Desalination and Water Treatment
www.deswater.com

doi: 10.5004/dwt.2023.29677

299 (2023) 203–211
July

Risk prediction method of shallow sea oil removal based on fuzzy analysis

Yue Shen

College of Geophysics, China University of Petroleum, Beijing 102249, China, email: sy202204@126.com

Received 14 December 2022; Accepted 18 May 2023

ABSTRACT

For the purpose of improving the risk prediction accuracy of shallow sea oil development in the oil industry, fuzzy clustering is used to classify the reservoirs, and the reservoirs that are easy to develop are obtained, get the genetic algorithm-neural network (GA-NN) model. The result is that the overall performance of the GA-NN model is better, and it has a higher prediction risk prediction accuracy. Compared with the neural network (NN) model and the genetic algorithm (GA) model, the GA-NN model has the highest net present value, and the optimal solution effect is better; as the amount of iterations increases, the memory and time occupied by the three models also gradually increase, but the GA-NN model consumes the least amount of memory and takes the least time, which saves computer memory and has a faster prediction speed. Under the same amount of iterations, the GA-NN model is 0.52 s faster than the GA model and 0.78 s faster than the NN model. The GA-NN model gets the optimal solution faster than the other two models. The error of the GA model tends to stabilize when it is trained for 37 times, and the error value remains around 1.5, that is, the GA model has the fastest stabilization speed and the smallest error. The highest prediction accuracy of the three models is 77.2%, 86.3%, and 96.3%, respectively. The result is that the overall performance of the GA-NN model in the risk prediction of shallow sea oil development is better and has a higher accuracy, which can reduce the cost of oil development and the risks in the process, and is conducive to obtaining the maximum profits.

Keywords: Fuzzy clustering; Shallow sea oil; Development risk; Genetic algorithm-neural network; Prediction

1. Introduction

The main purpose of oil development is to get maximum profits, but there are great risks in the process of oil exploitation. Maximize profit while reducing risk [1]. Because the easy-to-exploit oil is decreasing day by day, the cost of oil development is getting higher and higher, so it is very important to carry out risk analysis in the process of oil development. The risk analysis of shallow sea oil development needs to first classify the shallow sea oil. In the past oil classification process, the threshold value of a common classification parameter is usually used as the classification standard based on the experience of managers, and the classification is obtained by referring to other parameters. The results are used as the basis for formulating oil development plans [2]. With the increase of parameters, the classification method cannot comprehensively consider all the oil parameters, so the shallow sea oil development plan formulated is unrealistic and will bring serious economic losses. Using fuzzy clustering analysis method to classify shallow sea oil can transform the classification from single parameter boundary to multi-parameter comprehensive analysis, and classify the oil according to the degree of approximation, which makes the classification results more scientific and reasonable [3]. Then, according to the classification results, the shallow sea oil that is easy to develop is selected, and the mathematical model of development risk is constructed by combining genetic algorithm and neural network. The risk prediction of shallow sea oil development based on fuzzy cluster analysis can enhance the accuracy and efficiency of prediction, and the results are feasible and objective, which has certain reference value for the risk prediction of shallow sea oil development.

2. Related works

Shallow sea oil is an important oil and gas resource, but its development costs are high and risks are high. Classifying oilfields and assessing development risks based on scientific and rational methods is crucial for improving oil recovery and the economic benefits of oil development. Using fuzzy clustering to classify oil fields can get more reasonable classification results, and using genetic algorithm and neural network to build a mathematical model of development risk can accurately predict the development risk of shallow sea oil. Scholars in many fields have carried out a lot of research on classification and prediction combined with fuzzy cluster analysis, genetic algorithm and neural network, and have achieved a lot of research results.

He et al. [4] designed an adaptive interval type II fuzzy clustering method for the problem of fuzzy uncertainty in the classification of land cover by remote sensing images. Based on the interval value symbol modeling, it can better analyze the classification uncertainty description, and enhance the separability and classification accuracy between different classes. Aiming at the shortcomings of the current clustering analysis algorithm in data mining applications, Guo et al. [5] proposed a fuzzy clustering algorithm suitable for large data sets. The result is that the proposed algorithm can obtain good outcomes. In order to enhance the recognition rate of diabetic retinopathy, Rajesh et al. [6] designed an image recognition method based on fuzzy clustering. The result is that the use of this recognition method can enhance the detection accuracy, which is conducive to the timely diagnosis and treatment of diabetic retinopathy. For the purpose of solving the problem of low classification accuracy of big data information in teaching ability evaluation, Chen [7] designed a teaching ability evaluation algorithm based on big data fuzzy *K*-means clustering and information fusion. The result is that the proposed method can enhance teaching ability evaluation. accuracy, and promote the rational application of teaching resources. For the purpose of solving the problem that the standard Conceptual Clustering Method (CCM) algorithm cannot handle continuous data, Li et al. [8] designed a learning model based on fuzzy concepts, and constructed a new fuzzy concept learning framework. The result is that the method can enhance the classification performance and has a certain effective sex.

