

Identification of the Quaternary low gas-saturation reservoirs in the Sanhu area of the Qaidam Basin, China

Li Xiongyan^{1,2}, Li Hongqi^{1,2,3*}, Zhou Jinyu⁴, He Xu³, Chen Yihan^{1,2} and Yu Hongyan^{1,2}

¹ State Key Laboratory of Petroleum Resource and Prospecting, China University of Petroleum, Beijing 102249, China

² Key Laboratory of Earth Prospecting and Information Technology, China University of Petroleum, Beijing 102249, China

³ Department of Computer Science and Technology, China University of Petroleum, Beijing 102249, China

⁴ Exploration and Development Institute, Changqing Petroleum Company, Shaanxi 710021, China

© China University of Petroleum (Beijing) and Springer-Verlag Berlin Heidelberg 2011

Abstract: Low gas-saturation reservoirs are gas bearing intervals whose gas saturation is less than 47%. They are common in the Quaternary of the Sanhu area in the Qaidam Basin. Due to the complex genesis mechanisms and special geological characteristics, the logging curves of low gas-saturation reservoirs are characterized by ambiguity and diversity, namely without significant log response characteristics. Therefore, it is particularly difficult to identify the low gas-saturation reservoirs in the study area. In addition, the traditional methods such as using the relations among lithology, electrical property, physical property and gas bearing property, as well as their threshold values, can not effectively identify low gas-saturation reservoirs. To solve this problem, we adopt the decision tree, support vector machine and rough set methods to establish a predictive model of low gas-saturation reservoirs, which is capable of classifying a mass of multi-dimensional and fuzzy data. According to the transparency of learning processes and the understandability of learning results, the predictive model was also revised by absorbing the actual reservoir characteristics. Practical applications indicate that the predictive model is effective in identifying low gas-saturation reservoirs in the study area.

Key words: Sanhu area, Qaidam Basin, low gas-saturation reservoir, decision tree, support vector machine, rough set, predictive model, identification

1 Introduction

A number of typical shallow biogenic gas reservoirs exist in the Quaternary of the Sanhu area in the Qaidam Basin. After several years' research, significant progress has been achieved in understanding the genesis and distribution of the biogenic gas reservoirs, caprocks and reservoir characteristics (Wei et al, 2005; Wang et al, 2007; Zhu and Kang, 2005; Zhu et al, 2006; Guo et al, 2008; Cheng et al, 2008). In recent years, exploration practices have shown that the low gas-saturation reservoirs with gas saturation less than 47% exist widely in the region, accounting for about one third of the total gas reserves. Because of their complicated genesis, low gas-saturation reservoirs do not have significant micro-features, and their log response characteristics show ambiguity and diversity. Consequently identifying low gas-

saturation reservoirs turns out to be difficult. Decision tree (DT), support vector machine (SVM) and rough set (RS) techniques are the logical classification methods which can be utilized to handle the logical structure. Compared with artificial neural networks (ANN), Bayesian networks (BN) and genetic algorithms (GA), the process is more clear, the rule is more comprehensible and the human-computer interaction function is stronger. However, they are seldom used in reservoir evaluation and fluid identification (Sheremetov et al, 2007; Chen et al, 2008). Based on the definition and characteristics of low gas-saturation reservoirs, we used the decision tree, support vector machine and rough set methods to construct a predictive model of low gas-saturation reservoirs, which can well and truly identify low gas-saturation reservoirs. The actual application of decision tree, support vector machine and rough set techniques is significant to the exploration and development of low gas-saturation reservoirs in the Quaternary of the Sanhu area in the Qaidam Basin.

*Corresponding author. email: hq.li@cup.edu.cn

Received October 26, 2009

2 Definition and characteristics of low gas-saturation reservoirs

2.1 Definition

In the study area, because of the presence of bound water and free water, the gas bearing interval with gas saturation less than 47% is named low gas-saturation reservoir. After many years of study in this region, the figure 47% was obtained and it can not be universally applied to other areas. As shown in Fig. 1, when the gas saturation increases to 47%, the gas relative permeability is high and the water relative permeability becomes very low. It indicates that the water is entirely bound water at this moment. When the gas saturation further increases, the water relative permeability is almost unchanged.

