ISSN: 2710-0723

Intelligent Method for Lithology Determination in Drill-following Formations based on GA-GRNN Model

He Zhang*, Jinhua Yang

School of Mechanical Engineering School Southwest Petroleum University, Chengdu, 610500, China

*Corresponding Author

Abstract

In oil and gas drilling operations, formation rock lithology identification is of great significance to the whole oil and gas drilling operation process. Aiming at the problems of poor real-time performance and insufficient model accuracy caused by the poor data quality of traditional lithology identification methods in the drilling process, this study proposes an intelligent identification method based on the optimization of generalized regression neural network (GRNN) by genetic algorithm (GA), which firstly performs the dimensionality reduction of the logging element data by correlation analysis and principal component analysis, and then searches the global optimal values of the smoothing parameters of generalized regression neural network (GRNN) by genetic algorithm (GA) to train the optimal identification model. The experimental results show that the recognition accuracy of the GA-GRNN model is improved by 12.1% and 7.5% to 92.9% compared with the BP neural network and standard GRNN model, respectively., and the field application verifies the engineering applicability of this method in real-time lithology recognition in X oilfield with the accuracy of 94% in the recognition of complex formations, which provides a new technological means for intelligent drilling decision-making.

Keywords

GA-GRNN Model; Lithology Intelligent Identification; Correlation Analysis; Principal Component Analysis.

1. Introduction

With the depletion of tectonic oil and gas resources, tight sandstone reservoirs have gradually become the focus of attention of oil and gas engineers. In order to effectively explore and develop such reservoirs, accurate identification of lithology is the key to reserve calculation and is an important task. Traditional reservoir identification methods mainly include petrophysical templates and crossplot techniques based on core or log responses. However, these methods tend to be expensive and time-consuming, and may be hampered by low signal-to-noise ratios (SNR) and prominent non-linear data.

In recent years, elemental logging, drilling logging and other drilling stratigraphic recording technologies have been developing rapidly, providing a large amount of reliable data for rock lithology identification of drilling stratigraphic rocks in the field, while intelligent algorithms are also widely used in the field of lithology identification with the development of artificial intelligence technology. In 2015, M. A. Sebtosheikh et al.[1] used Support Vector Machine method (SVM) and optimised the parameters of the SVM using a network search calculation algorithm to normalize the stratigraphic lithology set and select the attributes before identifying and predicting the stratigraphic lithology set of a non-homogeneous carbonate salt reservoir in Iran. In 2017, Sasmita Sahoo et al.[2] compared the results of traditional artificial

ISSN: 2710-0723

neural network based on gradient descent momentum training, artificial neural network method based on genetic algorithm and artificial neural network method based on Genetic Algorithm for identifying the formation lithology of complex multilayered aquifers in India, and the results showed that artificial neural network method based on genetic algorithm can better identify the formation lithology better. In 2018, Camila Martins Saporetti et al. [3]used six machine learning methods such as K nearest neighbour (KNN), Decision Tree (DT), Support Vector Machine (SVM), Random Forest (RF), and Multi-Layer Perceptron (MLP) to classify the geological carbonatite-siliciclastic lithology data of the South Provence Basin, of which Decision Tree (DT), Support Vector Machine (SVM), and Random Forest (RF) have better recognition results. In 2019, Solomon Asante-Okyere et al. [4]used K-means and Gaussian Mixture Model (GMM) to construct a more accurate and fast Gradient-Booster Machine (GBM) lithology identification model for stratigraphic lithologies in the southern basin of the South Yellow Sea, which improves the identification accuracy of mudstones and siltstones. In 2022, Soumi Chaki et al. [5]proposed a probabilistic neural network (PNN) model to identify the lithology of four types of formations: sandstone, muddy sandstone, sandy shale, and shale, using seismic attributes and logging data from boreholes.

The methods proposed in the above literature have some improvement for the lithology recognition accuracy, but the parameters are too few to extract the lithology information most accurately or are not applicable to the formation with drilling, in order to improve the accuracy of lithology recognition, this study utilizes Mg, Al, Mn, Fe and other recording elements for data fusion, analyzes the correlation of the data, and utilizes the principal component analysis method for feature extraction and fusion of the elemental data, and combines the generalized Regression neural network and genetic optimization algorithm are combined to establish an intelligent lithology identification model for multi-source information stratum rocks.

