

Horizontal Gas Well Liquid Accumulation State Determination based on Ensemble Learning Algorithm

Yu Fu, Qinghui He

Petroleum Engineering School, Southwest Petroleum University, Chengdu, China

Abstract: With the depletion of reservoir energy, liquid accumulation in horizontal wells of the Sulige gas field has become an increasingly severe issue. This study proposes a GBDT algorithm-based predictive model using pure on-site data mining, which overcomes the limitations of traditional models in considering complex wellbore structures and the coupling effects of multiphase flow. Through an analysis of the liquid accumulation state in 25 gas wells in the block, the model achieves an accuracy of 84% in determining the wellbore liquid status, demonstrating higher accuracy compared to traditional models. It provides a more accurate prediction method for addressing gas well liquid accumulation issues.

Keywords: Gas Well Liquid Accumulation; Critical Liquid-carrying Flow Rate; Machine Learning; Drainage Gas Recovery.

1. Introduction

After fracturing production in horizontal gas wells, as production time increases, the well's energy depletes further, resulting in liquid accumulation that leads to reduced production rates, shortened production time, and even shutdowns. It has been found that liquid accumulation in the horizontal section is one of the main factors limiting the maximization of gas well productivity. Accurately calculating the critical liquid-carrying flow rate is crucial for determining the liquid accumulation state in the wellbore. Extensive research has been conducted by scholars worldwide on this topic. Currently, the models used to calculate the critical liquid-carrying flow rate in gas wells include the Turner droplet model, Coleman model, liquid film model for vertical wells; the Belfoid model, Li Yingchuan impact oscillation model, Shekhar model for inclined wells; and the stratified flow model, K-H wave model, and Wang model for horizontal wells.

However, gas well liquid accumulation involves complex coupling problems between two-phase fluid flow in the wellbore and two-phase fluid seepage in the reservoir. The factors considered and the mechanisms involved in current traditional models for predicting the critical liquid-carrying flow rate are relatively limited and fail to comprehensively reflect the mechanism of liquid accumulation as a result of the joint action of two-phase fluids in the formation and wellbore. To overcome the limitations of these models and achieve accurate and efficient determination of the liquid accumulation state in gas wells, this study conducts a comparative analysis of existing liquid-carrying models, summarizes the main factors influencing liquid accumulation in the wellbore, establishes relationship models between different factors and the gas well liquid accumulation state using data-driven approaches, and carries out research on gas well liquid accumulation state prediction based on ensemble learning algorithms. The effectiveness and practicality of the proposed method are validated using horizontal wells in the Sulige block as an example.

2. Horizontal Well Liquid Carrying Models

Due to the influence of the trajectory of horizontal wellbore and complex flow patterns, traditional critical liquid-carrying flow rate prediction models such as Turner and liquid film models are not applicable for liquid accumulation determination in horizontal wells in the Sulige block. Currently, many scholars have made modifications to the liquid carrying models for horizontal wells by considering factors such as liquid production rate, well inclination angle, and varying build-up rates. The critical liquid-carrying models have been widely used in engineering for many years, but they still have theoretical limitations. A comprehensive analysis of the true mechanism requires considering the coupling of unstable pipe flow in the wellbore and unstable seepage in the reservoir. Additionally, the critical liquid-carrying flow rate is also related to the temperature and pressure distribution along the well depth. Therefore, using traditional critical liquid-carrying models to determine the liquid accumulation status in gas wells can result in significant prediction errors.

3. Establishment of Liquid Accumulation State Model based on Ensemble Learning Algorithm

3.1. Introduction to Ensemble Learning Algorithms

Ensemble learning is a machine learning technique that combines multiple base learning algorithms to construct a more powerful model for improved prediction accuracy and robustness. Ensemble learning can be applied to both classification and regression problems and has wide-ranging applications in many fields.

Here are several common ensemble learning algorithms:

(1) Voting: Voting is one of the simplest ensemble learning algorithms. It combines the predictions of multiple base models through voting or averaging to make the final prediction. Voting can be used for both classification and regression problems.

(2) Stacking: Stacking involves building multiple layers of

models for ensemble learning. Firstly, the training data is divided into several subsets, with each subset used to train different base models. Then, the predictions of these base models are used as inputs to train a meta-model for the final prediction.

(3) Boosting: Boosting is an iterative ensemble learning algorithm. It constructs a strong classifier by training a series of weak classifiers (e.g., decision trees). Each weak classifier attempts to correct the errors made by the previous classifier. Common boosting algorithms include AdaBoost, Gradient Boosting Trees, and XGBoost.

