Application of a Mixed Kernel in the Oil Water-flooded Layer Identification

Fuhua Shang[1], and Xue Zhang[1] ,and Tiejun Zhao[2]

1. Daqing Petroleum Institute, Computer&Information Technology College, Daqing, China
2. Harbin Institute of Technology, Computer Science and Technology College, Harbin, China

Abstract:
A B-spline kernel combined with RBF is developed, a mixed kernel is obtained. By analyzing the structure of the logging signal characteristics, the method is used to automatically identify the water-flooded status of oil-saturated stratum. The experimental results show that the mixed kernel has high recognition accuracy with the advantages of the short running time.
Key words:
Mixed kernel, Recognition with Water-flooded, B-spline kernel..

1. Introduction

Water-flooded layer identification of oil-saturated stratum is a difficult task during the middle-late period of the development[1]. To solve such problems, in addition to developing logging evaluation methods for water-drive reservoir, it's also meaningful to find new methods of water-flooded layer identification. Water-flooded layer identification is mostly based on log-curves which reflect physical and chemical properties of the stratum[2]. Each of the wells has information of logging curves. It is a difficult problem of geological analysis in oil field how to carry out automatically identifying by using the information.

During the oil field geological analysis, we can gain a group of oil-saturated stratum information, and the information of each oil field is a group of log-curve signals. Log-curve is a group of signals gained by measuring underground features (such as: underground resistivity). That is a group of curves of functions. The injected water will lead to changes of formation from water salinity(electricity of water-flooded, natural radioactivity) in the reservoir when the oilfield is water-flooded. At this point, the higher level of the water is, the lower resistivity will be. The amplitude and its decrease of Microelectrode at the level of water-flooded become comparatively smaller, and the value natural gamma in logging curve becomes higher. Those features are evidences which evaluate the water-flooded trustily[3]. Our task is using these signal functions to gain the properties of the underground oil field.

In Ref.[4], a learning algorithm of process neural networks was proposed on the basis of the expanded function orthogonal, which solves the problem of continuous time function feature. And it leads to the learning process formula of neural networks by means of the expanded orthogonal function. It has been successfully applied to the mechanical breakdown diagnosis and the oil exploitation simulation.

The transitional classification is based on the classification of vector set[5][6]. When SVM is used, the kernel plays a direct role. With dynamic programming technology, the kernel of inner products between string images that is used to calculate the special high-dimensional space has been developed. The method of kernel is used to classify structure of the object, the category can be (not very) general but a very efficient way to achieve kernel string.

On the basis of identifying signals of logging curves, this paper is mainly to interpret a group of logging signals, that have been turned into multiple feature vectors. Based on the idea of data structure, by analysing logging signal structure, it is proposed the mixed kernel that is based on B-spline kernel.

In Ref.[1], a B-spline-Based SVM Model was proposed, and applied to identification of water-flooded layer and enhanced the generalization ability of the model. But in practical application, the signal data of water-flooded must be transformed first and then identified in the model referred In Ref.[1]and[4], as a result, running efficiency will be influenced. The experimental results show that the mixed kernel has high recognition accuracy with the advantages of the short running time compared with other classification methods.

2. Kernel

Early kernel theory was that it used Mercer theorem to explain the kernel into the plot of Hilbert space. In 1964, Aizerman, Braceman, Rozonener[7] and others introduced

Manuscript received January 5, 2008
Manuscript revised January 20, 2008

this idea in the field of machine learning, but it wasn't used until Boser, Guyon and Vapnik put it in the application in a literature of SVM. Kernel can be used in all kinds of real-world input spaces, and people are constantly trying to use kernel for different purposes. As for the kernel for tree [8], the text[9] and so on.

2.1 Kernel

In pattern recognition, a classifier is designed according to the features extracted from training samples. Feature extraction is equivalent to the use of appropriate transformation, that is, mapping samples from the input space into the feature space $x \in R^l \xrightarrow{\phi} x^\phi \in F$, $x^\phi = \phi(x) = [\phi_1(x), \phi_2(x), ..., \phi_L(x)]^T$, the mapping $\phi : \mathrm{R}^l \mapsto F$ could be linear or nonlinear. The set of training samples is

$$X^\phi = \{(x_1^\phi, y_1), (x_2^\phi, y_2), ..., (x_N^\phi, y_N)\}$$

in the feature space. The dimension of feature space-L couldn't be too large, or it would lead to disaster of dimensions; and when the dimension is too large, the time spent in calculating $\phi(x)$ and the space storing x^ϕ will increase.