2. Intelligent Method of Lithology Recognition in Drill-following Stratum based on GA-GRNN Model

2.1. Generalised Neural Network

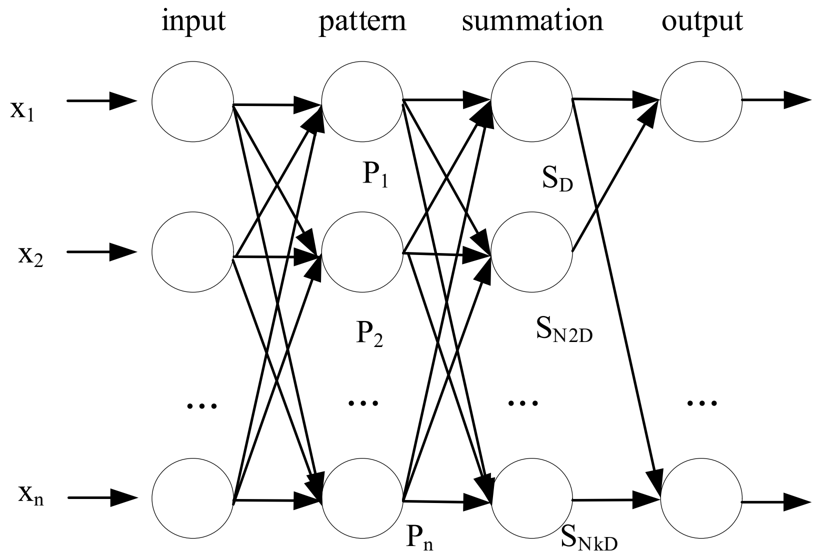

Figure 1. GRNN model structure diagram

Generalised regression neural network (GRNN)[6] is a network model based on radial basis neural network, which requires fewer samples, has high nonlinear approximation ability and high stability, and is widely used in the field of pattern recognition. The structure of generalised

ISSN: 2710-0723

regression neural network mainly includes input layer, pattern layer, summation layer and output layer. As shown in Figure 1, the network training sample input dimension is N, and the input sample vector can be expressed as X=[x_1,x_2,...,x_n], and the network training sample output dimension is K, and the output vector can be expressed as Y=[y_1,y_2,...,y_k].

The input layer of the training samples in the input layer is the number of samples as input parameters, n. The number of input parameters is equal to the number of neurons, and the input layer is able to pass the training set data directly to the pattern layer.

A pattern layer neuron is a radial basis neuron, which is a nonlinear function. The radial basis neuron receives the input vector X directly from the input layer, and then calculates the Euclidean distance dt between the input vector X and the centre vector C of this neuron, C=[c_1,c_2,...,c_n]. The formula for the Euclidean distance dt is shown in equation (1):

$$dt = \|X - C\| = \sqrt{\sum_{i=1}^{n}(x_i - c_i)^2} \tag{1}$$

The Euclidean distance is then multiplied by a threshold representing the sensitivity of the neuron and transmitted to the activation function of that neuron with the equation R(n):

$$R(n) = e^{-n^2} \tag{2}$$

The activation function of radial basis is also called radial basis function, and radial basis function is commonly used as Gaussian kernel function, then the activation function formula can also be expressed as:

$$R(X - C) = \exp\left(-\frac{(X-C^T)^T(X-C^T)}{2\sigma^2}\right) \tag{3}$$

The output vector of the pattern layer is P=[p_1,p_2,...,p_n].

The summation layer contains two types of neurons, one type of neuron does an arithmetic summation process on all the neuron outputs of the previous layer to get the output:

$$S_n = \sum_{j=1}^{n} p_j \tag{4}$$

Another neuron does weighted summation of all neuron outputs of the previous layer with neuron connection weights b_j and the output is:

$$S_{bn} = \sum_{j=1}^{n} b_j p_j \tag{5}$$

The output dimension in the output layer is the number of neurons and also the number of output parameters in the training sample k. The output of the output layer is:

$$y_k = \frac{S_{bn}}{S_n} \tag{6}$$

2.2. Genetic Optimisation Algorithm

Genetic Algorithm (GA) [7]is an algorithm proposed by J.H. Holland. The algorithm is the basis of intelligent optimisation algorithms through the three basic operations of chromosome

ISSN: 2710-0723

selection, crossover and mutation and thus the simulation of recombination, mutation and mutation behaviours of genetic genes.