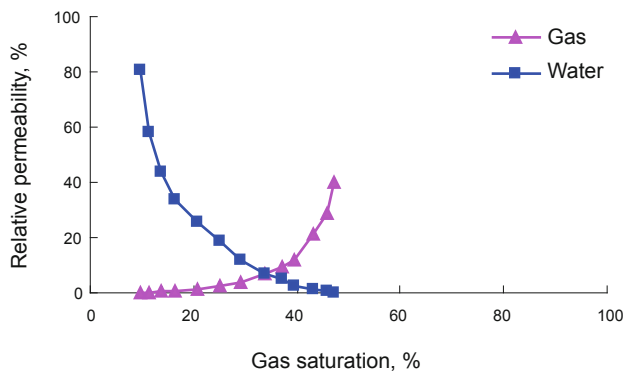


Fig. 1 Relative permeability of gas and water in low gas-saturation reservoir

2.2 Characteristics

In the Qaidam Basin, the cold climate and high salinity environment have slowed the degradation of organic matter. Besides, the dark shale provides a sufficient gas source, and the combination of reservoir and cap rock forms a favorable accumulation condition. Therefore, the “source below and reservoir above” as well as “self-source-reservoir” dynamic accumulation pattern has occurred in the Quaternary of the Sanhu area (Dang et al, 2004; Zhao et al, 2008). The genesis of low gas-saturation reservoirs can be divided into three

types: structure–deposition, diagenesis–compaction and accumulation. Different types of genesis lead to the diversity of low gas-saturation reservoirs, which results in complex micro-features and unclear log response characteristics.

The lithology of low gas-saturation reservoirs is mainly argillaceous siltstone and argillaceous fine siltstone. The porosity is generally from 25% to 40%, and the permeability is primarily from 1 mD to 10 mD, so a low gas-saturation reservoir is high porosity and middle-low permeability reservoir. The pore type is simple, mostly primary pores and a few secondary pores. The primary pores are mainly primary intergranular pores and micropores, while the secondary pores include dissolved pores and fractures. The reservoir thickness is small, and the single layer thickness is basically from 0.5 m to 2 m, accounting for about 70%. The reservoirs whose thickness is more than 5 m account for only about 15%. The content of clay minerals, content of shale, and formation water salinity are high. Carbonate and pyrite exist widely in the study area, as shown in Table 1. The montmorillonite and andreattite clay minerals have additional conductivity. High shale content leads to the development of micropores. Illite will expand when encountering water, and will block micropores. So it will adsorb more bound water, resulting in a higher bound water content in low gas-saturation reservoirs. Because of high formation water salinity, the concentration of conductive ions is high. Formation water with high concentration of conductive ions generates a well-developed conductive network in the pore channels of the rock. When the distribution and content of pyrite meet specific conditions, it will severely influence the apparent resistivity of formation. All the factors mentioned above will significantly decrease the apparent resistivity of the formation. However, due to the low conductivity of carbonate, the apparent resistivity of formation will increase. Additionally, because of the occurrence of thinly interbedded sands and shales and the low resolution of the well logging, the measured resistivity can not adequately reflect the actual formation information. Therefore, the resistivity of low gas-saturation reservoirs has a wide varying range, from 0.27 Ω·m to 1.61 Ω·m. The response characteristics of three porosity curves and the apparent resistivity of formation can not be regarded as a true reflection of gas bearing reservoirs.

Table 1 Micro-features of low gas-saturation reservoirs

Types	Clay mineral content				Shale content, %	Formation water salinity, g/L	Carbonate content, %	Pyrite content, %
	Kaolinite, %	Chlorite, %	Illite, %	Andreattite, %				
Max	13.00	28.00	64.00	46.00	44.48	264.34	37.20	66.90
Min	5.00	10.00	39.00	5.00	21.59	111.66	6.20	0.80
Average	9.39	15.82	52.06	22.67	39.54	157.29	14.44	15.81

3 Method of identifying low gas-saturation reservoirs

3.1 Fundamental concepts

A decision tree is a typical classification method,

which was originally proposed by J. Ross Quinlan in 1986 (Quinlan, 1986). Its algorithm has been improved by later researchers. It adopts a local searching approach to learn and analyze data sets, and uses the loss function with cross-validation as the score function. At last, the simple binary tree

structure is obtained, which will be applied in classification and prediction. The decision tree method is composed of tree construction and tree pruning. A top-down recursive divide-and-conquer method is used to construct a tree. The information gain, gain ratio and Gini index are the evaluation parameters. The algorithm selects an appropriate splitting attribute, and decides the next split node by analyzing other attribute values. The training set is recursively partitioned into smaller subsets as the tree is built. This process will be repeated until the division stops. In the tree pruning including pre-pruning and post-pruning, a statistical measure is applied to cut off the least reliable branches, which results in faster classification, as well as improvement of the classification and prediction capabilities of decision tree (Han and Micheline, 2007; Ian and Eibe, 2006).

