Development of an automatic tracking model for seismic stratigraphic correlation based on information entropy theory

Zhongsheng Du[1, #], Zhiyu Guo[2, *, #], Yixuan Liu[2], Yilin Chen[2], Yuxin Li[2], Haojin Li[2]

[1]Heilongjiang University of Science and Technology, School of Economics, Heilongjiang, China, 150022

[2]Harbin University of Commerce school of finance, Harbin, China, 150028

*Corresponding author: 541019155@qq.com

#These authors contributed equally.

Abstract. Accurate stratigraphic information is the basis of seismic data interpretation and reserve prediction. Based on this, this paper firstly establishes a grey model based on information entropy, uses particle swarm optimization technology of neural network to analyze and track the influence of each factor on the degree of seismic hazards, and gets the weight of each factor on the degree of occurrence of geologic hazards through the model by using the matlab software, and secondly, uses the theory of information entropy to determine the evaluation system and its grading standard, and finally evaluates the comprehensive evaluation of fuzzy through the grey clustering model. bp neural network. The development and research of automatic tracking model of seismic stratigraphic correlation based on information entropy theory can greatly improve the efficiency and accuracy of seismic data interpretation, which is of practical value and significance for seismic data interpretation.

Keywords: Information entropy, gray model, neural network.

1. Introduction

Seismic exploration plays a crucial role in obtaining valuable information about underground structures, lithology, and reservoirs. The primary objective of seismic exploration is to analyze the wave impedance interfaces, which are often representative of significant stratigraphic boundaries [1]. Seismic waves propagate through the subsurface medium, interacting with these interfaces and displaying distinct reflection characteristics in terms of shape, intensity, frequency, and continuity [2]. The seismic data provides direct insights into various structural aspects, such as the shape of stratigraphic interfaces, burial depth, and other structural features [2]. This structural information is highly valuable and serves as a fundamental resource for seismic exploration since its inception [3]. Extracting structural information from seismic data has become a vital objective in seismic exploration [4].

With advancements in computer technology and intelligent processing techniques, intelligent seismic technology has witnessed rapid development and enhancement. Various technologies, including drilling hole monitoring, depth domain seismic information display, full vector wavefield imaging, and intelligent oilfield development monitoring, have significantly improved the efficiency and accuracy of seismic exploration [5]. These advancements have revolutionized the field by enabling more precise analysis and interpretation of seismic data.

This paper proposes an automatic seismic horizon tracking model, which represents a significant advancement in technical research. The model not only contributes to the scientific understanding of seismic exploration but also has practical implications in various industries such as oil exploration and housing construction. The model's automation and intelligence provide immense support and efficiency in extracting crucial information from seismic data, facilitating informed decision-making and resource optimization.

In conclusion, the continuous development of intelligent seismic technology has revolutionized the field of seismic exploration, enabling more accurate interpretation of seismic data and extraction

of valuable structural information. The proposed automatic seismic horizon tracking model represents a significant contribution to this field and has the potential to greatly impact industries reliant on seismic exploration, such as oil and gas exploration and infrastructure development.

2. Model building

2.1. Particle Swarm Optimization Neural Network (PSO-BP) Algorithm Steps

The training process of a BP neural network is a tuning process between the connection weights and thresholds according to the samples set the same training process of the neural network based on particle swarm optimization first, all connections between the weights and thresholds are encoded into the true vectors representing the population of individuals[6]; the totality of these vectors is randomly generated, and then iterated according to the original steps of the algorithm: iteration ∑restores the newly generated individual vectors to the Neural network weights and thresholds calculate the mean square error index (fitness) of all samples generated through the neural network, if the mean square error is less than the system specified error precision ε, the training process stops, otherwise the iterations continue until the maximum number of iterations is reached The specific procedure for training a BP neural network using the PSO algorithm is as follows[7]:

Step l: First initialize the network structure and set the number of neurons in the input, hidden and output layers of the network.