By introducing kernel function, the problem can be simplified. Kernel function substitutes inner product in the original space, which is equivalent to that through a mapping, where data is mapped to a high dimensional feature space defined by one kernel function. Selecting a kernel function is defining a high-dimensional feature space. However, in practical application, if a kernel function is chosen at will for a particular classification, poor performance of the kernel function model may be caused and correct classification[10] may not be carried out.

2.2 B-spline kernel

B-spline kernel is frequently used, and there is limited nodes used in this paper[11].

N-p-(p non-negative integer) B-spline kernel function that has limited nodes is defined on $R_n \times R_n$, which is obtained from one-dimensional p-spline function expansion. Assuming the set of nodes in the given one-dimensional space is $N = \{t_1, t_2, ..., t_m \subset R\}$, and the corresponding one-dimensional p-spline kernel function is as follows:

$$K(x, x'; t_1, t_2 \cdots, t_m) = \sum_{i=1}^{m} (x - t_i)_+^p (x' - t_i)_+^P , \quad \forall x, x' \in R \qquad (1)$$

Among them, $x_+^p = \begin{cases} x^p, x > 0 \\ 0, x \le 0 \end{cases}$, x may be $(x - t_i)$ or $(x' - t_i)$ and not a specific one in formula (1).

N-dimensional spline kernel function is defined by one-dimensional spline kernel K_I. Assuming the set of nodes in the given n-dimensional space is $N = \{t_1, t_2, .., t_m\} \subset R^n$, where $t_1 = \{t_{11}, ..., t_{1n}\}^T$, ..., $t_m = \{t_{m1}, ..., t_{mi}\}^T$, $x = ([x]_1, ..., [x]_n)^T$, $x = ([x']_1, ..., [x']_n)^T$, so the n-p- spline kernel function is defined as follows:

$$K(x, x') = K(x, x'; t_1, ..., t_m) = \prod_{i=1}^{n} K_1([x]_i, [x']_i; t_{1i}, ..., t_{mi}) \qquad (2)$$

3-band B-spline kernel function that has limited nodes is used in the experiment, and the dimension of spline kernel is determined according to the experimental data.

2.3 Mixed kernel

The choice of the kernel and parameters has greater impact on the classification results. Smits, and others[11]proposed the mixed kernel. The kernel combines with a number of different kernel functions and has better characteristics, which is the basic principles of mixed kernel

Assuming that K_1, K_2 are kernel functions based on $X \times X$, $X \subseteq R^n$, $a \in R^+$, $f(\cdot)$ is a real function based on X, $\phi : X \to R^N$, K_3 is a kernel function based in $R^N \times R^N$, then the following function could be obtained:

$$K(x, x') = K_1(x, x') + K_2(x, x') , \qquad (3)$$

$$K(x, x') = K_1(x, x')K_2(x, x') , \qquad (4)$$

$$K(x, x') = K_1(\phi(x))K_2(\phi(x')) , \qquad (5)$$

$$K(x, x') = aK_1(x, x') , \qquad (6)$$

$$K(x, x') = f(x)f(x') \qquad (7)$$

is kernel. It is proved that the mixed kernel satisfied the Mercer theorem thoroughly.

2.4 Construction of mixed kernel based on the B-spline

In the experiment of this paper, it is found that the effect of classification of a single kernel is poor. Generally, it will be better if we use the mixed-kernel[13]. The mixed-kernel based on B-spline and RBF constructed through the construction principle of mixed-kernel has achieved good results.

