Rock typing from cored intervals to all wells with method of decision tree

Xinglong WANG¹, Chen LI¹, Peiyuan CHEN²
¹CNOOC Research Institute Ltd., Beijing 100028, China; ²CNOOC International Ltd., Beijing 100028, China
wangxinglong527@qq.com

Abstract. Rock typing is a fundamental work in geology study of oil and gas field. Generally, it is difficult to define a quantitative relationship to expand the classification from cored intervals to all wells with the traditional method based on geologic origin analysis, when rock of complex pore structure is confronted. On the basis of conventional core analysis and well logs, the authors propose a workflow to extend rock typing from cores to all wells with the method of decision tree. In the study case of F oilfield from Middle East, the method and workflow not only predict the rock type on wells, but also get satisfying precision of permeability impetration. Therefore, when data used to quantify pore structure is scare and traditional rock typing method is impossible, the method and workflow founded on decision tree could be a universal alternative.

1 Introduction
Rock classification is a fundamental work in reservoir characterization and even in the entire filed development. Generally, it is easy to obtain a good classification result with the help of geologic origin analysis when dealing with conventional sandstone reservoirs[1,2]. However, in some conditions, this methodology would be not successful any more. On one hand, when the study object is rock of complex pore structure and advanced logging data like NMR and lab analysis data like CT are few, it is impossible to carry out this mission. On the other hand, after rock classification is done on cored intervals, it is uneasy to find a definitive and quantitative expression between type and other features with the knowledge of geology and physical laws. Thus, it is required to explore a method to expand the classification from core wells to uncored wells. In order to deal with such a circumstance, machine learning like decision tree algorithm would have their strengths.

2 Background
The study object is a Cretaceous interval consisted of marine bioclastic limestone in F oilfield, Middle East. Effected by sedimentation and diagenesis, such as compaction, cementation and corrosion, original intergranular pore, interparticle and intraparticle dissolution pore are preserved in the subsurface rock[3,4]. This pore association result in the complex geometry of pore structure. In the aspect of physical property, the permeability of target interval is mainly less than 100mD, characteristic of low permeability. Especially, it shows a much poor mathematical relation between porosity and permeability. Samples of the same porosity have permeability ranging from 0.01mD to 1000mD.
3 Database
Conventional core analysis data are rich, while logging and lab data that can be used to describe the pore structure system is scare in the study case. As to limestone of complex pore structure, a large number of data such as capillary pressure analysis, nuclear magnetic resonance, microscopic thin sections should be collected and utilized. It would be best to have massive enough data to cover rock types of the whole interval. However, the study case is short of data mentioned above, only conventional logging curves and physical analysis on cores are gathered. Specifically, 561 samples of physical property are conducted across several wells, original logs and interpreted logs like GR, AC, SW and PHIE are conducted on each well (Table 1). Dataset composed of these physical analysis and well logging are ready to be explored and mined by decision tree algorithm.

| Lab analysis                          | Logging data                                      |
|---------------------------------------|---------------------------------------------------|
| 561 samples of conventional core analysis | 14 curves including original logging, logging interpretation, and logging arithmetic: GR, RD, RS, DEN, CNL, AC, PHIE, SW, VCL, log(RS), log(RD), log(RD)-log(RS), RD-RS, RD/RS |

4 Methodology
Decision tree is a heuristic algorithm that has tree-like structure and simulates human’s judgment and decision[5–9]. This algorithm is intuitive and easy to comprehend. A decision tree has four elements: root node, terminal nodes, intermediate nodes, and direction lines[9] (Figure 1). Root node is the initial node where the whole tree begins. Root node is unique and is characterized by no input line and only output lines. Terminal nodes is at the end of a tree and it has no output line and only one input line. A tree has many terminal nodes. The other nodes are intermediate nodes and are in the middle part of a tree. An intermediate node has both one input line and no less than one output lines. Direction lines are unidirectional, and are the judgment foundation when higher level node splits into lower level one. There are two keys when a decision tree is designed: (1) judgment criteria of whether to split or terminate, (2) pruning strategy.

Whether to split or terminate is based on the quickest decrement of information purity. Information purity is measurement of uncertainty or complexity of information contained in a dataset[9]. Splitting the dataset with decision tree is essentially a process of decreasing the information purity of the dataset. The faster the drop of purity, the more efficient the classification.

Due to extremely high classification accuracy of the trained dataset, it is likely to get an over-fitting classification result, which leads to the low accuracy of dataset except the trained one. In order to relieve over-fitting, it is necessary to prune the tree.
The most common pruning strategies used are pre-pruning and post-pruning[10]. Pre-pruning is a process carried out when decision of whether to split is made, and it uses the extra validation set to judge if the split improve the performance of the decision tree model. If it fails to improve, the node is stamped as terminal node; or else, the node will be split. Post-pruning process needs no extra dataset. It gets a complete decision tree first, and then from bottom to top checks every node: if replacing the subtree with its terminal node could improve the classification precision, replacing the subtree with its terminal node is acceptable; or else, unacceptable. Finally, a well pruned tree is obtained.

5 Workflow of typing on wells
Based on the theory above, workflow and steps are devised to make an expansion of rock typing from cored intervals to all wells.

Step 1: Dataset extraction and cleaning. After the cored interval is classified, more than ten logging curves are used as source from which features are extracted (Table 1). Because we have no idea which features are the most valuable ones before data mining, the more curves extracted, the much more possible the good results. After extraction, the null and invalid samples are supposed to be dealt with carefully. A simple rule here: if the samples are massive, just remove the samples containing null and invalid features; or else, that is, each sample is extraordinary precious, prudent examine and replacement.

