Scholar Publishing Group
Machine Learning Theory and Practice
https://doi.org/10.38007/ML.2023.040107
ISSN 2790-0983 Vol. 4, Issue 1: 52-60

Prediction Method of Sand Body Lithology Based on Machine Learning Algorithm and Dual Optimization of Attribute Characteristics

Xiangru Hou*

Department of Information Engineering, Heilongjiang International University, Harbin 150025, China

houxiangru@hiu.net.cn

**corresponding author*

Keywords: Machine Learning Algorithm, Attribute Characteristics, Sand Body Lithology, Prediction Method

Abstract: With the development of science and technology, the prediction of earthquake disasters in China has made some progress. For some complex environments, target recognition gradually depends on machine learning algorithms. In order to accurately detect oil and gas fields, this paper intends to analyze and study the lithology of sand bodies before and after the earthquake. This paper intends to use machine learning algorithm to study the prediction method of sand body lithology, in order to improve the accuracy of identification and prediction. In this paper, experimental design and algorithm comparison are mainly used to analyze the scientific and technological use of sand body lithology prediction methods. The experimental results show that the ELLA algorithm proposed in this paper performs well in the classification and prediction of sand body lithology, and the error is less than 15%. Therefore, lithology prediction of sand body based on attribute characteristics can be calculated by ELLA algorithm.

1. Introduction

Economic development has caused serious damage to the ecological environment. In order to alleviate this problem, the state has issued a series of policies, one of which is to use machine learning algorithm and decision tree technology to predict sand lithology[1-2]. On the one hand, it can improve the exploration speed of oil and gas fields, on the other hand, it can increase economic benefits and reduce pollution. Through a series of measures to achieve the goal of protecting ecological balance and sustainable developmentp[3-4]. Therefore, the state began to strengthen the regulation of sand and gravel, and achieved results.

The deep underground roadway is in a complex coupling environment of high ground stress and particle ratio, The mechanical properties of roadway surrounding rock will appear obvious weakening phenomenon Gongqin Y, so the influence of rock damage effect should be considered in the research process of the stability of roadway surrounding rock under the "particle condition". Through fish language, the influence law of particle rate on rock mechanical properties is introduced into FLAC3D numerical simulation software, which can further analyze the influence law of rock damage effect on the stability of surrounding rock [5] 。 The horizontal disturbance of shield tunneling to the slate stratum is greater than the axial one, and the maximum lateral displacement of each measuring point is about 1.1~2.6 times of the axial one. According to the change rule of soil stress with the distance from the cut to the section, Jing LJ divided the stress disturbance of shield tunneling to the slate formation into four stages: slow reduction, fluctuation, slow change and recovery. Due to the different degree of shield disturbance, the segments of the soil above the tunnel and the soil on both sides are different [6]. The tight thin sandstone natural gas reservoir of the Upper Paleozoic in the Longdong area in the southwest of the Ordos Basin is deeply buried, with low porosity and permeability, and small difference in wave impedance between sandstone and mudstone. This paper conducts petrophysical analysis based on logging data, selects the improved Wang Z model suitable for the area and optimizes the model parameters, establishes the quantitative interpretation chart of petrophysics under different lithology, porosity and gas saturation conditions, and optimizes the sensitive elastic parameters of the reservoir. The pre-stack elastic parameter inversion and intersection analysis should be used to predict the reservoir and gas bearing capacity [7]. The above research and analysis are not correct and need to be improved.

In this paper, we first study the lithologic identification method, which involves the common wavelet transform and cluster analysis. Secondly, the seismic attribute of sandstone is described. Then, the lithologic prediction method of roof sandstone is proposed. Finally, by building a deep learning model for attribute feature optimization, the application of five different algorithms in sand body lithology is analyzed by using experimental simulation, and a conclusion is drawn.

2. Sand Body Lithology Prediction Based on Machine Learning Algorithm and Dual Optimization of Attribute Characteristics

2.1. Lithology Identification Method

Wavelet transform is used to segment and refine the signal at multiple scales to achieve multi-resolution analysis of the signal. Wavelet transform is used for stratigraphic sequence distribution and lithology identification. Wavelet transform can be used to divide lithofacies. However, wavelet analysis is difficult to separate signal from noise in the case of low signal-to-noise ratio [8-9].

Cluster analysis is a process of classifying sample data sets according to their similarity. In the research of structural carbon identification methods, the longest distance method is used for classification. By calculating the distance between two pairs of all samples in two categories, take the maximum distance as the distance between two categories, and then combine the nearest two categories, and so on, until all samples are classified into one category. This ensures a high degree of similarity between samples of the same category and samples of different categories [10-11].

