Forecasting Method of Crude Oil Output Based on Adaboost_BP

Hongtao Hu [1], Rong Hui [1] and Xin Guan [2] +

[1] School of Computer Science Xi'an Shiyou University, Xi'an, China

[2] Research Institute of Petroleum Exploration & Development-LangFang, Langfang, Hebei, China

Abstract. In order to improve the prediction accuracy of BP neural network, Adaboost algorithm and BP neural network are combined to propose the Adaboost_BP prediction model. The prediction model firstly preprocesses the training data set, which selects different parameters of BP neural network to construct several BP neural networks. Then, the test data set and its distribution weight are initialized. According to the Adaboost algorithm, the prediction error of each BP weak predictor is calculated and the distribution weight of the test data set is adjusted according to the error, and the weight of the predictor is obtained. Finally, all weak predictors are combined to generate a strong predictor. Based on the original data of oilfield block from 1994 to 2016, the results show that the prediction of Adaboost_BP prediction model is more accurate than the single BP neural network model, and achieves good predictive result, which validates the effectiveness of the proposed model.

Keywords: crude oil output, adaboost algorithm, BP neural network, strong predictor.

1. Introduction

Crude oil output forecasting is an important work in oilfield development. Accurate forecasting result is conducive to improve the efficiency of mining. Currently, there have been extensive researches conducted on crude oil output around the world, such as proposed grey theory [1], support vector machine [2], time series method [3] etc. However, these methods have lower prediction accuracy, therefore many scholars introduced BP neural network into the crude oil output forecasting, and established a BP forecasting model which based on the back propagation of the neural network. However, the BP neural network algorithm is a weak learning algorithm which has over-fitting characteristics and easily falls into local optimization limitations. Hence, by training weak predictors to improve the prediction accuracy, the Adaboost algorithm is put forward to construct BP weak predictors as a strong predictor.

Aiming at improving lower prediction accuracy of BP neural network, Adaboost algorithm is combined to BP neural network algorithm to construct the Adaboost_BP neural network prediction model in this paper. Then apply Adaboost_BP prediction model to forecast the crude oil output. Through the simulation experiments, the effectiveness is verified for the prediction model.

2. Bp Prediction Model and Adaboost Algorithm

2.1. The Principle of BP Neural Network

BP neural network is a multi-layer feed forward algorithm which can learn and store large number of input-output mode mapping relationship [4]. The network structure is composed of input layer, output layer and hidden layers. Before network training, the connection weights for each layer are assigned with a random number between [0, 1]. Because of the random settings, the convergence rate is difficult to control so that the prediction value may turn into the local optimal solution, rather than the global optimal solution.

+ Corresponding author: Hongtao Hu. Tel.: + 86-029-88382713; fax: +86-029-88382710.
  E-mail address: huhongtao@xsyu.edu.cn.

2.2. The Principle of Adaboost Algorithm

Adaboost algorithm is an iterative algorithm. By generating different weak predictors for the same training set training, these weak predictors are linearly combined to construct a strong predictor with strong generalization ability. The core idea of Adaboost algorithm is to improve the sample weights of poor training effect in training set and the weights of weak predictor with large prediction error, and reduce the sample weights of good training effect in training set and the weights of weak predictor with small prediction error, to improve the prediction accuracy of weak predictors [5].

3. Adaboost_BP Neural Network Prediction Model

3.1. BP Neural Network Parameters Setting

Multiple neural networks are constructed with different parameters of BP neural network [6]. The constructor function is: net=newff (P,T,S,TF,BTF,BLF,PF,IPF,OPF,DDF), where P and T are matrix, S is hidden layer nodes, TF is transfer function, BTF is training function and BLF is learning function. PF is performance analysis function, IPF and OPF are processing function, DDF is verification data partition function. Generally, different neural networks are generated by setting these parameters: S, TF, BTF, BLF. The specific parameters setting process:

- Determine the number of hidden layer nodes: the formula is: $S < n-1$, $S = \sqrt{m+n} + \alpha$, S=lbn, where n is input neurons value, m is output neuron value and α is a constant factor between [0, 10].
- The transfer function mainly used: tansig, purelin, hardlims, logsig.
- Training function mainly used: traingd, trainlm, traungda, traingdm, traingdx.
- Learning function mainly used: learngd, learngdm.

