Screening and Database Establishment of Factors Influencing Multi-thin Layer Fracturing in Straight Wells

Qinxi Tang [1], Binghong Xie[2], Huohai Yang[3, *]

[1]Sulige Gas Field Branch of the Great Wall Drilling Company in CNPC, Panjin, China

[2]Shuna Gas Field, Southwest Oil and Gas Field Co., Ltd., PetroChina, Luzhou, China

[3]Petroleum Engineering School, Southwest Petroleum University, Chengdu, China

* Corresponding Author Email: SF444805@163.com

Abstract. In this paper, we discuss how to construct and optimize a straight-well multi-thin-layer fracturing model in the field of petroleum exploration by means of influencing factor screening and database establishment, combined with data mining techniques. First, we identify and screen out the key factors affecting the fracturing effect, such as lithology, formation thickness, interstitiality and so on. Next, a comprehensive database is established to integrate the data of these key factors, including geological information, formation data, and fracturing operation records. On this basis, we apply data mining techniques to establish a straight-well multi-thin-layer fracturing model. For model selection, we considered various algorithms such as decision tree, support vector machine, and neural network, and conducted tests to determine the optimal solution. During the model training process, cross-validation, feature engineering and parameter tuning are used to continuously optimize the model performance to ensure that it can accurately predict the fracturing effect. The experimental results show that by utilizing the screened key factors and the established database, the trained data mining model performs well on the test set and is able to effectively predict the fracturing effect of straight wells with multiple thin layers. This approach not only improves the accuracy of prediction, but also provides a systematic analysis and decision-making tool for oil exploration. Future research directions include further enriching the database information, exploring more complex data mining models, and enhancing the interpretability of the model to make the results easier for engineers to understand and apply. This will provide more accurate and reliable decision support for straight-well multi-thin-layer fracturing work.

Keywords: Fracturing; influencing factors; database; straight wells; data mining.

1. Introduction

In recent years, unconventional oil and gas development has gradually become the focus of global attention. As a kind of unconventional oil and gas resources[1-3], the exploitation of tight gas is of great significance. The United States is the country with the earliest development and the most successful development and utilization of tight gas industry in the world, and its tight gas reservoirs are characterized by large thickness of gas layer, high abundance and condensate-containing oil[4-5], and the gas wells have high final cumulative gas production, and the development benefits are good. After nearly 20 years of efforts, China's tight gas development has made great achievements, built China's largest gas field in terms of production scale - Sulig gas field, and formed a series of low-cost development technology for tight gas reservoirs[6-7] . The extraction of tight gas can meet the growing energy demand and reduce the dependence on traditional energy sources, which is conducive to the diversification of the energy structure and ensure energy security. Therefore, the extraction of tight gas is of great significance for energy supply, technological progress, economic development and environmental sustainability. It is of great significance to combine reservoir physical properties, rock mechanics, and fracturing construction factors to clarify the degree of influence of the fracturing scheme on the production capacity of a single well, and to complete the screening and correlation analysis of the influencing factors of multi-thin-layer fracturing in straight wells, and to establish a database.

2. Screening of influencing factors and database establishment

2.1. Establishment of principles for the selection of evaluation indicators for the fracking database

The selection of evaluation indicators in the establishment of the fracking database should be based on the objectives and needs of the database construction, and the operability, reliability, comprehensiveness, generalization, comparability, independence, practicality and applicability of the indicators should be taken into consideration to ensure the scientificity and effectiveness of the evaluation indicators. The main considerations here are comprehensiveness, generalizability, and independence. The evaluation of post-pressing output effect is a direct reflection of the post-pressing effect, which is generally characterized by the daily output, which is greatly affected by the production system and production conditions, such as the effect of production differential pressure.