As an important part of petrophysical analysis, fluid replacement is one of the important tools for fluid identification and reservoir quantitative prediction. Fluid displacement is a petrophysical method to simulate P-wave and S-wave velocities of various saturated fluids under special reservoir conditions such as temperature, pressure, porosity and rock matrix type [12].

Attribute technology combines multiple seismic attributes and can directly predict logging attributes, thus highlighting effective information in seismic data. Reduce the diversity of inversion results and avoid model dependence.

In seismic lithologic interpretation, multiple seismic attributes are often used for linear regression to obtain the calculated logging curve value:

$$\mathrm{W}(v) = \psi_0 + \psi_1 X_1(v) + \psi_2 X_2(v) + \Lambda + \psi_m X_m(v) \tag{1}$$

Wherein, W (v) represents logging curve value and $X_o(v)$ represents seismic attribute. Each weight factor can be calculated by the least square method:

$$F^2 = \frac{1}{M}\sum_{o=1}^{M}(W_o - \psi_0 - \psi_1 X_{1o} - \psi_2 X_{2o} - \Lambda - \psi_m X_{mo}) \tag{2}$$

Using linear fitting seismic attributes to predict logging attributes is a point to point mapping. There are many kinds of seismic attributes, so it is particularly important to screen the combination of seismic attributes that are sensitive to the target prediction value. The prediction error of attribute combination on target logging curve is inversely proportional to the number of attributes. The more attributes, the smaller the error. However, too many attributes will improve the degree of curve fitting, but will weaken the prediction ability of neural network, so that excessive matching occurs [13].

2.2. Seismic Attribute Sandstone

The main purpose of attribute feature selection is to determine that various types and grades of sand bodies can best meet the classification conditions through calculation and analysis. Before extracting attribute information, we should first study what kind of data samples the object is. This paper mainly compares the thickness, depth and diameter of sand body in the data, as well as the water content under different geological structure types to determine the optimal classification method. In this paper, we take two parameters as basic parameters. In addition, the two indicators of attribute and distance can be used to jointly determine the optimization results [14].

Through the analysis of borehole connection profile and advanced seismic simulation in the actual working area, it is found that when the thickness of the river sandstone reservoir is less than the thickness of the seismic data. The seismic response of different overlying sandstone reservoirs is a single waveform, but their amplitude, waveform and frequency characteristics are different, resulting in different seismic attributes. On the basis of optimizing seismic sensitive attribute set and fluvial sandstone stacking model, the statistical correlation between seismic attribute values and sandstone reservoirs under different stacking conditions is systematically analyzed, and a multi-dimensional seismic sensitive attribute network map is established. Before the research on multi-attribute decision-making, it is necessary to enumerate the basic characteristic indicators such as sand body thickness, water content and porosity contained in a large number of different geological feature data sources [15].

In the prediction of lithologic and geological parameters, in addition to determining the influencing factors (sand body thickness, water saturation and void ratio), other natural factors such as topography, climate conditions and human activities should also be considered. Therefore, this paper uses the classification attribute based on the decision tree model to establish attribute association. Various influencing factors need to be taken into account in each attribute selection process. Therefore, we selected several parameters as the main basis to determine the weight value and analyze and predict it. Finally, the most suitable attribute vector is selected from the comprehensive weight coefficients.

2.3. Prediction Method of Roof Sandstone Lithology

The lithology of sandstone is mainly characterized by the development, distribution and evolution of joints and fissures, but it is heterogeneous in rock properties, that is, thin and uneven in thickness. The most common one is that there are many fractures in the fault fracture zone. Secondly, the rock deformation and shear failure lead to the poor lithology of sandstone and fracture development. The lithology of roof sandstone is breathable. Permeability coefficient, bedding property of surrounding rock and structural type are the research directions. The study of permeability refers to the measurement and analysis of porosity of different rock mechanical strength. In the same thickness range, due to the different physical and chemical composition and structural characteristics of rocks, they form various and complex formations with different characteristics. The most common is thin quartz sandstone, which is characterized by strong water permeability and good compactness. The main physical property of roof lithology is bedding structure. It has great influence on sandstone porosity, permeability and water permeability. Its mechanical properties, to a certain extent, determine the macro tectonic role of the rock skeleton. Lithology is formed by long-term weathering of rocks, accompanied by sedimentary environment and affected by natural factors such as climate conditions. The grain size minerals in sandstone mainly include calcareous, Triassic stone like quartz, feldspar and carbonate plate crystal chip powder. These substances play an important role in rock mass stability.

