Gas Logging Data Normalization Processing Based on Rough Set Theory

Wu Xiao-yu, Zhang Fang-zhou, Li Qiang, Sun Rui-xue, Sun Chun-yu
College of Computer and Information Technology, Northeast Petroleum University, Daqing, Heilongjiang 163318, China.

Abstract
As everyone knows, gas logging data is affected by many factors, such as geological factors and drilling factors, The factors cause difference of gas logging data in different regions, and improve the difficult of gas logging data processing. So, selecting the parameters of gas logging data is very important. Aiming at the RBF neural network algorithm manages the gas logging data has instable and other faults, this thesis presentsa normalizationmethod based on rough set theory, using this method improves the training speed of RBF neural network algorithm manages the gas logging data. In order to verify the feasibility of this method , this thesis uses the gas logging data from Liao he oil field. Throughing the experimental results, thismethod can effectively improve the RBF neural network algorithm to manages the gas logging dataspeed.

1. Introduction

Gas logging is the basic method on oil and gas exploration, this method can directly acquirethedata of Oil and gas drilling engineering[1-4]. With the continuous improvement of the exploration degree on Liaohe oil field and exploration targets from the simple to the complex,Gas logging is accompanied by the progress of oil and gas field exploration and development needs and technology and gradually developed a logging technology while drilling, mainly collected during drilling in oil, gas and water display information and related parameters of the project is an important part of oil and gas exploration technology[5].Gas logging is at the wellhead of formation of oil and gas shows continuous monitoring technology, found in the fractured oil and gas, light oil and gas reservoir, condensate reservoir, continuity and sensitivity advantages of logging technology[6-7].It is one of the very important means of the discovery and evaluation of gas reservoirs. But as the serious impact on gas logging data from drilling, the next single, the sampling analysis, core drilling, drilling change, change of displacement, residual gas, mud property, degasser in fluid volume changes, the use of gas logging data interpretation layer standard and chart is difficult to establish, the traditional RBF neural network method for gas logging data processing oftraining time is longer, slower convergence speed defects, this thesis presents a based on rough set theory of the gas measured logging data normalization processing method, in order to improve the training speed of RBF network.

2. RBF neural network

RBF neural network is one of the high efficiency feed forward neural network, it has the characteristicsoffast learning speed,simple structure and so on.Below will be a detailed introduction to the network model of the RBF neural network.

RBF neural network is a kind of three layer static forward network, and the topological structure is divided into three layers [8-10]asit is shown in Figure 1, respectively input layer, hidden layer and output layer. The input layer transfers the signal to the hidden layer, using the Gauss function describes it, but the output layer only uses the Linear functiondescribes it. The hidden layer node

Manuscript received July 5, 2015
Manuscript revised July 20, 2015

function is mainly for the local response of the input signal, when the input signal transfers to the middle of the function, it will Produce larger output in hidden layer node.

So, the Radial basis function is a kind of Global approximation function, it has the Characteristic of fast learning speed.The Radial basis function are commonly used as the Gauss function, which is expressed as:

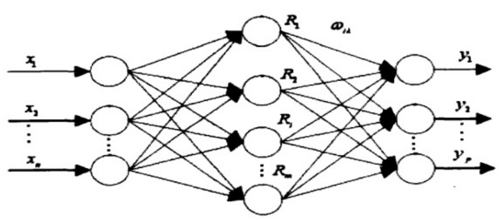

Figure 1

, i=1,2,3 The is theoutput of the hidden layer of the I node in the Network structure; X is input sample; Ciis the center of the i layer node with the same X dimension; Si is the Standard length; i is the Layer node variable; m is the number of hiddenlayer node. In the Radial basis function network, every hiddenlayer node has their own radial basis center vectorCi.Ciis T i=1,2,3,n. n is the Center number of the Network. The Net input of the hidden layer nodeis . In the radial basis function network the distance of the X of the input sample is represented by the distance of the center vector of the hidden layer. The output of each hidden layer node is [11-12].In other network hidden layer training is on the weight matrix regulation containing, but for the radial basis function network is different, in the radial basis function network is mainly for each hidden layer nodes choose the appropriate center vector. In the radial basis function network, the input layer and output layer are respectively realizedX→ αi（x）Nonlinear mapping and αi（x）→ykLinear mapping.

