The Establishment of a Drilling Rate of Penetration Prediction Model Based on GA-BP

Li Qi1,a, Zhang Wei2,b, Xie Cong3,c

1Xi’an Shiyou University, Xi’an Shiyou University, Yanta District, Xi’an City, Shaanxi Province, China
2Xi’an Shiyou University, Xi’an Shiyou University, Yanta District, Xi’an City, Shaanxi Province, China, 86-18792430016
3Xi’an Shiyou University, Xi’an Shiyou University, Yanta District, Xi’an City, Shaanxi Province, China, 86-13152195220
aliqi@xsyu.edu.cn, b425808315@qq.com, c13152195220@163.com

ABSTRACT—The prediction of ROP is the key to optimization and control of drilling engineering. The lack of application of actual engineering data in existing theoretical models makes it difficult to meet field requirements. This paper establishes a new model for predicting ROP, which combines artificial intelligence algorithms and neural networks. First, the wavelet filtering method is used to reduce the noise of the measured data at the drilling site, and the input parameters of the ROP prediction model are optimized according to the mutual information correlation analysis to reduce the model redundancy. Secondly, the genetic algorithm (Genetic Algorithm) is used to optimize the initial weights and thresholds of the BP neural network, thereby establishing a new ROP prediction model. The final results show that GA-BP has stronger convergence and search ability and better calculation accuracy.

1. INTRODUCTION
Intelligent drilling is one of the important links in the construction of intelligent oilfields. The application of new information technologies such as big data and machine learning to oil drilling will help promote the realization of intelligent drilling. Zhao Ying and others established an offshore drilling ROP prediction model based on an extreme learning machine, which is to monitor the ROP in real time and optimize drilling parameters to achieve early warning and effective prevention of drilling accidents, thereby improving drilling efficiency[1]. Su Xinghua and others designed and implemented a mechanical drilling rate prediction model based on the GBDT algorithm. At the same time, compared with machine learning algorithms such as SVM, LR, and KNN, the algorithm has a high accuracy rate and provides a scientific reference for improving the drilling rate[2]. Liu Shengwa, Sun Junming and others proposed an implementation method of ROP prediction model for directional wells based on artificial neural network technology. Under the conditions of sufficient data volume and high data quality, the prediction model constructed by neural network can be more improved. Make predictions efficiently[3]. Wang Wen, Liu Xiaogang and others analyzed various factors affecting the deep ROP and established a deep ROP prediction neural network, and verified them based on examples, which proved that the prediction results of the neural network are consistent with the actual ROP. Good[4].

The prediction of ROP is one of the core issues in the intelligent drilling process. In recent years, data-driven SVR (Support Vector Regression), ANN (Approximate Nearest Neighbor), GBDT (Gradient Boosting Decision Tree), etc., have been used to carry out ROP prediction research. However, these models have their own limitations. Therefore, this paper proposes a new ROP prediction model based on wavelet filtering and genetic algorithm.
Boosting Decision Tree) and other algorithms have been widely used in ROP prediction\cite{5-6}. However, these algorithms are easy to fall into the local optimum and the prediction results are poor in stability. On this basis, it is necessary to develop a ROP prediction model based on a hybrid algorithm. This paper proposes a BP (Back Propagation) neural network ROP prediction model based on GA (Genetic Algorithm) optimization. The model uses wavelet filtering method to denoise the data, and uses mutual information for correlation analysis to determine model input parameters. The simulation results show that the GA optimized BP neural network has stronger search ability and convergence, and has higher prediction accuracy.

2. BP neural network and genetic algorithm

2.1. BP neural network

BP neural network is currently one of the most active front-end research in the field of artificial intelligence \cite{7}. BP neural network is a typical hierarchical multi-element network with input layer, multiple hidden layers and output layer. The topology of the neural network is shown in Figure 1.

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{bp_neural_network.png}
\caption{Topological structure of BP neural network}
\end{figure}

2.2. GA algorithm

The genetic algorithm is inspired by Darwin’s theory of evolution. The genetic algorithm has a larger search range for the optimal solution, and the genetic algorithm is more robust. Genetic algorithm is a simulation of biological genetic evolution mechanism, and gradually developed into an optimization algorithm that searches for extreme values at random.

The specific steps of GA algorithm are as follows.