(4) Bagging: Bagging constructs multiple base learners by randomly sampling with replacement from the original training set. Each base learner is trained independently, and their predictions are combined through voting or averaging to obtain the final prediction result. Random Forest is a common bagging algorithm.

3.2. Data Acquisition and Preprocessing

By the end of 2023, a total of 147 horizontal wells have

Table 1. Sample of Block Data Standardization

Well No.	Oil Pressure	Casing Pressure	Daily Liquid Production	Daily Gas Production	Wellhead Temperature	Well Inclination Angle	Vertical Depth	Liquid Accumulation Status
1	0.115	0.578	0.104	0.741	0.280	0.171	0.142	1
2	0.276	0.377	0.417	0.586	0.257	0.367	0.574	1
3	0.355	0.687	0.684	0.297	0.177	0.714	0.687	0
4	0.847	0.125	0.247	0.471	0.771	0.378	0.417	1
.....								
125	0.348	0.748	0.639	0.238	0.947	0.787	0.372	1
126	0.413	0.479	0.471	0.412	0.828	0.324	0.747	0
127	0.789	0.537	0.489	0.367	0.147	0.374	0.489	0

3.3. Establishment of Liquid Accumulation Status Prediction Model

Due to the limited number of collected samples, ensemble learning algorithms demonstrate better predictive performance on small-sized datasets. Therefore, the random forest and gradient boosting decision tree (GBDT) algorithms from ensemble learning are employed to learn from the field sample data and establish the model for determining the liquid accumulation status in gas wells.

Table 2. Hyperparameters for Random Forest and GBDT

Model	Hyperparameters
Random Forest	n_estimators=120, bootstrap=true, max_depth=12, min_sample_leaf=1, min_sample_split=2
GBDT	max_depth=24, n_estimators=100, learning_rate=0.1

This study evaluates the predictive models using the accuracy metric (ACC), which ranges from 0 to 1. A value closer to 1 indicates a better fitting effect of the model.

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + FP + TN} \quad (1)$$

Among them, TP represents the true positive cases that are correctly predicted as diabetic; TN represents the true negative cases that are correctly predicted as non-diabetic; FN represents the false negative cases that are wrongly predicted as non-diabetic when they are actually diabetic; FP represents the false positive cases that are wrongly predicted as diabetic when they are actually non-diabetic.

To prevent overfitting, the average accuracy from 5-fold

been put into production in a certain block of the Sulige field, among which 138 wells are low-yield and low-pressure horizontal wells. Dynamic production data and actual liquid accumulation status were collected from these 138 horizontal wells. Parameters such as oil pressure, casing pressure, daily liquid production, daily gas production, wellhead temperature, well inclination angle, and vertical depth were determined as input parameters for the liquid accumulation model, while the output parameter was the liquid accumulation status of the gas wells, with "0" indicating liquid accumulation and "1" indicating no liquid accumulation.

Data samples were processed to handle missing values and outliers, resulting in a final dataset of 127 gas wells for the ensemble learning modeling study. Additionally, there are variations in the magnitude of different parameters, and this variation can amplify the impact of certain parameters on the results. To mitigate this issue, min-max normalization was applied. A portion of the standardized data is shown in Table 1 below:

cross-validation is used as an estimation of the learning algorithm's precision. The results of the random forest and GBDT algorithms for predicting the liquid accumulation status in the Sulige Block gas wells are shown in Figure 1.

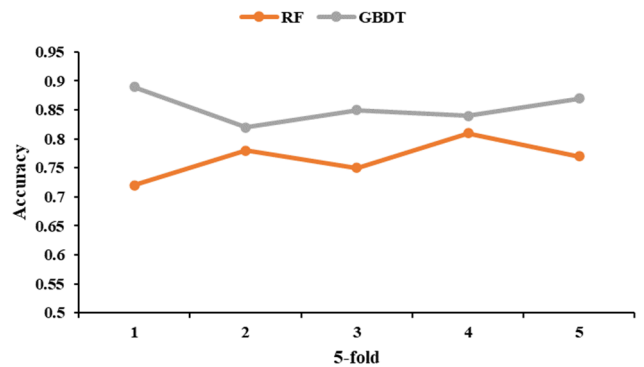


Figure 1. Prediction Results of Fluid Accumulation Model

Based on the established prediction models, the average accuracy of the random forest model is 0.766, while the GBDT algorithm achieves an average accuracy of 0.854. Therefore, the GBDT algorithm is selected to establish the liquid accumulation status prediction model.