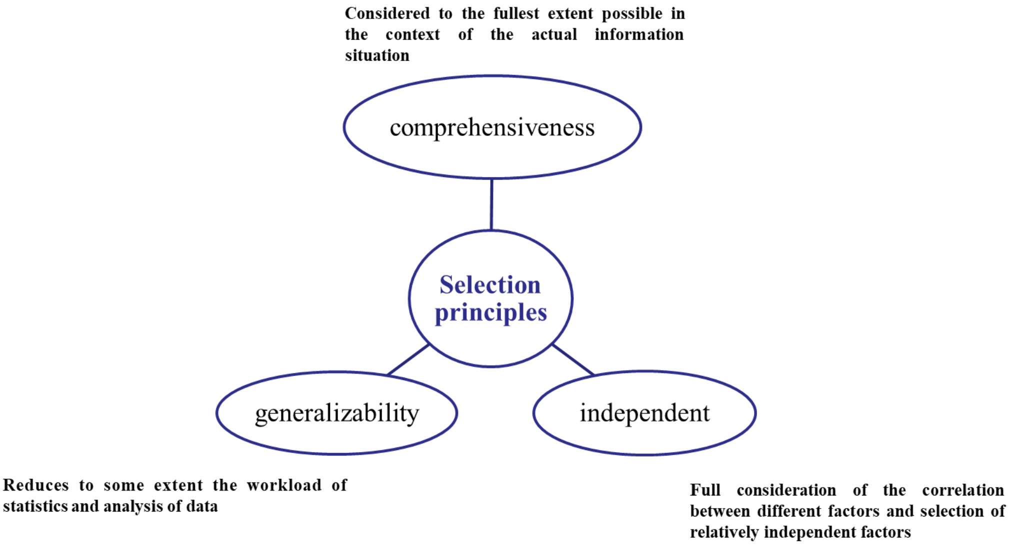

Fig. 1 Principles of selecting evaluation indicators for fracking database

2.2. Establishment of a database to evaluate sources of indicators

Establish a comprehensive and effective database evaluation metrics system to ensure a comprehensive assessment and analysis of database performance, quality and value.Selection of sources: reservoir physical parameters, rock mechanics parameters and fracturing construction parameters.

Table 1 Establishment of a table of evaluation indicators for databases

Select source	Selection results
Reservoir Physical Parameters	Effective reservoir thickness, porosity, permeability, oil saturation, mud content, pore structure index, acoustic time difference, neutron porosity, density, resistivity, thickness of coarse sandstone, thickness of fine conglomerate, thickness of small medium conglomerate, thickness of large medium conglomerate
Rock mechanical parameter	Young's modulus, Poisson's ratio, brittleness index, minimum horizontal principal stress, vertical stress, horizontal principal stress difference
Fracturing Construction Parameters	Pre-liquid ratio, construction displacement, sand addition, liquid nitrogen, average sand ratio

2.3. Establishment of a fracking database

2.3.1 Formation of a set of databases

The data of 5138 well sections of 1606 wells have been completed, and each well section contains 120 parameters related to reservoir physical properties, rock mechanics, fracturing construction, etc., forming a set of database with 120 parameters. Among them, there are 612 wells and 1,357 well sections in Su-10 block, 677 wells and 1,656 well sections in Su-11 block, and 317 wells and 2,125 well sections in Su-53 block. Among them, 220 horizontal wells are mainly concentrated in Su 53 block.

2.3.2 Establishment of sub-databases

In response to the need for correlation analysis, the production blocks were processed statistically separately. Since there were only 30 entries after screening the data of Su 53 block, the number of samples was seriously insufficient and it was difficult to be divided into layers, the data were analyzed by combining the data of the 3 blocks with a total of 850 production wells, and more accurate results were obtained by increasing the training set to establish 9 sub-databases.

2.3.3 Screening of low-producing wells completed

Using single-day gas production as a measure, wells with daily gas production less than 5,000 m3 were screened as low-producing wells, and a separate database of low-producing wells was established, containing 70 low-producing wells.

3. Data mining model building and optimization

3.1. Raw data discretization

The dimensionlessization of raw data is an important step in data preprocessing, and its main purpose is to eliminate the differences in the magnitude of different variables, so that the data are more comparable in the analysis. The physical significance of the original data in the database is very different, the scale is not the same, and the order of magnitude is also different, and the direct use of the original data set will affect the effective degree of the later data mining model. Before data mining, it is necessary to pre-process the data with different scales and orders of magnitude, and transform the data into comparable data with the same scales and orders of magnitude that can be used for mathematical operations. For the fracturing database with small sample size, on the one hand, it is necessary to effectively eliminate the influence of the scale and order of magnitude, and on the other hand, it is necessary to try to retain the difference information between the variables. Based on the comparison of the characteristics of commonly used data discretization methods, the benefit type of polarization method is finally selected for data discretization.

Table 2 Characteristics of commonly used data dimensionless methods

Commonly used data dimensionless methods	Specificities	Note
Polarization method	Effectively eliminates scale and order of magnitude effects	Extremes in the data need to be treated with caution
Standardized methodology	Currently more used in multivariate synthesis analysis	For cases where the raw data is normally distributed
Standard deviation method	Ensuring comparability of data while retaining information on the degree of variability of variables	A variant of the standardized approach, where variables with a greater degree of variability in their values have a greater impact on the combined analysis
Averaging method	Some retention of information on the degree of difference in the values of the variables	/

3.2. Modeling the Pearson correlation coefficient

The Pearson's coefficient is a value between -1 and 1 that is used to describe the tendency of two linear sets of data to change and move together. The overall correlation coefficient ρ is defined as the ratio of the covariance between two variables X, Y and the product of the standard deviation of both. The overall correlation coefficient ρ is defined as the ratio of the covariance between two variables X, Y and the product of the standard deviation of both, as follows:

$$\rho_{X,Y} = \frac{\operatorname{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{E\left[\left(X - \mu_X\right)\left(Y - \mu_Y\right)\right]}{\sigma_X \sigma_Y}$$