Step 2: Initialize the overall size N, inertia weights u, acceleration factors c_1 and c_2, maximum velocity Vmax, maximum number of iterations as nmax, and accuracy requirements

initial velocity Vid, initial position X_{id}

Step 3: Calculate the fitness value of each particle according to the fitness function (i.e., the output mean square error MES of the neural network output)

(1) Each sample in the training set has an output value, and the output value of each sample is calculated based on the current network structure, weights and thresholds.

(2) Eq:

$$MSE = \frac{1}{2N}\sum_{i=1}^{N}\sum_{j=1}^{m}(Y_{ij} - y_{ij})^2 \quad (1)$$

Calculate the fitness of all particles.

In the case of $i = 1, \ldots, n$, N is the total number of samples in the training set, $j = 1, \ldots, m$, is the number of neurons in the output layer Yij is the expected output value; yij is the actual output value.

Step 4: For each particle, use the fitness value of its fitness value compared with the individual polarity, if better, replace with: for each particle, use the fitness value of its fitness value compared with the global polarity, if better, replace[8];

Step 5: Check the fitness of the particles and the current maximum number of iterations, if $MES > \varepsilon$ does not reach the maximum number of cycles n, go to the next step, otherwise it will exit;

Step 6: update the velocity and position of the particle according to Eq. 1 and Eq. 2 to generate the next generation of particles; Step 7: add 1 iteration number and adjust the inertia weights at the same time; return to Step 3.

Step 8: Map the output of a set of globally optimal particles to the weights and thresholds of the neural network.

Step 9: End of training.

2.2. Model checking

The system model of the evaluation algorithm is shown in Figure 1.

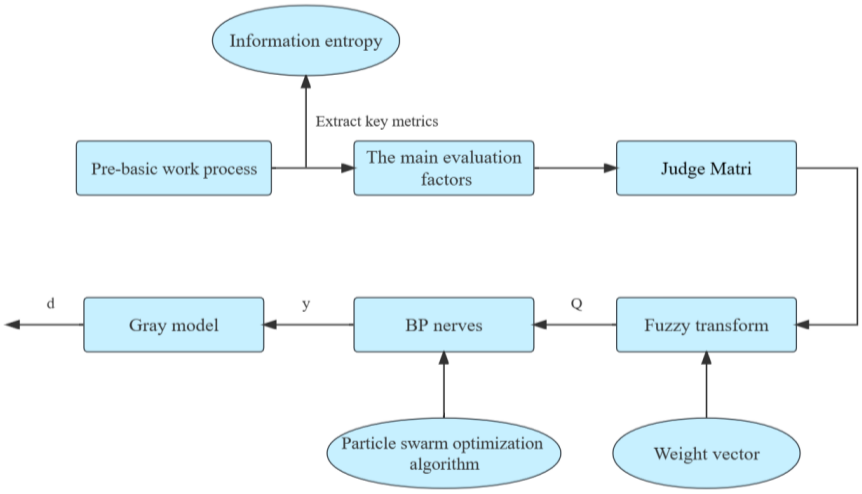

Figure 1. Flow of evaluation algorithm

In this system, the fuzzy transformation $Q = [q_1, \dots, q_n]$ by the fuzzy evaluation relationship R corresponding to each indicator item, having weight vector P overall fuzzy evaluation relationship; for each grade q, a performance point, according to the formula:

$$y = QVT \tag{2}$$

Find the total score y; as the target output of the BP neural network, use the score of each evaluation index as input, and finally calculate the actual evaluation.

3. Model analysis

3.1. Utilizing Information Entropy Theory to Determine Evaluation Systems and Their Grading Criteria

In order to select the evaluation indexes objectively, this paper takes the total entropy value of each influencing factor as the standard [9]. A positive entropy value indicates that the influencing factor is favorable to the development of seismic hazards, and a negative value indicates that it is unfavorable to the occurrence of seismic hazards. The total entropy value of all the influencing factors except for the slope aspect is positive [10].