4. Model Application

The gas well liquid accumulation status in Block 25 is determined using the liquid accumulation prediction model based on the GBDT algorithm. Additionally, four critical liquid carrying capacity models (Wang-Yuejie model, Wang

model, Li-Li model, and Zhou-Chao model) are used to calculate the overall wellbore critical liquid carrying capacity. The maximum value among all critical liquid carrying capacities is compared with the actual gas production to determine the wellbore's liquid accumulation status. The predicted liquid accumulation results from the ensemble learning-based model and the four critical liquid carrying capacity models are compared, and the prediction accuracy of each model is presented in Table 3.

Table 3. Comparison of Liquid Accumulation Prediction Accuracy among Various Models

Models	Accurate prediction count		Prediction accuracy
	Liquid accumulates well	No liquid accumulation well	
Li Li Model	9	4	52%
Wang Model	8	6	56%
Wang Yuejie Model	11	5	64%
Zhou Chao Model	9	5	56%
GBDT Model	14	7	84%

According to the comparison of liquid accumulation prediction accuracy in Table 3, it can be observed that among the models used to determine liquid accumulation in gas wells based on the critical liquid carryover model, the model proposed by Li Li shows the poorest predictive performance for liquid accumulation in the Sulihe Block's horizontal wells. The model developed by Wang Yuejie exhibits the highest prediction accuracy, albeit only reaching 64%, which falls far short of the required precision. This discrepancy is attributed to the variations in gas well conditions across different blocks, which makes the transfer application of the critical liquid carryover model unsuitable for the Sulihe Block. In contrast, the GBDT ensemble model, established based on actual production data from the Sulihe Block, achieves an accuracy of 84% in predicting gas well liquid accumulation. This accuracy surpasses that achieved by using the critical liquid carryover model, enabling effective guidance for assessing gas well liquid accumulation and selecting drainage and gas recovery processes in the operational areas of the Sulihe

Block.

5. Conclusion

The current horizontal critical liquid carryover models, which are modified based on the Turner droplet model and liquid film model, do not fully reflect the complex wellbore structure and multiphase flow coupling in tight gas horizontal wells. Consequently, they often suffer from low prediction accuracy when specific blocks.

The use of ensemble algorithms for predicting liquid accumulation in wellbores does not require prior theoretical knowledge. Analysis of liquid accumulation examples from 25 wells in the Block demonstrates that the GBDT algorithm establishes the most accurate gas well liquid accumulation prediction model, with a precision of 84%. This approach facilitates the early detection of gas well liquid accumulation and enables timely decision-making regarding drainage and gas recovery processes.

References

- [1] Wang Yuejie, Zhang Hongyou. A new method for predicting the critical liquid carryover of natural gas slanted wells[J]. Oil and Gas Chemical Industry, 2022, 51(04): 69-74.
- [2] Wang Rui, Ma Yun, Dou Liangbin, Cheng Jiarui, Zhang Ningsheng. Research progress on critical liquid carryover and liquid accumulation models for water-producing gas wells[J]. Science Technology and Engineering, 2019, 19(33): 10-20.
- [3] Zhou Chao, He Zuqing, Fu Daoming, Luo Xuan, Liu Huanle, Sun Zhiyang. Full wellbore critical liquid carryover model for shale gas horizontal wells[J]. Oil Drilling and Production Technology, 2021, 43(06): 791-797. DOI: 10.13639/j. odpt. 2021. 06. 016.
- [4] He Yingxu, Guo Chunqiu, Zhang Lixia, Wang Fenghui, Shan Yunpeng, Zhang Ben. A new method for predicting the critical liquid carryover of natural gas slanted wells[J]. Daqing Petroleum Geology and Development, 2021, 40(04): 87-93. DOI: 10.19597/j.issn.1000-3754.202006029.
- [5] Li Li, Zhang Lei, Yang Bo, Yin Yin, Li Dengwei. Method for predicting critical liquid carryover in natural gas slanted wells [J]. Oil and Gas Geology, 2012, 33(04): 650-654.
- [6] Breiman, Leo. Bagging Predictors[J]. Machine Learning, 1996, 24 (2): 123-140.
- [7] Dietterich, T. G. Ensemble methods in machine learning: Multiple Classifier Systems: First International Workshop, MCS 2000 Cagliari.