Switching decision of sliding and rotating modes in compound directional drilling based on random forest

Lijuan Fan[1,3,4], Hao Li[5,6], Hongchao Wei[5], Wangnian Li[2,3,4,5], Chengda Lu[2,3,4,⋆], and Min Wu[1,2,3,4]

[1] School of Future Technology, China University of Geosciences, Wuhan 430074, China
[2] School of Automation, China University of Geosciences, Wuhan 430074, China
[3] Hubei Key Laboratory of Advanced Control and Intelligent Automation for Complex Systems, Wuhan 430074, China
[4] Engineering Research Center of Intelligent Technology for Geo-Exploration, Ministry of Education, Wuhan 430074, China
[5] CCTEG Xi'an Research Institute (Group) Co. Ltd., Xi'an 710077, China
[6] China Coal Research Institute, Beijing 10013, China

Abstract. Compound directional drilling consists mainly of a combination of sliding and rotating modes. Incorrect switching between the two modes can lead to large trajectory deviations. Therefore, this study considers the switching of two modes as a classification problem and investigates the use of random forest to predict mode switching. Through linear interpolation and feature importance analysis of drilling parameters, seven decision variables with high relevance to the problem were extracted. The random forest model of switching decision between sliding and rotating modes were established. Compared with other classification models, random forest demonstrates significant superiority in solving the classification problem of the two modes in compound directional drilling.

Keywords: random forest · sliding mode · rotating mode · compound directional drilling.

1 INTRODUCTION

Compound directional drilling technology plays an important role in resource exploration, especially in coal bed methane and coal mining[1]. Its principle is to use a rotating disc or top drive device in conjunction with downhole sliding inclined drilling tools, to use sliding inclined drilling tools for directional sliding mode in the directional inclined section, and to use drill strings and inclined drilling tools to rotate at the bottom of the hole in a stable inclined section. Accurate control of drill bit direction is crucial in directional drilling, as it directly

⋆ Corresponding author: Chengda Lu (luchengda@cug.edu.cn)

affects drilling efficiency, cost, and final coal production. Sliding and rotating are two key operations for controlling the direction of the drill bit during the drilling process. The former uses high-pressure media to drive a screw motor at the bottom of the hole to rotate the drill bit to cut and break rock, effectively reducing torque loss, and achieving low drilling pressure, high torque, and high-speed drilling. The latter transfers the large torque generated by the rotation of the drilling rig through the drill string to the drill bit for rock fragmentation, but there are problems such as large mechanical energy loss and complex force on the drilling tool. By effectively utilizing alternating sliding and rotating modes of operation, overall drilling performance is improved without compromising stability and efficiency.

At present, research on compound directional drilling technology mainly focuses on drilling mechanism and on-site application, force analysis and numerical simulation of screw motor combination, drilling trajectory control technology, and drilling tool optimization[2]. Xu et al. introduced and analyzed the principle and characteristics of compound drilling technology, and analysed the influence of factors such as drilling pressure on the regulation of drilling trajectory in compound directional drilling[3]. Zhong et al. proposed a novel wellbore trajectory design in sliding mode to reduce total drilling time and improve economics[4]. In sliding drilling process, the conventional tool face angle control is very laborious and time-consuming. Wang et al. proposed a model-based optimized control method to facilitate tool face angle setting thereby changing the tool face angle in sliding mode[5]. Chu et al. proposed that the ratio of sliding and rotating modes in drilling in compound directional drilling depends on the stacking rate of the sliding drill and the rate of change in inclination and azimuth of the composite, independent of the target zone parameters and well diameter fluctuation rate[6]. In practice, it is difficult to grasp the timing of mode switching in compound directional drilling, and there are no definite rules for the factors that affect the timing. Therefore, switching decision method based on experience and rules like the results aforementioned are no longer suitable for solving problems the switching problem between two drilling modes.

Thus, this paper investigates a switching decision method based on random forest between sliding and rotating modes. Through in-depth analysis and intelligent processing of drilling parameters, classification models of four algorithms, namely random forest[7], decision tree[8], neural network[9] and support vector machine (SVM)[10], are trained. By comparing the accuracy of the four classification models, random forest was chosen as the main algorithm studied in this paper, which provides a new way to solve the trajectory prediction problem of compound directional drilling.