The support vector machine method was introduced in 1963 by Vladimir Vapnik (Cortes and Vapnik, 1995). It is a very promising classification method and can replace artificial neural networks. Its algorithm can construct a maximum marginal hyperplane in the training sample with limited data. The maximum marginal hyperplane can separate two types of data sample as much as possible, while maximizing the difference between them. The support vector machine finds the maximum marginal hyperplane through nuclear techniques in high-dimensional space, so the data sample must be projected onto the high-dimensional space. Consequently, it has a strong capacity for distinguishing two types of attributes and a good ability to generalize the high-dimensional sample properties (Han and Micheline, 2007; Ian and Eibe, 2006; Shi, 2008; Cortes and Vapnik, 1995; 6FK QNRSHWQ 3QDW □

The rough set method was first described by the Polish scholar Zdzislaw Pawlak in 1982. It is a useful data analysis method to deal with imprecise or fuzzy data. By analyzing inaccurate, inconsistent and incomplete data, the rough set can discover hidden knowledge and reveal underlying rules. The rough set method is based on the knowledge reduction aims to classify and maintains the strong classification capability. The basic concepts of rough set theory include the basic knowledge block, the indistinguishable relation,

the upper approximation, the lower approximation and the positive region. In principal, attributes are divided into condition attributes and conclusion attributes in the database. Then according to the attribute value, the tuples will be divided into the corresponding subset. At last, the decision rules will be generated based on the relationships between the upper approximation and the lower approximation in the subsets of condition attributes and conclusion attributes (Han and Micheline, 2007; Ian and Eibe, 2006).

3.2 Modeling ideas

With different types of reservoirs as the class labels, we can use abundant information to classify various reservoirs. That is the essence of using logging, core and well testing data to identify low gas-saturation reservoirs. When the algorithm obtains the optimal learning model, the fluid in unknown reservoirs can be predicted. This is a nonlinear, complex and high-dimensional classification problem. The decision tree, support vector machine and rough set methods can all gain the classification rules by analyzing a group of non-order, non-rule and complicated instances, as shown in Fig. 2. Compared with the black box model of artificial neural networks and Bayesian networks, they have relatively transparent processes and easily understandable rules. In addition, they can also clearly show the sensitivity of each attribute.

Therefore, the logging, core and well testing data can be considered as the attributes, which are analyzed and compared by different methods, to decide their sensitivities. Due to the ambiguity of logging information in low gas-saturation reservoirs, the traditional three porosity curves and resistivity property are not necessarily the most sensitive attributes. We use decision tree, support vector machine and rough set methods to process respectively a variety of attribute combinations based on the most sensitive attributes. The rationality and accuracy are used to evaluate the analysis results of different parameter combinations. Besides, the integration of effective rules and reduction of invalid rules are indispensable processes. As a result, we will obtain the optimal attribute combination and corresponding predictive model of low gas-saturation reservoirs.

