ISSN: 2664-9640

Production Capacity Prediction of Fractured Vertical Wells Based on BP Neural Network

Jiaxi Zhang

College of Computer Science, Sichuan University, Chengdu, China

Abstract

Accurate prediction of production capacity for tight gas fractured wells is an important prerequisite for making informed development decisions. Addressing the limitations of traditional methods in terms of assumptions and lack of historical data, as well as the complex nonlinear relationship between geological and engineering parameters, this paper takes the Sulige tight gas field in the Ordos Basin as an example. It utilizes Pearson correlation analysis to determine the dominant factors and weights that influence the post-fracturing production of 253 tight gas vertical wells in the Sulige area. Then, combining geological and engineering parameters, a production capacity prediction model for tight gas vertical wells is established based on the BP neural network algorithm. The results show that the average error of the predicted production capacity of tight gas vertical wells based on the established model using production well data from the Sulige area is 11.47%. This model enables accurate and efficient prediction of post-fracturing production capacity, providing important scientific basis for the economic development of tight gas fields.

Keywords

Neural network, Pearson coefficient, fracturing vertical wells, production capacity prediction, tight gas.

1. Introduction

The Sulige gas field in the Ordos Basin is the largest onshore gas field in China, characterized by "four lows": low porosity, low pressure, low permeability, and low abundance in tight sandstone gas reservoirs. Hydraulic fracturing is an effective method to enhance the production of tight sandstone gas reservoirs, and production capacity prediction is a key scientific issue for achieving economic development of tight gas reservoirs. Accurate prediction of production capacity for vertically drilled tight gas wells after hydraulic fracturing is an important prerequisite for making informed development decisions, and it holds significant importance for the exploration and development of the Sulige gas field in the Ordos Basin. Currently, the methods for evaluating production capacity in low-permeability tight gas reservoirs, both domestically and internationally, mainly fall into two categories: analytical models based on complex mathematical formulas derived from production equations, and numerical models that utilize production dynamic data for capacity prediction. These methods are primarily based on theoretical models, requiring idealized assumptions and parameters that are not easily obtainable. Moreover, during the early stage of exploration and production testing, there is a lack of historical fitting data, making it impractical to apply theoretical models for production prediction. Additionally, due to the comprehensive influence of geological and engineering parameters, there exists a complex nonlinear relationship between geological parameters, fracturing engineering parameters, and production of tight gas vertical wells. Conventional linear regression methods have limited accuracy in predicting production capacity.

While simple models such as linear regression often have better interpretability and are easier to understand in terms of their internal workings, their predictive capabilities are often limited and unable to model the inherent complexity within the dataset, such as feature interactions. In contrast, black-box models such as neural networks typically exhibit higher accuracy and stronger model generalization capabilities. In recent years, some scholars have utilized artificial neural networks to address parameter prediction problems that arise from the complexities and uncertainties of actual oilfield production processes. In the development of tight gas fields, tackling the nonlinear relationship between numerous geological and engineering parameters and tight gas well production is one of the significant challenges of the BP neural network. In this paper, we first calculate the weights of geological and engineering factors that affect post-fracturing production capacity using Pearson correlation analysis. Then, by employing the BP neural network algorithm and leveraging data mining techniques, we directly incorporate geological and engineering parameters, surpassing the limitations of traditional theoretical models, to establish a production capacity prediction model for tight gas vertical wells in the Sulige area, aiming to improve the efficiency and accuracy of capacity prediction.

2. Data Acquisition and Preprocessing

2.1. Data Source

The collected raw data consists of 287 fractured vertical wells in the Sulige gas field. The parameters influencing production capacity include porosity, permeability, gas saturation, effective thickness, average total hydrocarbon, reservoir pressure, and six geological parameters. Additionally, there are six fracturing construction parameters: flowback rate, pumping rate, proppant concentration, total fluid volume, total proppant volume, and nitrogen usage. The target parameter is the cumulative gas production after one year of fracturing.

