Research on high performance computing of power system based on machine learning algorithm

Jilin Chen[1]
China Electric Power Research Institute
East Road of Xiaoying 15#, Haidian District ,Beijing
chenjilin@epri.sgcc.com.cn

Yong Chen[2]
Platform Research Office of power system department
China Electric Power Research Institute
East Road of Xiaoying 15#, Haidian District ,Beijing

Zhonghua Guo[3]
China Electric Power Research Institute
East Road of Xiaoying 15#, Haidian District ,Beijing

Weijiang Qiu[4]
Platform Research Office of power system department
China Electric Power Research Institute
East Road of Xiaoying 15#, Haidian District ,Beijing

***Abstract*—The change of power system operation mode can cause the change of the power equipment parameters, operation mode of the calculation with the power grid is gradually increasing with the increase of complexity. In order to solve the heavy workload and repetitive tasks , this paper, the machine learning algorithm is used to study and practice the high performance calculation of power grid operation mode. The linear regression algorithm, the logical regression two classification algorithm and the deep neural network algorithm are compared and analyzed, and the performance evaluation index is given.**

Keywords-component; machine learning; power system operation mode;high performance calculation

I. INTRODUCTION

With the rapid development of power grid construction and the significant expansion of power grid scale, the equipment model is becoming more and more complex, the simulation scale is further expanded, and the calculation amount and adjustment content of power grid operation mode are also growing in scale. If the operation mode personnel only rely on manual mode to adjust the trend, the adjustment process will become very cumbersome, and the adjustment work is often carried out according to experience, with low efficiency and certain randomness, so it is difficult to ensure the correctness. How to reduce the complexity and repeat of adjusting the operation mode of power grid has become an urgent problem.

Nowadays, science and technology have made a great breakthrough in how to use computers to deal with big data effectively. For example, the distributed system efficiently stores and manages large quantities of data, and the parallel system greatly accelerates the feasibility of some complex algorithms. Machine learning has made substantial progress with the development of this wave of computing technology, and is now widely used in finance, power, business, market, factory and other major fields. Especially in the field of Internet, it has the most profound influence[1]. For example, Google has developed tensorflow platform, which is an open-source algorithm library using data flow graphs, mainly for machine learning and deep neural network. Its subordinate architecture is flexible, which enables computing needs to be developed on a variety of platforms, including desktop computers, servers, GPU nodes, mobile devices, etc.

Based on the advantages of tensorflow architecture and machine learning algorithm, this paper studies the high-performance calculation of power grid operation mode, and gives the performance evaluation indexes of linear regression, logical regression and neural network algorithm.

II. POWER SYSTEM OPERATION MODE DATA

At present, PSASP software developed by China Electric Power Research Institute is mainly used in the calculation of the operation mode of the national dispatching system. The input of model parameters adopts the traditional parameter card mode, and there is a typical internal logic relationship between model parameters, that is, one characteristic quantity is adjusted, which has a relatively obvious impact on other quantities and results. In a large provincial power grid, the model data is often tens of thousands of lines, and the model parameter card contains dozens of characteristic quantities. The amount of two-dimensional calculation matrix data formed horizontally and vertically is very large. The adjustment of model parameters for business needs, workload and difficulty are also exponential growth. According to the practical application demand of power grid mode calculation, this paper uses deep learning technology to construct empirical value model of mode calculation characteristic parameter adjustment, realizes fast prediction of power flow calculation results after characteristic parameter adjustment based on tensorflow platform, and recommends the best combination of parameter configuration in the second step for mode calculation personnel after providing characteristic parameters in each step, so as to help the calculators of power system improve their work efficiency. Table 1 is part of the document of parameter card generator for parameter input and debugging of power flow calculation by power simulation PSASP software.

This work is supported by Science and Technology Project of State Grid:Research and Development of High Performance Computing Technology for Power System Applications.

978-1-7281-9837-8/20/$31.00 ©2020 IEEE
DOI 10.1109/CIBDA50819.2020.00053

TABLE I. POWER FLOW CALCULATION SOFTWARE PARAMETER

Valid	*Generator name*	*Node name*	*Type*	*P (kW)*	*Q (kw)*	*V*	*Angle (°)*
1	BUS5	BUS5	PQ	4.3	3.34	1	0.21
1	BUS1	BUS1	Slack	1.1	2.08	1	0.45
1	BUS6	BUS6	PV	-0.01	0.37	1	0.25
1	BUS7	BUS7	PQ	6.1	2.6	1	0.31
1	BUS4	BUS4	PV	2.25	3.52	1	0.11
1	BUS8	BUS8	PV	3.1	4.1	1	0.14
1	BUS2	BUS2	PV	3.06	2.12	1	0.32
1	BUS3	BUS3	PV	2.08	1.24	1	0.28

