convergence, low search efficiency.

On the basis of these studies, this paper combines BP neural network improved by gradient optimization with fuzzy C-means clustering particle swarm optimization algorithm, and applies this algorithm to transformer fault diagnosis. The experimental results show that the optimized CPSO-BP neural network algorithm can improve the accuracy of transformer fault diagnosis, and the gradient optimization can reduce the training time. Meanwhile, the algorithm has high accuracy, strong stability and good classification effect.

2. BP neural network

2.1. Principle of BP neural network
BP neural network, one of the most widely used neural network models, is a multi-layer feedforward network trained by error reverse algorithm [7]. Structurally, BP neural network has input layer, hidden layer and output layer. The number of neurons in the input layer is the same as the dimension of the input data, and the number of neurons in the output layer is the same as the number of data to be fitted. The hidden layer can be expanded into multiple layers, and the number of neurons and layers in the hidden layer are set according to the actual situation and experience. There is no connection between the same layer, but the front and rear layers are fully connected. Such an architectural arrangement ensures the information processing ability of BP neural network.

2.2. Improvement of BP algorithm
The algorithm used in the process of reverse feedback of BP neural network is usually summarized as the training algorithm of BP neural network. The iterative update of weights and thresholds in reverse feedback directly determines the training speed, convergence and training accuracy. Gradient optimization algorithm is a method to obtain the optimal value of parameters in the defined model during the training process. At present, the popular gradient optimization algorithms include stochastic gradient descent (SGD) algorithm, Adagrad algorithm, RMSProp algorithm, Adadelta algorithm and Adam algorithm [8].

SGD algorithm is the most basic gradient optimization algorithm, but when the sample data type is sparse data or the features are not obvious, SGD algorithm is easy to converge to local optimal [9]. The RMSProp algorithm performs well in processing non-convex function trained neural networks, and the computation is simplified, but the learning trajectory may converge to a local convex saddle point. Adam algorithm can be seen as the revised RMSProp algorithm with momentum, is essentially a momentum item RMSProp, it is estimated based on the first moment of the gradient and the second order moment to estimate the dynamic adjustment of each parameter vector, after threshold correction, the vector each iteration has a certain scope, makes the parameter is more stable, solved the need to determine the initial momentum factor and fall into the local optimal problem, improve the strength of the neural network training, its formula is as follows:

$$\begin{align*}
m_t &= \mu \cdot m_{t-1} + (1-\mu) \cdot g_t \\
n_t &= \nu \cdot n_{t-1} + (1-\nu) \cdot g_t^2
\end{align*}$$

(1)
When Adam gradient descending method is used to improve BP algorithm, the convergence speed is faster and the learning effect is more effective. Moreover, it can correct the problems existing in other optimization techniques, such as the disappearance of learning rate, too slow convergence or the large fluctuation of loss function caused by parameter update with high variance. To a certain extent, it effectively makes up for the deficiency of neural network.

3. Principle of CPSO

Combining Particle Swarm Optimization (PSO) with neural network, the global search efficiency and convergence speed of neural network can be greatly improved through continuous Optimization of initial weights and thresholds. The improved fuzzy C-means clustering particle swarm optimization algorithm has stronger global optimization ability, and the initial weights and thresholds of BP neural network are optimized to improve the accuracy of transformer fault diagnosis. In the particle swarm optimization algorithm, each particle is a solution in the solution space.

The particle swarm optimization algorithm determines its own optimal position through individual optimal solution and global optimal solution. The updated formula is [10]:

\[
\begin{align*}
\vec{v}_{iD}^{t+1} &= \omega \cdot \vec{v}_{iD}^{t} + c_1 \cdot r_1 \cdot (\vec{p}_{iD}^{t} - \vec{x}_{iD}^{t}) + c_2 \cdot r_2 \cdot (\vec{g}_{iD}^{t} - \vec{x}_{iD}^{t}) \\
\vec{x}_{iD}^{t+1} &= \vec{x}_{iD}^{t} + \vec{v}_{iD}^{t+1}
\end{align*}
\]

(2)

\[
\begin{align*}
\vec{p}_{iD}^{t+1} &= \begin{cases} 
\vec{p}_{iD}^{t} & f(\vec{x}_{iD}^{t+1}) \geq f(\vec{p}_{iD}^{t}) \\
\vec{x}_{iD}^{t+1} & f(\vec{x}_{iD}^{t+1}) < f(\vec{p}_{iD}^{t})
\end{cases}
\end{align*}
\]