3.2. The Process of Adaboost_BP Algorithm

The process of Adaboost_BP neural network prediction model is as follows: firstly divides original data into BP training data set and Adaboost_BP test set, pre-processed data set. Secondly by selecting different parameters, several BP weak predictors generated. Then initializes the test data and distribution weights of Adaboost_BP, and all test data are assigned with the same distribution weights [7-8]. Next calculate the prediction error of each weak predictor and according to prediction error to adjust the distribution weight of test set, simultaneously to obtain weights of weak predictors. Finally, combine these weak predictors linearly to a strong predictor so as to improve the prediction accuracy of the weak predictors. The process of Adaboost_BP algorithm flow is shown in Fig .1.
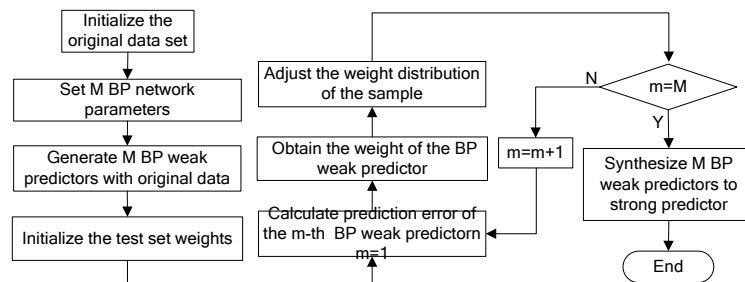

Fig. 1: The process of Adaboost_BP algorithm

3.3. Algorithm Steps of The Adaboost_BP

The algorithm steps are as follows:

- Initialize the original data set: Divide original data into BP neural network training data set and training test set as well as normalize original data.
- Set BP neural network parameters: Determine the neural network structure and initialize the weight, and generate m BP weak predictors by selecting different network parameters.
- Initialize the test set weights: All test data are assigned with the same weights in the initialization state. The initialize process is shown in equation (1), where N represents test sample numbers and $W_{1,i}$ is a weight of the i-th sample data.

$$w_{1,i} = \frac{1}{N}, i = 1,2,...,N \tag{1}$$

- Calculate prediction error of the m-th BP weak predictor: The iterations with m=1, 2,…,M, m BP neural networks are used as the m weak predictors of Adaboost_BP algorithm. Calculate the prediction error of each sample of different weak predictors. The formula is shown in equation (2), where $G_m(x)$ represents weak predictor prediction sequence and y_i is the desired value. And when the weak predictor generates prediction errors, the Parameters I is assigned with the value of one, otherwise is assigned with zero.

$$e_m = \sum_{i=1}^{N} W_{m,i} I(G_m(x) \neq y_i) \tag{2}$$

- Obtain the weight of the weak predictor: During the training process according to the prediction error of weak predictors sequence $G_m(x)$. The weight of weak predictors is calculated in equation (3).

$$\alpha_m = \frac{1}{2} \ln \frac{1-e_m}{e_m} \tag{3}$$

- Adjust the weight distribution of the sample: According to the prediction sequence $G_m(x)$, adjust the distribution weight $W_{m+1,i}$ of the test data. The adjustment formula is represented as equation (4).

$$w_{m+1,i} = \frac{w_{m,i}}{Z_m} \exp(-\alpha_m y_i G_m(x)), i = 1,2,...,N \tag{4}$$

- Strong predictor: After M iterations, the M weak predictors and the corresponding weights of weak predictors are combined linearly to construct strong predictor. The strong predictor G(x) is formed in equation (5).

$$G(x) = sign\left(\sum_{i=1}^{M} \alpha_m G_m(x)\right) \tag{5}$$

4. Experiment Prediction Model

4.1. Original Data Collection

Crude oil output is the most important indicator of oil development. There are many data related to crude oil output, which can be used to forecast the crude oil production data. A total of 276 monthly data from January 1994 to December 2016 are extracted from an oil production database in an oil field, and what can be learned is that the main influencing factors of crude oil output is oil output electricity, gathering electricity, water injection electricity, water injection. The above influencing factors are used as input conditions for the Adaboost_BP prediction model, and crude oil output is used as output. These production data is processed to obtain the original data set, as shown in TABLE I.