III. ALGORITHM ANALYSIS

A. Two classification model algorithm of logical regression

According to the data of PSASP software parameter model, the applicability of the algorithm is analyzed, and the conclusion is a typical binary classification application scenario of logistic regression. Before carrying out the corresponding research, the first step is to extract the feature quantity (more) from the model file. The second step is to select about 8 general feature quantities of the actual project, then obtain the adjustable range agreed by the feature quantity, establish the model simulation data body, and establish the algorithm model. Based on the practical application of power grid calculation, when the model is put into training, this paper selects the SVM model in the logic regression algorithm to carry out research and experiment. The core idea of classification is to find a hyper-plane to segment the positive and negative samples in a given sample set containing positive and negative examples, and separate the hyper-plane. For a classifier, G (x) > 0 and G (x) < 0 represent two different categories {+ 1, - 1}, which shows that the sample of power grid operation mode is suitable for SVM algorithm[2].

SVM is determined by an important training sample (support vector). Therefore, the SVM classification problem can be described as the constrained optimization problem of linear classification.

$$\min_{w,b} \quad \frac{1}{2}|w|^2 \tag{1}$$

$$s.t. \quad y_i\left(w\cdot x_i+b\right)-1\geq 0 \tag{2}$$

Introducing Lagrange multipliers for every inequality (constraintLagrange multiplier) α i ⩾ 0,i=1,2,N ； the Lagrange function

$$L\left(w,b,a\right)=\frac{1}{2}|w|^2-\sum_{i=1}^{N}a_i\left[y_i\left(w\cdot x_i\right)+b-1\right] \tag{3}$$

According to Lagrangian duality, the original constrained optimization problem can be equivalent to the minimax dual problem

$$\max_{a}\min_{w,b} \quad L\left(w,b,a\right) \tag{4}$$

if L(w,b,α) is partial derivative of W, B and is equal to 0, then

$$\frac{\partial L}{\partial w}=w-\sum_{i=1}^{N}a_i y_i x_i=0 \quad \Rightarrow \quad w=\sum_{i=1}^{N}a_i y_i x_i \tag{5}$$

$$\frac{\partial L}{\partial b}=\sum_{i=1}^{N}a_i y_i=0 \quad \Rightarrow \quad \sum_{i=1}^{N}a_i y_i=0 \tag{6}$$

$$\max_{a} \quad -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}a_i a_j y_i y_j\left(x_i\cdot x_j\right)+\sum_{i=1}^{N}a_i \tag{7}$$

Equivalent to optimization problem:

$$\min_{a} \quad \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}a_i a_j y_i y_j\left(x_i\cdot x_j\right)-\sum_{i=1}^{N}a_i \tag{8}$$

$$s.t. \quad \sum_{i=1}^{N}a_i y_i=0 \quad a_i\geq 0, \qquad i=1,2\cdots,N \tag{9}$$

In most cases, the training samples are not separable because of noise or special points. Therefore, a more general learning algorithm is needed. SVM model has many unique advantages in solving small sample, nonlinear and high-dimensional pattern recognition and can be extended to other machine learning problems such as function fitting.

B. Deep neural network algorithm

The power of each generator is regarded as a neuron, and a large number of data constitute the neural network. Through continuous training and learning to adjust the weights and offsets of neurons, the convergence of the calculation results of power grid is predicted. The input of neural network needs data set and data set mark. Then according to the data set and corresponding mark, the deep neural network algorithm fit a function parameter which can express the data set, and get the fitting function[3]. The task process of neural network algorithm includes modeling and training, as shown in Figure 1.

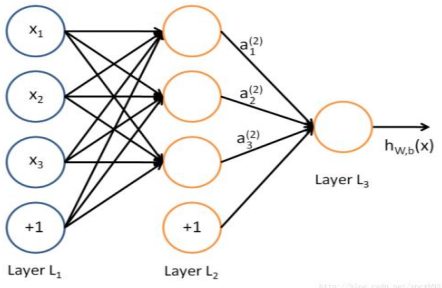

Figure 1. Neural network algorithm structure

- Softmax loss function.

In statistics, loss function[4-6] is a measure of loss and error, the simplest of loss function is variance.The function is

$$h_\theta(x^{(i)}) = \begin{bmatrix} p(y^{(i)}=1 \mid x^{(i)};\theta) \\ p(y^{(i)}=2 \mid x^{(i)};\theta) \\ \vdots \\ p(y^{(i)}=k \mid x^{(i)};\theta) \end{bmatrix} = \frac{1}{\sum_{j=1}^{k} e^{\theta_j^T x^{(i)}}} \begin{bmatrix} e^{\theta_1^T x^{(i)}} \\ e^{\theta_2^T x^{(i)}} \\ \vdots \\ e^{\theta_k^T x^{(i)}} \end{bmatrix} \tag{10}$$