Almedallah and Walsh [9] designed a comprehensive model for the problem that surface facilities and wellbore trajectories are not jointly considered during the development of offshore oil and gas fields, and designed a comprehensive model that comprehensively considered the optimization of wellbore trajectories and surface facilities in the development of shallow water offshore oil and gas fields. Shu et al. [10] designed a high-resolution frequency band broadening processing technology in view of the fact that the dual-sensor seismic data obtained by conventional means cannot meet the requirements of thin reservoir seismic description and oil-bearing fluid identification. The result is that this technology can enhance the accuracy of seismic data. The signal-to-noise ratio and resolution have a high fluid identification coincidence rate, which can supply reference value for oilfield exploration and development. For the purpose of enhancing the accuracy and speed of wave prediction, Wang et al. [11] designed a back propagation (BP) neural network prediction model optimized by a thinking evolution algorithm, which combined the local search ability of the BP neural network and the global search ability of the thinking evolution algorithm. The result is that the model established by the study can enhance the prediction accuracy and prediction time. For the purpose of solving the problem of unsatisfactory classification accuracy of the algorithm in the identification of basalt tectonic environment, Ren et al. [12] designed an enhance genetic algorithm and an optimized neural network coupling identification method, using the classification accuracy as a fitness function. The result is that the proposed method enables adjustment of piece count data, optimizes unknown parameters, and enhances classification accuracy. Aiming at the problem that the creep error of piezoelectric ceramics affects its positioning accuracy, Fan et al. [13] designed a method for predicting the creep of piezoelectric ceramics based on the back-propagation neural network of improved genetic algorithm, and built a prediction model. The result is that the maximum absolute error predicted by the model is less than 0.2 μm, and the maximum creep error is less than 1.5%, which can meet the creep prediction requirements of piezoelectric ceramics. For the purpose of improving the prediction accuracy of limestone sample cohesion, Khandelwal et al. [14] constructed a prediction model combining improved genetic algorithm and artificial neural network. The result is that the model can better predict the cohesion of rock.

The above content is research on fuzzy cluster analysis, genetic algorithm, neural network, and oil and gas resources development by scholars in different fields, and they have achieved good results. However, there are few studies that combine cluster analysis, genetic algorithm and neural network to predict the risk of shallow sea oil development. Therefore, this study will combine fuzzy clustering to classify oil fields, and use genetic algorithm and optimized neural network to construct development risk mathematics model to reduce the risk of shallow sea oil development and enhance economic benefits.

3. Construction of risk prediction model for shallow sea oil development based on fuzzy analysis

3.1. Mathematical model design of shallow sea oil development risk

Oil is one of the most important energy sources in the development of human society at present and in the next few decades. There are still huge quantities of oil to be exploited around the world. At present, the world's annual proven oil reserves are very limited, and the development rate of large oilfields is decreasing year by year [15]. For the purpose of facilitating the classification of a large amount of reservoir data, a fuzzy multivariate statistical analysis system platform is constructed based on the classical statistical dynamic clustering analysis method and the fuzzy statistical clustering analysis method. Cluster analysis refers to a statistical method for classifying research objects based on certain characteristics of the research objects and combining mathematical tools. In the process of analysis, for the purpose of avoiding comparing data of different dimensions and orders of magnitude together unreasonable phenomena occur, and the original data needs to be transformed [16]. Assuming that the original observation data matrix is X, the expression is shown in Eq. (1).

$$X=\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n2} & x_{n2} & \cdots & x_{np} \end{bmatrix} \tag{1}$$

where n, p represent the amount of samples and the amount of variables, respectively, and are $x_{ij}(i = 1, 2, \ldots, n; j = 1, 2, \ldots, p)$ used to describe the observed value of the j-th variable of the i-th specimen. Data can be exchanged using centering, normalizing, and normalizing transformations. After that, some statistical characteristics of random variables or measured data are obtained, and statistical laws are found based on statistical characteristics, while the statistical values of random variables in some aspects are the numerical characteristics of random variables, generally including mathematical expectation, variance, covariance and correlation coefficient [17]. Assuming a random variable $X = \{x_1, x_2, \ldots, x_n\}$, a probability distribution $P = \{p_1, p_2, \ldots, p_n\}$, Eq. (2) can be obtained.

$$E(X)=\sum_{k=1}^{n} x_k p_k \tag{2}$$

The mathematical expectation expressed by Eq. (2) X is also the mean value of X, which is one of the important digital characteristics of the random variable of mathematical expectation. Meanwhile, the variance of X is shown in Eq. (3):

$$V(X)=E\left[\left(X-E(X)\right)^2\right]=\sum_{k=1}^{n}\left(x_k-E(X)\right)^2 p_k \tag{3}$$

The variance of X can also be expressed as σ_X^2, so $\sigma_X=\sqrt{V(X)}$ is the mean variance or standard deviation of X. The variance can explain the dispersion degree of X value relative to its average value of $E(X)$, which is also an important numerical feature of random variables. Assume that the binary random variable $X = (X_1, X_2) = \{(x_{11}, x_{12}), (x_{21}, x_{22}), \ldots, (x_{n1}, x_{n2})\}$, the joint probability distribution of this variable is $P_X = \{p_1(x_{11}, x_{12}), p_2(x_{21}, x_{22}), \ldots, p_n(x_{n1}, x_{n2})\}$, and the mathematical expectation is $E(X_1)$ and $E(X_2)$, then the covariance between X_1 and X_2 is shown in Eq. (4).

$$\mathrm{Cov}(X_1,X_2)=E\left\{\left[X_1-E(X_1)\right]\left[X_2-E(X_2)\right]\right\} \tag{4}$$

The covariance between X_1 and X_2 is generally recorded as σ_{12}, so Eq. (5) can be obtained:

$$\sigma_{12}=\sum_{i=1}^{n}\left[x_n-E(X_1)\right]\left[x_{i2}-E(X_2)\right]p_i(x_{i1},x_{i2}) \tag{5}$$

Similarly, covariance is also an important digital feature of binary random variable X, which can describe the relationship between the components of X. Assuming that the binary random variable $X = (X_1, X_2) = \{(X_{11}, X_{12}), (X_{21}, X_{22}), \ldots, (X_{n1}, X_{n2})\}$ has an average value of $E(X_1)$ and $E(X_2)$, the variances are $V(X_1)$ and $V(X_2)$, respectively, and both variances are greater than 0, the correlation coefficient or standard covariance between X_1 and X_2 is shown in Eq. (6).

