Research on fast identification model of water-flooded layer in old oilfield-- Taking Xingbei area of Daqing Oilfield as an example

Xiaoyu Shang

No.4 Oil Production Plant in Daqing Oilfield Co Ltd., Daqing 163511, China

Abstract: After long-term water flooding development in old oilfields, oil layers are generally flooded. Accurate and rapid recognize the water flooding layer is the key to later infill well layout and development plan adjustment. In this paper, taking Xingbei area of Daqing Oilfield as an example, on the basis of clarifying the characteristics of the water-flooded layer curve, through logging curve optimization, data preprocessing and algorithm model optimization processes, a rapid identification model of water-flooded layer suitable for this block is established. The results show that the HAC, CAL, RLLS and RMG curves with hidden duplicate information can be removed through the correlation screening of logging curves and the importance score of the tree model, which can reduce the amount of data calculation. When the four algorithms are used to identify the flooding level of each layer, the recognition rate of the XGboost algorithm can reach up to 95.45%; the reliability of this result has been confirmed in the model verification process (87.89%), which further shows that the model can be used to identify Xingbei area flooded.

Key words: oil field; water-flooded layer; logging curve; data processing; algorithm model.

1. Introduction

The continuous development of oil and gas resources in old oilfields makes the water cut of oil Wells up to 99%, and the current situation of flooding is serious [1]. Setting infill well is one of the main means to improve the development effect of old oil fields, which can control the water cut of oil Wells and improve the recovery rate at the same time [2-3]. Therefore, it is of great practical significance to quickly identify the distribution characteristics of water-flooded layer for the arrangement of infill well, the adjustment of development plan and the improvement of water flooding efficiency in old oil fields. At present, the water-flooded zone identification of sandstone reservoir mainly relies on logging curve identification technology [4]. Reservoir waterlogging characteristics are closely related to sedimentary microfacies, sedimentary rhythm and other factors of sand bodies. The heterogeneity within and between reservoir layers has corresponding changes in the degree of waterlogging, which is reflected in the logging curve [5]. For example, the amplitude of deep and shallow lateral resistivity decreases, while the amplitude of acoustic curve increases. Therefore, the key to identify the characteristics of flooded layer is to excavate the hidden information of logging curve quickly and accurately. Traditional methods generally identify and classify aquifer levels based on the undulatory characteristics of

logging curves in flooded zones combined with intersection graph [6], deletion graph analysis [7] and other methods, which are time-consuming and have low accuracy. With the development of artificial intelligence technology, data mining has been applied in the identification of flooded layer. Yang Mingren et al. used adaptive enhancement algorithm to identify the water-flooded layer characteristics of Changqing tight sandstone reservoir [8]. Li Jianping et al. used quantum neural network technology to automatically identify the flooded layer and improve the accuracy of the flooded layer identification [9]. Of course, there are also logging technologies such as C/O ratio spectral logging [10] and C/H ratio logging that can directly identify the characteristics of the flooded layer [11]. Although the recognition accuracy is higher, the cost is relatively high. Therefore, based on the data mining technology, this paper proposes an intelligent classification method combining the optimization of the corresponding curve of the water-flooded layer, data standardization processing and XGboost algorithm to realize the rapid identification of the water-flooded layer in the old oilfield in Xingbei area of Daqing.

2. Classification of waterlogged layer in Xingbei area of Daqing

Xingbei area in Daqing is located on the secondary structure of Daqing placantide in the central depression of Songliao Basin, and the main oil and gas reservoirs (Saertu and Tuohua oil reservoirs) are typical fluvial delta facies sedimentary system [12]. After more than 40 years of development, the current comprehensive water cut of the reservoir has exceeded 90%[13]. At present, it is urgent to clarify the distribution characteristics of the water-flooded reservoir and establish a rapid identification method, so as to provide a basis for the setting and adjustment of the infill-well in the later stage. During reservoir flooding, the injected fresh water continuously dilutions the original formation water and increases the formation water saturation, resulting in the continuous decrease of the formation resistivity, density and other logging curves, while the acoustic time difference and spontaneous potential curves continue to increase. According to the degree of flooding and the corresponding characteristics of logging curve, the water-flooded layer in Xingbei area of Daqing has been divided into three levels by using movable water saturation: low water flooding, medium water flooding and high water flooding (Table 1). In this paper, according to the existing waterlogged layer classification standard, the algorithm classification calibration is carried out in the later stage.