The assumption function of softmax outputs a k-dimension column vector, and the number of each dimension represents the probability of that category. The cost function is

$$J(\theta) = -\frac{1}{m}\left[\sum_{i=1}^{m}\sum_{j=1}^{k} 1\left\{y^{(i)}=j\right\} \log \frac{e^{\theta_j^T x^{(i)}}}{\sum_{i=1}^{k} e^{\theta_j^T x^{(i)}}}\right] + \frac{\lambda}{2}\sum_{i=1}^{k}\sum_{j=0}^{n}\theta_{ij}^2 \tag{11}$$

To minimize the cost function, simple and effective gradient descent can also be used. It should be mentioned that in the program implementation, batch random gradient descent is generally used, that is, msgd, that is, the gradient and update parameters are calculated only after each batch sample is traversed, and a batch generally has dozens to hundreds of individual samples.

- Gradient descent optimizer

In order to reduce the number of iterations, feature scaling is introduced. This method is applied to gradient descent in order to speed up the execution of gradient descent[7]. The core idea is to standardize the value of each feature, so that the value range is approximately - 1 ⩽ x ⩽ 1. The common method is mean normalization, i.e., [x-mean (x)] / std (x).

- Activation function

In order to solve the problem that not all models are linear separable, we can make linear changes or introduce nonlinear factors to solve the problem that linear models cannot solve. Using relu activation function, the principle is that softplus (x) = lg (1 + ex) uses sigmoid and other functions. When calculating the activation function (exponential operation), the calculation amount is large. When calculating the error gradient by back propagation, the derivation involves division, and the calculation amount is relatively large. However, the calculation amount of the whole process is saved, the interdependence of parameters is reduced, and the occurrence of over fitting problem is alleviated.

IV. APPLICATION PRACTICE AND ALGORITHM COMPARISON

A. Two classification model algorithm of logical regression

Build the original model based on the two classification model algorithm of logical regression, and try to visualize the data, compress it into two dimensions, as shown in Figure 2.

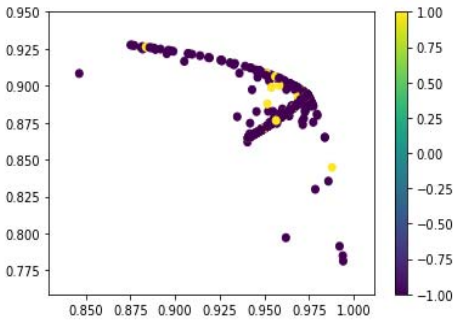

Fig.2 Data compression processing

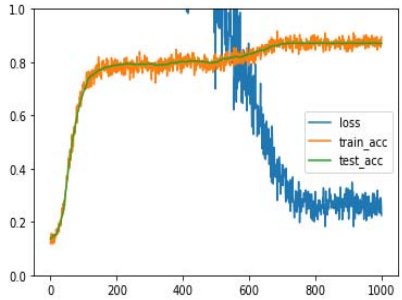

Fig.3 SVM model effect diagram

The SVM model of linear kernel function is established according to the actual needs, and the training data is divided by the k-fold cross validation method[8]. The training results are as follows: training accuracy, test accuracy, loss curve. The accuracy of the test set is 0.872. Finally, the weight W and B corresponding to the minimum loss value obtained by the model are stored in the best_w.npy and best_w.npy files, as shown in Figure 3.

According to the given 50000 data in advance, randomly take 5000 sample data as the prediction data of the training generation model and store them in the pre data.csv file, the remaining 45000 as the training samples and store them in the train data.csv file. The generator parameters are the features of training. We use F1 to express the parameters of each generator. N engines are: F1, F2 , FN as the training feature data, convergence values 1 and - 1 as the result identification. Through the analysis of data by pandas tool, it is found that 24800 data convergence values of training samples are - 1,25200 data convergence values of 1, the proportion of positive samples to total training samples is 51%, the proportion of negative samples to total training samples is 49%, and the sample level is balanced,

B. Deep neural network algorithm

According to the given 50000 pieces of data, first read the data and standardize the format of the data, then randomly take 2 / 3 of the data as the prediction data of the training generation model, and 1 / 3 of the data as the test data of the model. Set the input format of the deep learning model, and set the input layer as the full connection layer with 128 neurons, the activation function as relu, and the number of neurons in the hidden layer as 64, 64,128, 64, 32,

32, respectively. Set the probability of the whole network neuron disconnection as 0.8. Second, set the output to 2D, use softmax classification, generate the model and set the model to use gradient descent optimizer, and the learning rate is 0.01. Finally, compile the model and use the loss of mean square function.after the model is established, the data of the model file is read, and the data order is disordered. Each time, 100 pieces of data are read randomly and in batches. After 500 times of training, the loss and accuracy of acc tend to be stable, and the model converges.