2.2. Data Preprocessing

The data used in this study is derived from the actual production of the Sulige gas field. Due to variations in data recording among different blocks and the presence of missing or outlier values in the actual production data, direct training is not feasible. Therefore, prior to using machine learning to predict post-fracturing production capacity, data preprocessing operations, such as data cleaning, are necessary to achieve higher prediction accuracy.

Currently, there are two main approaches for handling missing values: directly deleting samples or features with missing values or filling in the missing values. In this study, features with missing values that account for more than half of the original data are removed. For features such as porosity, reservoir pressure, and gas saturation that exhibit varying degrees of missing values, features with missing values that account for more than half of the original data are removed, while for other geological parameters with missing values, the corresponding feature means are used for imputation.

The presence of outlier values in the dataset can also affect model prediction accuracy. In this study, an outlier detection method based on box plots is employed. In a box plot, outliers are typically considered as values greater than the upper quartile plus 1.5 times the interquartile range or values less than the lower quartile minus 1.5 times the interquartile range. The upper quartile represents the value below which 75% of the data falls, while the lower quartile represents the value above which 25% of the data falls. The interquartile range is the difference between the upper quartile and the lower quartile.

After handling outlier and missing values, a total of 253 usable samples are obtained to establish the dataset for predicting post-fracturing production capacity in the Sulige tight gas field.

3. Identification of Post-Fracturing Production Capacity Influencing Factors Based on Pearson Correlation Analysis

3.1. Pearson Correlation Analysis

Pearson correlation analysis is performed on the determined q influencing factors. Pearson correlation is used to assess the strength of the linear relationship between two continuous variables. Its purpose is to remove parameters that exhibit high linear correlation with each other and reduce the dimensionality of the data. Let X and Y represent the sample data, and the calculation formula for Pearson correlation is as follows:

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}} \tag{1}$$

After calculating the Pearson correlation coefficient matrix, a correlation matrix heatmap can be created to visualize the magnitude of the correlation coefficients between the parameters.

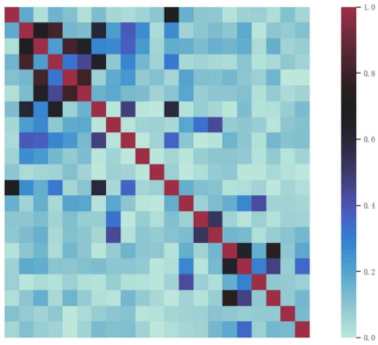

Figure 2.1 Heatmap of Correlation Coefficients

Table 2.1 Correlation Coefficients between Factors and Production Capacity

Parameters	Coefficient	Parameters	Coefficient
Porosity	0.756	Permeability	0.453
Gas Saturation	0.820	Effective Thickness	0.661
Average Total Hydrocarbon	0.734	Reservoir Pressure	0.314
Return Ratio	0.597	Displacement	0.379
Sand Ratio	0.612	Total Fluid Volume	0.624
Total Sand Volume	0.791	Liquid Nitrogen Usage	0.479

3.2. Model Indicator Selection Based on Embedded Methods

Embedded methods are a technique that allows the algorithm to autonomously determine which indicators to select, while simultaneously training the model. As the selected indicators become optimal, the model's accuracy improves. To identify the optimal indicators, the Support Vector Machine (SVM) algorithm was employed. The sorted parameters were fed into the SVM model to predict the post-fracturing production capacity.

The impact of different numbers of indicators on the production capacity prediction results is illustrated in the figure below:

ISSN: 2664-9640

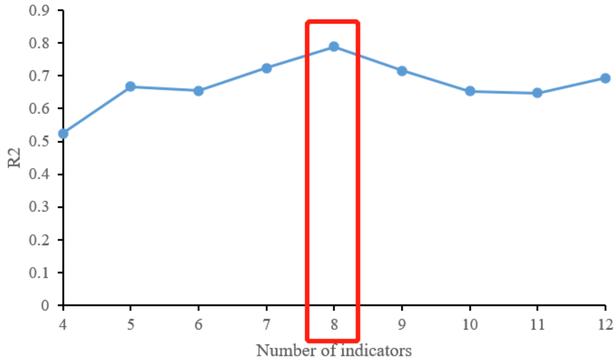

Figure 2.2 Relationship Between Number of Indicators and Model Correlation Coefficient

From the graph, it can be observed that the model's correlation coefficient is highest when the number of indicators is 8. Based on the above analysis, the following indicators were selected as inputs for the prediction model: gas saturation, total proppant volume, porosity, average total hydrocarbon, effective thickness, total fluid volume, proppant-to-fluid ratio, and flowback rate.