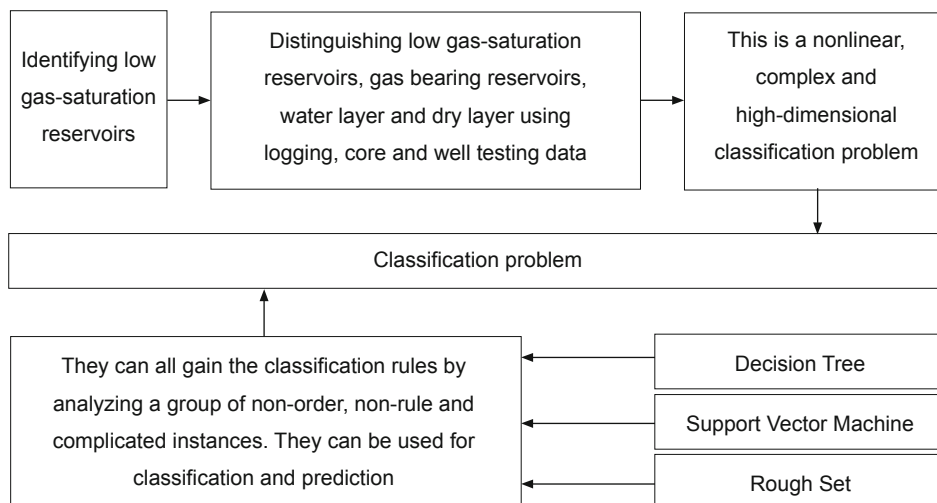


Fig. 2 Feasibility analysis of identifying low gas-saturation reservoirs using DT, SVM and RS

3.3 Modeling process

In the study area, we select 86 key wells and 336 well testing intervals including 146 low gas-saturation reservoirs (LGS), 50 gas bearing reservoirs (GB), 118 water layers (WL) and 22 dry layers (DL). There are a great number of low gas-saturation reservoirs, which have great potential for exploitation. The information of logging and core data in well testing intervals is expressed by 16 parameters, namely spontaneous potential relative reading (ΔSP), natural gamma ray (GR, unit: API), caliper (CAL, unit: cm), acoustic velocity (AC, unit: $\mu s/m$), compensated neutron porosity (CNL, unit: %), densilog (DEN, unit: g/cm^3), deep investigation laterolog relative reading (ΔRD), shallow investigation laterolog relative reading (ΔRS), microlaterolog relative reading ($\Delta RMLL$), deep investigation induction log (ILD, unit: $\Omega \cdot m$), medium investigation induction log (ILM, unit: $\Omega \cdot m$), porosity (POR, unit: %), permeability (PERM, unit: mD), reservoir quality index (RQI, unit: μ/m), displacement pressure (P, unit: MPa), shale content (SH, unit: %). As the mud resistivity has serious impact on the laterolog, we used their relative values. The sensitivity of main parameters to fluid properties is $P > ILD > AC > DEN > CAL > ILM > GR$. The

physical parameters including POR, PERM, RQI and SH have no effect on the fluid recognition. It indicates that there are no significant differences between the micro-structure of low gas-saturation reservoirs and that of other reservoirs. The small displacement pressure of low gas-saturation reservoirs results in an undersize hole. Consequently, CAL becomes one of main parameters in identifying different fluids. With the seven core parameters as well as others, we comprehensively utilized all the logging information to identify low gas-saturation reservoirs, gas bearing reservoirs, water layers and dry layers.

The results of processing various parameter combinations with the decision tree, support vector machine and rough set methods are shown in Table 2. Decision tree and rough set methods are better than the support vector machine in the accuracy of modeling. With regard to the parameter combination, none of the parameter combinations has a distinct advantage. Additionally, in Table 2, the accuracy is the overall accuracy to identify all fluids, while the accuracy of identifying low gas-saturation reservoirs requires further analysis. Because of some false IF-THEN rules produced by the decision tree and rough set methods, the assistance of

Table 2 Processing results of different parameter combinations

No.	Parameter combinations	Accuracy		
		DT	SVM	RS
1	DEN, AC, P, ILD	76.68%	61.66%	60.89%
2	DEN, AC, P, ILD, ILM	82.90%	62.18%	63.89%
3	CAL, DEN, AC, P, ILD, ILM	86.01%	82.05%	74.50%
4	CNL, DEN, POR, PERM, RQI, P, ILD, ILM	84.97%	82.05%	77.51%
5	AC, CNL, DEN, POR, PERM, RQI, P, ILD, ILM	86.53%	71.50%	90.05%
6	GR, AC, CNL, DEN, POR, PERM, RQI, P, ILD, ILM	86.53%	71.50%	88.05%
7	ΔSP , GR, AC, CNL, DEN, POR, PERM, RQI, P, ILD, ILM	88.08%	71.50%	78.06%
8	ΔSP , GR, CAL, AC, CNL, DEN, POR, PERM, RQI, P, ILD, ILM	88.60%	71.50%	75.50%

the support vector machine method and domain knowledge is essential. We further analyze the predictive model of all reservoirs based on parameter combinations from 3 to 8. Eight rules of identifying low gas-saturation reservoirs are shown in Table 3.