$$p(X_1,X_2)=\frac{\mathrm{Cov}(X_1,X_2)}{\sqrt{V(X_1)V(X_2)}} \tag{6}$$

where $p(X_1, X_2)$ is a dimensionless quantity, which can also be written as p_{12}. Fuzzy cluster analysis is the application of fuzzy mathematics for cluster analysis, and its mathematical basis is fuzzy set theory. However, fuzzy clustering analysis methods are generally based on the maximum and minimum paradigm, and most of the obtained clustering analysis results are local optimal solutions, which will increase the arbitrariness and make it impossible to accurately grasp the clustering results. Cluster analysis method for classification. The first-order matrix plays an important role in fuzzy statistics, and its main statistical characteristics include correlation coefficients such as fuzzy sets, fuzzy mathematical expectations, and fuzzy variances. Because there is only a small amount of abnormal data in the classification process, the amount of fuzzy iterations increases significantly, and the fuzzy variance, fuzzy covariance, and fuzzy correlation coefficient calculated at the end have strong anti-interference ability, and the calculation results can more accurately reflect the real situation of the actual observation data.

Assuming that the measured data Y is a set of samples to be clustered, including n samples and p indicators, $x_{ij}(i = 1, 2, \ldots, n; j = 1, 2, \ldots, p)$ is used to describe the observation value of the j-th variable in the i-th experiment, as shown in Eq. (7).

$$Y=\begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1p} \\ y_{21} & y_{22} & \cdots & y_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ y_{n1} & y_{n2} & \cdots & y_{np} \end{bmatrix} \tag{7}$$

After getting the model, first set the initial value of the fuzzy cluster center. The value of the j-th variable of the i-th cluster center is described by $\tilde{z}_{ij}(i=1,2,\ldots,n;j=1,2,\ldots n)$, that is, n samples to be clustered are set as the initial value of n fuzzy cluster centers, that is, $\tilde{Z}=X$. Then calculate the fuzzy cluster center, which needs to calculate the deviation, the mean deviation, the membership of the deviation and the fuzzy cluster center in turn. $d_{ij}(l)$ $(i = 1, 2, \ldots, n; j = 1, 2, \ldots, p; l = 1, 2, \ldots, n)$ is used to represent the deviation of the j-th variable of the i-th sample from the l-th center, so a matrix of n rows and p columns can be used to represent the deviation data, as shown in Eq. (8).

$$d_{ij}(l)=\left|x_{ij}-\tilde{z}_{ij}\right| \tag{8}$$

Use $\tilde{d}_j(l)(j=1,2,...,p;l=1,2,...n)$ to describe the mean deviation of the j-th variable relative to the l-th center, and a matrix with p columns in one row can be obtained to represent the mean deviation data, as shown in Eq. (9).

$$\bar{d}_j(l)=\frac{1}{n}\sum_{i=1}^{n}d_{ij}(l) \tag{9}$$

$d\ \mu_{ij}(l)$ is used to describe the deviation membership of the j-th variable of the i-th sample relative to the l-th center, so a matrix of n rows and p columns can be used to represent the deviation membership data, as shown in Eq. (10).

$$d\mu_{ij}(l)=e^{\frac{d_{ij}(l)}{\bar{d}_j(l)}} \tag{10}$$

From this, the fuzzy clustering center can be obtained, and the fuzzy clustering center set can be defined, as shown in Eq. (11).

$$\tilde{z}_{ij}=\frac{d\mu_{ij}(l)\cdot x_{ij}}{\sum_{y=1}^{n}d\mu_{rj}(l)} \tag{11}$$

where $i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, p$; $l = 1, 2, \ldots, n$; $r = 1, 2, \ldots, n$. Then calculate the maximum allowable distance between two fuzzy clustering centers, measure the distance with Euclidean distance, and calculate the Euclidean distance of centers i and l, as shown in Eq.(12).

$$\Delta_i(l)=\sqrt{\sum_{j=1}^{p}\left(\tilde{z}_{ij}-\tilde{z}_{lj}\right)^2} \tag{12}$$

Eq. (12) is used to describe the Euclidean distance between two centers. T means the maximum distance allowed between two fuzzy clustering centers, as shown in Eq. (13).

$$T=\frac{\sum_{i=1}^{n}\sum_{l=1}^{n}\Delta_i(l)}{2n^2} \tag{13}$$

Then merge them in turn, and calculate the distance between the l-th center and the i-th center. If it is shown in Eq. (14),

$$\Delta_i(l)<T \tag{14}$$

Then the l-th center will be merged to the i-th cluster center. If $\Delta_i(l) \geq T$, it will not be merged; Repeat the merging step until all the fuzzy clustering centers are merged. Finally, calculate the merged fuzzy cluster center. If there are k centers in the m category after merging, describe the center that is merged into the k category. The expression of the k category center after merging is shown in Eq. (15).

$$\tilde{z}_{kj}=\frac{\sum_{k=1}^{m}\tilde{z}_{ij}(k)}{m} \tag{15}$$

where $i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, p$; $k = 1, 2, \ldots, m$. So far, the fuzzy cluster analysis is completed. The flowchart of the fuzzy statistical cluster analysis algorithm is shown in Fig. 1.