Table 1: Crude Oil Output and Influencing Factors

Time	Oil Output Electricity /10^4 kWh	Gathering Electricity /10^4 kWh	Water Injection Electricity /10^4 kWh	Water Injection /10^4 m^3	Crude oil Output /10^4 ton
199401	389.29	65.32	146.87	18.78	3.29
199402	335.39	48.63	101.54	10.15	2.75
199403	253.55	52.09	102.06	14.69	3.13
199404	247.72	44.32	110.12	12.69	3.02
⋮	⋮	⋮	⋮	⋮	⋮
201609	482.64	70.89	156.98	13.89	4.94
201610	492.21	72.41	154.77	16.21	5.13
201611	507.28	102.69	156.00	16.47	4.55
201612	592.52	103.73	154.71	17.07	4.46

4.2. Experiment Parameters Setting

By changing the number of hidden layer nodes, transfer function, training function and learning function constantly, 10 different BP neural networks are generated. Through the simulation experiment, oil output electricity, gathering electricity, water injection electricity, and water injection are taken as the input and the crude oil output is used as the output of the forecasting model. The network structure is shown in Fig .2. A total of 266 sets of data were obtained as training samples and the last 10 groups were used as test samples to establish the Adaboost_BP prediction model, and the validity of the model was validated.
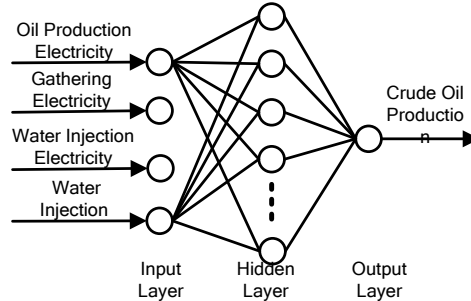

Fig. 2: BP neural network structure

4.3. Experiment Condition

This paper applies Matlab [9] language to complete the program, and the sample data are normalized by equation (6) so that the values are distributed between [0, 1]. Then by changing the various parameters of the BP network structure, we generate 10 different neural networks which are used as weak predictors in Adaboost_BP prediction model,where $y_{max} = 1$, $y_{min} = -1$, the normalized formula is as equation (6).

$$y_i = \frac{(y_{max} - y_{min})*(x_i - x_{min})}{(x_{max} - x_{min})} + y_{min} \tag{6}$$

5. Experiment Results and Analysis

In simulation experiment, firstly we used 10 sets of test data as input values in BP neural network to get prediction results, and calculated the absolute error and the relative error of each sample. Secondly, we used prediction results as input data of Adaboost_BP algorithm, calculated prediction error of weak predictors and adjusted distribution weights of test data set meanwhile got the predictor weights. Finally, we got 10 sets of strong prediction results. We compared both prediction results of the Adaboost_BP and the traditional neural network algorithm with the true output of the sample. In conclusion, the results show that the predicted results of Adaboost_BP are closer to the real values. Hence the prediction effect of Adaboost_BP is superior to the traditional neural network, which improves the accuracy of crude oil output forecast.

According to the experiment, 10 sets of prediction values are obtained by Adaboost_BP strong predictor. Meanwhile, the traditional neural network also generated 10 sets of weak prediction values, we calculated the absolute error and relative error of the two algorithms. The data analysis results are shown in TABLE II.

Table 2: Crude Oil Output and Influence Factors

Serial Number	Desire Values	BP Neural Network Algorithm			Adaboost_BP Algorithm		
		Prediction Values	*Absolute Error*	*Relative Error*	*Prediction Values*	*Absolute Error*	*Relative Error*
1	5.03	6.14	1.11	0.22	5.85	0.82	0.16
2	4.86	6.26	1.40	0.29	5.33	0.47	0.10
3	5.13	5.82	0.69	0.13	5.26	0.13	0.03
4	5.06	6.27	1.21	0.24	5.16	0.10	0.20
5	5.22	5.83	0.61	0.12	4.92	0.30	0.06
6	5.23	5.86	0.63	0.12	4.98	0.25	0.05
7	4.94	6.38	1.44	0.29	5.15	0.21	0.04
8	5.13	6.01	0.88	0.17	4.94	0.19	0.04
9	4.55	6.11	1.56	0.34	5.20	0.65	0.14
10	4.46	6.25	1.79	0.40	5.40	0.94	0.21