Because of the complex geological environment where low gas-saturation reservoirs are formed, the identification rules of different types of low gas-saturation reservoirs are quite different, as shown in Table 3. Therefore, the identification rules of low gas-saturation reservoirs are complicated, involving multiple parameters of logging information. It also verifies the fact that only relying on the traditional threshold values to identify low gas-saturation reservoirs is not enough. The eight rules are used to deal with 86 wells in the study area and the processing results are compared with the well testing conclusions. In the total 146 low gas-saturation reservoirs, 144 can be identified with the accuracy of 98.63%. As each rule is a necessary, but not sufficient, condition of the corresponding reservoir, the rules should be synthesized in practical application, in order to obtain better recognition effect.

Table 3 Identification rules of low gas-saturation reservoirs

No.	Identification rules	Number of eligible reservoirs				Accuracy
		LGS	GB	WL	DL	
1	$2.28 < DEN \leq 2.38, AC > 389.52, 0.04 < P \leq 1.18$	26	0	0	0	100%
2	$CAL < 21.42, 2.06 < DEN \leq 2.28, 0.04 < P \leq 0.22$	19	0	0	0	100%
3	$CAL \leq 21.86, DEN > 2.31, P > 0.22$	12	0	1	0	92.31%
4	$CAL \leq 21.40, AC > 381.08, 0.04 < P \leq 1.18$	29	0	0	0	100%
5	$AC \leq 400.95, 2.28 < DEN \leq 2.37, GR > 80.33, P \leq 0.22$	10	0	1	0	90.91%
6	$AC > 502.77, DEN > 2.38, ILM \leq 1.07$	19	1	2	0	86.36%
7	$DEN \leq 2.28, GR \leq 100.07, 0.54 < ILD \leq 1.24, ILM \leq 1.07$	16	2	0	0	88.89%
8	$CAL \leq 21.79, GR > 101.93, 0.67 < ILM \leq 0.78$	13	0	1	0	92.86%

Due to the complex genesis of low gas-saturation reservoirs, the same parameters have entirely different threshold values in various low gas-saturation reservoirs. The most sensitive parameter to the fluid properties is displacement pressure (P), appearing respectively in 1, 2, 3, 4, and 5 rules. It has significant difference in different rules, as shown in Fig. 3. The displacement pressure of low gas-saturation reservoirs is distributed from 0.01 MPa - 1.68 MPa (two main ranges are 0.04 MPa - 0.22 MPa and 0.22 MPa - 1.18 MPa and their accuracy are 67.21% and 53.16%). In gas bearing reservoirs and water layers, the displacement pressure values have no significant difference. It shows that the displacement pressure value has no threshold value in low gas-saturation reservoirs. Other parameters have similar

situations. In the study area, on account of various genesis of low gas-saturation reservoirs, the traditional threshold values do not exist.

4 Applications

The integrated identification rules were applied to process the actual logging data in the study area on the basis of the low gas-saturation reservoir characteristics. Fig. 4 is the processing result of well A. LSG is the parameter to identify low gas-saturation reservoirs, and WL is the parameter to recognize water layers. When the value of LSG and WL is 1, the corresponding interval is the low gas-saturation reservoir and water layer, respectively. In Fig. 4, LSG shows the first and second layers are low gas-saturation reservoirs and WL indicates the third layer is the water layer. The gas production rate is 30,095 m³ per day.

The processing results of 86 key wells and 6 new wells indicate that the overall accuracy is 88.60% and the accuracy of identifying low gas-saturation reservoirs is 91.70%. The reason is that the predictive model of low gas-saturation reservoirs using the combination of decision tree, support vector machine and rough set methods is better than that of water layer and gas layer. Compared with the traditional interpretation method, the predictive model adds 276 low gas-saturation reservoirs, whose effective thickness is 303.48 m.