Sand body lithology prediction is mainly based on geological historical data, while artificial neural network and entity recognition methods have high accuracy for data information. But the artificial neural network technology can make up for the above shortcomings. This paper mainly uses decision tree classification model (ANSYR) and neural network algorithm to predict the lithology of sand body. Due to the mutual restriction and interdependence between different attribute influencing factors, it is necessary to analyze them according to their characteristics when processing data. Collect the original data and input the attribute values into the parameter pre-processor. After determining the relationship between variable sets through parameter matching algorithm, the decision table model, evaluation rule base and final scheme selection output results are established. According to the constraint conditions, the most optimal sand body properties are obtained by using the supervised classification standard difference method to calculate the weight.

In this paper, through the prediction of different sandstone lithology, a two-dimensional thickness model is established. The 2D sequence combination is obtained by combining the checkpoint data with the rock interface of adjacent blocks. This method can obtain 3D stratum without considering geological factors and tectonic stress. At present, the most commonly used algorithm is the traditional thin longitudinal beam calculation algorithm. The algorithm is mainly based on empirical formula to derive the three-dimensional rock density distribution map, discrete medium mechanical equation and lithologic parameters to predict the thickness change of sandstone, and determine the continuity of rock interface in each fault zone through numerical simulation.

3. Deep Learning Model for Attribute Feature Optimization

3.1. Feature Representation Method Based on Deep Learning

In the deep learning model, the process of generating neural network weights through pre training is unsupervised. The constrained Boltzmann machine is a generative energy probability model consisting of two-layer nodes. The hidden layer nodes and visible layer nodes are fully connected, but the nodes in the layer are not connected. The training process of RBM network is an unsupervised learning method. The hidden layer nodes are used to reproduce the visible layer nodes by connecting the weight values, so the trained hidden layer node values can be regarded as a feature representation of the visible layer. The model learning process is shown in Figure 1:

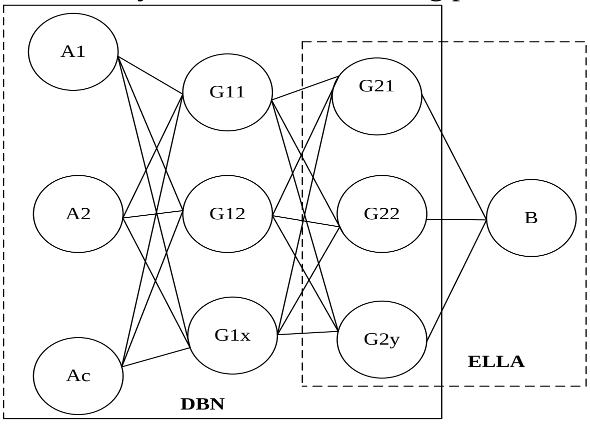

Figure 1. Model Learning Process

The ELLA model makes the data after feature extraction more similar to the same distribution by adding the feature extraction model of unsupervised deep learning. This not only helps the machine learning system transfer knowledge between tasks with different distributions, but also makes the input feature representation after feature extraction more suitable than the original data as the input of subsequent classifiers or linear regression learners.

3.2. Experimental Evaluation Method

Five contrast algorithms are used in this experiment. The first algorithm is the single task learning algorithm STL, which serves as the benchmark algorithm for single task learners. The second algorithm is DG-MTL, a non-intersecting task grouping algorithm, which is used as the benchmark of multi task algorithm. The third algorithm is efficient learning algorithm ELLA. The fourth algorithm is a learning algorithm LLDR based on discriminant representation. The fifth algorithm is the multi task promotion algorithm OMB.

3.3. Experimental Setup

All algorithms are tested on 40 training and test data blocks randomly generated in each database, and cross validation is used on the training data to determine model parameters. In order to verify that the HLLA model can still maintain good performance even when the number of tags is small, we set up 8 experiments. The whole experiment process strictly follows the setup of the machine learning system. In the evaluation and test phase, different tasks of each database enter the machine learning system in a certain order. When a new task arrives, the machine learning algorithm performs feature learning and model learning on the training set, and then predicts the test set. The final result is the average value of the evaluation index values of all tasks.