3. Normalization method based on Rough Set Theory

3.1 Rough set theory

Assuming the information system is K=(U, A), The domain U is a collection of samples, the Sample attribute set is A=C∪D, C is the attribute set of input variables in the sample. D is the attribute set of output Va is Each attribute a ∈ A Attribute value. Definition 1 (not distinguish relation) given information system k = (U, a), and for each subset B define an equivalence relation ind (b), the indiscernibility relation is referred to as, that is 2IND (b) = {(x, y) is well a in B (a (x) = a (y)} (1) among them, [x]B said sample x u a and B equivalence classes. Definition 2 (upper approximation and the lower approximation and boundary) given information system k = (U, a), X u is a collection of a sample, a B is the set of attributes, is B - approximation and upper approximation B and B boundary for B (x) ={x epsilon U:[X]B x}; B (x) x}. B-(X)={ x∈U:[x]B∩X≠∅} ; BNR(X)=R-(X)-R-(X).

3.2 Data normalization method

Using sample input as condition attributes, output as the decision attribute, each attribute discretization by the information system of the decision table after attribute reduction and rule reduction, respectively, are simplified decision table.

Rule 1: For information systems K= (U, A), the decision attribute set is defined as D={d}, Vd= 1, 2,... , R, the decision not to be the same as the decision of the value of Vd should be in the same class.

Rule 2: For information systems K= (U, A), the conditional attribute set is defined as C={c1, C2, C3,...... , cn}, the average energy of the P class sample is

$$Wp=\frac{1}{m}\sum_{j=1}^{m}\sum_{i=1}^{n} V^{P_j\ c_i\ 2}, p=1,2,3\ldots\ldots, r$$

V is the value of the attribute values corresponding to the CI input of the J class of the P class.

Rule 3: For the information system K= (U, A) condition attribute set is defined as C={c1, c2,... , cn} specifies the distance between the P Class J Q and the K class

$$dp^{j}q^{k} = \left(\sum_{i=1}^{n} (V^{p_j}{}_{c_i} - V^{q_k}{}_{c_i})^2\right)^{\frac{1}{2}} \quad p\neq q,$$

p,q=1,2,3,4,……r ; j=1,2,3,4, ……, m ; k=1,2,3,4, ……t ; V is the class of P of the j samples of C,, I input corresponding to the attribute value; V Q class of the K samples of C,,, I input corresponding to the attribute value.

Rule 4: P Of the j samples and Q in all samples is the minimum distance definition for DP Q =min DP Q; P = q; P, q = 1, 2,3,4...... R; j=1, 2, 4, 3,...... M; k=1, 2, 4, 3...... T;

Rule 5: the original input sample press type to carry out the expansion and preprocessing:

$$yp^{j}i = xp^{j}i*\left(1+\frac{S1}{dp_{j}q}\right) \qquad Wp>Wq$$

$$yp^{j}i = xp^{j}i*\left(1-\frac{S1}{dp_{j}q}\right) \qquad Wp<Wq$$

p≠q ; p, q=1, 2,3,4⋯⋯r ; j=1, 2, 3, 4, ⋯⋯m ; k=1, 2, 3, 4⋯⋯t ;

Rule 6: the input of the original sample is normalized by theenergy type.

$$Zi(l)=\frac{y_i(l)}{\max\limits_{l}(y_i(l))}$$

yi(l) isthe i of the sample L is represented by the extension, Zi(l) is the i input of the sample L is expressed by the normalized sample.

4. Gas Logging Data Normalization Processing

4.1 The analysis of Gas logging datainfluence factors

Gas content depends on the degassing apparatus. As the degassing apparatus is using for separating the gas from drilling fluid. At present we usually use the basic electric degassing device. It is the degassing efficiency is not high, but also by the drilling fluid displacement, the height of the liquid level and the influence of external factors such as wind speed.

