**Paper ID:2011209**

# Quantitative automatic recognition method research for logging facies based on support vector machine

Chen Kegui, Wang Dahai

Southwest Petroleum University, College of earth science and technology, xindu road no.8, Xindu district of Chengdu, 610500, Sichuan, China

***Wang Dahai, E-mail: 775761158@qq.com***

## Abstract

Sedimentary microfacies research is a very important foundation in the exploration of subtle reservoirs in the continental strata.Traditional research methods integrated cores and well logging information for qualitative identification, but coring has limited information and high costs.Therefore, quick and accurate identification of sedimentary microfacies with logging data is important in the process of exploration and development of oil and gas fields. Based on the vector machine theory, this paper established a precise quantitative identification using conventional logging data. A model was established by : first, use traditional method to identify the depositional microfacies with core data of several key wells. Then, study the logs that most reflect deposition micro-phase through curve variations, contact relationship and smoothness to describe the reservoir. Last, use box dimension number and shale volume to establish geological constraint. The support vector machine (SVM ) is then used for automated classification and identification. A study was made for a section of Sulige gas field. The reservoir is composed of tight sandstone with low gas saturation. The section is classified as braided River sedimentary environments with low porosity and permeability, high heterogeneity, and complex pore structure. The actual application of this method in the study of Sulige tight gas reservoir shows promising results, with sedimentary facies identification accuracy of up to 90%, and improvements of the quantitative methods of logging facies identification.

## Introduction

Currently, logging facies research, with the development of logging methods and interpretation, has gradually moving towards the high-precision, automation and intelligent direction.

No classical mathematical functions can clearly describe the Complexity of geological phenomena. Fractal theory was developed,whose the important ideas are self-similarity and fractional dimension.Therefore,it avoids multiplicity and inaccurate characterization by using fractal theory to analyze and study the geometry of a logging curve well logging parameters of logging facies of geological constraints.

Support vector machine (support vector machines,SVM) is machine learning method which solves nonlinear and multidimensional pattern recognition problems in small samples, and has many unique advantages used for the ares of pattern recognition, image processing, and face recognition. This paper uses the principle of vector machine method for logging precise identification and application to sedimentary facies research.

## Theory

The Support vector machine is a data mining method based on statistical learning theory, and it is very successful in dealing with pattern recognition, Classification, discriminant analysis, and many other issues. SVM has been used in defense, economy, agriculture And other disciplines, and can be generalized to prediction And evaluation.SVM has a simple structure, fast learning, and optimal solution with only minimal advantages which can be applied to well log studies.The main idea is: each sedimentary facies can be represented by a set of eigenvectors,certain samples of facies types by nonlinear transformations are mapped to a higher-dimensional space,find an optimal classification face between the microfacies in the high-dimensional feature space,which makes the category both guarantee the classification accuracy and maximize the space on either side of it.Constructing decision functions, and identifing unknown samples of facies types enable logging facies identification. The basic principles are presented in Figure 1

## Work flow

Step1：Build a sample data set

Select the core wells in the region as training data set and got Training sample set { （$\mathbf{x}_i$, $y_i$），i=1,2，…,m},m Is the sample number, and expected outputsyi $y_i \in (+1，-1)$.

Step2：Derived from characteristics of well logging parameters

First is to construct of characteristic parameters of support vector samples Vector $\mathbf{x_s}=(x_1, x_2,，…,x_{12})$, Which $x_s$ as Logging facies characteristics after filtering parameters. This thesis selects in conventional log sequence,the best curves, such as natural gamma (GR) and combined with Sonic curve (AC) to reflects the region's sedimentary microfacies characteristics.Quantitative descriptions of morphological characteristics of logging curves can be derived from the curved shapes, top and bottom contact, the degree of smoothness, etc. shown in Table 1.

1)Well logging curve shape

① Relative gravity (geometric centroid ):

Relative weight can be used to reflect changes in the curve shape, and illustrate the particle sequence

② Mean square amplitude :

Mean square amplitude of log response characteristics, reflect the curve data overall volatility, which can predicate the sedimentary environment and the sedimentation disturbance.

③ amplitude to thickness ratio :

The ratio of the difference between the value of the shale baseline segment logs and the thickness of the curve can reflect the lithology change rate.

④ The Average Median

The average medians reflects the concentrative degree of the main amplitude of the primary curve.

2)The relationship between the top and bottom

① mutation amplitude difference
② Average Slope

3)Smoothness

① The second derivative of the logs envelope

Envelope shape may reflect the strength of the hydrodynamic energy and deposition rate changes in the deposition process.so the second derivative of the logging curve segment maxima sequence fitting curve denote the concave and convex of envelope.the average slope of the curve can further differentiate product before.For the reduction curve, the minimum sequence and the method used to determine the envelope convexity.

② the rate of the curve tooth

The number of logs sawtooth can reflect intermittent deposition,stability of depositional environment. Firstly construct difference sequence,if the adjacent differential value becomes reversal and neighboring absolute difference is not less than a certain magnitude difference,then considered a sawtooth.

