On the BP neural network and support vector machine in the role of oil production forecast

Li Xiaofeng, Gu Xinglong

Abstract: *The accurate prediction of oil production, the oil companies to develop a reasonable production plan, avoid blind investment, important conditions for achieving sustainable development. In order to improve the accuracy of production forecasts, the paper proposes a BP neural network model, detailing its basic principles and algorithms and the concrete process of this model, and finally apply it to predict job in the oil industry. After the model of learning and training simulation to predict actual results of the comparison show that the model can well achieve the desired effect, it is important to the oil production forecasting. Oilfield proposed system modeling theory based SVM, and the original application-dual algorithm to solve SVM quadratic programming problems. SVM is used to predict the oil production wells, forecasting examples show that the maximum relative error of generalization 5.611%, very close to the predicted value of the actual output of oil wells; compared with other prediction methods, the prediction model has high prediction accuracy. Using BP neural network and support vector machine to predict the oil production and the advantages and disadvantages of the two methods[1].*

Keywords: *BP neural network support vector machine hidden layer nodes primal - dual algorithm Oilfield Production Forecast*

I. INTRODUCTION

Neural network (Neural Network) and SVM (Support Vector Machines, SVM) is the representative method of statistical learning.

Neural networks can be considered and support vector machines are derived from the perception machine (Perceptron). Perceptron is a linear classification model (1958) by the Rosenblatt invention. Perceptron linear classification is valid, but the reality of classification is usually non-linear.

Neural networks and support vector machines (including kernel method) are nonlinear classification model. In 1986, Rummelhart and McClelland invention neural network learning algorithm Back Propagation. Later, Vapnik et al proposed support vector machine in 1992. A neural network is a multi-layer (usually three) non-linear model, support vector machine using nuclear techniques to convert the non-linear problem into linear problem.

Manuscript received March 26, 2016

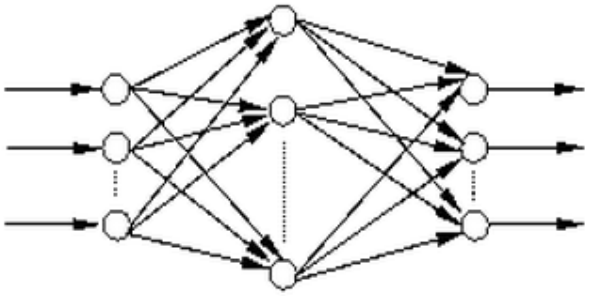
Li Xiaofeng, School of Sciences,Southwest Petroleum University,Chengdu,Sichuan 610500,China

Gu Xinglong, School of Sciences Neijiang Normal University,Neijiang,Sichuan 641000, China

II. THE BASIC IDEA

2.1 The basic idea of BP Neural Network

BP (Back Propagation) network in 1986 by a team of scientists led by Rumelhart and McCelland proposed, according to a former back propagation algorithm to train multi-layer feedforward network, it is one of the most widely used neural network model. BP network can learn and store a large amount of input - output mode mapping without prior reveal mathematical equations that describe the mapping relationship. It's learning rule is to use the steepest descent method, by reverse spread to constantly adjust the network weights and thresholds and the minimum sum of squared error of the network. BP neural network topology comprises an input layer, hide layer and an output layer.



Pictrue 1 BP neural network schematic diagram

2.2 Support Vector Machine basic idea

(1) which is designed for finite samples, to achieve the structural risk minimization: between the complexity of a given data accuracy and approximation function approximation to seek a compromise in order to obtain the best generalization ability;

(2) It is the ultimate solution is a convex quadratic programming problem, in theory, will get the global optimal solution to solve the problem of local extrema in the neural network method can not be avoided;

(3) it is practical problems through non-linear transformation to convert the high dimensional feature space constructed linear decision function in a high-dimensional space to achieve the original non-linear decision function space, cleverly solved the dimension of the problem, and to ensure the It has good generalization ability, and has nothing to do with the complexity of the algorithm sample dimension.

