Proceedings of the 2018 International Conference on Machine Learning and Cybernetics, Chengdu, China, 15-18 July, 2018

A DEEP LEARNING LOAD FORECASTING METHOD BASED ON LOAD TYPE RECOGNITION

JINGJIE YANG[1], QIANG WANG[2]

[1]Liaoning Jianzhu Vocational College, Liaoyang, Liaoning province. 111000, P. R. China
[2]Beijing City University, Beijing,100083, P. R. China
E-MAIL: jshdw@126.com

Abstract:

In this paper, a short-term load forecasting model based on load type recognition is established. According to the characteristics of electric power load, the load data can be divided into three types: peak load, valley load and normal load. During the forecasting process, the support vector machine recognition model is used to determine load type on the predicted time point. After then, load samples of the same type are selected as the training sample to establish deep learning forecasting model. Finally, the trained model is applied to forecasting the load value. The experimental comparison shows that its prediction accuracy is higher than that of other models.

Keywords:

Short term load forecasting; Support vector machine; Deep belief network (DBN); Load type recognition; Deep learning

1. Introduction

Short-term power load forecasting is very important to the economical and safe operation of power system. Traditional load forecasting methods include time series model, regression analysis model and so on. These models are relatively simple and perform well when load sequences are of high stability. Because they only use historical load values, it is difficult for them to achieve high prediction accuracy when they deal with the complicated and rapid change trend of short-term power load [1,2].

With the development of artificial intelligence, artificial neural network model, support vector machine model and so on [3-6] have been widely applied to short-term power load forecasting. Due to their good fitting ability of nonlinear functions, these models can deal with the nonlinear relationship between the influencing factors and the load output with good performance. On the basis of these methods, some improved algorithms and combination algorithms increase the accuracy of load forecasting. For example, some intelligent optimization algorithms such as genetic algorithm, particle swarm algorithm are applied to estimate model parameters [7-9]. However, all these models are based on the shallow machine learning algorithms and their cognitive ability to the potential characteristics of a large number of nonlinear load data is insufficient.

The adaptive learning ability used in depth study machine can effectively solve the above problems. The deep belief network (DBN) load forecasting models based on these ideas illustrate their effectiveness in load forecasting [10,11]. DBN is consisted of many restricted Boltzmann machines (RBM) stacked together. As an effective treatment method for characteristics data, RBM is able to solve problems of high dimensional, complex and nonlinear power system load forecasting. At the same time, while solving the DBN parameters, the training set is divided into small batch data to calculate and improve the training efficiency.

In this paper, the support vector machine is used to determine the load type of history data. According to the load data of the actual sample, the load type are divided into three categories: peak load, valley load and normal load. Firstly, the support vector machine (SVM) recognition model is used to determine the type of predicted load; then, the load samples of the same type are selected as the training sample to set up DBN prediction model. In this paper, the DBN and SVM are applied together. The experimental comparison shows that the prediction accuracy is higher than that of other models.

2. Deep Belief Network

DBN is a deep network learning algorithm proposed by Hinton to deal with high dimensional and large scale data problems, such as image feature extraction and collaborative filtering. The short-term load forecasting structure model based on DBN method is shown in Fig. 1. The input layer includes history load data, meteorological

978-1-5386-5214-5/18/$31.00 ©2018 IEEE

factors and date types.

DBN is a neural network model composed of multiple RBM stacks and the network structure should be trained before it is applied in load prediction. The purpose is to determine the connection weight and the neuron bias. The training process includes pre-training and reverse trimming. Firstly, the pre-training process will train each layer of RBM separately with the unsupervised greedy algorithm and ensure that the feature vectors can retain the feature information as much as possible when they are mapped to the next layer[12]. The pre-training process can provide a good initial value for the entire DBN network. The traditional BP algorithm of neural network is used to fine-tune the parameters so that the model converges to the best advantage.

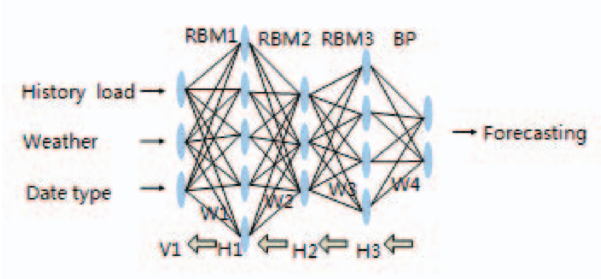

FIGURE 1. Load forecasting structure model based on DBN

Fig.1 is a DBN of three layers of RBM, a single RBM (such as RBM1) is a symmetry and random neural network without self-feedback which includes an hidden layer and a visible layer, neurons of the same layer have no connections, and neurons between the layers are connected by weightings. V1 is the visible layer which is connected to the observed data; H1 is the hidden layer, which is used to extract the valid characteristics of the input data; W1 is the connection weight between the visible layer and the hidden layer. The neurons in the network are only inactive and active, usually in binary 0 and 1. RBM is a kind of model based on energy, v_i is the state of neuron i in the visible layer, its corresponding offset value is a_i; h_j is the state of neuron j in hidden layer, its corresponding offset value is b_j; w_{ij} is the weight which connect neuron i and j, then, the energy of RBM determined by (v,h) can be represented as:

$$E(v,h|\theta)=-\sum_{i=1}^{n}a_i v_i-\sum_{j=1}^{m}b_j h_j-\sum_{i=1}^{n}\sum_{j=1}^{m}v_i w_{ij} h_j \tag{1}$$

where $\theta=(w_{ij},a_i,b_j)$ is the parameter of RBM, and n and m are the neuron quantity for the visible layer and the hidden layer.

The joint probability distribution of (v,h) can be obtained by the energy function as the follows:

$$p(v,h|\theta)=\frac{1}{Z(\theta)}\exp(-E(v,h|\theta)) \tag{2}$$

where $Z(\theta)=\sum_v\sum_h \exp(-E(v,h|\theta))$ is normalization factor.

For a training set with N samples, parameter θ can be obtained by the maximum logarithmic likelihood function of the learning sample.

$$\theta^*=\arg\max_\theta L(\theta)=\arg\max_\theta \sum_{n=1}^{N}\log p(v^n|\theta) \tag{3}$$

where $p(v,h|\theta)=\frac{1}{Z(\theta)}\sum_n \exp(-E(v,h|\theta))$ is the likelihood function of the observed data V.

In the course of training, as the calculation of normalized factor $z(\theta)$ is of great complexity, Gibbs and other sampling methods are adopted. Hinton proposed a fast learning algorithm based on contrastive divergence (CD) to train network parameters, which improved training efficiency and promoted RBM development [13]. Firstly, the CD method calculate the valued state of neurons in hidden layer by visible layer neurons vector. Secondly, the neurons state of visible layer is rebuilt through the hidden layer neurons; then, the valued state of neurons in hidden layer is calculated by the visible layer neurons again after the reconstruction which can obtain the new state of hidden layer neurons.

Since the activation states of each neuron in the same RBM layer are independent of each other, the activation probability of neuron j in the hidden layer can be calculated according to the state of the visible layer.

$$p(h_j=1|v,\theta)=\frac{1}{1+\exp\left(-b_j-\sum_i v_i w_{ij}\right)} \tag{4}$$

According to the reconstruction of the hidden layer, the activation probability of neuron i is.

$$p(v_j=1|v,\theta)=\frac{1}{1+\exp\left(-a_i-\sum_j h_j w_{ij}\right)} \tag{5}$$

Therefore, the maximum value of logarithmic likelihood function is solved by stochastic gradient ascending method, the calculation criterion for each parameter variation is:

$$\Delta w_{ij}=\langle v_i h_j\rangle_{data}-\langle v_i h_j\rangle_{recon} \tag{6}$$

$$\Delta a_i=\langle v_i\rangle_{data}-\langle v_i\rangle_{recon} \tag{7}$$

$$\Delta b_j = \langle h_j \rangle_{data} - \langle h_j \rangle_{recon} \tag{8}$$

where $\langle * \rangle_{data}$ is the distribution of the original observation data, $\langle * \rangle_{recon}$ is the distribution after reconstruction. Considering learning rate ε, the parameter updating criterion is :

$$w_{ij}^{k+1} = w_{ij}^{k} + \varepsilon \Delta w_{ij};$$
$$a_i^{k+1} = a_i^{k} + \varepsilon \Delta a_i;$$
$$b_j^{k+1} = b_j^{k} + \varepsilon \Delta b_j$$

3. DBN-SVM forecasting Model

Short-term load forecasting is a multivariable regression problems, the actual load is the output of regression function, the corresponding influencing factors such as historical load and meteorological information and date type are inputs.

In this paper, the traditional support vector machine is used to process the historical data. Based on statistical learning theory, robust regression theory and duality planning theory, a learning algorithm called SVM (SVM) is proposed by Vapnik. Unlike the traditional neural network model, the support vector machine (SVM) algorithm realize the structural risk minimization principle. It also minimizes empirical risk and the VC dimension, thus make the actual risk even more small. Support vector machines have good generalization performance, high fitting accuracy and global optimum. It was originally applied to pattern recognition to find the best decision rules for generalization. With the introduction of the insensitive loss function, the support vector machine has shown obvious superiority in solving the nonlinear regression problem, and has begun to receive more and more attention.