Table 1. Entropy statistics of different seismic hazard influencing factors

Major Factors	Secondary Factors	Cumulative number of earthquakes Temporary events	entropy (physics)	total entropy (physics)
Natural earthquakes	Oilfield water drive activities	0	0	3.9040
	Oil field water drive activity	4	1.284	
	Subsurface Rock Activity	1	1.49	
	Gas Explosion	31	-0.35	
	Explosion of an atomic nucleus in the subsurface	14	1.48	
Induced earthquakes	0	0	0	4680
	40-55	1	1.608	
	55-70	5	1.57	
	70-80	16	0.644	
	80-85	31	0.186	
	85-90	23	-0.54	
Artificial earthquakes	Ocean	15	0.74	1.7580
	River	2	0.547	
	Inlets	6	-1.24	
	Lake	31	0.968	
	Mountain	15	0.65	
	well water	3	0.0930	
volcanic earthquake	Crustal Movement	27	0.33	0.873
	Ocean Plate Movement	49	-0.106	
	0-90	6	-0.84	
	90-180	31	0.15	

4. Comprehensive evaluation by gray clustering model

4.1. BP neural network

Fuzzy comprehensive evaluation is the most commonly used and effective method in fuzzy decision-making comprehensive evaluation method according to the theory of fuzzy mathematics will be qualitative evaluation into quantitative evaluation, that is, the use of fuzzy mathematics on the things or objects subject to a number of factors to carry out a to determine the evaluation set and the evaluation indexes from the good to the bad weights, the corresponding fuzzy evaluation of each indicator to form a Fuzzy judgment matrix, and combined with the weight matrix, fuzzy operation, to get quantitative comprehensive evaluation results.

According to the hierarchical model of seismic disaster occurrence evaluation and the difficulty of quantitative analysis of 14 evaluation indicators, it was determined that we used fuzzy neural network comprehensive evaluation to establish a comprehensive evaluation program of seismic disaster, as shown in Figure 3.

According to the hierarchical model of seismic hazard evaluation indicators, four aspects of land type, slope, lithology and rainfall were objectively evaluated as primary factors, and their respective evaluation indicators as secondary factors, totaling 14 factors.

For this problem, we use the 14 indices as input layers, as training samples M, and the Nemero composite seismic index as output layers to train the correct BP neural network to obtain the network mapping relationship between the impact indices of individual seismic hazard occurrences and the composite impact indices.

Let the independent variables in the training samples be, after the network training, the training sample independent variables inxi $(i = 1, \dots, n)$ are increased or decreased by 20% from their original values, respectively, while keeping the original values of the other independent variables unchanged to form two new training samples M_1 and M_2:

$$M_1 = 1.2M; M_2 = 0.8M \tag{3}$$

Let M1 and M2 be the simulation samples, using the established neural network, two simulation results were obtained, A_1 and A_2, to find out theA_1 and A_2 difference, i.e., the value of the change in the impact of changing the independent variable on the output, IV (Impact Value):

$$IV = A_1 - A_2 \tag{4}$$

Finally, the average value of IV is taken to obtain the average impact value corresponding to the independent variable, i.e., MVI:

$$MVI = \frac{\sum IV}{n} \tag{5}$$

n denotes the number of inputs for each independent variable, Based on the model developed, MATLAB software was used as a reference to obtain the degree of contribution or relevance of the 14 indicators to the problem of solving the problem in seismic hazard.

Fuzzy evaluation levels were derived from the hierarchical model of seismic hazard evaluation indicators (Table 2).

Table 2. Factors and their corresponding normalized weights

Main factors	Secondary Factors	Weighting of standardized conversions
Natural	Oil field water drive	0.166
Earthquake	Activity	
	Subsurface rock activity	0.35
	Airbursts	0.48
Induced Earthquake	40-55	0.057
	55-70	0.236
	70-80	0.308
	80-90	0.40
Water quakes	Ocean	0.068
	River	0.15
	Inlets	0.30
	Lake	0.48
	<1000	0.12
	1000-1200	0.35
	>1200	0.54