2.2.1. Select operation

The fitness ratio method is commonly used when performing selection operations, which is usually called the roulette method. The selection operation refers to selecting the best from the old group with a certain probability and creating a new group. If the individual is more adaptable, the individual is more likely to be selected. Assuming one of the chromosomes \( x \) in the population, the fitness value corresponding to the chromosome is \( f_i \), that is, the probability that the chromosome is selected to be inherited from the next generation can be expressed by the following formula:

\[
F_i = k_f / f_i
\]

\[ p_i = F_i / \sum F_i \]  

(1)  

(2)
2.2.2. Cross operation
Cross operation refers to selecting two random individuals from the original population, and exchanging the genetic information of the selected individuals to produce a more excellent new individual. The cross method used in this paper is an arithmetic cross operation. If two random individuals are both parents and they are paired, they usually choose the cross point $\alpha$ with a certain probability to perform the cross operation. The specific expression is as follows [8]:

$$\begin{align*}
x'_a &= x_a (1-h) + y_a h \\
y'_a &= y_a (1-h) + x_a h
\end{align*}$$

(3)

In formula (3), $h$ usually takes a value between $[0,1]$.

2.2.3. Mutation operation
The mutation operation refers to the random selection of a single individual from the population to mutate one or some parts of its chromosome in order to make the offspring of the population better. The more commonly used mutation operation is the non-uniform mutation operation. Non-uniformity means that the evolutionary algebra is different, and the amount of chromosome variation is uneven. Suppose you find a chromosome $X=(x_1,x_2,\ldots,x_k,\ldots,x_n)$ from the population. After mutation, a new chromosome $X'=(x'_1,x'_2,\ldots,x'_k,\ldots,x'_n)$ is obtained according to the following mutation operation method.

$$x'_i = \begin{cases} x_i + (d_i - x_i) \times f(e) & \zeta_i > 0.5 \\ x_i - (x_i - c_i) \times f(e) & \zeta_i \leq 0.5 \end{cases}$$

(4)

$$f(e) = \zeta_2 \left(1-e/e_{\text{max}}\right)^2$$

(5)

In formulas (4) and (5), $x_i$ the upper bound is $d_i$, the lower bound is $c_i$, and the value range of $\zeta_i$ and $\zeta_2$ is $[0,1]$, $e_{\text{max}}$ which is the largest evolutionary algebra of the chromosome, and $e$ is the current evolutionary algebra of the chromosome.

Based on the BP neural network optimized by the GA algorithm, the BP neural network prediction model is used to randomly generate the initial weights and thresholds of $[0,1]$; secondly, the model error is taken as the optimization target, and the GA algorithm is used to calculate the fitness of the error result. When the calculation accuracy or iteration requirements are met, the GA optimized weights and thresholds are output, then a new BP neural network is trained.
3. GA-BP ROP prediction modeling

3.1. Data preprocessing

3.1.1. Wavelet filtering processing
The commonly used filtering processing methods in data preprocessing include Wiener filtering, Kalman filtering, matched filtering and wavelet filtering.

3.1.1.1 Wiener filtering
Wiener filtering is the best filtering for a stationary random process, in which the state parameters of the system are stable and unchanging. There are certain limitations in practical applications: it is not suitable for the filtering of non-stationary random processes, it needs to use sampling data at all times, the required data storage capacity is large, the calculation amount is large and the Wiener-equation has no solution.

3.1.1.2 Kalman filtering
Kalman filter is not only suitable for stationary random processes, but also for non-stationary random processes. It expresses the state transition of the system by state equations, and replaces the linear
equations in Wiener filtering with fixed-dimensional matrix operations recursive. It overcomes a series of limitations of Wiener filtering and has been successfully applied.

### 3.1.1.3 Matched filtering

Matched filtering does not belong to the state estimation of the system, but belongs to the statistical detection of the signal. Matched filtering is different from general filtering methods. Its purpose is not to better recover the signal waveform, but to maximize the output signal-to-noise ratio at a certain decision time T, thereby effectively detecting the signal.

### 3.1.1.4 Wavelet filtering

Wavelet filtering is a frequency domain filter. Its characteristic is to separate the signal and noise in frequency, suppress the noise outside the useful signal band, and pass the useful signal, but it cannot suppress the noise that occupies the same frequency band as the useful signal. The wavelet filter transform is suitable for the spectrum analysis of time-varying signals, and can display the characteristics of signal frequency changing with time.

This paper uses wavelet filtering. There are various interference factors in the complex drilling environment, resulting in a certain error between the measured value of the parameter data received by the measuring instrument and the actual value. In this paper, wavelet transform is used for denoising processing, which has higher frequency resolution and lower time resolution in the low frequency part, and higher time resolution and lower frequency resolution in the high frequency part, which can effectively remove white noise generated in the measurement equipment. Wavelet filtering will analyze the signal to do wavelet transform as follows:

\[
W_f(\tau,a) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} X(t) \cdot \varphi\left(\frac{t-\tau}{a}\right) dt
\]

In the formula, \(a > 0\) is the scale factor, which realizes the expansion and transformation of the basic wavelet; \(\tau\) is the translation factor, which realizes the translation and transformation of the basic wavelet on the time axis.