The genetic algorithm is a simulation of the laws of biological evolution and is a global optimisation seeking algorithm. The steps of the genetic algorithm are as follows:

(1) Coding. Binary coding of the problem to be studied as a chromosome;

(2) Population initialisation. Randomly generate the initial population;

(3) Calculation of fitness value. Calculate the fitness value of each individual in the population;

(4) Individual selection. Select the best individuals with the top fitness value in the population and delete the individuals with the bottom fitness value;

(5) Chromosome crossover. Make random parts of chromosomes to produce crossover to transpose and increase the average fitness value of the population, and perform a random search globally;

(6) Gene mutation. To produce changes in the smallest units that make up a gene according to a certain probability, to perform a random search locally;

(7) Repeat steps (3) to (6) until the final condition is satisfied to approximate the global optimal solution.

(8) Output the optimal solution.

2.3. Establishment of Intelligent Recognition Model of Lithology in Drill-Following Formation Based on GA-GRNN Model

The flow of the GA-GRNN lithology intelligent judgement model of the formation with drilling constructed in this study is as follows, the flowchart is shown in Figure 2:

(1) Data analysis: correlation analysis and principal component analysis.

(2) GRNN network construction: initialize the structure of the generalized regression neural network and determine the number of nodes in the input, hidden and output layers.

(3) GA parameter setting: set the parameters of the genetic algorithm, such as population size, iteration number, crossover probability, mutation probability and so on.

(4) Initial population generation: randomly generate the initial population, each individual represents a smoothing factor of GRNN.

(5) Adaptation Evaluation: Apply the parameters corresponding to each individual to the GRNN, use the training data for training, and calculate the adaptation value from the test data.

(6) Selection operation: select the good individuals based on the fitness value for generating the next generation.

(7) Crossover operation: perform crossover operation on the selected individuals to generate new individuals.

(8) Mutation operation: perform mutation operation on the newly generated individuals to increase the diversity of the population.

(9) Judgment of termination condition: Judge whether the termination condition is satisfied, if it is satisfied then stop iteration, otherwise return to step (5).

(10) Determination of optimal parameters: Select the parameters corresponding to the individuals with the best fitness values as the final parameters of the GRNN.

(11) Model Evaluation: Construct the GRNN model using the finalized parameters and evaluate it using the test data.