2 Process Description and Characteristics Analysis

By studying the technology and characteristics of compound directional drilling, it can be understood that the uncertainty of parameters and different operating conditions have a significant impact on the switching decision problem.

2.1 Process Description in Compound Directional Drilling

The key to the technology of compound directional drilling lies in the manual control of the borehole trajectory, with a focus on the timing of transitioning between the sliding and rotating modes during drilling operations. During the sliding mode, manual real-time continuous control of the borehole trajectory bending direction can be achieved by adjusting the tool face angle of the bottom hole motor. During rotating mode, due to the continuous rotation of the bottom hole motor tool face, manual control of the borehole trajectory cannot be realized. However, by analyzing the lateral force of the bottom hole drilling tool under compound drilling conditions, the deviation of the borehole trajectory can be determined. Based on this deviation, the appropriate drilling method can be selected, during drilling, efforts should be made to utilize the bending characteristics of compound directional drilling for borehole trajectory control[11].

2.2 Characteristics Analysis

The drilling parameters generated during the actual directional drilling process are diverse, excellent datasets require clear data and statistical analysis to obtain. Possible sensor failures and operational errors may result in a large number of missing and abnormal values. Reasonable interpolation, filling, and removal should be carried out to ensure the integrity of the data. There may be a certain correlation between drilling parameters, and exploring their regularities can streamline the dataset and reduce the generation time of the model.

The influence of complex geological formations and varying drilling conditions makes it difficult to address switching decisions through rules and mechanisms[12]. Moreover, this approach is incompatible with the trend towards intelligent and automated development. Although fuzzy comprehensive evaluation methods are frequently utilized in switching decision during drilling processes, the outcomes are often subject to the subjective judgments of decision-makers, lacking objectivity. As the number of evaluation factors increases and the fuzziness of relationships complicates, the computational complexity of fuzzy comprehensive evaluation methods escalates, leading to issues of high computational load and time consumption. Additionally, these methods are highly sensitive to fuzzy input information, demanding high-quality input data to avoid inaccuracies caused by errors or uncertainties.

3 Frames for sliding and rotating modes switching

The flowchart of the switching decision between sliding and rotating modes in compound directional drilling is shown in Fig. 1. Due to severe missing values in the original dataset, linear interpolation method is used to fill in the missing values in the original dataset. The importance assessment method is used to select decision variables with high relevance for research, aiming to reduce the dimensionality of the dataset and the complexity of the model. The new dataset is trained using the random forest algorithm to obtain the optimal prediction model. The trained model is saved, and the model is loaded when new decision variables are input. The accuracy function of the random forest is then utilized to predict sliding or rotating mode. It is also crucial to choose the appropriate