After load pattern is identified, the sample points of the same load type are selected as the training samples of DBN, so as to determine the prediction model.

3.1. Set up a sample set

The sample set $\{(\overline{X_i}, y_i)\}_{i=1}^{n}$ can be set up according to the load value of N continuous time point and its related influencing factors at that time. $\overline{X_i}$ is the set of influencing factors, and y_i is the load value at i time point, the set of influencing factors:

$$\overline{X_i} = \{L_i^{i-1}, L_i^{i-2}, \cdots, L_i^{i-6}, T_i, M_i, W_i, H_i\} \tag{9}$$

where L_i^{i-1} is the actual load value at the time of $i-1$; L_i^{i-2} is the actual load at the time $i-2$; T_i is the temperature; M_i is the humidity at time i ; W_i is the week type; H_i is holiday type. All these factors are at i time point.

Week type is calculated as the follows:

$$W_i = w/7 \tag{10}$$

where $w = 1,2,3,4,5,6,7$

Holiday type is calculated as the follows:

$$H_i = \begin{cases} 1 & holiday \\ 0 & others \end{cases} \tag{11}$$

Load type as the output of SVM is assigned as follows: peak load type is 0; normal load type is 1/2; the valley load type is 1. All data are normalized and processed to prevent data overflow.

3.2. Load type classification

According to the actual load, the historical load is divided into three categories: peak load, normal load，and valley load. The details are as the follows:

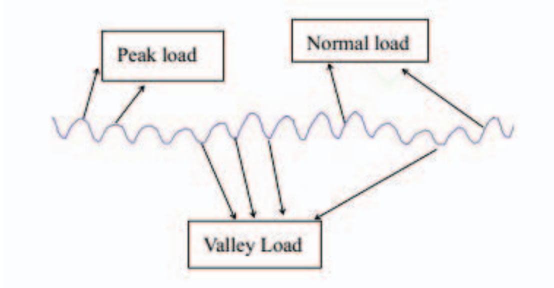

FIGURE 2. Load type classification

For a sample set $\{(\overline{X_i}, t_i)\}_{i=1}^{N}$, t_i is the load type value, peak load type is 0; normal load type is 1/2; the valley load type is 1. This sample set can be divided into three subsets: peak load set P, normal load set N and valley load V as the follows:

$$\begin{cases} P = \{(\overline{X_i}, 0)\}_{i=1}^{m} \\ N = \{(\overline{X_i}, 0.5)\}_{i=1}^{n} \cdots \quad m+n+l=N \\ V = \{(\overline{X_i}, 1)\}_{i=1}^{l} \end{cases} \tag{12}$$

where m , n , l is the sample quantity of each subset.

3.3. SVM training

According to the above load classification method for training SVM, the influence factors are input and the load type value is output. All data is normalized before input.

ISMO algorithm is used in the training process, and the training samples are selected according to the above method. Input layer has 10 nodes, output layer 1, kernel function is Gaussian, expression is as follows:

$$K\left(\widetilde{X}_i, \widetilde{X}_j\right) = \exp\left[-\frac{\left\|\widetilde{X}_i - \widetilde{X}_j\right\|}{\sigma}\right] \tag{13}$$

σ is the width parameter of the Gaussian kernel. According to experience, other parameter values are selected as follows: $c = 10$, $\varepsilon = 0.01$, $\sigma = 1$, $\tau = 10^{-5}$

3.4. Select the training sample of DBN

After training, SVM can predict the load type of the prediction point. The sample subset with the same type is selected as the training sample of DBN.

3.5. DBN training and forecasting

For peak load sample set, a DBN network with three hidden layers (RBM1, RBM2 and RBM3) is created. Its structural parameters are 10-8-5-3-1, 10 input neurons in the visible layer, and RBM1, RBM2 and RBM3 contains 8, 5, 3 neurons respectively, the output layer of DBN has only 1 neuron. The learning rate of RBM is set to 1 and the number of iterations is 1000 times.

For normal load sample set, a DBN network with two hidden layers (RBM1 and RBM2) is created. Its structural parameters are 10-8-5-1, 10 input neurons in the visible layer, RBM1 and RBM2 contains 8, 5 neurons respectively, the output layer of DBN has only 1 neuron. The learning rate of RBM is set to 1 and the number of iterations is 1000 times.

For valley load sample set, a DBN network with three hidden layers (RBM1, RBM2 and RBM3) is created. Its structural parameters are 10-8-5-3-1, 10 input neurons in the visible layer, and RBM1, RBM2 and RBM3 contains 8, 5, 3 neurons respectively, the output layer of DBN has only 1 neuron. The learning rate of RBM is set to 1 and the number of iterations is 1000 times.

Input the influence factor of the prediction point into the trained DBN whose output is to predict the data, and the whole process is carried out in the following computer configured: EMS is 256MB, hard disc 46G.