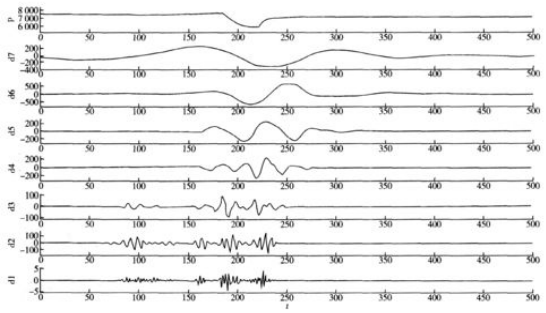
Currently, SVM algorithm in pattern recognition, regression estimation, probability density function estimation and other aspects of the application, and the algorithm in efficiency and accuracy over traditional learning algorithm.

III. OILFIELD PRODUCTION FORECAST

3.1 BP neural network to predict oil production

(1) input layer nodes selected

Stress Test logging curve obtained is pressure time-varying data, it needs to do wavelet transform in time domain is equivalent to the pressure conversionTo the frequency domain, see Figure 1. Using different resolution wavelet transform analysis of the data shows the details of the zoom function, a resolution in Fig. 1 values, respectively2, 4, 8, 16, 32 ..., as a increases, the system becomes smooth and the corresponding wavelet lack characteristics. This example uses a single pressure test curve.According to the calculation, and the resolution of a = 2, to obtain wavelet coefficients sampling sequence over time, so as input data input node, and select eight input nodes[2].



Pictrue 2 Wavelet coefficients of the original pressure curve and different resolutions

(2) hidden layer node selected

In order to improve the accuracy of network training, by increasing the number of nodes increases the number of layers than the method of approach is much simpler, so the network design is as little as possible.

Select the best number of nodes under the premise of layers. However, the number of hidden layer nodes selected is appropriate, in theory, does not have a clearly defined. Too little hidden layer nodes, does not reflect the value of the relationship between the subsequent precursor values, fault tolerance is poor, modeling is not sufficient; too much can lead to learning for too long, generalization diminished capacity, network forecasting ability. Therefore, according to the neural network theory theorems Kolmogorov theorem, hidden layer nodes can refer to the following formula: m = (p + q) 0 .5 + a, where m is the number of nodes in the hidden layer, p is the input layer nodes, q is output layer nodes, a constant between 1-10. After repeated training, hidden layer nodes at 16.

(3) Select the output layer node

This example is oil and gas production forecast, therefore select two output nodes. And the desired output sample is known, which consists of historical data given.

(4) network training parameter

The example of a company log data as training samples, the data do normalization pretreatment and extract 2/3 data used to train the network, and the remaining 1/3 of the data used to test the model's accuracy on. Use Matlab7 .0 Neural Network Toolbox software for the network model learning and training, the training period is 1000 times, the training target 0.001.

(5) Analysis

After the success of network training, training simulation output and the actual output sample better consistency, the correlation coefficient was 0.96, indicating a good train network. Also verify that the sample simulation output and the actual output the same consistent, indicating that the resulting strong training network generalization ability, so use this model to predict oil and gas. Forecast results and the actual value of the verification samples are compared in

Table 1: Predicted and actual value comparison table

Month	Actual output/10^4t	Predictive value/10^4t	Relative error/%	Month	Actual output/10^4t	Predictive value/10^4t	Relative error/%
1	28.10	24.088	12.143	17	14.32	13.788	3.715
2	30.09	26.852	10.760	18	15.90	15.414	3.055
3	28.60	25.320	11.467	19	17.97	16.23	9.081
4	22.94	20.638	10.036	20	15.13	14.628	3.320
5	16.68	17.511	-4.984	21	12.31	12.32	-0.081
6	22.13	19.774	10.646	22	11.55	12.14	-5.109
7	25.23	22.383	11.284	23	14.96	13.783	7.871
8	23.04	21.221	7.895	24	14.96	14.708	1.683
9	19.55	17.627	9.838	25	15.13	13.83	8.590
10	14.39	15.427	-7.208	26	11.00	11.624	-5.669
11	19.33	17.458	9.683	27	10.03	10.407	-3.760
12	21.90	19.756	9.792	28	10.35	10.29	0.558
13	20.80	19.523	6.138	29*	9.35	9.8747	-5.611
14	19.58	17.674	9.735	30*	9.43	9.8979	-4.962
15	15.07	14.858	1.406	31*	11.13	11.498	-3.311
16	13.23	13.226	0.028				