Step 2: Features selection. Moderate and appropriate features should be chosen from the features obtained above. Less features would possible result in under-fitting, while excess features are likely to cause over-fitting[11]. In this case, the dataset is explored and analyzed in IBM SPSS Modeler, and 5 features are selected. According to the importance sequencing, they are interpreted: porosity PHIE, water saturation SW, density DEN, clay content VCLNEW and shallow lateral resistivity RS, successively (Figure 2).
Step 3: Comparison of different types of decision tree algorithm. In this case, compared with CHART and others, C5.0 decision tree is the best tree. C5.0 is the improved version of C4.5 which is among the most classic machine learning algorithms[9].

Step 4: Tree pruning. In IBM SPSS Modeler, pruning is done by adjusting the parameter pruning severity or expected noise. In this study case, a simple model is selected, that is, the pruning severity is set to 0.75. In reality, pruning is a repetitive process in which the effectiveness of interpreted permeability and geological setting should be taken into consideration.

Step 5: Model arrangement. All wells are laid out with the pruned decision tree and get a so-called interpreted rock type curve. By utilizing porosity, rock type curve and permeability-porosity expression, permeability curve is calculated.

6 Application effects

On the theory of flow zone index (FZI)[12], the rock in this case is classified into 7 types beforehand. According to the value of FZI, they are named as RT1–RT7 in descending order. In the term of samples, precision of these types ranges from 0.63 to 0.91(Table 2). The total precision is 0.84, showing a proper fitting that is neither over-fitting nor under-fitting. The type RT1 and RT7 have low precision, which is caused by sample shortage of these types. The difference between logarithm value ΔlogK of predicted and measured permeability is 0.22 which also shows that the precision is low. Also due to the shortage of samples, ΔlogK of RT1 is 2.21, much higher than the overall ΔlogK. The other types, especially RT4, RT5 and RT6, have little error.

In addition to fitting precision of type and error ΔlogK of permeability, another important merit to judge the final classification results: predicted or interpreted permeability are not supposed to hop and leap on the curve (Figure 3).
Table 2 Prediction accuracy of types and statistics of ΔlogK

| Type code | RT1 | RT2 | RT3 | RT4 | RT5 | RT6 | RT7 | Total |
|-----------|-----|-----|-----|-----|-----|-----|-----|-------|
| Sample count | 8   | 34  | 52  | 99  | 203 | 130 | 35  | 561   |
| Prediction accuracy | 0.63 | 0.91 | 0.85 | 0.79 | 0.90 | 0.85 | 0.63 | 0.84  |
| Range of ΔlogK | 0.60~3.30 | -0.45 | 0.04 | -0.60 | -1.03 | -0.88 | -1.17 | -1.17  |
| Average of ΔlogK | 2.21 | 1.67 | -1.00 | 0.37 | 0.03 | -0.29 | -0.24 | 0.22  |

7 Conclusion
By designing reasonable workflow, decision tree is applicable in the process of typing rock of complex pore structure. On the basis of physical property analysis and logging data, the method and workflow employed here could not only expand typing from cored intervals to all wells, but also get satisfactory permeability interpretation. Decision tree could be an alternative and perhaps robust method that can be applied in rock typing.

Reference
[1] Bloch S and Helmold K P 1995 Approaches to predicting reservoir quality in sandstone AAPG Bull. 79 97–115
[2] Meng W, Lv Z, Tang Y and Feng M 2013 Reservoir permeability prediction based on sandstone texture classification J. China Univ. Pet. 37 1–6, 13
[3] Aqrawi A A M, Thehni G A, Sherwani G H and Kareem B M A 1998 Mid-Cretaceous Rudist-Bearing Carbonates of the Mishrif Formation: AN Important Reservoir Sequence in the Mesopotamian Basin, Iraq J. Pet. Geol. 21 57–82

[4] Deng Y, Guo R, Tian Z, Tan W, Yi Y, Xu Z, Xiao C, Cao X and Chen L 2016 Geologic features and genesis of the barriers and intercalations in carbonates: A case study of the Cretaceous Mishrif Formation, West Qurna oil field, Iraq Pet. Explor. Dev. 43 136–44

[5] Breiman L, Friedman J H, Olshen R A and Stone C J 1984 Classification and Regression Trees (Belmont: Wadsworth International Group)

[6] Quinlan J R 1986 Induction of Decision Trees Mach. Learn. 1 81–106

[7] Quinlan J R 1993 C4.5: Programs for Machine Learning Mach. Learn.

[8] Lemon S C, Roy J, Clark M A, Friedmann P D and Rakowski W 2003 Classification and regression tree analysis in public health: Methodological review and comparison with logistic regression Ann. Behav. Med. 26 172–81

[9] Zhou Z 2016 Machine learning (Beijing: Tsinghua University Press)

[10] Quinlan J R 1993 C4.5: programs for machine learning (Morgan Kaufmann Publishers Inc.)

[11] Zhou W and Zhang H 2019 IBM SPSS Modeler 18.0 Data Mining Definitive Guide (Chinese Edition) (Beijing: People’s Posts and Telecommunications Press)

[12] Jude O. Amaefule, Mehmet Altunbay, David G. Kersey and Dare K. Keelan 1993 Enhanced Reservoir Description: Using Core and Log Data to Identify Hydraulic (Flow) Units and Predict Permeability in Uncored Intervals/Wells Annual Technical Conference and Exhibition of the Society of Petroleum Engineers (Houston: SPE) pp 205–20