The expression of the mixed kernel used in this paper is:

$$K(x,x') = \lambda\prod_{i=1}^{n} K_1([x]_i,[x']_i;t_{1i},...,t_{mi}) + (1-\lambda)\exp(-\|x-x'\|^2/\sigma^2),\quad \lambda\in(0,1).$$

Where $\prod_{i=1}^{n} K_1([x]_i,[x']_i;t_{1i},...,t_{mi})$ is the B-spline kerkel, $\exp(-\|x-x'\|^2/\sigma^2)$ is a Gauss RBF kernel; is a parameter to regulate the impact of two former kernel on the mixed kernel. When λ =0, mixed kernel degenerates to Radial Basis kernel function; when λ equals to1, the mixed kernel function becomes B-spline function. λ is regulated to make mixed kernel function adopt to different data distribution, which is equivalent to the integration of prior knowledge of specific issues when making choice of kernel function.

Based on the hybrid Kernel Support Vector Machine, the decision-making function is

$$f(x) = \mathrm{sgn}(\sum_{sv} a_i y_i K(x,x_i) + b)$$

Where x_i is the support vector; x is the signal input vector; y_i is the output vector corresponding to supporting vector; $K(x,x')$ is the kernel; b is a constant; $W_i = y_i$, W_i is the right value; SV is the support vector.

In practical application, the sample needs for B-spline transform, so that it can be put into the learning-Machinery. B-spline kernel based on the Hilbert space which emerged after the B-spline function transform. Therefore, it was not needed to do the B-spline transform for the sample in the application of the B-spline kernel and the mixed kernel.

3. Analysis of the Experimental Results

3.1 Selection of log signal and training sample

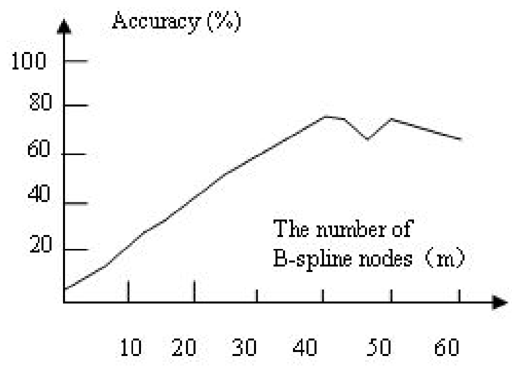
During oil field geological analysis, the method of making cored well is used to know about water-flooded layer condition, geological analysis according to the character of subterranean rock-core from cored well. But it is very expensive to drill a cored well, so the cored well data are limited. Our task is to reflect geometrical peculiarity of actual curve by the received data from limited log-curve information, and then derive the correct rules and make use of it for the future analysis of geology.

From the limited data of oil-saturated stratum, 110 samples for training are chosen, Thereinto, the degree of water-flooding is divided into, and non(0)15, weak(1)40, middle(2)15, strong(3)40water-flooded layer, the number in bracket is the value of output. The remaining samples are used for testing.

Five representative parameters are chosen (thick, deep-lateral resistivity, spontaneousness potentio, interval transit time, puny potentio, lithology of stratum) as input by expert experience and test analysis. All features are log signals except the lithology of stratum. The lithology of stratum is divided into mudstone (0), microclastic rock (1), siltstone (2), politic siltstone (3), silty mudstone (4) and pedocal(5), the number in bracket is the value of the output. The five parameters, in addition to lithology is a value, other features are time function.

3.2 Selection of the number of B-spline nodes in kernel

First, it is decided the number of B-spline nodes (the variable m in formula 1). The result showed in Fig.1 indicates a balanceable accuracy can be gained when the number of B-spline nodes increases to a certain number. The number of spline nodes is 40 in the experiment.



Figuer1 Rule of B-spline tansform node number

3.3 Analysis of the result

Training the samples by SVM training machine which has studied well, the accuracy is 90.5%, and the accuracy is 78.1% when training the measuring data. It is a good result in Water-flooded layer identification when used mixed kernel in Table 2.

Among them, the time is the all which is from input to output, but B-spline SVM entire system running time includes the uptime for B-spline transform and the uptime for the support vector machine. The time for the B-spline transform is slower by nearly an order of magnitude than the inner Plot computing of the kernel.

Under the same condition, the mixed kernel system has the shortest running time compared with different sample identification algorithms, such as neural networks, B-SVM, wavelet-SVM and so on. The result is showed in Table 3.