ISSN: 2710-0723

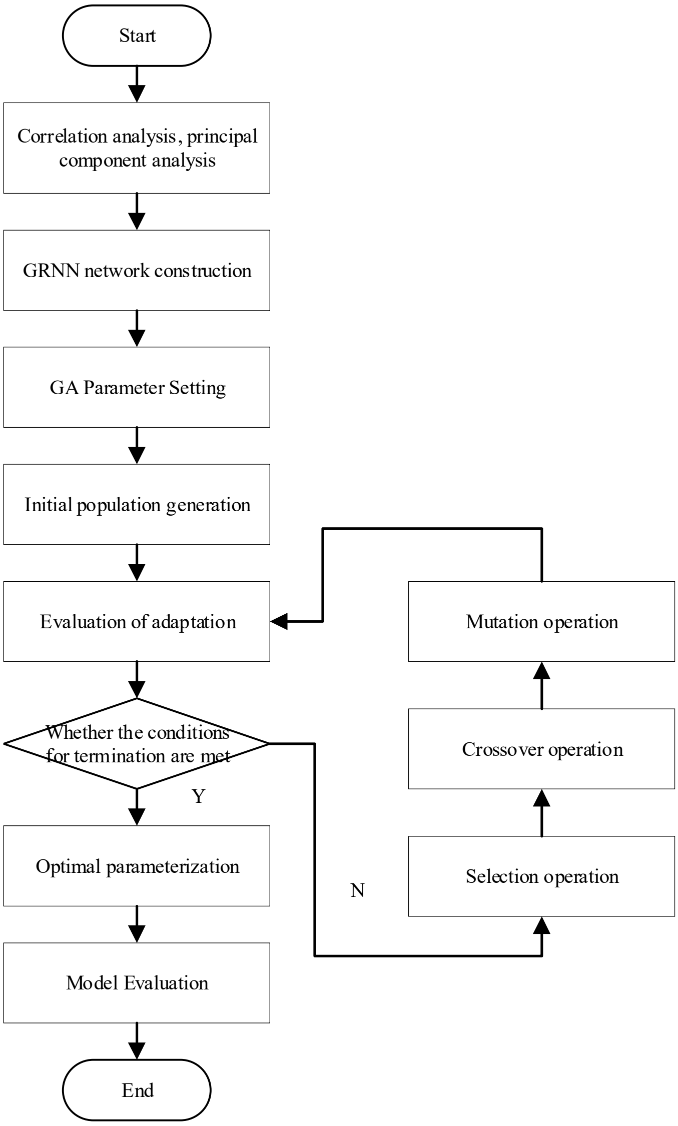

Figure 2. GA-GRNN model flowchart

3. Experimental Results and Analysis

In this paper, the source area of the research data is a dataset containing logging elements from a well site in the tectonic zone of a depression area in the Bohai Bay Basin, and initially 10 Na, Mg, Al, Si, P, S, Cl, K, Ca, and Ba were selected as the characteristic parameters. The rock types were identified as fine sandstone, conglomerate-bearing fine sandstone, sandy conglomerate, conglomerate-bearing medium sandstone, and siltstone, and the labels were set to 1,2,3,4,5, respectively.

3.1. Pearson Correlation Analysis

In this paper, we use Pearson correlation coefficient analysis method to analyze the correlation of recorded well element data, Pearson correlation coefficient can determine whether two variables are correlated with each other or not, the coefficient value ranges from [-1,1]. The heat map of elemental correlation coefficients is shown in Figure 3.

ISSN: 2710-0723

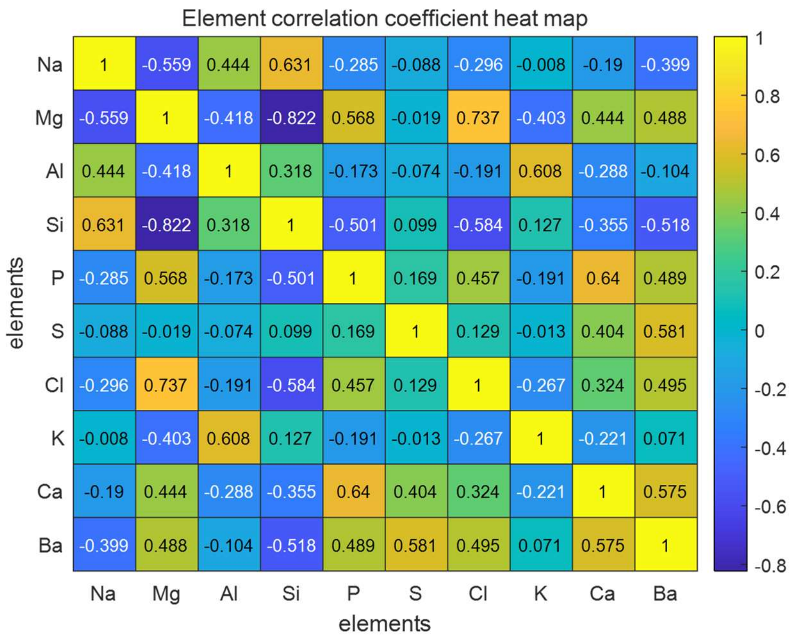

Figure 3. Heat map of element-related systems

Pearson correlation coefficient is negative value means negative correlation, positive value is positive correlation, and the closer to 1 the more significant correlation. For example, Na and Mg have a more significant negative correlation, and Na and Si have a more significant positive correlation.