3.3. Establishment of linear regression model

When building a linear regression model, attention needs to be paid to the linear relationship between the variables, the selection and interpretation of the variables, the hypothesis testing of the model, and the fit and predictive accuracy of the model. Also, the use of interaction terms, polynomial terms, and variable transformations can be considered to improve the fit and predictive ability of the model. For multiple features in the sample, establish a multiple linear regression model, mapping the linear relationship between the input feature matrix X and label y by constructing a prediction function, using the constructor to find out the parameter vector of the model, and judging the size of the correlation of each factor by calculating the regression coefficient of each parameter.

$$\hat{y} \sim \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots \ldots + \theta_n x$$

Expression in quantized form

$$\hat{y} \sim h_\theta(X) = \theta^T \cdot X$$

3.4. Machine Learning Model Optimization

3.4.1 Minimum absolute contraction and selection operator

The Least Absolute Shrinkage and Selection Operator (LASSO) is a method for feature selection and regression analysis.LASSO achieves feature selection by restricting the regression coefficients so that most of them become zero. To limit the impact of the multicollinearity problem and prevent model misalignment due to overestimation of the parameter w caused by multicollinearity, in order to limit the degree of regularization, the loss function is used to control the parameters as follows.

$$\min_\omega \frac{1}{2n_{\text{samples}}} \| X\omega - y \|_2^2 + \alpha \| \omega \|_1$$

3.4.2 Multiple nonlinear regression

Multiple nonlinear regression is the process of describing the relationship between multiple independent variables and the dependent variable by fitting a nonlinear equation model. Unlike linear regression, a nonlinear regression model cannot be represented as a simple linear relationship between the dependent and independent variables. Has a high degree of interpretability, fitting parameters to each feature to understand what each feature does for the label. Retains the nature of linear regression as a linear model and is not prone to overfitting.

4. Summary

After screening and organizing the influencing factors of straight well multi-thin-layer fracturing, we can draw the following conclusions:

(1) Fracturing design parameters are one of the most important influencing factors in straight-well multi-thin-layer fracturing, including the selection of parameters such as volume of construction fluid,

type of fracturing fluid, and viscosity of fracturing fluid, which directly affects fracturing effect and production efficiency.

(2) The nature of rock fractures and geological characteristics of the formation are the key influencing factors of multi-thin-layer fracturing in straight wells, which have an important impact on the formation and expansion of fractures, and require detailed geological exploration and analysis of fracture characteristics.

(3) The performance of fracturing fluid and construction technology are not to be neglected in straight well multi-thin-layer fracturing, and the standardization and refinement of the construction process will directly affect the fracturing effect and the smooth progress of downhole operation.

In order to better optimize and maximize the benefits of straight well multi-thin-layer fracturing technology, a database containing the above key influencing factors is established, and through the monitoring, analysis and optimization of these factors, the success rate and productivity of fracturing technology can be improved to provide effective technical support and decision-making basis for oil and gas field development and production.

(4) Data preprocessing is a crucial step in the process of data mining, which can help us clean the data, deal with missing values, deal with outliers, etc., to ensure that the model established has a high-quality data base.

(5) Parameter tuning can help us find the optimal combination of model parameters and improve the prediction accuracy of the model.

References

[1] WANG Cheng. Diagenesis and cementation of the deep tight gasreservoir and their influences on the physical properties in EastDaqing Placanticline [J]. Petroleum Geology & Oifield Development in Daqing,2019,38(4):143-151.

[2] WU Bailie ,ZHOU Jianliang,CAO Yanfeng, et al. Key parameteroptimization of horizontal well multi-stage multi-cluster fracturingin tight gas reservoir [J].$pecial 0il & Gas Reservoirs , 2016,23(4):127-130,157.

[3] ZHANG Dazhi,CHU Lilan,ZHOU Xiang,et al. Diagenesis anddiagenesis facies of tight gas reservoir of Shahezi Formation , inXujiaweizi Fault Depression of North Songliao Basin [J]. Journalof Jilin University(Earth Science Edition),2021,51(1):2234.

[4] HU Suyun, ZHU Rukai, WU Songtao, et al. Profitable explora-tion and development of continental tight oil in China[J]. Petro-leum Exploration and Development, 2018, 45(4):737-748.

[5] U.S.Energy Information Administration. Outlook for shale gas and tight oil development in theU.S.[EB/OL].(2013-05-21)[2021-12-011].https://www.eia.gov/pressroom/presentation.

[6] SUN Longde, FANG Chaoliang, LI Feng, et al. Petroleum explo-ration and development practices of sedimentary basins in Chinaand research progress of sedimentology[J]. Petroleum Explora-tion and Development, 2010,37(4):385-396.

[7] TANG Dahai, TAN Xiucheng, WANG Xiaojuan, et al. Tight gasaccumulation elements and favorable zone evaluation of XuiiaheFormation in Sichuan Basin[J]. Special Oil & Gas Reservoirs,2020,27(3):40-46.