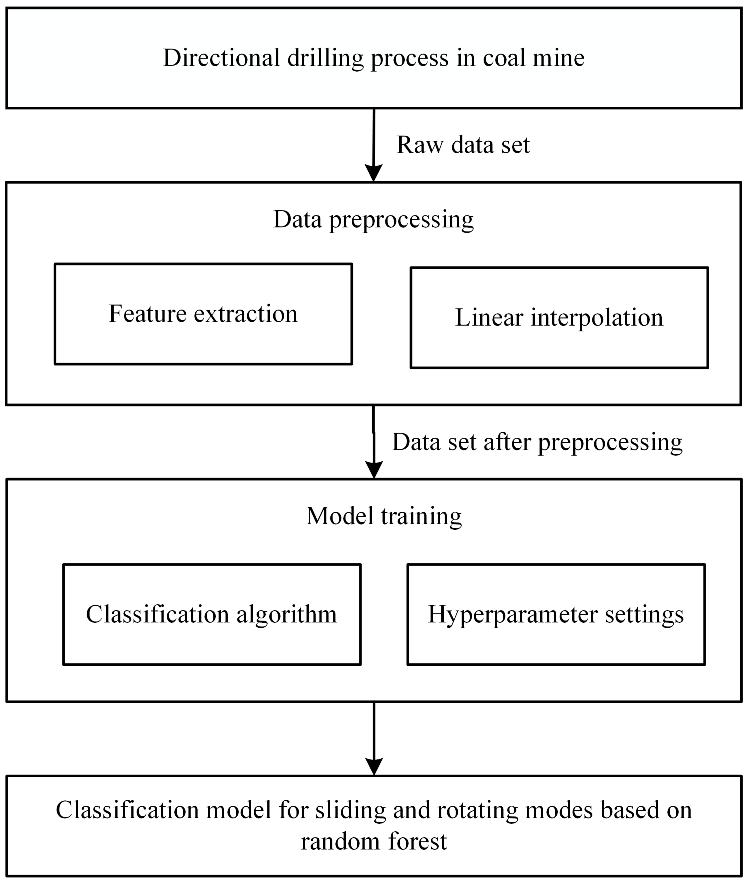

Fig. 1: Flowchart of switching decision between sliding and rotating modes

dataset before model training to assess its impact on the model. By using linear interpolation and importance assessment to preprocess the original dataset, an excellent dataset that is more suitable for research can be obtained. The linear interpolation method will not change the distribution characteristics of the original data, and can maintain the overall distribution shape of the data. This helps to avoid data distortion or the introduction of noise. Doing an importance assessment[13] of the dataset allows for the extraction of feature values that have a large impact on the model's performance, which can simplify the model and improve its generalization. The idea of feature importance assessment in random forest is to calculate how much each feature contributes to each tree in the random forest, take the average value, and finally compare the contribution between the features.

Random forest is an ensemble learning method commonly used for classification and regression tasks. It consists of multiple decision trees, each trained independently and with randomness referenced in the growth process. During the construction of each decision tree, bootstrap sampling is used to randomly select training samples. Secondly, when each node splits the features, only a part of the features selected at random is considered. These two types of randomness ensure the diversity of each decision tree. When making classification predictions, the random forest votes on the predictions of each decision tree, and then selects the category with the most votes as the final prediction result. The detailed process of random forest construction is shown in Fig. 2.

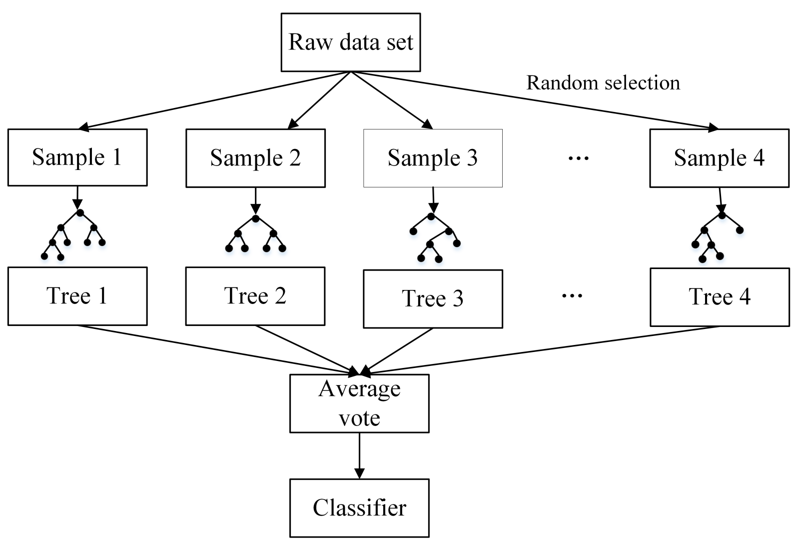

Fig. 2: Generation process of random forest

4 Case study

The data for this experiment comes from 15 sets of variables from 11 boreholes in Sangshuping No.2 well. A dataset composed of variables is divided into two parts: 70% training set and 30% testing set.

4.1 Data Preprocessing

When the original data set collected in the field is observed, the difference between adjacent data is very small. According to the data collected from Sangshuping No. 2 well, the data loss is very serious. In order to ensure the integrity of the data, the linear interpolation method is used to make up the missing values. From Fig. 3, the blue curve is a curve composed of the original data, and the red points are newly inserted values using interpolation, which fills in the white space between the data points to create a smooth curve or surface. This can help reduce noise or oscillations in the chart and make trends in the data clearer.