3.2. Principal Component Analysis

Principal Component Analysis (PCA)[8] compresses the data and extracts the characteristic information of the data. The principal component variance table is shown in Table 1:

Table 1. Total variance interpretation table

principal component	Initial eigenvalue		
	total	Percentage of variance	Cumulative/%
1	7.069	41.583	41.583
2	2.722	16.014	57.597
3	1.766	10.386	67.983
4	1.489	8.756	76.739
5	0.998	5.868	82.608

The total variance interpretation table shows that the cumulative variance percentage of the first five principal components reaches 82.608%, which can basically restore the feature information of the original data, and the dimension of the input data is reduced from 10 dimensions to 5 dimensions, which greatly improves the efficiency of network training and learning.

3.3. Lithology Intelligent Recognition Model Results and Analysis

Genetic algorithm is used to determine the smoothing factor of GRNN, set the population as 50, the number of iterations as 100, and the range of smoothing factor is set as (0.01,10), which is

ISSN: 2710-0723

gradually adjusted according to the validation results. In order to verify the validity of the model, 800 samples are now used for the model input data, which are divided according to the training set: test set as 7:3. The recognition results of the GA-GRNN model are compared with those of the GRNN model and BP neural network to calculate the model recognition accuracy. The comparison of the recognition results is shown in Figure. 4,5,6.

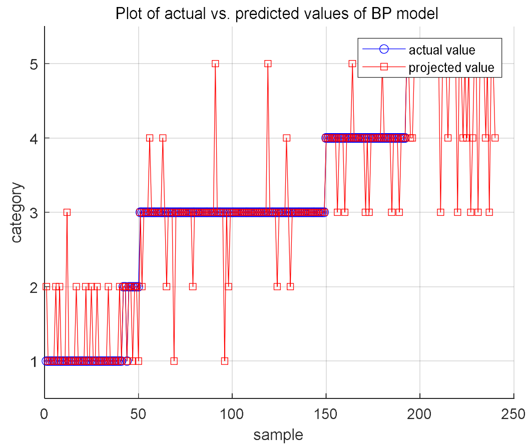

Figure 4. BP model result chart

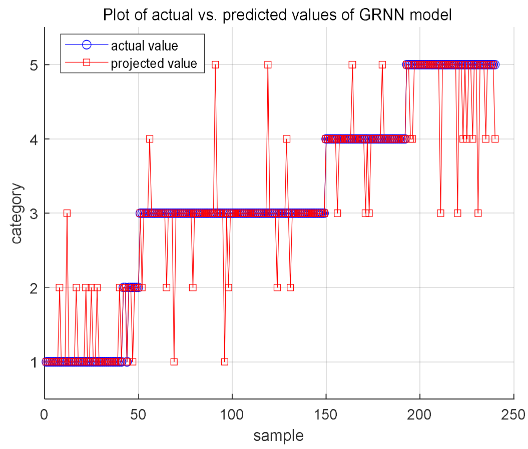

Figure 5. GRNN model result chart

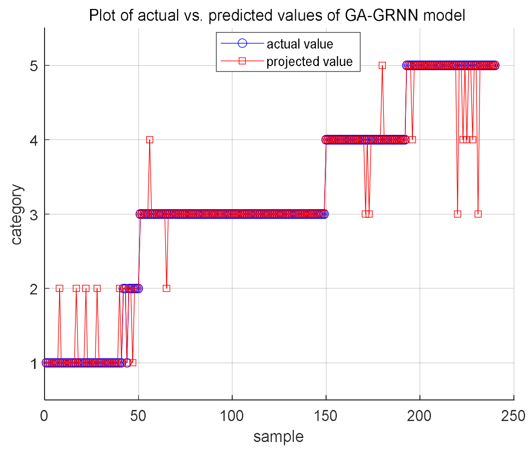

Figure 6. GA-GRNN model result chart

ISSN: 2710-0723

Table 2. Comparison table of model accuracy

Model	Accuracy(%)
BP	80.8%
GRNN	85.4%
GA-GRNN	92.9%

As can be seen from the figure and table, the recognition accuracy of the GA-GRNN model is 92.9%, while the accuracy of the BP and GRNN models is 80.8% and 85.4%, respectively. Therefore the GA-GRNN model is more effective for lithology recognition.