4. Experimental Results and Analysis of Machine Learning Algorithm in Sandstone Lithology Detection

4.1. Classification Accuracy Index of Five Comparison Algorithms on Sandstone Lithology Detection Data

This paper reflects the experimental results of five contrast algorithms with different training set sizes on the multitask database of sand body lithology identification and prediction. The data in the table are the classification accuracy indicators of the five algorithms, and the higher the value, the better the representativeness. See Table 1 for details:

Table 1. Results of the five comparison algorithms on the sandstone lithology detection data

	20%	40%	60%	80%
STL	0.622	0.698	0.719	0.732
DG-MTL	0.646	0.702	0.718	0.735
ELLA	0.674	0.741	0.758	0.765
LLDR	0.665	0.714	0.735	0.751
OMB	0.661	0.709	0.728	0.746

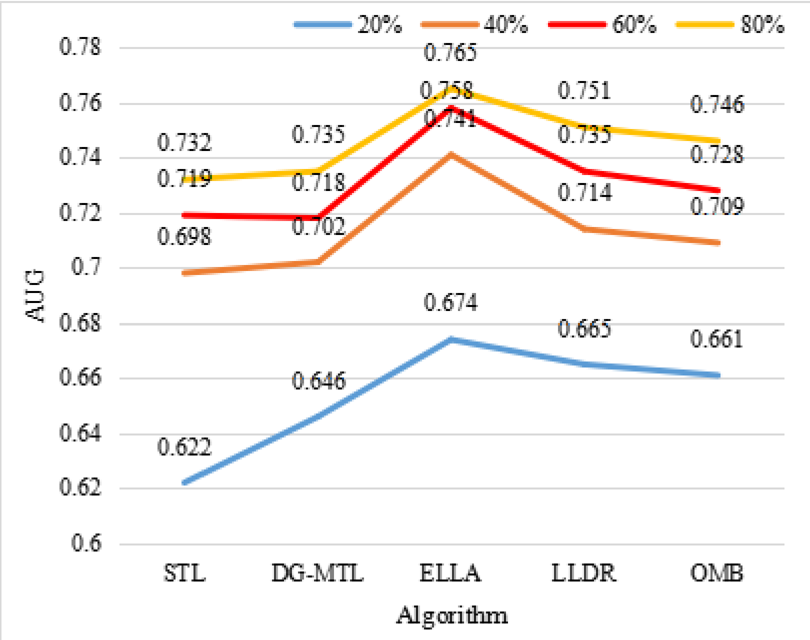

Figure 2. Results of the five comparison algorithms on the sandstone lithology detection data

As shown in Figure 2, we can find that the ELLA algorithm proposed in this paper performs best in all cases. At 20%, 40%, and 60% key points, the classification accuracy of ELLA algorithm is 5.3%, 4.3%, 3.9% higher than STL algorithm, 0.9%, 2.7%, 2.3% higher than LLDR algorithm, and 1.3%, 3.2%, and 3% higher than OMB algorithm respectively. It can also be seen that the HLLA algorithm can still maintain good evaluation performance when the amount of data is small.

4.2. Regression Error of Five Comparison Algorithms in Sandstone Lithology Prediction

This paper reflects the experimental results of five correlation algorithms with different training set sizes on the multitask database of sand body lithology prediction. The regression error (RMSE) indicators of various algorithms under various conditions are shown in Table 2. The smaller the regression error value, the better the performance.

Table 2. Regression error of sandstone lithology prediction by the five comparison algorithms

	STL	DG-MTL	OMB	LLDR	ELLA
20%	13.83	14.23	12.26	12.22	11.31
40%	12.1	13.99	10.98	13.99	10.57
60%	11.38	13.06	10.37	10.71	10.31
80%	10.98	12.64	10.31	10.53	10.15

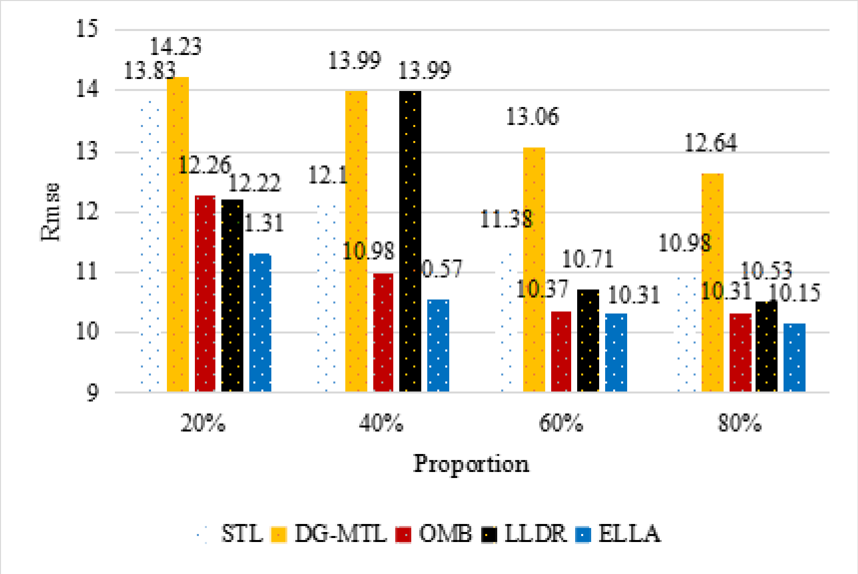

Figure 3. Regression error of sandstone lithology prediction by the five comparison algorithms

As shown in Figure 3, we can find that the ELLA algorithm model is the optimal algorithm in almost all cases, especially when the number of tags is less, the gap between ELLA and other algorithms is greater. The ELLA algorithm requires less tag data and faster model learning, which can better meet the timeliness requirements of lifelong learning.