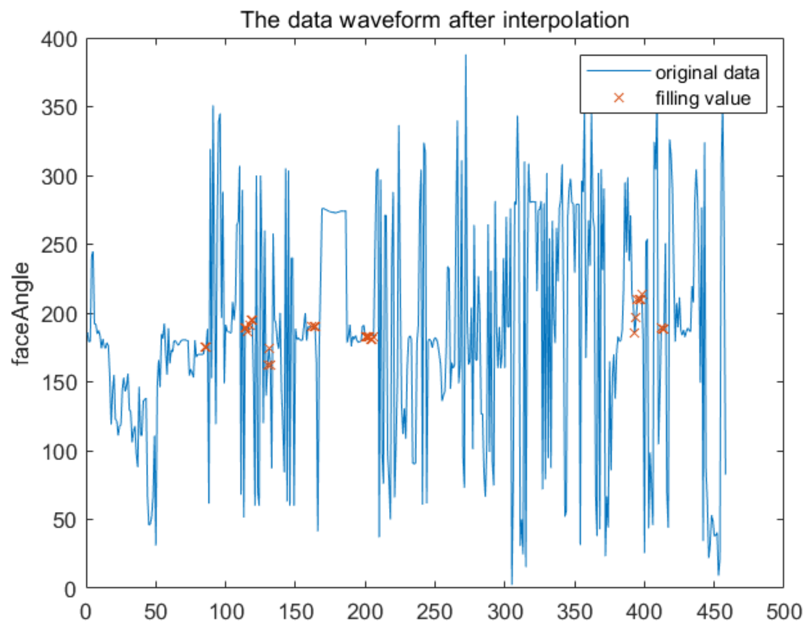

Fig. 3: Tool face angle after linear interpolation

In the feature importance assessment of this study, the out-of-pocket data importance of a feature X, for each decision tree, select the corresponding out of bag data to calculate the out-of-bag data error, and record b1; Noise interference is randomly added to feature X of all samples of data out of bag outside the bag (which can randomly change the value of samples at feature X) and the error of data outside the bag is calculated again, which is denoted as b2. If there are n trees in the forest, then the importance of the feature X(c_i) is calculated by

$$C_i = \sum(b2 - b1)/n. \tag{1}$$

Table 1 shows the importance of features based on random forest. It can be seen from the figure that features such as azimuth Angle, azimuth deviation, up and down displacement deviation and left and right displacement deviation are of high importance. Selecting 7 features that have the greatest impact on decision making to form a small sample dataset and reconstruct the decision model based on random forest. The main purpose of selecting important feature values is to reduce the training time and the complexity of the model by reducing the dimension of features on the premise of maintaining the correctness of the decision.

Table 1: Importance of seven characteristics

	Eigenvalue	Significance
1	Tool face angle	0.5015
2	Azimuth deviation	0.6543
3	Inclination deviation	0.5401
4	Up and down displacement deviation	0.8064
5	Left and right displacement deviation	0.5412
6	Feed pressure (loaded)	0.5589
7	Stratigraphic description	0.6876

4.2 Accuracy of Classification Results

In the classification problems based on random forest, the common evaluation method is confusion matrix, which is mainly used to compare the classification results with the actual measured values, and the accuracy of the classification results can be displayed in a confusion matrix. In the binary classification problem, accuracy in the confusion matrix is chosen as the index to evaluate the model. The formula for calculating accuracy[13] is given by

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN}. \tag{2}$$

It is assumed that the classification results in binary classification problems have positive and negative classes. True Positive (TP) : Indicates the true class. The true class of the sample is a positive class, and the result of model recognition is also a positive class.False Negative (FN) : Indicates a false negative class. The true class of the sample is a positive class, but the model recognizes it as a negative class. False Positive (FP) : Indicates the false positive class. The true class of the sample is a negative class, but the model recognizes it as a positive class. True Negative (TN) : Indicates the true negative class. The true class of the sample is a negative class, and the model recognizes it as such.