4. Field Application Results and Analysis

The trained discrimination model is applied to the lithology identification of the formation between 3500-3700m in well X. Every 4 meters is identified in turn, i.e., there are 50 points to be identified, and the identification results are shown in Fig.

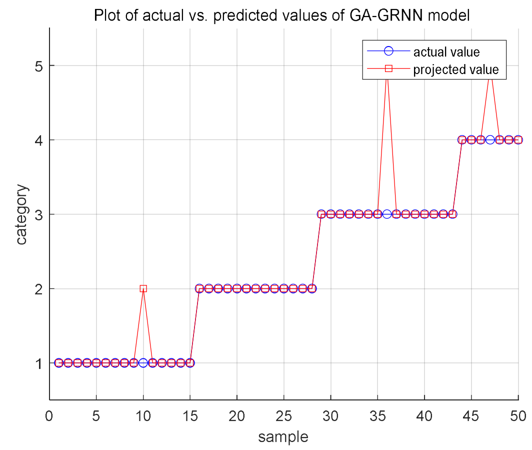

Figure 7. GA-GRNN model application result chart

As can be seen from the figure, out of 50 recognition points, only 3 points are recognized incorrectly, so the recognition accuracy is 94%, which meets the field application requirements.

5. Conclusion

(1) Preliminarily select the elemental data related to rock types, and use correlation analysis and principal component analysis to reduce the dimension of input data from 10 to 5 dimensions, and the model structure is simplified.

(2) Using genetic algorithm to find the optimal smooth factor of the generalized neural network to get the optimal model, and test the test set, and compare the model recognition results with the recognition results of the BP neural network and the unoptimized GRNN model, and the results show that the GA-GRNN model has the highest prediction accuracy.

(3) When the model is applied to the field lithology recognition, the accuracy rate is achieved, which meets the practical application requirements of lithology recognition in the formation with drilling.

References

[1] Sebtosheikh M A, Salehi A. Lithology prediction by support vector classifiers using inverted seismic Attributes data and petrophysical Logs as a new approach and Investigation of training data set size

ISSN: 2710-0723

effect on Its performance in a heterogeneous carbonate reservoir[J]. Journal of Petroleum Science & Engineering, 2015, 134 ,143-149.

[2] Sahoo S, Jha M K. Pattern recognition in lithology classification: modeling using neural networks, self-organizing maps and genetic algorithms[J]. Hydrogeology Journal, 2017, 25 (2), 311-330.

[3] Camila Martins Saporetti,Leonardo Goliatt da Fonseca,Egberto Pereira,Leonardo Costa de Oliveira. Machine learning approaches for petrographic classification of carbonate-siliciclastic rocks using well logs and textural information[J]. Journal of Applied Geophysics,2018, 155, 217-225.

[4] Solomon Asante-Okyere,Chuanbo Shen,Yao Yevenyo Ziggah,Mercy Moses Rulegeya,Xiangfeng Zhu. A Novel Hybrid Technique of Integrating Gradient-Boosted Machine and Clustering Algorithms for Lithology Classification[J]. Natural Resources Research,2019, 29 (4), 2257-2273.

[5] Santos D, Roisenberg M, Nascimento M. Deep Recurrent Neural Networks Approach to Sedimentary Facies Classification Using Well Logs[J]. IEEE Geoscience and Remote Sensing Letters, 2021, 99:1-5.

[6] Research on GRNN Prediction Model for Acoustic Quality of Yueqin Based on Wood Vibration Characteristics.[J].Forest Engineering,2024,Vol.40(4): 160-167.

[7] JIANG Yujin;YANG Qi;YANG Mao;LI Bo;YAN Zhongjiang.Fairness guarantee algorithm of high-density WLAN based on neural network and genetic algorithm[J].Xibei Gongye Daxue Xuebao,2023,Vol.41(5): 887-894.

[8] Nassim Nicholas TalebCA1;Pierre Zalloua;Khaled Elbassioni;Haralampos Hatzikirou;Andreas Henschel;Daniel E. Platt.Informational rescaling of PCA maps with application to genetic distance[J].Computational and Structural Biotechnology Journal,2025,Vol.27: 48-56.