The influence on the interpretation of the evaluation is in the following areas:

1.the amount of gas extraction is insufficient, resulting in a group of points is not correct;

2.the change of the atmospheric environment, so that the extraction efficiency of different;

3.gas accumulation in the degassing device, resulting in abnormal high (gas test value of more than 100%) of the gas plane.

4. The air is sucked into the trap, which has a negative effect on the measurement of the non hydrocarbon gases (CO_2 and N_2);

5.the degassing device is far, making a long response time.

Through the analysis of the influencing factors and to the site using the gas detector test, the table below is to test data and results (table 3-1 and 3-2): 3000 ml of 1% standard gas sample experiment, as shown in table 3-1 this table reflects the three instruments at different points of time delay, from time difference can be seen, they use the sample pump and pipelines are not the same. We know that the drilling fluid continuous degassing device out by pumping gas samples to identify the internal instrument. Degasser off gas speed can not meet the sample pump pumping speed, most of the time is to add a certain amount of air from the air inlet of the deaerator, in order to meet the sample pump as the pumping capacity. The whole process is degasser prolapse gas through the sample pump swabbing, adding dilution air, then the shunt sampling to total hydrocarbon and chromatographic analysis. So, gas logging instrument of the detected gas content, and not degasser disengages from the drilling fluid gas percent content, but have added a certain amount of airContent. As a result, the differences in the results measured by the instrument are shown in table 3-2.

Table3-1 A standard gas sample time 3000ml, 1%

Mud logging team	Instrument type	Sample pump	Degassing device	Pipeline delay
Tenth team	SK-II	1'25"	3'03"	4'
Ninth team	SK-II	50"	1'02"	59"
Fifth team	SK 2000c	1'16"	5'23"	2'57"

Table3-2 A standard gas sample gas 3000ml, 1% of the measured value

Mud logging team	Instrument type	Total hydrocarbon value			Group score	
		Sample pump	Degassing device		Sample pump	Degassing device
Tenth team	SK-II	50mv	30mv	40	30mvl	25mv
Ninth team	SK-II	55mv	20mv	63.6	25mv	10mv
Fifth team	SK 2000c	140mv	100mv	28.6	120mv	85mv

From the above analysis, the following conclusions can be obtained:

(1) Also vary with an instrument to collect data, this difference is mainly from the environment and people, such as: using a sample of pipeline volume caused pipeline time delay is not the same, gas logging value is also different.

(2) Different instrument difference is relatively large, from table 3-1, 3-2 in the data table can clearly see the difference.

4.2 Normalization method of gas logging data

The normalized method steps are as follows: (1) the input variables of the sample are the condition attributes, the output quantity is the decision attribute, and the information system decision table is obtained. (2) attribute reduction and rule reduction are performed respectively, and the simplified decision table is obtained. (3) the normalized processing of the raw gas log sample according to the rule 1 to rule 6. (4) the RBF neural network is trained by using the normalized learning sample. (5) the RBF neural network is verified by using the normalized test sample.

Case analysis

The data of gas logging data were processed by using the normalized processing method, including the parameters of the whole hydrocarbon and the components (C1, C2, C3).

Table 3-2-1 and table 3-2-2 are the data table of the total hydrocarbon value of in the Liaohe oil field. Figure 3-2-1 and figure 3-2-2 is a comparison chart of the total hydrocarbon value of the gas test data in Liaohe oil field. From the whole hydrocarbon data is in disorder and disorder of the whole hydrocarbon data before processing. The oil and gas water layer can be divided into oil and gas layer basically after the normalization process, and the effect is very obvious.