### 3.1.2. Mutual information correlation analysis

Commonly used in correlation analysis are covariance and covariance matrix, correlation coefficient, simple regression and multiple regression and mutual information correlation analysis.

#### 3.1.2.1 Covariance and covariance matrix

Covariance is used to measure the overall error of two variables. If the trends of the two variables are the same, the covariance is positive, indicating that the two variables are positively correlated; if the trends of the two variables are opposite, the covariance is negative, then It shows that the two variables are negatively correlated; if the two variables are independent of each other, then the covariance is 0, which means that the two variables are not correlated. The calculation formula of the covariance is as follows:

\[
COV(X,Y) = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{n-1}
\]

Covariance can only perform correlation analysis on two sets of data. When there are more than two sets of data, a covariance matrix is required. The following is the calculation formula of the covariance matrix of three sets of data \(x, y, z\).

\[
C = \begin{bmatrix}
COV(x,x) & COV(x,y) & COV(x,z) \\
COV(y,x) & COV(y,y) & COV(y,z) \\
COV(z,x) & COV(z,y) & COV(z,z)
\end{bmatrix}
\]
3.1.2.2 Correlation coefficient
The correlation coefficient is a statistical indicator that reflects the closeness of the relationship between variables, and the value interval of the correlation coefficient is \([-1, 1]\). -1 means that the two variables are completely negatively correlated, 1 means that the two variables are completely linearly correlated, and 0 means that the two variables are not correlated. The closer the data is to 0, the weaker the correlation. The calculation formula of the correlation coefficient is as follows:

\[
\gamma_{xy} = \frac{s_{xy}}{s_x s_y}
\]

Where \(\gamma_{xy}\) represents the sample correlation coefficient, \(s_{xy}\) represents the sample covariance, \(s_x\) represents the sample standard deviation of \(x\), and \(s_y\) represents the sample standard deviation of \(y\).

3.1.2.3 Simple regression and multiple regression
Regression analysis is a statistical method to determine the relationship between two or more variables. Regression analysis is divided into univariate analysis and multivariate analysis according to the number of variables. Two variables use simple analysis, and two or more variables use multiple analysis.

The simple regression equation is shown below, where \(y\) is the dependent variable, \(x\) is the independent variable, \(k\) is the slope of the equation, and \(b\) is the intercept of the equation.

\[
y = kx + b
\]

\[
b = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}
\]

The multiple regression equation is as follows:

\[
y = b + k_1x_1 + k_2x_2 + \cdots + k_x x_x
\]

3.1.2.4 Mutual information correlation analysis
Mutual information is a method to measure the correlation between the characteristic values of data. Since most of the drilling parameters are non-linear relationships, in order to better analyze the drilling speed parameters, the text adopts mutual information correlation analysis.

In order to facilitate the numerical characterization of the relationship between non-linear related variables and realize quantitative statistics, this paper introduces mutual information correlation analysis \([9]\). Mutual information is a common method used to measure the non-linear relationship between variables. When \((X, Y) - p(x, y)\), the mutual information between variables \(X\) and \(Y\) is defined as:

\[
MI(X;Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \cdot \log \frac{p(x, y)}{p(x) \cdot p(y)}
\]

Where \(p(x, y)\) is the joint probability distribution function of \(X\) and \(Y\); \(p(x)\) and \(p(y)\) are respectively the marginal probability distribution function of \(X\) and \(Y\).

3.2. ROP prediction modeling
The GA-BP ROP prediction modeling process is shown in Figure 3.

The GA-BP ROP prediction process is divided into three parts as a whole. The first part is data preprocessing, which mainly includes data preparation, data cleaning and correlation analysis. This part is mainly to sort the field data, reduce noise and correlation analysis, and screen the input parameters of BP neural network; the second part is GA optimization BP Neural network, using GA algorithm to optimize the initial BP neural network parameters, find the best weights and thresholds according to the error results; the third part establishes the GA-BP drilling rate prediction model, and uses the second part output results to create a new BP network topology. The structure and the results of mutual information screening are used as input parameters to train the BP neural network. If the model accuracy
meets the design requirements, the model is output, otherwise the iteration is repeated until the accuracy meets the requirements to output the model.