③ curve teeth midline

Teeth centerline can use slope K of tooth midline midline and dip $\theta$ to characterize.

4)Geological factors

① box dimension

this paper introduces the fractal dimension as a geological constraints parameter to improve the accuracy of the use of log data to identify sedimentary microfacies.

The fractal dimension of well logs and formation of stratum structure and lithology related , stratigraphic texture and lithology are more complex, the stronger heterogeneity,the larger dimension of logging curve;stratigraphic texture and lithology are more simple,the reservoir is more homogeneous,the logs fractal dimension is smaller. analysis window take the most appropriate length of 1-2 meters.

② shale content

Shale content can reflect the size of the sedimentary energy and distance from the deposition material source.Shale content of different sedimentary facies were significantly different, logging curve shape also will be change.

Step3：Kernel options

Low-dimensional linear indivisible by nonlinear maps to a high-dimensional feature space can achieve linear separable, if the immediate adoption of the technology in high-dimensional classification or regression, nonlinear mapping functions and parameters, in the form of feature space dimension, and so on. By using Kernel function $K(\mathbf{x}_i, \mathbf{x})$, the calculation of support vector machines ingeniously solves the inner product operation in a high-dimensional space, avoids the dimensions of disaster. Kernal function value is the similarity of $\mathbf{x}_i$ and $\mathbf{x}$. The selection of the appropriate kernel function is the key to the success of nonlinear classification. In situation where probability of classification is unknown, this paper uses RBF kernel function (RBF):

$$K(\mathbf{x}_i, \mathbf{x}) = e^{-r\|\mathbf{x}_i-\mathbf{x}\|^2}$$

get the best Gaussian kernel parameter r is 0.1，the penalty parameter c is 10.

Step4：Constructing and seeking optimal classification face

Classification based on Lagrangian relaxation method for solving optimal problem can be expressed as a constrained optimization problem. It is the generalized optimal separating surface of SVM

Step5：Optimal Decision

The best decision used for identification of sedimentary microfacies functions satisfying the constraints is:

$$y = sign\{\sum_{i=1}^{n} a_i * y_i K(\mathbf{x}_i \bullet \mathbf{x}) + b*\}$$

Finally, use one against one method for multi-classification, and divide the data into two groups, to build $C_n^2 = \frac{n(n-1)}{2}$ Vector machine, and finally "vote" for classification

## Application example

For example Well s77-23-37, by using the Logging facies recognition for Under $Box_8$ and Hill[23] segment .the results are presented in Figure 2and Figure 3. Layer bit 3 automatically recognition for River stranded deposition, where expert explained for heart Beach deposition. The layer is located in the braided River alternating deposition stage, with complex rock transition and thin deposition layers. With this method, the thin layers are precisely recognized, which shows the superior application of the method in this research blocks. Other layer automatic recognition results also agree with interpretation by experts. A program is developed using the methods above to automatically identify facies using well logs in the sulige area. The rate of success is around 90%.

## Conclusions

Through this study, comprehensive and accurate quantitative description of log facies characteristic feature is an important basis for the automatic recognition of log facies. Application of this method to get accurate automatic facies identification using conventional logging data, significantly reduce costs for coring.This automatic recognition method has the following advantages and disadvantages:

(1) Using RBF kernel function, SVM Classification techniques of automatic identification of logging facies is highly accurate and practical.

(2)For non-homogenous geology, introduce box-counting dimension and mud content which can reflect the true sedimentary characteristics of tight sandstone reservoir, and enable precise identification of thin layers.

(3)Deficiencies of the method is the depositional environment and thickness of geological layer have to be known in order to apply. Facies recognition of other sedimentary environments still need to be tested, but there are some significant values for the identification of non-homogeneous stratigraphic facies.

## References

1. Pierson SJ.Geological logging data analysis [M].Beijing: Petroleum Industry Press, 1984.

2. CRISTIANINI N, TAYLOR J S. Introduction to Support Vector Machines [M].Li Guozheng, translated.Beijing: Electronic Industry Press, 2004.

3. Xie Heping, Xue Xiuqian.Fractal applications mathematical foundations and methods [M].Beijing: Science Press, 1997. 4. D.P. Wei, and H.A. Spikes "The Lubricity of Diesel Fuels", Wear, 111, 217-235, 1986.