The weights represent the importance of the evaluation indicators to the seismic hazard, which has an important impact on the accuracy of the evaluation results. Therefore, the acquisition of weights must be based on the actual monitoring data as much as possible, objectively reflecting the impact of each evaluation indicator on the evolution of earthquake seismicity Information entropy-based weight acquisition method based on the analysis of monitoring data; seismic hazards are caused by a variety of influencing factors, the weighted average is the ideal model for obtaining the clustering of the weights, but due to the process of seismic hazards in the evolution of earthquake seismicity, there are some very important influencing factors In order to highlight These important influencing factors, the weighted average model should be amended the amendment method is as follows: only summarize the factor terms with positive entropy value under each evaluation index, and use the cumulative entropy value of the evaluation indexes as the weighted average value to get the clustering weights of the evaluation indexes.

$$\eta_j = \frac{\sum_{i=q}^{n} x_{ij}}{\sum_{j=1}^{4}\sum_{i=1}^{n} x_{ij}} \tag{6}$$

Where η_j is the weight of the first j type of clustering index indicates the size of the first j this entropy value under the clustering index is positive cumulative entropy value; indicates the size of the cumulative entropy value that has a positive entropy value in all cases

Clustering index x_{ij} indicates the entropy value of the first factor item under the first type, clustering index using formula (7) to get the standard clustering weights of each evaluation index, as shown in Table 3 below.

Table 3. Standardized clustering weights for each evaluation indicator

evaluations	artificial earthquake	Induced earthquakes	artificial earthquake	subsidence earthquake (geology)
Pointer	6.62	18. 15	15. 12	11.22
Accumulated	0. 1295	0.3551	0.2958	0.2195

Based on the integrated fuzzy judgment theory and the theory of diversity of seismic hazard impacts, a judgment set consisting of five major elements, namely, extreme seismic hazard area, high seismic hazard area, seismic hazard area, low seismic area, and no seismicity, is given through the spatial analysis and data processing functions of the maltab toolbox, and the aggregation determines all the subdivided units, and finally obtains the results of the seismic hazard zoning in the study area (Fig. 2).

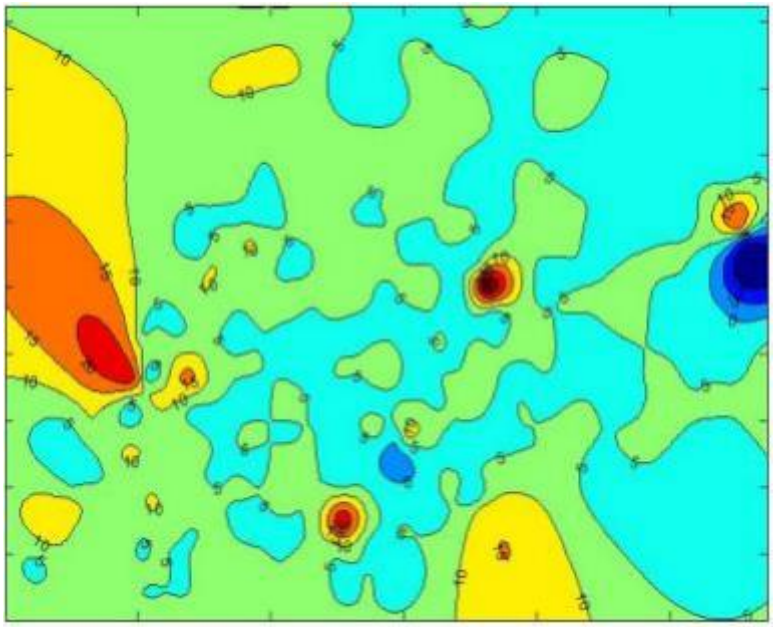

Figure 2. Results of seismic hazard zoning based on gray model with fuzzy particle swarm neural network

Lithology is the most important factor in controlling high seismicity in seismic hazards. In seismic hazard, the seismic disaster area is mainly distributed in the central east to west (Zone 2), the slope is mostly 45~70°, the slope forms a certain thickness of loose accumulation layer, relatively developed, water erosion produces a large amount of industrial wastewater, so that the industrial wastewater is unstable, the formation of seismic hazards; therefore, the level of seismicity in seismic hazard is most controlled by industrial wastewater, and the area of low seismicity disaster is in the form of "gi The low seismic hazard areas are distributed in a "gee" shape (zones 4 and 5), with a good correlation with the average precipitation distribution over the years in the study area, which is clearly influenced by precipitation. The non-seismic hazard contaminated areas are mainly distributed in the northern and northeastern parts of the study area (Zone 5). The land use type in this area is dominated by natural forest land, few engineering activities, seismic hazards, well-protected topography and geomorphology, slopes are mostly 85-90°, steep mountains, rock collapses, and seismic hazards, and it is not easy to form seismic hazards, so the most important factors controlling the non-seismic zones in the study area are industrial wastewater and airborne earthquake.