Forecast oil production	Gas production forecast	Actual oil production	Actual gas production
110.000	34001.944	110	34002
109.999	34001.779	110	34002
109.958	34001.936	110	34002
109.945	34001.935	110	34002

Table 2 A mouth actual production wells in Shengli Oilfield comparison with predicted values

3.2 Support vector machine to predict oil production

In some wells in Shengli Oilfield as an example, using the original ~ dual algorithm, tested in Matlab. The first 28 months of production wells[4].

As the amount of training data, after three months of production forecast. Based on experience, the choice of design parameters of SVM £ === 0.01, C-200, a 0.05.

Prediction data obtained with the actual output of the comparison are shown in

As seen from Table 2, in the early stages of training, the relative error is large, up to 12.143, but in the forecast period, the maximum relative error is 5.611 generalization, the predicted value can be a good track actual production of the wells. Hara prediction model dual algorithm applied to the development of different wells, versatile, simplifying forecasting process.

IV. Comparison of the two methods in terms of Oilfield Production Forecast

Prediction for the oil industry, the introduction of artificial neural network in which the BP neural network model. The results show: the model good performance, high prediction accuracy. After training in the network, the prediction becomes very simple, just enter the data by the wavelet transform, quickly available predictions.

While in the practical application of the model also has some deficiencies, the size of the network model and algorithm, including the layers of the network, each node number, connection, excitation type of the function and so on, these are the neural network used in petroleum the main obstacle to predict the industry, need to go through repeated training.

SVM is a sound theoretical basis for novel small sample learning methods. It basically does not involve probability measure and the law of large numbers, etc., so different from the existing statistical methods, greatly simplifies the usual classification and regression problems. SVM final decision function is only determined by a small number of support vectors, the computational complexity depends on the number of support vectors, rather than the dimension of the sample space, which avoid the "curse of dimensionality" in a sense[6]. However, SVM algorithm is difficult to implement a large-scale training sample, and is usually caused by the predicted data is huge. Since the SVM algorithm is the use of quadratic programming to solve the support vector, and solving quadratic programming will involve the calculation of a matrix of order m (m is the number of samples), when the storage and compute the number m of the matrix will be a great time to spend a lot of when the machine's memory and computing time.

Summary

The simulation can be drawn, both in the form of somewhat similar, but in fact very different. In short, the neural network is a "black box" optimization goal is based on empirical risk minimization, easy to fall into local optimum training results is not very stable, generally require large sample; and support vector machines have a rigorous theoretical and mathematical foundations, based on structural risk minimization principle, generalization ability than the former, global optimization algorithm, for small sample statistical theory, support vector machines for large sample storage and computing will be a great waste of time.

Reference

Stanislav Kaleta , Daniel Novo tny , Peter Sincak .Prediction Sy stems Based on FIR BP Neural Netwo rks[J] .Springer-Verlag ,2001 .29-34.

Zhoucai Lan, Liu Min BP neural network in the prediction of oil production [J] Wuhan University of Technology, 2009 (3): 125-129.

Zhao forecast based on oil and gas production history of BP neural network [J] Southwest Petroleum Institute, 1998 (2): 23-26.

Li Zhuo, Liu Bin, Liu Tienan support vector machine in Oilfield Production Forecast [J] Northeast Petroleum University, 2005, 29 (5): 96-97.

Jin Pi Yan, Rui Yong BP Algorithm Improvement and Application of various algorithms [J] Nanjing University of Aeronautics and Astronautics, 1994 (S1):. 201-205.

Ren Baosheng, Zhao Ming, Liu Zhibin, and other indicators of oilfield development dynamic support vector machine [J] Petroleum Planning and Design, 2008, 19 (6): 12-15.