Table. 1 The division standard of water-flooded layer in sandstone reservoir in Xingbei area of Daqing Oilfield

| Flooding level	Mobile water saturation（%）	Classification of algorithms
Not flooded	＜12	0
Low water	12~24	1
Water flooded	24~36	2
High water	＞36	3

3. Rapid identification of flooded layer

3.1 Log curve optimization

The Wells in the study area contained 14 logs: Acoustic time difference (HAC), borehole compensated acoustic wave (BHC), hole diameter (CAL), density (DEN), gamma ray (GR), spontaneous potential (SP), 2.5m bottom gradient resistivity (R25), compensated neutron (CNL), eight lateral resistivity (RFOC), deep lateral resistivity (RLLD), shallow lateral resistivity (RLLS)), microgradient (RMG), micropotential (RMN) and flushing zone resistivity (RXO). In order to reduce the repeatability indication information of the logging curve in the flooded layer, the logging curve with hidden similarity information was firstly identified through the correlation screening of the logging curve, and then the similar logging curve was removed through the importance score of the tree model, so the logging curve was optimized and the amount of calculated data was reduced.

（1）Well log correlation screening

Correlation screening is an important data processing method to remove repeated computation and reduce computation time [14]. Correlation screening of well logging curves in the study area can remove one or more curves and improve the calculation efficiency. The correlation coefficient between two curves can be obtained by Pearson function calculation. The higher the correlation coefficient is, the information implied by two curves is about similar. Generally, a correlation coefficient greater than 0.8 is considered to be highly similar [14]. FIG. 1 shows the correlation analysis of 14 known logging curves, among which the strong correlation (R>0.8) includes: HAC and BHC, CAL and DEN, RLLD and RLLS, RMG and RMN
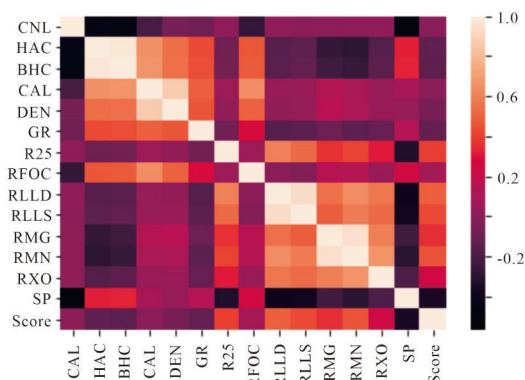
Fig. 1 Correlation screening of 14 logging curves in the study area

（2）Importance screening of logging curve tree model

In order to delete one of the similarity curves, we applied the tree model method to perform the importance score [15] to further optimize the logging curve. Figure 2 shows the scoring results, in which the corresponding degree of the flooded layer is good, and the score is often higher. We deleted the low scores from the well-correlated logs: HAC, CAL, RLLS, and RMG. Ten curves, BHC, DEN, GR, R25, RFOC, CNL, RLLD, RMN RXO and SP, were selected as the data mining research of flooded layer identification.
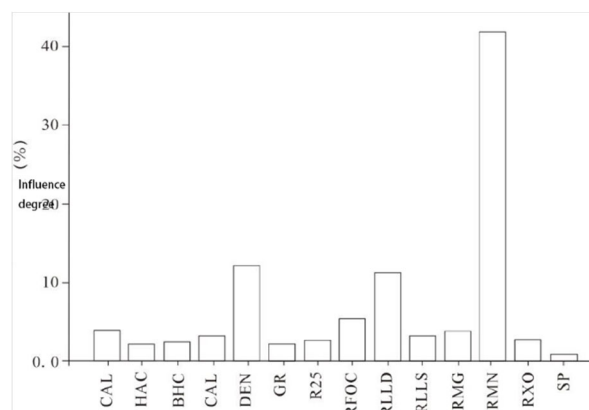
Fig. 2 Importance tanking of decision tree for 14 logging curves