4. Construction of Post-Fracturing Production Capacity Prediction Model Based on BP Neural Network

4.1. Basic Principles

The BP (Backpropagation) neural network is a self-learning method used for nonlinear fitting and modeling. It automatically adapts and determines the connection weights of each neuron based on the input training samples. After multiple training iterations, the neural network's weight values store the fitting information extracted from the sample dataset. Finally, by performing calculations using the input data and weights, the desired prediction values can be obtained.

4.2. Model Establishment

For the production capacity prediction model, a classic 3-layer neural network model was adopted. The 8 geological and engineering parameters related to production capacity were chosen as input parameters, resulting in 8 nodes in the input layer. The post-fracturing cumulative gas production after one year was selected as the output parameter, hence the output layer was set to have 1 node. Through multiple experiments, the number of nodes in the hidden layer was determined to be 16. The final topology of the constructed network is illustrated in Figure 3.1.

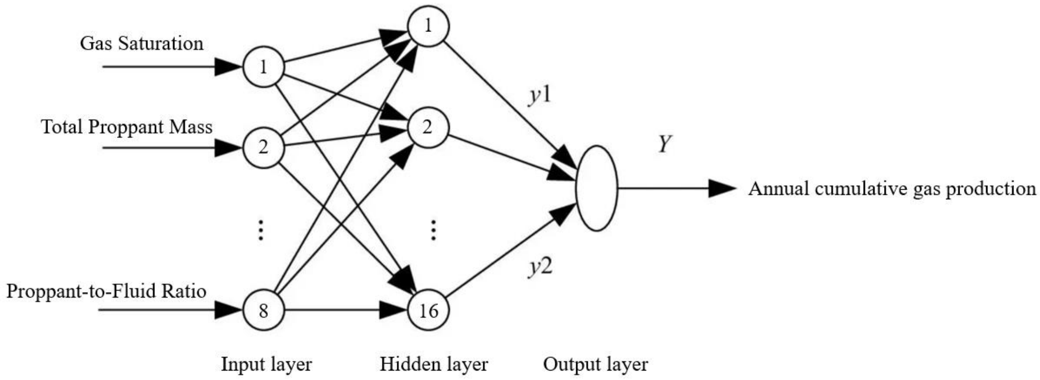

Figure 3.1 Structure of the BP Neural Network Algorithm

During the training process, it is necessary to set appropriate learning rate and iteration count. The learning rate determines the step size of parameter updates and should be chosen based on the specific situation. A learning rate that is too large may prevent the model from converging, while a learning rate that is too small may result in slow convergence. The iteration count represents the number of iterations in the training process and needs to be appropriately chosen to ensure accuracy while avoiding overfitting. Based on the model debugging calculations and experience, the maximum training count for this model was set to 10,000 iterations, with a desired accuracy of 0.00001. Considering the precision and stability requirements of the model, a learning rate of 0.05 was set.

In summary, a classical three-layer BP neural network model was selected. By combining the geological, engineering parameters, and production data of the already producing wells in the Sulige area, a production capacity prediction model for tight gas fractured vertical wells was constructed. By predicting the training data, the model achieved a correlation coefficient of 0.877 and a root mean square error of 2.348, indicating high training accuracy and good capability for production capacity prediction. The model was further validated using data from 12 wells, as shown in Figure 3.2.