(3)

\[
g_{iD}^{t} = \min \{ f(\vec{p}_{iD}^{t}), f(\vec{p}_{iD}^{t+1}),..., f(\vec{p}_{iD}^{t}) \}
\]

(4)

In the process of optimizing PSO algorithm, the fuzzy C-means clustering algorithm is a process of iteratively calculating the membership degree and cluster center of each output optimal particle. In this process, the fitness function is constantly minimized, that is, the Lagrange multiplier method is used to solve the unconditional extremum, the membership function and cluster class center searching function are obtained by taking the derivative of the membership degree and cluster class center respectively:

\[
J = \sum_{i=1}^{N} \sum_{j=1}^{N} u_{ij} (x_j - c_j)^2
\]

(5)

\[
\begin{align*}
\frac{\partial J}{\partial \vec{c}_i} &= \sum_{j=1}^{N} \sum_{i=1}^{N} \mu_{ij}^{\tau} (x_j - c_j)^2 c_i = 0 \quad \Rightarrow \quad c_j = \frac{\sum_{i=1}^{N} u_{ij} x_i}{\sum_{j=1}^{N} u_{ij}} \quad (6)
\end{align*}
\]

\[
\max_{\varepsilon} = \left\{ \left| u_{ij}^{t+1} - u_{ij}^{t} \right| < \varepsilon \right\}
\]

(7)

The implementation process of the optimized CPSO-BP neural network algorithm is as follows:

Step 1: Vectorize the weight of BP neural network as each individual particle in the particle swarm;

Step 2: Initialize each parameter and fitness function of particle swarm, learning factor c1, c2, Random number r1, r2,; The velocity and position of the j dimension of the particle at time t, and the mean square error function of the network is set as the fitness function of the particle;

Step 3: Update according to the PSO algorithm, and record the optimal particle individuals after the update; Repeat steps 2 and 3 for N times, and get N individual particles.

Step 4: Substituting N individual particles into the fuzzy C-means clustering algorithm to calculate the fitness function, and iterating. After the iteration, the output multi-dimensional vector results were used as the initial weights and thresholds of the neural network.

Step 5: BP neural network was created by using the weights and thresholds calculated in Step 5, each parameter was initialized, and data of m transformer oil chromatographic samples were extracted from the training set.
Step 6: Calculate the gradient for the target of the loss function as follows:

$$g_t = \frac{1}{m} \sum_{i=1}^{m} [L(f(X^i; \theta_t), y_i)]$$  \hspace{1cm} (8)

Step 7: Updating step size, calculating gradient first order matrix, second order matrix;
Step 8: Correct the first-order matrix and second-order moment. The specific calculation formula is as follows:

$$\begin{cases}
\hat{m}_t = m_t \frac{m_t}{1 - \mu_t} \\
\hat{v}_t = \frac{v_t}{1 - \nu_t}
\end{cases}$$  \hspace{1cm} (9)

Step 9: Update the parameters and update the formula as follows:

$$\Delta \theta_i = \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} \eta$$

$$\theta_i = \theta_{i-1} + \Delta \theta_i$$  \hspace{1cm} (10)

Step 10: Correct the connection weights and thresholds of each layer of BP neural network. When the training of the neural network reaches the stop criterion, that is, the final error is less than or equal to the set error, the training process ends. If not, return to Step 7.

4. Conclusion

As shown in Fig.1, the X-axis is the number of training cycles, the Y-axis is a numerical value. The blue line represents the multi-classification cross entropy of the test data set, the green line represents the average multi-classification cross entropy of each cycle, the orange line represents the diagnostic accuracy of the test data set, and the red line represents the average accuracy of each cycle.

With the increase of network training cycle, the accuracy rate of the test set diagnosis model increased rapidly in the second cycle, and the multi-classification cross entropy value decreased rapidly. The accuracy rate reached 96.2% in the fourth cycle, and the model rapidly converges and tends to be stable in the following training cycle. At the end of the training, the diagnostic accuracy of the test set reached 99.77%, and the multi-classification cross entropy was 0.0109. In the training process, the average multi-classification cross entropy of each cycle also decreases steadily, and the average accuracy of each cycle also increases, indicating that the training effect is good. After the convergence of the model, there is a small period of fluctuation, because the model automatically adds a slight disturbance to the training data set in this training cycle to improve the robustness of the model, reduce the over-fitting, and then improve the generalization ability of the model. After that, the model quickly recovered the high fault recognition rate, and the accuracy rate of the model remained above 80.73%, which proved that the model was not easy to be disturbed, had high stability and strong reliability.