Under the condition of the same sample, we identify the sample by using RBF and B-spline kernel respectively. Table 4 compares the identification time and the accuracy of measuring data and we can find that the mixed kernel's identification capability is more powerful.

4. Conclusion

Based on the analysis above, in this paper, we construct a mixed kernel function of B-spline kernel and RBF kernel. And we apply it to the identification of water-flooded layer. From the experimental results, the mixed kernel function's capability of classification is stronger than a single one. It can also greatly improve the efficiency of the operation of the whole system, and has some reference value in the problems of the research signal identification, time-varying pattern recognition system, system identification and simulation modeling.

Table 1: Condition of support vector

Parameter setting	Number of support vector	Nnumber of support vector of each kind			
		Strong	Middle	Weak	No
d＝3.0, g=0.25	343	46	110	41	146

Table 2 : Condition of training speed and accuracy

Parameter setting	Time（second）	The accuracy of training data (%)	The accuracy of measuring data (%)
d＝3.0, g=0.25	40	90.5	77.9

Table 3: System time of B-SVM、wavelet-SVM and Process neural networks

algorithm	Time（second）	The accuracy of measuring data (%)
Process neural networks	3360	73.4
B-SVM	1360	78.1
Wavelet-SVM	1150	77.8
Mixed Kernel	70	77.9

Table 4: Comparison of B-spline kernel, RBF kernel and Mixed kernel

Kernel	Time（second）	The accuracy of measuring data (%)
RBF	89	72.8
B-spline kernel	67	74.8
Mixed kernel	70	77.9

References

[1] Fuhua Shang, Xiujie Miao, Zhao-yu Wang. Automatic identifying algorithm of water-flooded zone based on B-SVM[C]. ICMIC.2006:4035-4039.

[2] Jinglong Lin, Qingguo Zhang, Yanjie Song,Shaohua Xu. Water-flooded formation well-logging analysis, Journal of DaQing Petroleum Institute, 25(2001)3, P.20-23

[3] Zhen Li, Miao Luo, Xi Yang, Zhenzhou Lin. Review of logging interpretation methods for water-flooded layer, Chinese Journal of Engineering Geophysics, 4(2001)3, P.288-294

[4] Shaohua Xu,Xingui He. Learning algorithms of process neural networks based on orthogonal function basis expansion. Chinese Jounal of Computers, 27(2004)5, P.645-650

[5] Vapnik,V. The Nature of Statistical Learning Theory.New York: Springer-Verlag,1995

[6] Tao,Q, Wang,Jiaqi, Wang Jue. Madaline and kernel technique in support vector machines. Technique Reports in Institute of Automation, Chinese Academy of Sciences, 2001

[7] M. Aizerman, E. Braverman, L. Rozonoer. Theoretical foundations of the potential function method in pattern recognition learning. Automation and Remote Control. 1964, 25:821~837

[8] Eom, Jae-Hong, Kim, Sun, Kim, Seong-Hwan, Zhang, Byoung-Tak. A tree kernel-based method for protein-protein interaction mining from biomedical literature. Lecture Notes in Computer Science, v 3886 LNBI, Knowledge Discovery in Life Science Literature - PAKDD 2006 International Workshop, KDLL 2006, Proceedings. 2006:42~52

[9] P Bruno, Z Nivio, W Meira. Set-based vector model:an efficient approach for correlation-based ranking. ACM Transactions on Information Systems. 2005, 23:397~429

Fuhua Shang : Received the D.E. degrees in Harbin Institute of Technology, Computer Science and Technology College in 2007.He has been a professor in Daqing Petroleum Institute, Computer & Information Technology College.Research interests are Artificial Intelligent and Machine learning.

[10] CRISTIANINI N, KANDOLA J, ELISSEEFF A, et al.On kernel target alignment[M].Cambridge:MIT Press,2002:367-374.

[11] Naiyang Deng, Yingjie Tian. A new method of data mining - Support Vector Machine, Science Press 2004.120-122

[12] SMITS,BARZILAY O, SHAHAVE R.On Global Local Mixed and Neighborhood Kernels for Support Vector Machines[J].Pattern Recognition Letters,1999,20:1183-1190.