Table3-2-1 total hydrocarbon values of gas test data were uncorrected

Gas data		Gas and water layer		Gas reservoir		Gas bearing stratum	
Porosity	Numerical	Porosity	Numerical	Porosity	Numerical	Porosity	Numerical
11.31	0.24	1.57	3.54	5.85	1.47	3.84	1.67
7.91	0.35	2.43	0.26	4.03	1.52	1.97	2.25

Continued table 3-2-1

5.85	0.47	3.55	4.31	7.87	1.49	4.25	2.38
5.89	0.61	8.93	5.24	10.12	1.54	7.98	2.44
5.97	0.71	7.65	0.98	7.97	1.66	7.85	3.86
7.77	5.61	5.44	3.67	3.98	2.25	3.77	4.35
4.53	7.42	6.53	0.84	2.15	3.87	4.01	4.46
11.13	11.54	9.83	5.29	1.23	3.94	3.86	6.26
8.53	9.78	4.32	0.78	4.13	5.65	4.24	7.75
6.01	10.12	5.96	1.23	8.07	7.56	5.94	8.34
8.15	9.78	6.12	5.54	5.84	7.98	6.23	8.42
7.54	9.34	11.48	8.94	5.97	8.66	9.84	9.17

Table3-2-2Total hydrocarbon value of gas test data is corrected

Gas data		Gas and water layer		Gas reservoir		Gas bearing stratum	
Porosity	Numerical	Porosity	Numerical	Porosity	Numerical	Porosity	Numerical
3.84	4.78	2.56	3.48	4.13	3.98	3.96	4.89
4.34	8.41	3.15	4.24	4.96	5.15	4.13	8.74
4.97	8.68	3.55	4.31	3.89	6.24	4.97	8.88
6.12	9.34	4.61	5.24	7.23	5.87	6.17	9.14
5.97	9.85	4.87	5.14	6.65	6.53	6.29	9.38
6.76	11.54	5.10	4.98	7.92	7.02	7.64	11.56
7.43	11.42	5.65	5.39	8.24	8.17	7.89	11.47
7.95	12.31	6.23	5.89	8.13	8.44	7.98	12.54
8.53	13.56	7.12	6.54	8.28	8.56	8.34	13.36
9.56	14.58	7.84	6.32	9.78	9.06	9.86	14.51
11.25	15.34	8.01	7.55	9.95	9.74	11.89	15.25
12.36	15.88	9.23	8.94	10.14	12.33	12.31	15.74