![Fig. 1 Changes of indexes in the training process of CPSO-BP-ADAM neural network](image-url)
Table 1. Comparison of accuracy rates under different neural networks

| Test times | SGD-BP accuracy | RMSProp-BP accuracy | CPSO-BP accuracy | ADAM-BP accuracy | CPSO-BP-ADAM accuracy |
|------------|-----------------|---------------------|------------------|------------------|-----------------------|
| Record 1   | 18.50%          | 98.83%              | 56.90%           | 99.20%           | 99.70%                |
| Record 2   | 18.47%          | 71.30%              | 68.83%           | 98.53%           | 99.77%                |
| Record 3   | 18.50%          | 71.60%              | 71.60%           | 96.83%           | 99.60%                |
| Record 4   | 18.45%          | 56.90%              | 71.30%           | 87.93%           | 99.27%                |
| Average    | 18.48%          | 74.66%              | 67.16%           | 95.62%           | 99.59%                |

This is shown in Table 1. that the optimization effect of ADAM optimization algorithm on BP neural network is better and the stability is higher. The CPSO algorithm has advantages in the process of optimizing the parameters of neural network and can show strong generalization ability in diagnosis, which can effectively improve the stability of transformer fault diagnosis and classification model.

Table 2. Comparison of accuracy rates under different neural networks

| Training cycle | Recall degree of training set | Accuracy of training set | Test set recall | Accuracy of test set | Test set F-1 |
|----------------|-------------------------------|--------------------------|-----------------|----------------------|--------------|
| 1              | 0.214000002                   | 0.774426997              | 0.273333341     | 0.598540127          | 0.375286045 |
| 5              | 0.981666684                   | 0.983962595              | 0.980000019     | 0.983277619          | 0.981636083 |
| 10             | 0.971000016                   | 0.973921776              | 0.980000019     | 0.98989898          | 0.984924628 |
| 15             | 0.994333327                   | 0.995328665              | 0.99333334      | 0.996655524          | 0.994991659 |
| 20             | 0.997333348                   | 0.997665882              | 0.99333334      | 1                    | 0.996655522 |

Classification indexes such as Recall, Precision and F-1 were introduced in this study to comprehensively evaluate the CPSO-BP-ADAM neural network algorithm. The closer the values of the three indexes are to 1, the better the clear effect will be. It can be seen from Table 2. That the performance of the CPSO-BP-ADAM model is continuously enhanced during the training process, and the diagnostic effect of the model on the test set is ideal after the training, and the model has high recognition rate and strong stability after the training.

References
[1] Hong Yu Harbin Engineering University Postdoctoral Workstation of Yunnan Power Grid Corporation Kunming, Intelligent Information Technology Application Research Association, HongKong. 2010 The 2nd International Conference on Industrial Mechatronics and Automation (ICIMA 2010) (Volume2) [C]. Intelligent Information Technology Application Research Association, HongKong: Intelligent Information Technology Application Society, 2010:6.

[2] Song Chaopeng. Application of IPSO-BP Neural Network Algorithm in Fault Diagnosis of Oil-immersed Transformer [J]. Mechanical and Electrical Engineering Technology, 2020, 49(12): 196-198.

[3] LUO Shan. Improved Genetic Algorithm to Optimize BP Neural Network Image Compression [J]. Shanxi Electronic Technology, 2021(01): 56-59.

[4] Lun Zeming. Improved Chaos Genetic Algorithm to Optimize BP Neural Network for Short-term Traffic Flow Prediction [J]. Computer Programming Skills and Maintenance, 2017(05): 18-20.

[5] Dou Xumeng. Research on Transformer Fault Diagnosis Based on APSO-BP Neural Network [D]. Inner Mongolia Agricultural University, 2020.

[6] Cheng Jiatang, Duan Zhimei, Xiong Yan, Ailey. Research on Fault Diagnosis Method of Transformer Based on QPSO-BP Model Based on DGA [J]. High Voltage Appliances, 2016,
52(02): 57-61.

[7] Tianqi Wang, Chao-Kai Wen, Hanqing Wang, Feifei Gao, Tao Jiang, Shi Jin. Deep Learning for Wireless Physical Layer: Opportunities and Challenges [J]. China Communications, 2017, 14(11): 92-111.

[8] LI Xingyi, YUE Yang. Overview of Gradient Descent Algorithms [J]. Software Engineering, 2020, 23(02): 1-4.