As can be seen from Fig. 1, it is necessary to first set the initial value of the fuzzy clustering center, and then calculate the dispersion, dispersion behavior and dispersion membership of the fuzzy clustering center to obtain a new fuzzy clustering center. Calculate the maximum distance between them. If it is $\Delta_i(l) < T$, the fuzzy cluster centers can be merged and calculated. If $\Delta_i(l) < T$ is not satisfied, the fuzzy cluster center needs to be recalculated. According to the process shown in Fig. 1, the shallow sea oil can be classified by fuzzy clustering. The system platform obtained by fuzzy multivariate statistical analysis can ensure the unsupervisedness of the clustering. Promote the effective development of shallow sea oil and enhance the exploitation efficiency.

3.2. Application of genetic algorithm and neural network optimization model

Oil extraction is full of risks and uncertainties, requiring precise assessment of various risks. At the same time, one of the main goals of oil development is to maximize profits. The realization of this goal requires improving oil recovery and reducing the cost of production operations. Risk analysis of shallow sea oil development refers to reducing risks and increasing profits in the process of shallow sea oil development [18]. Among them, the property of the reservoir is highly dependent on the well position, so the risk analysis of well layout has become the most important part of the risk analysis of the whole shallow sea oil development. The purpose of well placement risk analysis is to reduce well placement costs, enhance recovery or profit, and optimize well placement. In the process of well layout optimization, an algorithm that can meet the well position discontinuity is generally selected. Since the genetic algorithm can solve the optimization problem of a large amount

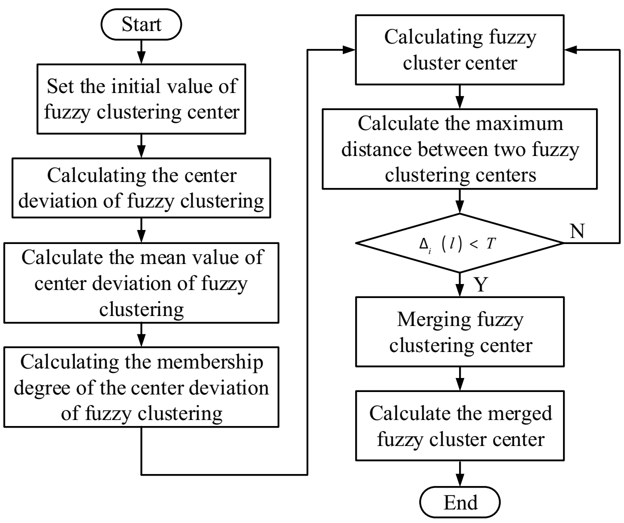

Fig. 1. Flow chart of fuzzy statistical clustering analysis algorithm.

of continuous or discrete decision variables, the genetic algorithm is generally selected to realize the well layout optimization in the shallow sea oil development system. Help the staff to analyze the risk situation of well layout. The flow chart of applying genetic algorithm to the risk analysis of shallow sea oil development to optimize well position is shown in Fig. 2.

As can be seen from Fig. 2, the well location coordinates (x, y) of the water injection well are determined by the decision variables, and (x, y) the value range is determined by the constraints; after the well layout optimization model is established, the objective function is set as the net present value, and the to-be-solved The target is the maximum value of the net present value. Among them, the genotype of the individual X and the search space of the genetic algorithm can be used as the individual encoding method of the feasible solution, and X then the decoding method is formulated according to the correspondence between the individual genotype $F(x)$ and the individual phenotype and the conversion method; then according to the objective function value X and individual fitness $F(x)$ The conversion criteria between the selected $F(x)$ quantitative evaluation methods. Next, the genetic operator is designed, and the specific operation methods of selection, crossover, and mutation operations are formulated, and finally the setting of the relevant allowable parameters of the algorithm is completed, involving crossover probability and termination algebra, etc. [19]. So far, the specific operation of well position optimization in the risk analysis of shallow sea oil development by genetic algorithm has been completed. The main application process of genetic algorithm in well layout optimization is shown in Fig. 3.

The genetic algorithm is used to optimize the well position and promote the analysis of the development risk of shallow oil fields. First, an optimization model should be established according to the description of the problem, that is, an objective function should be established, and then the genetic variables should be adjusted and the parameters of the genetic algorithm should be set. According to the setting of parameters, the initial population is randomly generated, and the population is prompted to perform crossover, mutation and selection calculations, and then the optimal population is obtained. Then, the obtained optimal group is used as the initial group of the next generation, and the cycle is performed according to this step until the optimal result is obtained. Well layout optimization needs to consider many parameters, such as reservoir structure, production parameters, well location, platform type and other parameters, as well as the uncertainty of reservoir geology, which makes it difficult to determine the objective function and constraints, and the optimization of well layout is difficult. Nonlinear and discontinuous problems also limit traditional methods. In order to realize the automation of well layout optimization process, combined with neural network and genetic algorithm to enhance the existing problems of well layout optimization.

Neural network is a mathematical model for processing information. It is connected by a large amount of neurons and involves a three-layer structure of input layer, hidden layer and output layer, which can solve complex classification problems. The genetic algorithm and the neural network influence each other, and the optimal parameters and the optimal discriminant effect are obtained through joint optimization. The overall structure of the genetic algorithm-neural network (GA-NN) model is shown in Fig. 4.