4. Experimental analysis

The model is applied to forecast the hourly load in Baoding city, Hebei province. The sample points are taken from the daily hourly load data from January, 1st, 2017 to December 31st, 2017, Totally, there are 8,760 sets data. Among them, the data between January 1, 2017 and November 31, 2017 are used for the training of the model and the remaining data are used to detect the model. The prediction accuracy is represented by relative error.

In addition to the SVM-DBN model in this paper, traditional DBN and SVM models are also used to predict the same data points so as to compare the predicted accuracy. The specific comparison results are shown in Table 1.

It can be seen from the followed table that the accuracy of SVM-DBN model introduced in this paper is higher than that of the simple DBN model and SVM model.

TABLE 1. Model accuracy comparison

TIME	SVM RE（%）	DBN RE（%）	SVM-DBN RE（%）
2017, 12, 1: 0:00	2.836	2.791	2.632
2017, 12, 1: 1:00	2.764	2.802	2.683
2017, 12, 1: 2:00	2.704	2.663	2.630
2017, 12, 1: 3:00	2.821	2.723	2.700
2017, 12, 1:4:00	2.832	2.735	2.616
2017, 12, 5: 0:00	2.850	2.710	2.600
2017, 12, 6: 0:00	2.760	2.890	2.630
2017, 12, 7: 0:00	2.910	2.920	2.800
2017, 12, 8: 0:00	3.010	2.990	2.900
2017, 12, 9: 0:00	2.970	3.110	2.861
2017,12,10: 0:00	3.212	3.192	3.000

5. Conclusion

In this paper, a short-term load forecasting model of SVM-DBN is established. Firstly, the model divides the sample points into peak load points, normal load points and valley load points. Secondly, the traditional support vector machine is used to predict the load type of the prediction point. Finally, the DBN is trained with the same point as the predicted point load type and the load prediction is carried out. This method can reduce the range of sample load and improve the accuracy of prediction. The actual application and comparison show that the model is superior to the simple SVM model and ANN model in the prediction precision.

References

[1] WAN Kun, LIU Ruiyu, “Application of Interval Time-Series Vector Autoregressive Model in Short-Term Load Forecasting”, Power System Technology, 2012，36(11)：77-81．

[2] Guan C，Luh P B，Michel L D，”Hybrid Kalman filters for very short-term load forecasting and prediction interval estimation”[J]．IEEE Transactions on Power Systems，2013，28(4)：3806-3817．

[3] Li P，Li Y，Xiong Q，”Application of a hybrid quantized Elman neural network in short-term load forecasting”[J] ． International Journal of Electrical Power & Energy Systems，2014，55(2)：749-759．

[4] Azadeh A ， Ghaderi S F ， Sheikhalishahi M ， ”Optimization of short load forecasting in electricity market of Iran using artificial neural networks”[J]．Optimization and Engineering，2014，15(2)： 485-508．

[5] Ceperic E，Ceperic V，Baric A．”A strategy for short-term load forecasting by support vector regression machines”[J] ． IEEE Transactions on Power Systems，2013，28(4)：4356-4364．

[6] Ge Shaoyun, Jia Ousha, Liu Hong. “Short-term power load forecasting based on genetic grey neural network model”, Power System Technology, 2012，36(1)：224-229．

[7] Wang Baoyi, Wang Dongyang, Zhang Shaomin, “Short-term distributed power load forecasting algorithm based on Spark and IPPSO_LSSVM”, Electric Power Automation Equipment, 2016，36(1)：117-122．

[8] Cheng R，Jin Y．”A social learning particle swarm optimization algorithm for scalable optimization”[J]．Information Sciences，2015，291(6)：43-60．

[9] Vahabie A H，Yousefi M M R，Araabi B N，”Mutual information based input selection in neuro-fuzzy modeling for short term load forecasting of Iran National Power System”[C]． IEEE International Conference on Control and Automation．IEEE，2007：2710-2715．

[10] Dedinec A，Filiposka S，Dedinec A，”Deep belief network based electricity load forecasting ： An analysis of Macedonian case”[J]．Energy，2016，115：1688-1700．

[11] Qiu X，Zhang L，Ren Y，”Ensemble deep learning for regression and time series forecasting”[C] ． IEEE SSCI 2014-2014 IEEE Symposium Series on Computational Intelligence．2014：1-6．

[12] Dragomiretskiy K ， Zosso D ． “Variational mode decomposition”[J]． IEEE Transactions on Signal Processing，2013，62(3)：531-544．

[13] Nair V，Hinton G E．”Rectified linear units improve restricted Boltzmann machines”[C]． International Conference on Machine Learning． DBLP，2010：807-814．