3.2 Log curve preprocessing

Standard deviation normalization is to conduct center translation transformation and dimensionless compression processing on logging curve data, so that each dimension of data has zero mean and unit variance, which can be used to eliminate the influence of large amplitude curve in logging curve [16]. It is divided into three steps :(1) calculate the mean value of the data in each dimension (calculate with all the data); (2) The mean value is subtracted from each dimension; (3) Divide each dimension of the data by the standard deviation of the data on that dimension. As can be seen from the comparison figure before and after treatment (FIG. 3), the change shape of the curve does not change, but the details of the change amplitude are reduced, which reduces the influence of abnormally high or low values.
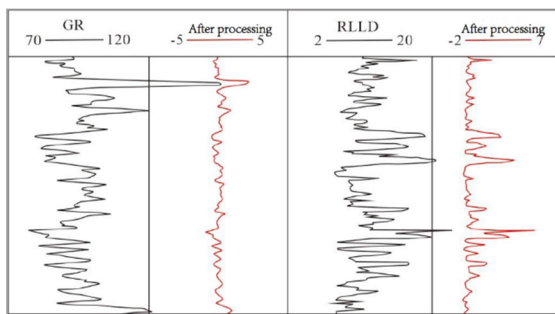

Fig. 3 The comparison diagram of logging curve of X5-1-2008 well before and after preprocessing

3.3 Classification and recognition model construction of flooded layer

A variety of classification and recognition models are applied here to achieve the purpose of algorithm optimization, which are as follows: Support Vector Machines (SVM) [17], eXtreme Gradient Boosting (XGboost) [18], Gradient Boosting Decision Tree (GBDT) [19] and Random Forest (RF) [20] algorithm. Firstly, 10 Wells were randomly selected, including 825 low-flooded Wells, 1156 medium-flooded Wells and 1086 high-flooded Wells, which had completed water-flooded zone identification and curve pretreatment. Then, logging datasets of the same flooding level are pooled and mixed. Finally, "cross validation" was applied in the actual calculation, and the flooded layer data was divided into training set (TRAIN, 90%) and test set (test, 10%). The classification and recognition calculation process was repeated for 10 times, and the average recognition rate was finally used as the evaluation standard of the model. All algorithmic procedures were completed on MATLAB 2010B (The Mathworks Inc., USA). The flow chart of the completed data processing module is shown in Figure 4.
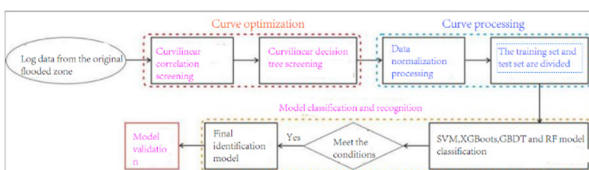

Fig. 4 The flow chart of data processing model

4. Evaluation and validation of identification result of flooded layer

4.1 Evaluation of identification result of flooded layer

We selected 14 logs for each well using correlation screening and tree model importance scoring. It can be seen from FIG. 1 that the correlation coefficient R2 greater than 0.8 is the same type of logging curves, such as HAC and BHC. Despite the differences in their implied information, their role in indicating the flood level is basically the same. In addition, it can be found from Figure 2 that similar curves, such as RLLS and RLLD, are more indicative of the flooded layer than those with shallow detection depth. Therefore, we finally deleted 4 kinds of curves, namely HAC, CAL, RLLS and RMG, and used the remaining 10 kinds of curves for flooded layer identification.

Table 2 is the comparison of the recognition results of SVM, XGboost, GBDT and RF algorithms when the selected flooded layer data of 10 Wells are used for model recognition. Through comparison, it is found that the recognition accuracy of all the classification algorithm models is high, and the recognition accuracy is more than 85%. Among them, XGboost has the highest accuracy, which can reach 95.45%. Therefore, XGboost algorithm is selected to establish a fast identification model when identifying the flooded layer level in Xingbei area based on logging curve. This is because XGboots algorithm is a novel tree learning algorithm that can deal with sparse data. It approximates the learning tree through distributed weighted histogram algorithm and reduces the contingency of sampling results combined with cross-validation [18], so it has good model prediction performance for water-flooded layer identification of such unbalanced sparse data.