It can be seen from Fig. 4 that GA-NN is constructed based on the improvement of the chromosome coding and fitness function of the genetic algorithm, combined with the external computing framework. GA-NN can improve the problem of easily falling into local optimal solution in well layout optimization, realize the automation of well layout

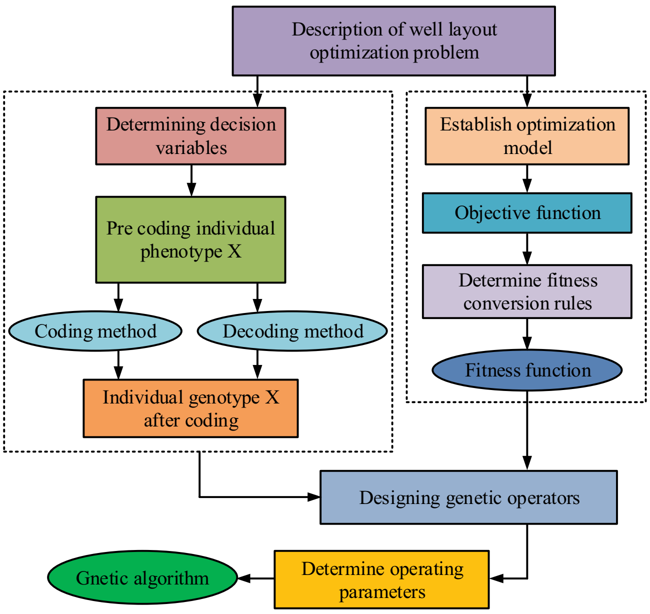

Fig. 3. Main application flow of genetic algorithm in well layout optimization.

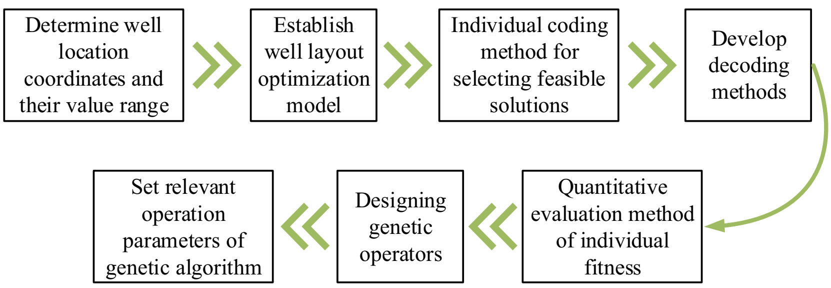

Fig. 2. Application of genetic algorithm in risk analysis of shallow sea oil development.

optimization process, and then optimize the risk prediction model of shallow sea oil development.

4. Analysis of prediction results of shallow sea oil development risk model

A total of 200 actual cases of shallow sea oil development at home and abroad were collected, and the development difficulty of the 200 oilfields was classified based on fuzzy clustering, and the shallow sea oilfields that were relatively easy to develop were obtained. On this basis, GA-NN is used to build a development risk mathematical model to predict the development risk of easy-to-develop oilfields, and the prediction effect of the GA-NN model based on fuzzy cluster analysis on the development risk of shallow sea oil is analyzed. For the purpose of increasing the diversity of comparison, the genetic algorithm (GA) model and the neural network (NN) model were constructed and predicted under the same experimental conditions, and the prediction effects of the three models were compared and analyzed. The parameter settings of the model are shown in Table 1.

In the actual oil reservoir resources, most of the oil reservoirs belong to heterogeneous oil fields. The well layout optimization of two easy-to-develop oil reservoirs is carried out. Both reservoirs A and B have 400 grids, and the optimization purpose is the coordinates of the water injection wells are fixed as (1, 20), (2, 17), and the oil production wells are optimized to maximize the net present value. Each time the three models are executed, a new production well location will be generated, and the net present value will also change accordingly. After many iterations, the optimal well positions (13, 2) and (17, 6) were found, and the maximum value of the net present value of the two reservoirs was obtained. The changes of the net present value of the three models in the optimization process are shown in Fig. 5.

Fig. 5a and b represent the iteration times and Net Present Value (NPV) changes of the three models in reservoirs A and B, respectively. In Fig. 5, with the addition of the amount of iterations, the net present value of the three models also increases; when the optimal solution is obtained, the net present value of the reservoir reaches the maximum; if the iteration continues, the net present value will be. Instead of increasing, it will decrease until the optimal solution is obtained. Among the three models, the GA-NN model has the highest net present value, and the optimal solution is better. The relationship between the increase of training times and the memory occupancy during the process of finding the optimal solution for the three models is analyzed and compared, and the consequences are shown in Fig. 6.

Fig. 6a and b, respectively represent the changes of the memory occupancy of the three models in reservoirs A and B with the addition of the amount of iterations. In Fig. 6, the memory occupied by the three models increases with the addition of the amount of iterations. Among them, the NN model occupies the most memory, and the GA-NN model occupies the least memory. For example, in reservoir A, when the amount of iterations is the 80th, the memory occupied by the NN model, GA model, and GA-NN model is 3,350; 2,750 and 1,880 kb, respectively, that is, the GA-NN model occupies the least memory and is more economical

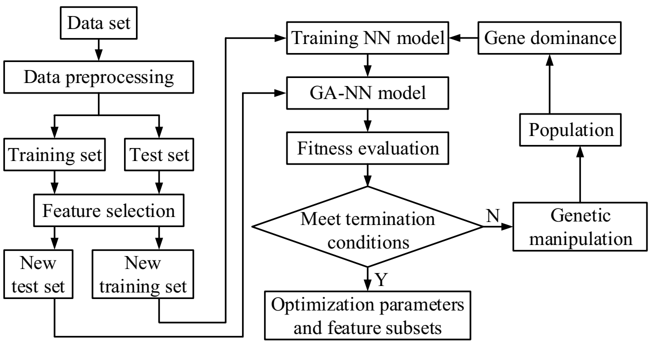

Fig. 4. Overall structure of genetic algorithm-neural network.