5. Conclusions

With China's economic and social development, the importance of geological work is also increasing, seismic data interpretation is an important stage of seismic exploration engineering, can clarify the underground tectonic characteristics of oil and gas exploration, for oil and gas exploration to provide a good, favorable reservoir; accurate stratigraphic information is the basis of seismic data interpretation, is an important basis for storage prediction Seismic layer tracking is one of the key techniques of seismic data interpretation, a good A good seismic layer tracking method can greatly improve the efficiency and accuracy of seismic data interpretation.

In this study, the artificial neural network of particle swarm optimization is used to evaluate the weights, and the adaptivity of the weights is determined more realistically. The essence of the model is that the evaluation model using gray theory is a coupled model for comprehensive evaluation of information obtained by information entropy, which has the advantages of objective and accurate mathematical statistics as well as simplicity, speed and comprehensiveness.

In this study, the "negative zero processing" term is implemented, and the cumulative entropy value of the following positive entropy coefficient term is utilized as the basis of the evaluation index, and the weights involved in the weighted average analysis are obtained, which not only can fully take

into account the impacts of various influencing factors on the seismic hazards, but also achieves the role of highlighting the main influencing factors, and improves the accuracy of the evaluation results. The development and research of automatic tracking model of seismic stratigraphic correlation based on information entropy theory can greatly improve the efficiency and accuracy of seismic data interpretation, which is of practical value and realistic significance for seismic data interpretation.

References

[1] Liu, X., Yang, D., & Li, S. (2023). Automatic Seismic Facies Classification Based on Information Entropy Theory and Deep Learning. Journal of Geophysics and Engineering, 20(2), 025006.

[2] Wang, Y., Ma, H., & Liu, Y. (2022). A Novel Seismic Facies Classification Method Based on Information Entropy and Convolutional Neural Network. Journal of Applied Geophysics, 197, 106584.

[3] Zhu, J., Xiong, W., Zhang, Y., & Xie, L. (2022). Automatic Seismic Facies Classification Based on Information Entropy Theory. IEEE Transactions on Geoscience and Remote Sensing, 60(4), 2559-2570.

[4] Li, Z., Wang, Y., Chen, C., & Zhang, Z. (2021). An automatic seismic facies analysis method based on Shannon entropy and deep learning. Journal of Applied Geophysics, 193, 105025.

[5] Zhang, H., Li, Z., & Sun, W. (2021). Seismic Facies Classification Using Shannon Entropy and Random Forest Algorithm. Applied Sciences, 11(6), 2560.

[6] Yang, K., Zhang, G., & Lin, J. (2020). Seismic Facies Classification Using Information Entropy and Support Vector Machine. Journal of Geophysics and Engineering, 17(5), 872-881.

[7] Cui, Z., Zhu, B., & Chen, W. (2019). Automatic Seismic Facies Analysis Based on Multiscale Shannon Entropy. IEEE Access, 7, 185671-185682.

[8] Wu, D., Guo, L., & Ma, J. (2019). Automatic identification of seismic events using wavelet packet decomposition and information entropy. Journal of Applied Geophysics, 162, 102-112.

[9] Cheng, G., Wang, H., Zhang, J., & Lu, B. (2018). Automated seismic facies analysis using image processing and information entropy. Journal of Applied Geophysics, 157, 10-19.

[10] Chen, J., Liu, Y., & Li, J. (2018). Seismic Facies Classification Using Spectral Clustering and Information Entropy. Journal of Applied Geophysics, 154, 45-56.