Table. 2 The comparison of calculation results of different classification algorithms

Algorithm model	Ordinary precision （%）	Cross validation accuracy （%）	The recall rate （%）
SVM	87.98	86.44	87.98
XGboost	95.45	95.30	95.45
GBDT	90.03	89.37	90.03
RF	95.04	94.80	95.04

4.2 Validation of waterlogged layer identification model

In order to further verify the reliability and universality of the selected model, we randomly selected an untrained well to identify its flood level. In the actual prediction process, we put the data into SVM, XGboost, GBDT and RF algorithm models respectively, and the classification results are shown in Figure 5. The classification results were 87.57% (SVM), 87.89% (XGboots), 87.02% (GBDT) and 86.35% (RF) respectively. The XGboost model still had the highest recognition rate. This result is consistent with the recognition result after the previous

model training. Therefore, we believe that the optimized XGboost model is most suitable for rapid identification of flooded layer level in Xingbei area.
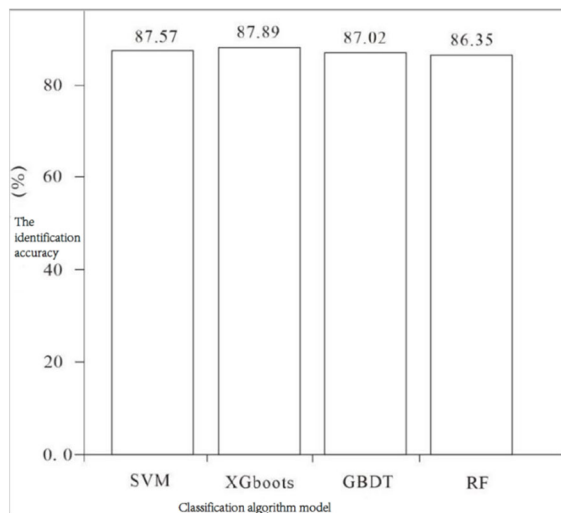
Fig. 5 The comparison of different model results during verification of water-flooded layer

hat the model building process here can be used as a reference for other blocks, but the established model may not be applicable to other blocks. Of course, if there is sufficient data training, the universality of the model will be higher. In addition, learning techniques such as data transfer [21] can also improve the universality of the model.

5. Conclusion

Based on the logging curve, this paper established the waterlogged layer identification model in Xingbei area through algorithm optimization, curve processing and classification model comparison. The conclusion is as follows:

(1) HAC, CAL, RLLS and RMG curves with implicit duplicate information were deleted by logging correlation screening and tree model importance scoring methods, and the remaining 10 curves were normalized by standard deviation to reduce the influence of abnormally high or low values.

(2) When identifying the flooding level, the recognition accuracy of SVM, XGboost, GBDT and RF is higher than 85%, and the XGboost algorithm has the highest recognition rate (95.45%).

(3) When verifying the established algorithm model, the difference between the results of the four algorithm models is small, but the highest accuracy is still the XGboost algorithm model (87.89%), which confirms the reliability of this model for identifying the flooded layer in Xingbei area.

References

1. Liu Yanhong, Luo Xingwang, Zhang Wu, Li Chen, Zhang Ruixue, Luo Qiang, Wang Bin, Zhang Yanmei. Study on logging identification method and influencing factors of waterlogged layer in tight sandstone reservoir. Special hydrocarbon reservoirs.

2. Ma Tao. Analysis of flume delta depositional system in continental large depression basin under arid climate conditions: A case study of Quan-4 Member of Fuxin Uplift belt in southern Songliao Basin [D]. Beijing: China University of Geosciences (Beijing), 2006.

3. Andrew D. Miall. Reconstructing the architecture and sequence stratigraphy of the preserved fluvial record as a tool Reservoir development: A reality check[J].AAPG Bulletin,2006,90(7): 989-1002.