Table 1
Experimental environment parameter setting

Number	Project	Size	Unit
#1	Memory	12	GB
#2	Programming tools	C++, Python	/
#3	Central Processing Unit	Intel Core i5-6500	/
#4	Operating system	Windows XP	/
#5	Computing equipment	Excel, SPSS 26.0	/

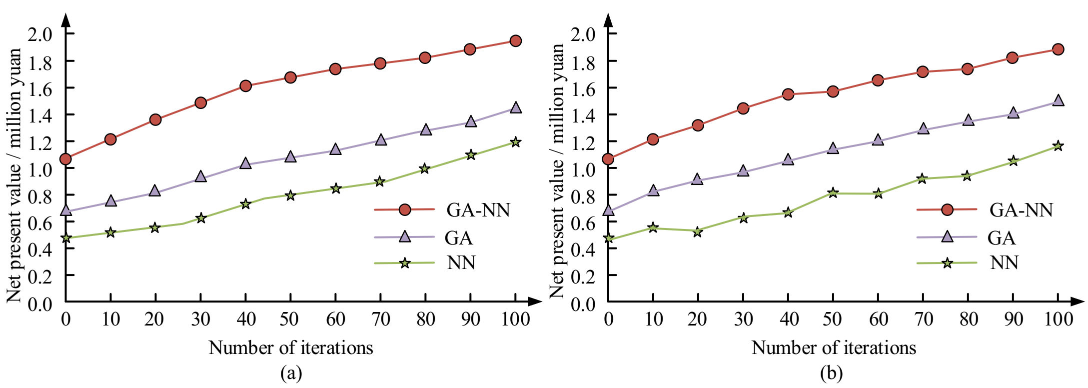

Fig. 5. Changes of net present value of the three models in the optimization process.

computer memory. The time spent in the iterations of the three models is compared and analyzed, and the consequences are shown in Fig. 7.

Fig. 7a and b, respectively represent the time changes of the three models in reservoirs A and B with the increase of the amount of iterations. In Fig. 7, with the increase of the amount of iterations, the time spent by the three models gradually increases, and the time spent by the three models is in descending order: NN model, GA model, and GA-NN model. For example, in reservoir B, when the amount of iterations is 50, the time spent by the NN model, the GA model, and the GA-NN model is 2.15, 1.89, and 1.37 s, respectively, that is, the GA-NN model is faster than the GA model. 0.52 s, which is 0.78 s faster than the NN model. The sample data of 200 oil fields are trained, and the training process of the three models is shown in Fig. 8.

In Fig. 8, the abscissa represents the amount of training times, and the ordinate represents the error. In Fig. 8, the error of the NN model tends to be stable when it is trained for 63 times, and the error value remains around 5; the error of the GA model tends to be stable when it is trained for 55 times, and the error value remains around 3.5; GA-NN When the model is trained for 37 times, the error tends to stabilize, and the error value remains around 1.5. That is to say, the training times of the three models are in descending order: NN model, GA model, and GA-NN model, that is, the GA-NN model obtains the optimal solution faster than the other two models, and the error value is also less than NN model, GA model. Finally, the GA-NN model, the GA model, and the NN model are compared for the accuracy of predicting the development risk, and the 10 oil fields are divided into one group, with a total of 5 groups. And repeat the training, and get the standard deviation, the consequences are shown in Fig. 9.

As can be seen from Fig. 9, among the three models, the GA-NN model has the highest prediction accuracy, the GA model is the second, and the NN model is the lowest, and the standard deviations of the three models are all within 0.3, indicating that the statistical difference is small. The highest

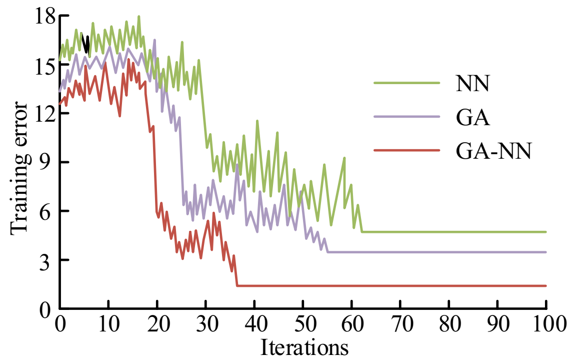

Fig. 8. Training process of three models.

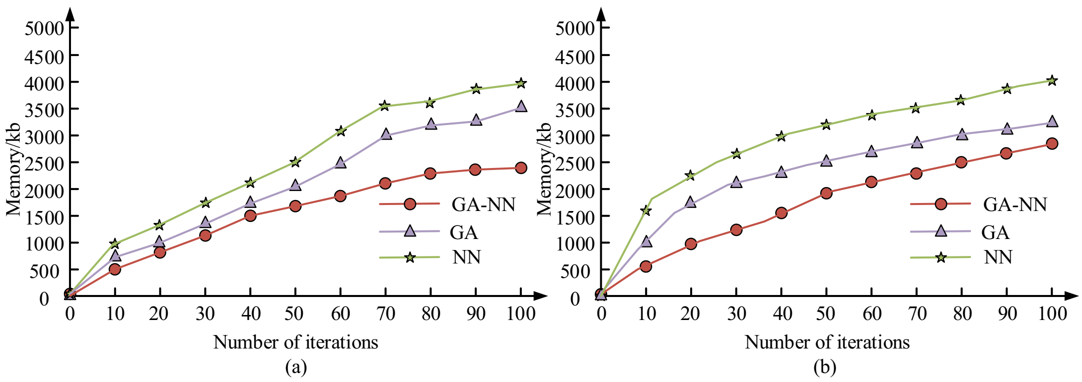

Fig. 6. Relationship between increase of training times and memory occupancy.