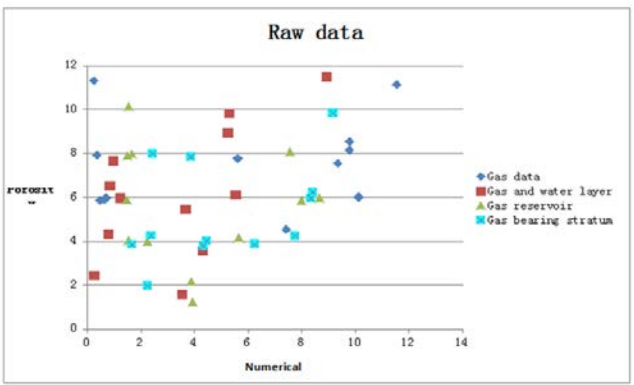

Figure 3-2-1 total hydrocarbon values of gas test data were uncorrected

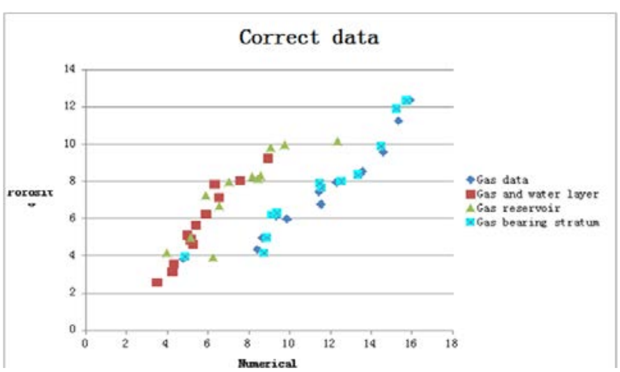

Figure 3-2-2 total hydrocarbon value of gas test data after correction

5. RBF Neural Network Training Data

5.1 The design of RBF neural network for gas logging data

(1)The selection of network parameters

RBF neural network is composed of input layer, hidden layer and output layer. Determine the number of input layer, hidden layer and output layer units, and the selection of initial weight is very important. According to the actual needs of this study, the number of nodes in the input layer, output layer and BP neural network are the same, and the number of hidden layer nodes and the number of nodes in the input layer are the same as that of the neural network. Since the system is nonlinear, the initial value of the study is to achieve the local minimum. Whether it can converge and the length of training time is very large. If the initial value is too large, the weighted input is easy to fall into the saturation region of the activation function, which makes the network adjustment process may fall into a standstill. So always want after the initial weights of each neuron output values are close to zero, which can ensure the weights of each neuron can be in their activation function changes the maximum point is adjusted for initial weight value is usually taken as the interval [1, 1] between the random number.

(2) transfer function selection

The transfer function of the hidden layer is the Gauss kernel function, and the transfer function of the output layer is linear function.

(3) input and output vector design

After obtaining input and output variables, we must deal with the normalized method which is mentioned in this paper. After the data of the normalized processing, it is easier to train and study the neural network.

5.2 RBF neural network training and simulation

RBF neural network method is used to train the data of gas logging data after normalization. MATLAB neural network toolbox of RBF neural network to create a function for the Newrbe (), the network is the creation of the process is the training process, the error has been created by the network 0. The most important parameter in the training process is the radial basis function of the distribution constant Spread. The greater the Spread, the more smooth the network's prediction performance. But it is not the bigger the better, too large Spread may lead to the calculation of the problem, where the first set to 1, and then to 0.2 of the interval. Different Spread values, the network training error is different. From the graph, we can see that the training time of the data training time is shorter than that of the normalized data set.

6. Conclusions

In this thesis, based on the theory of rough set, the method of normalized gas logging data processing is proposed, which is based on RBF neural network. Firstly, the rough set theory of original sample data structure reduction, according to the properties of certain sample input value of minimum distance between attribute values of other types of sample input sample need to determine the proportion of telescopic, the scaled sample normalization using samples normalized for neural network training. The simulation results show that the method can significantly shorten the training time and improve the efficiency of the RBF neural network processing gas logging data. The method is feasible.

References

[1] QiuYinan. The methodology of petroleum development geology (two) [J]..1996 petroleum exploration and development, 23 (3): 48-51.

[2] Luo handsome. Protection of oil reservoir in oilfield development and production [M]. Beijing, oil industry press.1996, 0 (8): 1-75M.Nuclear magnetic resonance and formation porosity [J]. Petrophysics

[3] Appel, 2004,45 (3): 296-307.W B Prison, S J.The Effect of Fractional Wettability on the Archie Saturation

[4] MorganAnnual Logging Symposium Midland, TX, May, Exponent[R].Trans.SPWLA.Fifth13-15, 1964 B. Sec.

[5] Liu Zhongyang, XuDongwei, Wang Ning, Zhang Haichuan. [J]. based on the neural network model of ozone concentration soft measurement study RBF Journal of Dalian University of Technology. 2010 (06)

[6] Yang Chunhua, Liu Junxia, Wang Yalin. Research on the mathematical model of the spiral classification process and application of [J].computer engineering and applications. 2010 (04)

[7] Zhou Yong, Hu Zhonggong. RBF neural network theory and its application in control [J]. Journal of Wuhan University of Science and Engineering 2007 (05).

[8] Dong Changhong.Matlab neural network and application [M]. Beijing: National Defense Industry Press, 2005Julong.Spectrum mapping in grey the- ory[J].The Journal of Grey System

[9] Deng, 2000 (2): 116-117J L.Greyforecastingmodel[M]Greysystem.Beijing:China Ocean Press 1988:54-69,

[10] DengJL.Properties of the grey forecasting model GM [M] Greysystem.Beijing:China Ocean Press

[11] Zhao Zhenyong, Wang Li, Wang Baohua, Yang Benjuan. Research on improved genetic algorithm [J]. (S2):.2006 (): 9-12

[12] Wang Wei, CaiLianhong. [J]. Computer Engineering, based on rough set theory, 2001,27 (5): 65-67.