4. Ju Wu, Shen Huilin, Yang Hong, et al. Research on logging evaluation method of water-flooded layer in conglomerate reservoir in junggar basin [J]. Progress in geophysics, 2009,24(3):974-980

5. Tian Zhongyuan, Mu Longxin, Sun Deming, et al. Study on logging characteristics and mechanism of conglomerate water-flooded layer [J]. Acta Petrolei Sinica, 2002, (6) : 50-55+3

6. Chen Ganghua, Liang Shasha, Wang Jun, et al. Application of machine learning AdaBoost.M2 algorithm in conglomerate fluid identification. Oil geophysical prospecting, 2019,54 (6) :1357-1362.

7. Chen Xin, Chen Kegui, Wang Zhaofeng, Dai Xiongjun, Fan Xiangyu. Logging response characteristics and waterlogging analysis of flooded zone in Aryskum oilfield. Progress in geophysics.

8. Mingren Yang, Huilin Shen, Sa Qu, Qipeng Sun, Limin Zhang, Shuming Xiao. Application of AdaBoost algorithm in water-flooded layer identification of tight sandstone. China offshore oil and gas (in Chinese), 2021,33(4):1673-1506

9. Li Jianping, Yang Du, Fan Yougui. Quantum AD hoc network model and its application in oilfield waterlogged layer identification [J]. Microcomputer applications, 2020, 36 (12): 9-11+15.

10. Wang B T, Gao Y F, Fan Y X, et al. Study on logging evaluation method of remaining oil in Lamadian Oilfield [J]. Journal of oil and gas technology, 2014, 36 (7): 84-88+6.

11. He Zhongsheng, Cui Zhigang, Chen Guanghui, et al. Evaluation of residual oil and water flooding degree using hydrocarbon ratio logging technology [J]. Journal of southwest petroleum university (science & technology edition), 2016,38 (5) : 41-49.

12. The white deer. Study on main controlling factors of Fuyang reservoir formation in Xingbei Oilfield, Daqing Placanticline, Northeast Petroleum University. 2015.

13. Zhang Haiying. Development strategy of Xingbei Oilfield in ultra-high water cut stage. Scientific management. 2019.3:256.

14. Sun Peipei, Zhang Ning, Zhao Linjing, Peng Yuanyuan, Liu Xijian, Liu Xiaohui, Zhai Yassen, Feng Meiqing. Effects of subinhibitory concentration of cinnamon essential oil on lipid homeostasis in

Candida albicans [J]. Chinese Journal of Analytical Chemistry,202,50(01):92-107.

15. Cao Taoyun. Research on variable importance based on random forest [J]. Statistics and Decision, 222,38(04):60-63.

16. Li Sihan, Wang Changquan, Wu Hua, Liu Tao, Zhao Xu. Normalization method of capillary pressure curve based on effective reservoir thickness [J]. China science and technology papers, 2020,15(05):593-598.

17. Zhang Junhua, Ren Xiongfeng, Zhao Jie, Tan Mingyou, Yu Zhengjun. Reservoir prediction method and application based on cross-validation support vector machine [J]. Science technology and engineering, 2020,20(13):5052-5057.

18. Yan Xingyu, Gu Hanming, Xiao Yifei, Ren Hao, Ni Jun. Application of XGBoost algorithm in logging interpretation of tight sandstone gas reservoir [J]. Oil geophysical prospecting,2019,54(02):447-455+241.

19. Gu Yufeng, Zhang Daoyong, Bao Zhidong, Feng Zhigang, Li Jinggong. Prediction of permeability by gradient lifting Decision Tree (GBDT) : A case study of tight sandstone reservoir in the west Chang 4+5 member of Jiyuan Oilfield [J]. Progress in geophysics (in Chinese),2021,36(02):585-594.

20. Wang Yan, Wang Ruogu, Wei Keying, Lu Haijiao, Xiong Xiaowei, Zhang Tianjie. Classification of tight reservoir based on random forest: A case study of the 8th Member of Eastern Hehe Formation in Yan 'an Gas Field [J]. Journal of xi 'an shiyou university (natural science edition),2021,36(06):1-8.

21. Yang C, Zhao H, Bruzzone L, Benediktsson JA, Liang Y, Liu B, Lunar Impact Crater Identification and Age Estimation with Chang 'e Data by Deep and Transfer learning. Nat Commun 2020; 11 (1).