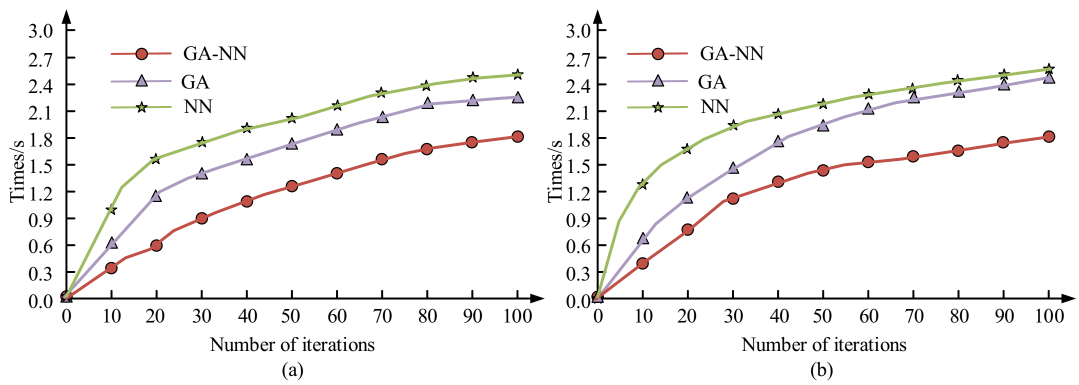

Fig. 7. Time spent in the iteration process of the three models.

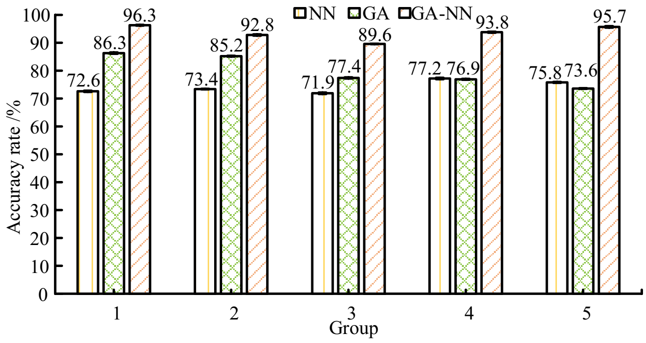

Fig. 9. Comparison of prediction accuracy of three models.

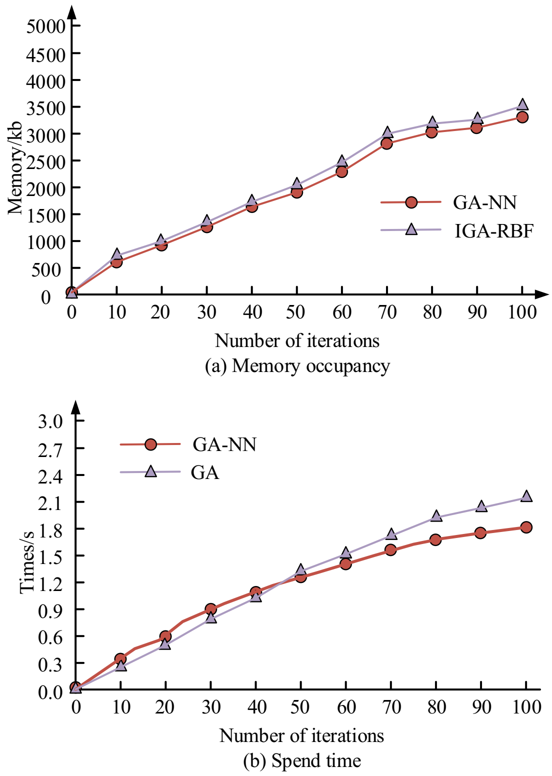

Fig. 10. Performance analysis of two models.

prediction accuracy of the GA-NN model, GA model, and NN model are 96.3%, 86.3%, and 77.2%, respectively, and the lowest prediction accuracy of the GA-NN model is 89.6%, which are higher than the highest of the two models. The prediction accuracy shows that the GA-NN model has a higher accuracy in the development risk prediction of shallow sea oil.

In addition, comparing the GA-NN model with the newer Radial Basis Function of Improved Genetic Algorithm (IGA-RBF) can further demonstrate the performance of the GA-NN model. Compare and analyze the increase of training times, memory occupancy and time spent in the two models, and the results are shown in Fig. 10.

As can be seen from Fig. 10, as the number of iterations increases, the memory occupied, and time spent by GA-NN and IGA-RBF increase. And the memory utilization of GA-NN is always lower than that of IGA-RBF; Before 45 iterations, GA-NN spent more time than IGA-RBF, and after 45 iterations, GA-NN spent less time than IGA-RBF, further indicating that GA-NN has better performance.

5. Conclusion

Oil is one of the most important energy sources in the development of human society at present and in the next few decades. Strengthening the management of oil reservoirs and improving the efficiency of oil exploitation are the key issues that human beings should pay attention to. However, oil extraction is full of risks and uncertainties, and various risks need to be accurately assessed in order to maximize profits and reduce the cost of production operations. Based on the fuzzy statistical cluster analysis method, it can promote the classification of reservoir data and formulate a more reasonable reservoir development plan. Combining neural network and genetic algorithm to enhance the problems existing in well layout optimization is helpful for the staff to analyze the well layout risks they are facing. The experimental results show that the GA-NN model occupies the least memory and takes the least time. Under the same amount of iterations, it saves 870 and 1,470 kb of memory compared to the comparison model GA model and NN model, and the time is 0.52 and 0.78 s faster, and it has a faster speed. The speed and minimum error of the optimal solution, as well as the highest risk prediction accuracy, the highest accuracy of GA-NN model is 96.3%, which is higher than 77.2% and 86.3% of NN model and GA model. It can be seen from the experiments that the overall performance of the GA-NN model constructed in this study is good, and it can provide a certain reference value for the prediction of the risk of shallow sea oil development, but it also has shortcomings. Appropriate improvements should be made to fuzzy clustering based on the actual situation of the oilfield, and a prediction model should be established in combination with more scientific theories to predict the recovery factor of Qianhai oilfield, and guide its subsequent development and management, so as to promote the effective development of shallow sea oil. And improve the mining efficiency, this is the direction of future research.