According to the data analysis in TABLE II, the average absolute error of the traditional neural network is 1.132, while the average absolute error of the Adaboost_BP prediction model is 0.406. The comparison of prediction results and desire values is shown in Fig.3. And the comparison of absolute error results is shown in Fig.4.
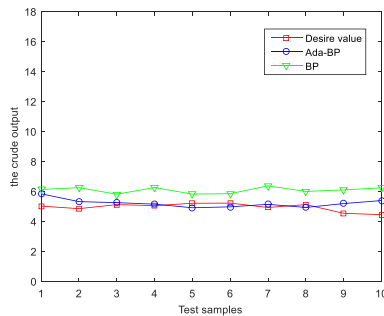

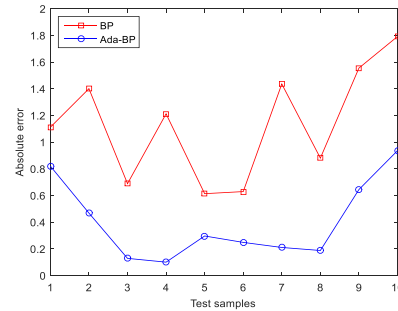Fig.3: Prediction results and desire values

Fig.4: Comparison of absolute error results

It can be observed from the absolute error comparison figure, that the absolute error of Adaboost_BP algorithm is lower than that of BP neural network. Thus the Adaboost_BP algorithm has higher prediction accuracy and better effect.

6. Conclusion

In order to resolve the limitation of over-fitting and local optimal solution of BP neural network, the paper proposed Adaboost_BP prediction model. According to simulation experiment of crude oil output, the Adaboost_BP algorithm prediction results show a lower absolute error than that of BP neural network, and the prediction accuracy is higher than that of BP neural network, which effectively improves prediction performance and has a certain application value in oil production system.

In the paper, there is still limitation in the prediction process. For example, the samples which are difficult to predict correctly may always be processed as the misclassified samples in the algorithm. This algorithm should be further studied in future research work.

7. References

[1] Xian Zhang, Shiping Li, Junqing Sun, Chao Tang. Dynamic Data Sequence Forecasting Base on Combined Model of Grey Neural Network. *Electronic Measurement Technology*. 2007, 30 (9): 60-61.

[2] Pan Ding, Fuzai Lv, Zhanqin Xiang. Intelligent Flaws Identification Method for Oil Casing Pipe Based on Wavelet Packet Decomposition and Support Vector Machine. *Journal of Iron and Steel Research*. 2012, **24** (5): 59-61.

[3] Shumway R H, and Stoffer D S. Time Series Analysis and Its Applications with R Examples. New York: Springer.2009.

[4] Lei Wang, Hongwei Gu, Hengshen Yao. Combined Forecasting Method of Oil and Gas Output Based on Artificial Neural Network. *Xinjiang Petroleum Geology*. 2005, 26 (5): 558-559.

[5] Zhigang Yao. Application of Improved Logistic Model in Crude Oil Output Forecast. *Petroleum Geology and Recovery Efficiency*. 2010, 17 (3): 93-94.

[6] Li Shuhui, Wunsch D C, Giesselmann M G, et al. Using neural networks to estimate wind turbine power generation. *IEEE Trans on Energy Conversion*. 2011, 6 (3): 277-281.

[7] Gaige Wang, Lihong Guo, Hong Duan, et al. Target threat assessment model and algorithm based on Elman - AdaBoost strong predictor. *Acta Electronica Sinica*. 2012, 5 (9): 903-905.

[8] Yijing Li, Haixiang Guo, Yanan Li, et al. a Boosting Based Ensemble Learning Algorithm in Imbalanced Data Classification. System Engineering-Theory & Practice. 2016, 36 (1): 191-193.

[9] Jinsong Liu, Qingyang Zhang, Xiaofeng Su, et al. Application of Matlab in automatic control theory experiments, *Experimental Technology and Management*. 2014, 31 (6): 139-140.