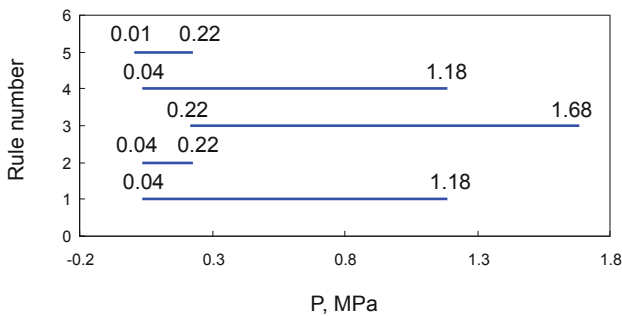


Fig. 3 Threshold values of P in low gas-saturation reservoirs

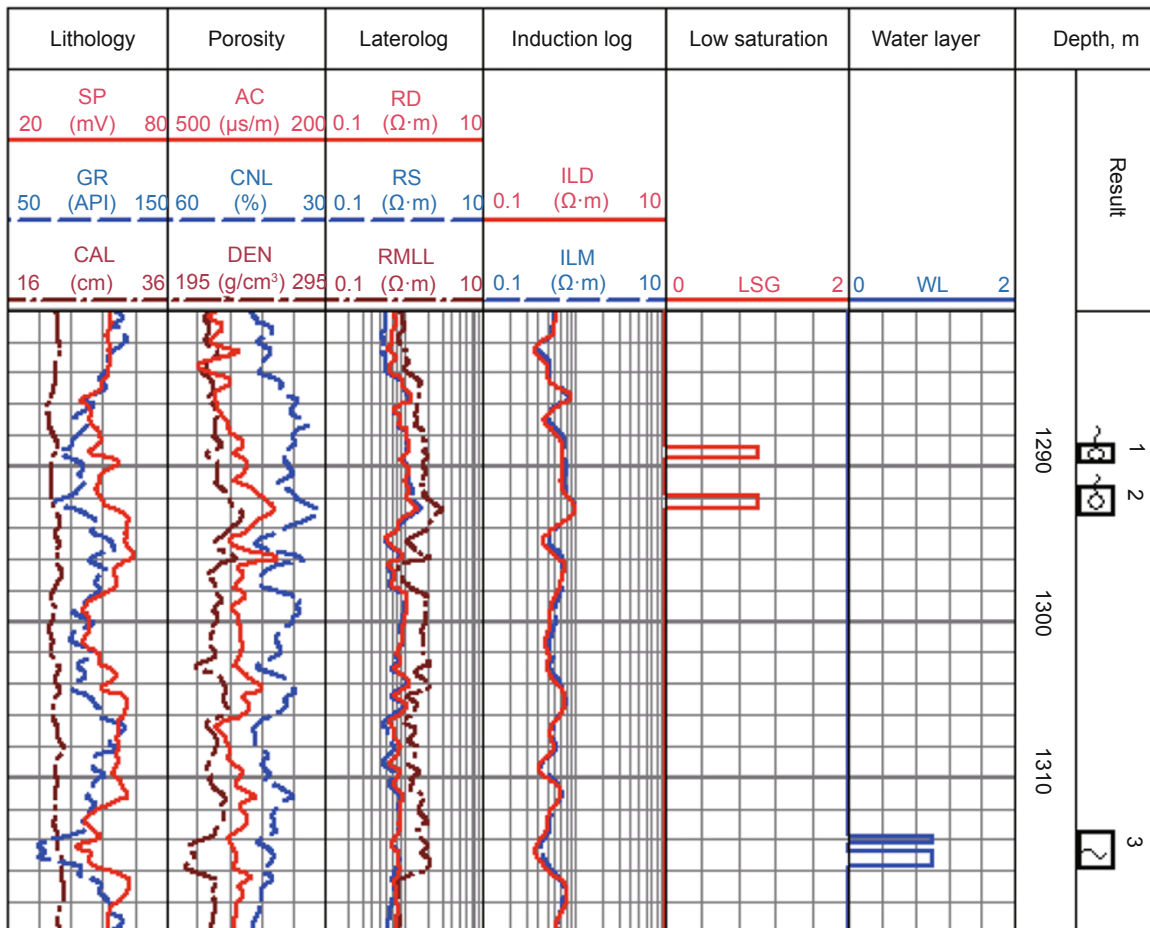


Fig. 4 Processing result of well A

5 Conclusions

1) Although the decision tree, support vector machine and rough set methods can quickly build a predictive model, a variety of parameter combinations can change the identification rules. There are also some false rules. Therefore, exhaustive parameter combinations based on the sensitive parameters are essential. Other methods in combination with domain knowledge can be used to remove the false rules.