Since Random Forest randomly selects training subsets and testing subsets, multiple experiments and cross-validation of the random forest model are usually performed to obtain stable mean values. To increase the accuracy of the experiments, the hyperparameters of the random forest are tuned to obtain the optimal combination of hyperparameters as follows: The number of decision trees is 250, and the depth of the tree is 3. Following from Fig. 4, it can be seen that the predicted values of the sliding and rotating modes based on the random forest model are close to the results of the real classification values in the test set. It can be concluded that the random forest algorithm proposed in this paper has good performance in the switching between sliding and rotating modes. With the confusion matrix in Table 2, using the accuracy formula, the prediction accuracy is 96.30% in 30% of the testing set, in which the classification result of the sliding mode is 97.00% correct, and the classification result of the rotating mode is 95.60% correct.

According to Table 3, we can observe the accuracy of the classifiers obtained after using the original data, the data after filling the missing values, and the data after feature selection as the dataset for model training, respectively. From the results, it is clear that all three steps have improved the accuracy of the model.

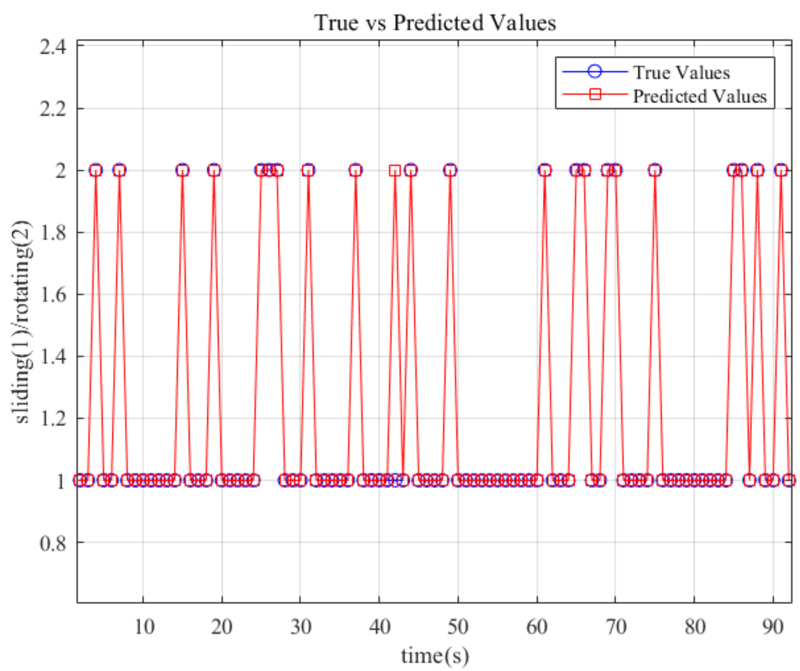

Fig. 4: Performance of the random forest model

4.3 Comparison and Analysis with Machine Learning Methods

Decision tree, SVM, neural network and other algorithms are widely used and representative in the field of machine learning and data mining, and are currently the mainstream algorithms in the field of prediction classification. Com-

Table 2: Confusion matrix diagram of random forest classifier

		Predict value		Total
		Sliding	Rotating	
True value	Sliding	291	9	300
	Rotating	11	239	250
Total		316	234	550

Table 3: Accuracy after data preprocessing

	Accuracy	Sliding	Rotating
Original data	89.01%	90.16%	88.86%
Filling missing values setting	92.76%	96.33%	89.20%
Feature selection	97.00%	95.60%	96.30%

pared with random forest algorithm, the above algorithms have their own advantages and applicable scenarios. By comparing them with random forests, it is possible to assess their strengths and weaknesses in processing complex data, generalization ability, model interpretation, etc., to gain a more comprehensive understanding of the performance and application value of these algorithms in specific tasks.

Four different algorithms were used to form classification prediction models. Table 4 shows the accuracy of the four different algorithms in predicting the switching of sliding and rotating modes. By comparison, we can conclude that the classification prediction results of random forests are the best.

Table 4: Performance comparison of random forest model with other models

Model name	Sliding	Rotating	Total accuracy
Random forest	97.00%	95.60%	96.30%
Neural network	89.81%	91.15%	90.48%
Decision tree	90.75%	87.81%	89.28%
SVM	81.71%	76.54%	79.12%

CONCLUSION

Based on the borehole data of Sangshuping No. 2 well, a classification prediction model of sliding and rotating modes based on random forest is established. Experiments show that the random forest algorithm is superior to other algorithms such as SVM and decision tree in switching decision of sliding and rotating modes. Random forest algorithm not only has a strong feature extraction ability but also can effectively classify and predict data sets, which provides a reliable method for sliding and rotating modes switch decisions. The future research direction will focus on exploring the application of a more effective random forest model in the field of compound directional drilling to further improve the accuracy and robustness of model prediction.