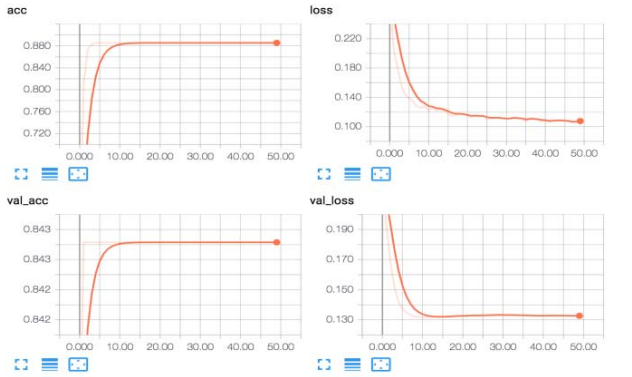

Fig.4 Deep neural network prediction and analysis result chart

C. *Comparative analysis of algorithm application*

The comparison of the above algorithms is as follows: logical regression two classification model is a single layer perceptron, which can only classify the linear separable data. The model is simple, easy to debug, more suitable for the input of obvious and easily controlled features, and the input feature vector is enough for the logical classifier to judge. From the perspective of simplicity, logical classification can be used; linear regression is a linear model, through a large number of data to find the implication relationship between parameters, a linear matrix equation is directly solved, it is likely to be unable to directly solve, there is a unique data set. Generally, it is necessary to turn the problem of Parameter Solving into the problem of minimum error, and get the closest solution, but the close solution in the practical problem is often far from the actual one;deep learning is a multi-layer perceptron, through the middle hidden layer to extract features, one layer of transformation, the original linear non separable data can finally become linear separable, so deep learning can deal with complex functions. However, the deep learning model is complex[9]. Only through continuous debugging can we get a more reasonable model. The model is complex and the debugging cost is high.

The loss function and optimizer are adjusted continuously by establishing the two classification model of logical regression, linear regression and deep learning model, and the accuracy is 0.87, 0.16 and 0.86 respectively. Because the whole data does not conform to the normal distribution and the accuracy is too low, the linear model is not suitable, and the accuracy of logical regression and deep neural network algorithm is not very different. However, by analyzing the characteristics of power grid model parameter data, it fully conforms to the characteristics of the two classification model. The features are obvious, the input is easy to control, and the input feature vector is enough for the logic classifier to judge[10].

V. *CONCLUSION*

Based on the advantages of tensorflow architecture and machine learning algorithm, this paper studies the high-performance calculation of power grid operation mode in practice. Using the parameter model of PSASP software for power grid operation calculation, it carries out the logical regression two classification algorithm and neural network algorithm for comparative analysis and gives the performance evaluation index, which has achieved some results, but also some deficiencies, such as insufficient data samples, etc. In the future, wen can improve the prediction of power grid operation mode by large amount of data and high performance technology.

REFERENCES

[1] SU Rui, ZHANG Xiaojie. Study of digit recognition based on BP neural network[J]. Computer Knowledge and Technology, 2013, 9(14): 3362-3365.

[2] Amraee T, Ranjbar S. Transient instability prediction using decision tree technique[J] . IEEE Transactions on Power Systems, 2013, 28(3) : 3028-3037.

[3] DING Shuo,WU Qinghui. Performance comparsion of function approximation based on improved BP neural network[J] Computer and Monetization, 2012(11):10-13, 17.

[4] GAO Qiang, YANG Wu, LI Qian. Fast training model for image classification based on spatial information of deep belief network[J]. Journal of System Simulation, 2015, 27(3): 549-558.

[5] He Jin, Fei Lifan. Further study on three dimensional Douglas-Peucker algorithm and its application to generalization of DEM[J]. Geomatics and Information Science of Wunan University, 2008,33(2):160-163.

[6] Li Bijun , Fu Hongjun , Hu Yangyu , et al . Design and implementation of automatic computation and report generation system for annual power grid operation mode[J]. Automation of Electric Power Systems, 2010, 34(4) : 112-115.

[7] Mitchel M T, Lei M, Su C J .A collaborative control system for mass customization manufacturing[J] . Annals of the CIRP , 1997 , 46(1):) : 373-376.

[8] State Grid Corporation of China . Smart grid dispatch technology support system design[M]. Beijing : State Grid Corporation of China, 2008.

[9] Al-Masri A, Kadir M, Hizam H, et al. A novel implementation for generator rotor angle stability prediction using an adaptive artificial neural network application for dynamic security assessment[J]. IEEE Transactions on Power Systems, 2013, 28(3) : 2516-2525.

[10] Carreras B A, Newman D E, Dobson I. North American blackout time series statistics and implications for blackout risk[J] . IEEE Transactions on Power Systems , 2016 , 31(6) 4406-4414