**Table 1：logging phase characteristic parameters of the study area.**

| Sedimentary microfacies | Form | | | | Contact between the top and bottom | | | Smoothness | | | Geological constraints | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Relative to the center of gravity G | The magnitude of the variance S | The ratio of the Width to thickness P | The average median Am | Mutation amplitude difference am (bottom) | Mutation amplitude difference am (top) | Average slope Kp | The second derivative of the envelope | Tooth rate R | Tooth midline K | Box dimension D | Shale content Vsh |
| High energy heart beach | 0.519 | 67.32 | 32 | 80.56 | 160 | 140 | 0.0041 | -0.6 | 0.091 | ≈ 0 | 1.092 | 3. |
| Low energy heart beach | 0.489 | 16.66 | 5 | 56.96 | 44 | 77 | -0.0415 | 0.48 | 0.093 | 0 | 1.117 | 20.16 |
| Channel fill | 0.49 | 49.42 | 13 | 81.68 | 100 | 86 | -0.0053 | -2.02 | 0.1 | ≈ 0 | 1.14 | 25.1 |
| Channel lag deposit | 0.475 | 27.29 | 25 | 58.9 | 140 | 100 | -0.012 | 5.84 | 0.08 | 0 | 1.078 | 11.06 |
| Floodplain | 0.536 | 23.08 | 6 | 136.76 | 120 | 20 | 0.0707 | -0.46 | 0.141 | 0 | 1.12 | 31.83 |
| Flooding swamps | 0.491 | 23.28 | 11 | 138.55 | 100 | 50 | -0.0314 | -2.86 | 0.2 | 0 | 1.115 | 69.2 |

**Figure 1 SVM principle of classification**

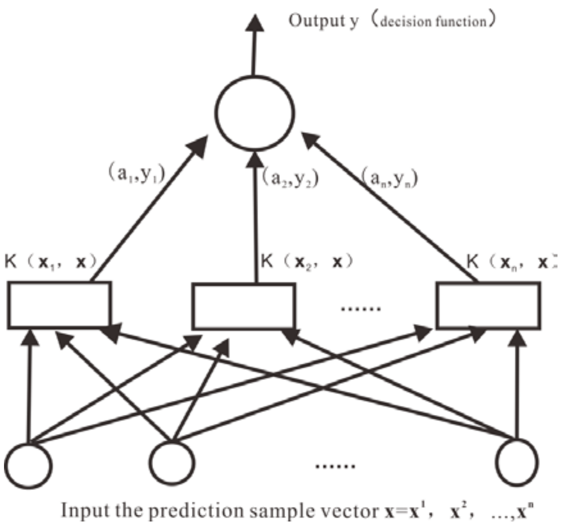

**Figure 2 s77-23-37 wells Under Box$_8$ section logging phase diagram**

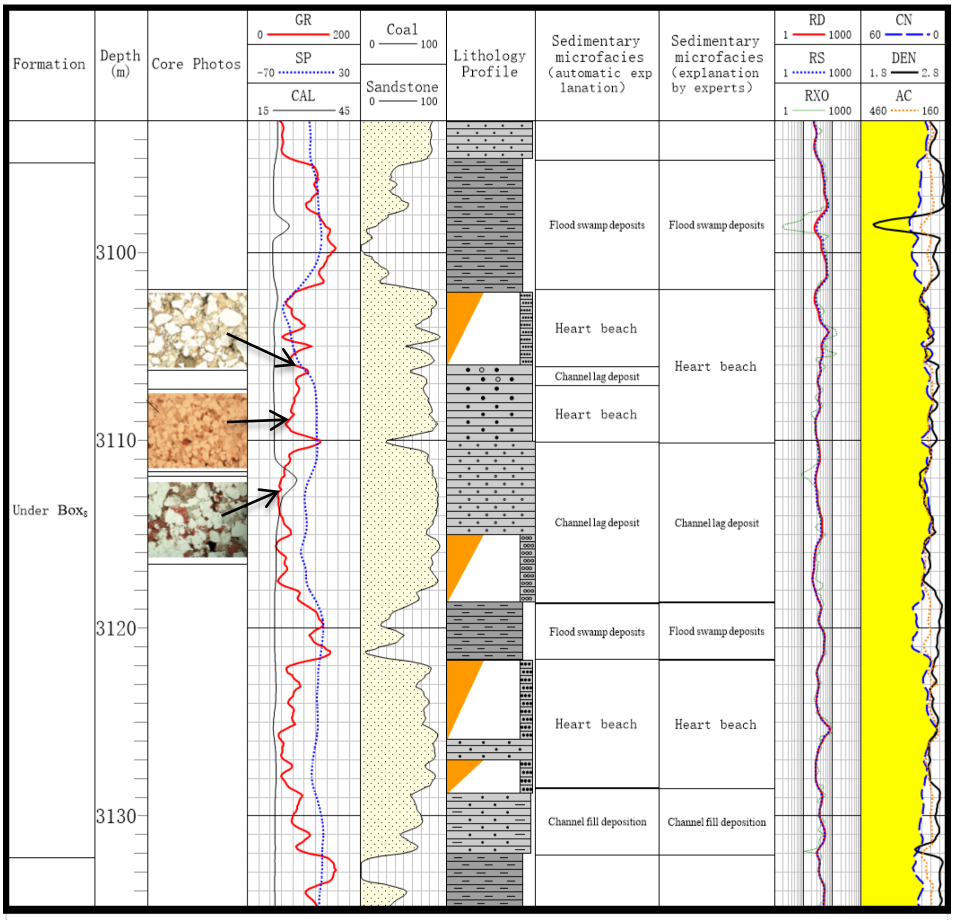


**Figure 3 logging phase diagram s77-23-37 Wells Hill$^{23}$ paragraph**