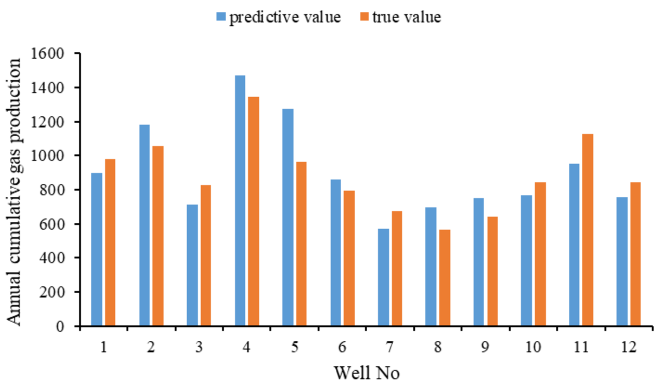

Figure 3.2 Training Data Error Comparison

4.3. Model Application

To further validate the applicability of the shale gas horizontal well volume fracturing production prediction model, the geological and engineering parameters of Well A, Well B, and Well C in the Sulige area were inputted into the software to conduct production prediction tests. The predicted results were then compared with the actual field-measured data, as shown in Table 3.1.

Table 3.1 Comparison of Predicted Data and Actual Data

Well No	Predicted Results (10^4m^3)	Actual Data (10^4m^3)	Error (%)
A	896	981	8.66
B	1179	1054	11.86
C	712	827	13.91
Average Error			11.47

From Table 3.1, it can be observed that the maximum relative error of the trained model for predicting production in tight gas fractured vertical wells is 13.91%, and the average error is 11.47%. This indicates that the production prediction model based on BP neural network can effectively capture the underlying patterns and relationships between the test production and various influencing factors. Moreover, the model exhibits a relatively small error rate in terms of prediction accuracy, providing an efficient, feasible, and reasonably accurate method for predicting production in tight fracturing operations.

5. Conclusion

(1) A classical three-layer BP neural network model was selected, and based on the factors affecting tight gas production, a production prediction model was established with seven geological and engineering parameters including gas saturation, total proppant volume, porosity, average total hydrocarbon, effective thickness, etc., as input layers, and one-year cumulative gas production as the output layer.

(2) By utilizing actual data from 253 fractured wells in the Sulige area as training samples, a BP neural network model was constructed to predict production in tight gas vertical wells with high accuracy. The model's generalization ability was validated using data from three actual wells, resulting in a maximum relative error of 13.91% and an average error of 11.47%. The model demonstrates flexibility in operation and high prediction accuracy. This data mining-based analytical approach provides a new perspective for production prediction in the Sulige area, improving the efficiency of production forecasting in gas wells.

References

[1] Li Xianwen, Wang Lili, Wang Wenxiong, et al. Innovation of key technologies and efficient development practice based on completion of small wellbore fracturing: A case study of tight gas reservoirs in the Sulige gas field[J]. Natural Gas Industry, 2022, 42(09): 76-83.

[2] Fu Suotang, Fei Shixiang, Ye Zhen, et al. Parameter optimization of horizontal wells in tight sandstone gas reservoirs[J]. Natural Gas Industry, 2018, 38(4): 10. (In Chinese)

[3] Anderson, Roger-N. "Petroleum Analytics Learning Machine" for optimizing the Internet of Things of today's digital oil field-to-refinery petroleum system[C]. IEEE, 2017.

[4] Temizel, Cenk, Canbaz, Celal-Hakan, Palabiyik, Yildiray, et al. A Comprehensive Review of Smart/Intelligent Oilfield Technologies and Applications in the Oil and Gas Industry[C]. Society of Petroleum Engineers (SPE), 2019.

[5] Yu Hongyan, Ding Shuaiwei, Gao Yanfang, et al. Application of artificial intelligence in improving the effectiveness of oil and gas field exploration and development[J]. Journal of Northwest University (Natural Science Edition), 2022, 52(6): 1086-1099.

[6] Liu He. Prospects of artificial intelligence application in petroleum exploration and development[J]. Journal of Intelligent Systems, 2021, 16(6): 985.

[7] Li Yanzun, Bai Yuhu, Chen Guihua, et al. New technology for shale oil and gas production forecasting based on artificial neural network method: A case study of the Eagle Ford shale oil and gas field in the United States[J]. China Offshore Oil and Gas, 2020, 32(4): 104-110.