5. Conclusion

This paper mainly focuses on the prediction of sandstone lithology, makes quantitative analysis of the results using limited data, and draws the following conclusions. When calculating the rock strength, the thickness of thin and thick zones and the variation of confining pressure under geological conditions are considered. The mechanical parameters and boundary treatment of thin wall rock stratum are studied by building a three-dimensional model with software and using simulated dense column. By comparing different methods, the error between the predicted value and the actual value of sandstone lithology of the two combined roof is obtained. The prediction method in this paper is not perfect enough. It needs to use the combination of discrete solution and coupled approximate solution to simulate the action of confining pressure on formation deformation under various geological conditions.

Funding

This article is not supported by any foundation.

Data Availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Conflict of Interest

The author states that this article has no conflict of interest.

References

[1] Xie Y. Facies-controlled reservoir prediction based on the seismic configuration. Geophysical Prospecting for Petroleum, 2021, 60(5):784-793.
[2] Fu G, Wang H. Prediction Method of Favorable Position in Oil-Gas Accumulation around Oil-Source Fault and its Application. Geotectonica et Metallogenia, 2019, 43(1):69-76.
[3] Zhai X, Ma L. Medium and long-term wind power prediction based on artificial fish swarm algorithm combined with extreme learning machine. International Core Journal of Engineering, 2019, 5(10):265-272.
[4] He X, Nie Y, Guo H, et al. Research on a Novel Combination System on the Basis of Deep Learning and Swarm Intelligence Optimization Algorithm for Wind Speed Forecasting. IEEE Access, 2020, PP(99):1-1.
[5] Sun F, Shi G. Study on the application of big data techniques for the third-party logistics using novel support vector machine algorithm. Journal of Enterprise Information Management, 2022, 35(4/5):1168-1184.
[6] Kanimozhi E D, Akila, Akila D. Journal of critical reviews 491 Journal of Critical Reviews An Empirical Study On Machine Learning Algorithm For Plant Disease Prediction. Critical Review, 2020, 7(5):491-493.
[7] Lee Y H. A Study on Analytical Machine Learning Method Applying Discretization and Hierarchical Clustering Algorithm. The Journal of Korean Institute of Information Technology, 2021, 19(1):55-61.
[8] Yu X, Wei X, Huang X, et al. Study on Tracking Control Strategy of Sandblasting System for Wind Turbine Blade Based on Recursive Least Squares Algorithm. Journal Of Engineering Science And Technology Review, 2020, 13(6):118-124.
[9] Zhang J, Yin X, Zhang G, et al. Prediction method of physical parameters based on linearized rock physics inversion. Petroleum Exploration and Development, 2020, 47(1):59-67.
[10] Gongqin Y E, Cao H, Gao Q, et al. Numerical Simulation Study On The Influence Of Particle Proportion On Rock Mechanics Characteristics. Journal of Geomechanics, 2019, 25(6):1129-1137.

[11] Jing L J, Li J B, Chen Y, et al. A case study of TBM performance prediction using field tunnelling tests in limestone strata. Tunnelling and underground space technology, 2019, 83(JAN.):364-372.

[12] Wang Z, Liu J. Application ofrock physics modeling in the prediction of tight sandstone reservoir - A case study of H block, Northern Erdos basin. Wutan Huatan Jisuan Jishu, 2019, 41(1):34-40.

[13] Ding K, Wang L, Wang W, et al. Study on the Development Height of Overburden Water-Flowing Fracture Zone of the Working Face. Geofluids, 2021, 2021(5):1-10.

[14] Sun Y M, Han X, Zhang D X, et al. Study on online soft sensor method of total sugar content in chlorotetracycline fermentation tank. Open Chemistry, 2020, 18(1):31-38.

[15] Gu C. Research on Prediction of Investment Fund's Performance before and after Investment Based on Improved Neural Network Algorithm. Wireless Communications and Mobile Computing, 2021, 2021(1):1-9.