4. Case analysis

4.1. Parameter setting
The data of this example comes from the operation data of a directional well in a block on site. The well is a production well. The designed well depth is 3800 meters, the actual drilling depth is 3720 meters, the completion method is perforation completion, the drilling cycle is 23 days, and the average ROP is 12.52 m/h.

Perform mutual information correlation analysis on the filtered operation data. Figure 4 shows the correlation analysis diagram of drilling parameters.
Among them, the ROP has a strong correlation with drilling parameters such as rotation speed, weight on bit, depth, well diameter, and drilling fluid consistency coefficient. The mutual information value with displacement and riser pressure is low. Therefore, the rotation speed, weight on bit, depth, well diameter and drilling fluid consistency coefficient were selected as input variables of the ROP prediction model, and the GA-BP ROP prediction model was established.

4.2. Experimental results
In order to verify whether the GA-BP neural network model is superior to other optimized intelligent models, this paper selects the standard BP neural network model and the BAS-BP neural network model for comparative analysis.

Figure 5 is a comparison diagram of the prediction results of the BP network model test set after multiple algorithm optimization.

It can be seen from Figure 5 that the fold line trend of the BP forecast value and the BAS-BP forecast value deviates greatly from the true value, while the GA-BP fold line trend deviates from the true value relatively small.

Calculate the goodness of fit of different models, and the results are shown in Table 1.
### Table 1  Comparison of evaluation indexes of different models

| Algorithm type | BAS-BP | BP   | GA-BP |
|----------------|--------|------|-------|
| $R^2$          | 0.9133 | 0.8957 | 0.9513 |

It can be seen from Table 1 that the goodness of fit refers to the overall fit of the regression equation. The maximum value is 1. When the value of is closer to 1, the better the fit of the regression line to the observed value is. The analysis of the results in Table 1 shows that the goodness of fit of GA-BP is closest to 1. Therefore, the GA-BP ROP prediction model proposed in this paper has better prediction accuracy than BP and BAS-BP models.

5. Conclusion

(1) Applying wavelet filtering method and mutual information correlation analysis to the prediction of ROP can effectively reduce noise interference and model redundancy, and simplify the calculation process.

(2) The application of the GA-BP neural network prediction model overcomes the shortcomings of the standard BP neural network such as poor stability, slow convergence, and easy to fall into local optimum. This method improves the performance of the network and the accuracy of the prediction results.

(3) The research in this paper proves the feasibility and effectiveness of the data-driven model in the application of drilling ROP prediction, and provides a better way to predict ROP.

REFERENCES

[1] Zhao Ying, Sun Ting, Yang Jin, et al. Drilling rate monitoring and real-time optimization of offshore drilling machinery based on extreme learning machine[J]. China Offshore Oil and Gas, 2019,31(06):138-142.

[2] Su Xinghua, Sun Junming, Gao Xiang, Wang Min. Research on Drilling Machine Drilling Rate Prediction Method Based on GBDT Algorithm[J]. Computer Application and Software, 2019,36(12):87-92.

[3] Liu Shengwa, Sun Junming, Gao Xiang, Wang Ming. Analysis and establishment of drilling speed prediction model of drilling machinery based on artificial neural network [J]. Computer Science, 2019,46(6A):605-608.

[4] Wang Wen, Liu Xiaogang, Dou Peng, et al. Prediction Method of Deep Drilling Speed Based on Neural Network [J].Oil Drilling & Production Technology,2018,40:121-124.

[5] Hamid Reza Ansari,Mohammad Javad Sarbaz Hosseini,Masoud Amirpour. Drilling rate of penetration prediction through committee support vector regression based on imperialist competitive algorithm [J]. Carbonates and Evaporites,2017,32(2).

[6] Liu Shengwa, Sun Junming, Gao Xiang, Wang Min. Analysis and establishment of drilling rate prediction model based on artificial neural network [J]. Computer Science, 2019, 46(S1): 605-608.

[7] Wu Fengbo, Zhao Pan, Lu Qiantong. Urban water consumption prediction based on GA-BP neural network [J]. Modern Electronics Technique, 2020,43(08):147-150.

[8] Sun Liwen, Liu Hai, Wang Haiyang, et al. Application of GA-BP Neural Network to the Research on the Interior Sound Quality in Accelerating Conditions [J]. Science Technology and Engineering , 2017,17(17):340-345.

[9] Yashuang Mu, Xiaodong Liu, Lidong Wang. A Pearson’s correlation coefficient based decision tree and its parallel implementation. Information Sciences, 2018, 435:40-58.