References

[1] J. Li, Y. Sun, L.L. Gong, N. Chai, Y.F. Yin, Multiattribute fuzzy decision evaluation approach and its application in enterprise competitiveness evaluation, Math. Probl. Eng., 2021 (2021) 1–11, doi: 10.1155/2021/8867752.

[2] P.S. Chen, Y.J. Zheng, L. Li, T. Jing, X.X. Du, J.Z. Tian, J.X. Zhang, M.Y. Dong, J.C. Fan, C. Wang, Z.H. Guo, Prediction of PM2.5 mass concentration based on the back propagation (BP) neural network optimized by *t*-distribution controlled genetic algorithm, J. Nanoelectron. Optoelectron., 15 (2020) 432–441.

[3] B. Gao, V. Balyan, Construction of a financial default risk prediction model based on the LightGBM algorithm, Int. J. Intell. Syst., 31 (2022) 767–779.

[4] H. He, H.H. Xing, D. Hu, X.C. Yu, Novel fuzzy uncertainty modeling for land cover classification based on clustering analysis, Sci. China (Earth Sci.), 62 (2019) 438–450.

[5] X. Guo, T. Chen, W. Huang, Y. Tang, Fuzzy clustering algorithm of transaction database based on big data, IPPTA: Q. J. Indian Pulp Paper Tech. Assoc., 30 (2018) 579–586.

[6] S.R. Rajesh, E. Kanniga, M. Sundararajan, A fuzzy clustering approach for micro aneurysm lesion detection in diabetic retinopathy, Int. J. Res. Pharm. Sci., 11 (2020) 503–511.
[7] Z. Chen, Using big data fuzzy *k*-means clustering and information fusion algorithm in English teaching ability evaluation, Complexity, 2021 (2021) 5554444, doi: 10.1155/2021/5554444.
[8] Y.L. Mi, Y. Shi, J.H. Li, W.Q. Liu, M.Y. Yan, Fuzzy-based concept learning method: exploiting data with fuzzy conceptual clustering, IEEE Trans. Cybern., 52 (2022) 582–593.
[9] M.K. Almedallah, S.D.C. Walsh, Integrated well-path and surface-facility optimization for shallow-water oil and gas field developments, J. Pet. Sci. Eng., 174 (2019) 859–871.
[10] N.K. Shu, C.G. Su, X.G. Shi, Z.P. Li, X.F. Zhang, X.H. Chen, J.B. Zhu, L. Song, Seismic description and fluid identification of thin reservoirs in Shengli Chengdao extra-shallow sea oilfield, Pet. Explor. Dev., 48 (2021) 889–899.
[11] W.X. Wang, R.C. Tang, C. Li, P.S. Liu, L. Luo, A BP neural network model optimized by Mind Evolutionary Algorithm for predicting the ocean wave heights, Ocean Eng., 162 (2018) 98–107.
[12] Q.B. Ren, M.C. Li, S. Han, Discrimination and comparison experiments of basalt tectonic setting based on improved genetic algorithm-optimized neural network, Earth Sci. Front., 26 (2019) 117–124.
[13] W. Fan, Y.Y. Lin, Z.S. Li, Prediction model of the creep of piezo ceramic based on BP neural network optimized by genetic algorithm, Electr. Mach. Control, 22 (2018) 91–96.
[14] M. Khandelwal, A. Marto, S.A. Fatemi, M. Ghoroqi, D.J. Armaghani, T.N. Singh, O. Tabrizi, Implementing an ANN model optimized by genetic algorithm for estimating cohesion of limestone samples, Eng. Comput., 34 (2018) 307–317.
[15] Y. Wang, Risk prediction of information leakage in new product development stage based on data driven model, Int. J. Prod. Dev., 25 (2021) 114–129.
[16] X.H. Li, R.C. Jia, R.R. Zhang, S.Y. Yang, G.M. Chen, A KPCA-BRANN based data-driven approach to model corrosion degradation of subsea oil pipelines, Reliab. Eng. Syst. Saf., 219 (2022) 108231, doi: 10.1016/j.ress.2021.108231.
[17] R. Zhang, J.H. Wang, S.J. Liu, Z.Z. Zhang, L. Ma, W.B. Meng, Z.Y. Wang, Y. Huang, H.X. Liu, F. Xu, Modeling and durability behavior of erosion–corrosion of sand control screens in deepwater gas wells, ACS Omega, 6 (2021) 23943–23951.
[18] P. Saravanan, A.K. Gupta, H.B. Zheng, M.K. Panigrahi, S.K. Tiwari, S.K. Rai, M. Prakasam, Response of shallow-sea benthic foraminifera to environmental changes off the coast of Goa, eastern Arabian Sea, during the last ~6100 cal yr BP, Geol. Mag., 157 (2020) 497–505.
[19] Y. Zhuang, Simulation of economic cost prediction of offshore oil pollution based on cloud computing embedded platform, Microprocess. Microsyst., 83 (2021) 103993, doi: 10.1016/j.micpro.2021.103993.