Acknowledgments. This work is supported by the National Natural Science Foundation of China under Grants 62373332 and 62273317, the Natural Science Basic Research Plan in Shaanxi Province of China under Grant 2023-JC-YB-341, the Technology Innovation Foundation Project of CCTEG Xi'an Research Institute (Group) Co. Ltd. under Grant 2022XAYKF01-01, the 111 project under Grant B17040, and the Fundamental Research Funds for the Central Universities, China University of Geosciences.

References

1. C. Ke and X. Song. "Control Design for Directional Down-Hole Drilling Using Dual-Heuristic Programming With a High-Order Dynamic Model," *IEEE Transactions on Control Systems Technology*, vol. 30, no. 3, pp. 1009-1020, 2022.
2. Z. L. Wang, Y. D. Jiang, J. Q. Song and M. N. Jia. "New approach for the digital reconstruction of complex mine faults and its application in mining," *International Journal of Coal Science & Technology*, vol. 9, no. 1, pp. 43, 2022.
3. J. P. Zhang and C. Xu. "Composite directional drilling technology in underground coal mine (in Chinese)," *Coal geology and exploration*, vol. 46, no. 4, pp. 202-206, 2018.
4. X. Y. Zhong, J. P. Liu, K. D. Chen, J. Q. Chen, N. Y. Wang and C. W. He. "A time-optimal wellbore trajectory design for slide drilling systems," *Structural and Multidisciplinary Optimization*, vol. 63, pp. 881-896, 2021.
5. Y. X. Wang, H. J. Ni, R. Shor and R. Wang. "A model-based optimization and control method of slide drilling operations," Journal of Petroleum Science and Engineering," *Journal of Petroleum Science and Engineering* , vol. 198, pp. 108203, 2021.
6. H. Z. Chu, X. Chao, K. L. Hu, Y. Guo, Y. C. Shi and B. Zhang. "Enhancing the benefit of slide steering drilling systems in horizontal wells of unconventional reservoirs," *Frontiers in Energy Research*, vol. 11, pp. 1267228, 2023.
7. D. A. Devyatkin and O. G. Grigoriev. "Random Kernel Forests," *IEEE Access*, vol. 10, pp. 77962-77979, 2022.
8. D. Zhang, M. Wu, C. D. Lu, L. F. Chen, Y. W. Wang and J. Hu. "Tube-Based Adaptive Model Predictive Control for Deviation Correction in Vertical Drilling Process," *IEEE Transactions on Industrial Electronics*, vol. 69, no. 9, pp. 9419-9428, 2022.

9. H. P. Fan, X. Z. Lai, S. Du, W. K. Yu, C. D. Lu and M. Wu. "Distributed Monitoring With Integrated Probability PCA and mRMR for Drilling Processes," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1-13, 2022.
10. Y. Zhou, C. D. Lu, M. L. Zhang, X. Chen, M. L. Wu and W. H. Cao. "A Novel Rate of Penetration Model Based on Support Vector Regression and Modified Bat Algorithm," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 5, pp. 6659-6668, 2023.
11. T. Zhang, J. Long, H. Lin, Z. Liu, Z. Ye and H. Zheng. "A Novel Feature Evaluation Method in Mapping Forest AGB by Fusing Multiple Evaluation Metrics Using PolSAR Data," *IEEE Geoscience and Remote Sensing Letters*, vol. 21, pp. 1-5, 2024.
12. Speiser and Jaime Lynn. "A comparison of random forest variable selection methods for classification prediction modeling," *Expert systems with applications*, vol. 134, pp. 93-101, 2019.
13. W. N. Li, M. Wu, K. Yao, Q. X. Li, C. D. Lu and S. Du. "Design of An Intelligent Control System for Underground Directional Drilling Trajectory in Coal Mines," *Proceedings of the 41st Chinese Control Conference*, Hefei, China, pp. 2586-2590, 2022.