2) If the optimal linear combination of multiple attributes could work as the derived parameter, it can not only greatly improve the performance of the predictive model, but also integrate the linear regression of single parameter and multiple parameters in the well logging evaluation. Consequently, choosing a splitting parameter is one of new development directions in the process of building predictive models.

3) There is no threshold value of low gas-saturation reservoirs in the study area. As a result, the traditional reservoir evaluation methods could not be applied to the low gas-saturation reservoirs with complex genesis mechanisms and special geological characteristics. New methods and techniques are effective to evaluate and identify the complicated reservoirs.

Acknowledgements

This work is partially supported by the National High Technology Research and Development Program (863 Program 2009AA062802). The authors would like to appreciate Penny Peng for her kind assistance and support.

References

- Chen D, Hamid S, Dix M, et al. Cooperative optimization-based dimensionality reduction for advanced data mining and visualization. *SPE117765*. 2008
- Cheng F Q, Jin Q, Lin H X, et al. Controlling effect of reservoir-cap combination quality on richment grade of biogas in the Sanhu area of the Qaidam Basin. *Chinese Journal of Geology*. 2008. 43(2): 333-346 (in Chinese)
- Cortes C and Vapnik V. Support vector networks. *Machine Learning*. 1995. 20(3): 273-297
- Dang Y Q, Zhang D W, Xu Z Y, et al. Sedimentary facies and biogenic gas pool of the Quaternary of the Sanhu area in Qaidam Basin. *Journal of Palaeogeography*. 2004. 6(1): 110-118 (in Chinese)
- Guo Z Q, Li B L, Zhang L, et al. Discussion on minimum closure for low-amplitude structural nature gas pool: A case study from the Sanhu area in the Qaidam Basin. *Chinese Journal of Geology*. 2008. 43(1): 34-49 (in Chinese)
- Han J W and Micheline K. *Data Mining Concepts and Techniques (Second Edition)*. Beijing: China Machine Press. 2007. 3 (in Chinese)
- Ian H and Eibe F. *Data Mining Practical Machine Learning Tools and Techniques (Second Edition)*. Beijing: China Machine Press. 2006. 2 (in Chinese)
- Platt J C. Fast training of support vector machines using sequential minimal optimization. *Advances in Kernel Methods: Support Vector Learning*. Cambridge, MA, USA: MIT Press. 1999. 185-208
- Quinlan J R. Induction of decision trees. *Machine Learning*. 1986. 1(1): 81-106
- Schölkopf B, Smola A J, Williamson R C, et al. New support vector algorithms. *Neural Computation*. 2000. 12(5): 1207-1245
- Sheremetov L, Cosultchi A, Batyrshin I, et al. New insights and applications of soft computing on analysis of water production from oil reservoirs. *SPE108702*. 2007
- Shi G R. Application of support vector machine to multi-geological-factor analysis. *Acta Petrolei Sinica*. 2008. 29(2): 195-198 (in Chinese)
- Wang J P, Peng S M, Guan Z Q, et al. Mudstone caprock's sealing mechanism of biogenetic gas reservoir of Quaternary in Qaidam Basin. *Journal of Southwest Petroleum University*. 2007. 29(6): 63-67 (in Chinese)
- Wei G Q, Liu D L, Zhang Y, et al. Formation mechanism, distribution feature and exploration prospect of the Quaternary biogenic gas in Qaidam Basin, NW China. *Petroleum Exploration and Development*. 2005. 32(4): 84-89 (in Chinese)
- Zhao W W, Zha M and Wu K Y. Relationship between unconformity and hydrocarbon accumulation in the eastern Qaidam Basin. *Acta Geologica Sinica*. 2008. 82(2): 247-263 (in Chinese)
- Zhu H Y, Ma L N, Chen J J, et al. Research on characteristics of reservoir bed in Quaternary of Sebei-1 Gas Field. *Natural Gas Industry*. 2006. 26(4): 29-31 (in Chinese)
- Zhu X M and Kang A. Characteristics and evaluation of the Quaternary reservoirs in Qaidam Basin. *Natural Gas Industry*. 2005. 25(3): 29-31 (in Chinese)

(Edited by Hao Jie)