Automatic Identification of Log-curve Formation Based on Wavelet-SVM

Fuhua Shang[1,2], and Tiejun Zhao[1] ,and Xiujie Miao[2]

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

Summary

An arithmetic based on wavelet transform and process SVM (Support Vector Machine) for automatically identifying log-curve formation is proposed. Wavelet transform can transform any log-curve to space vector which the experiment system requires, then use the theory of process SVM automatically to identify log-curve. The results of experiment indicate this arithmetic has good identification ability and strong generalization ability on occasion that the number of training swatch is limited.

Key words:

Log-curve, process SVM, wavelet transform.

1. Preface

Now, water injection to stratum is the popular method to produce oil in more and more oil fields. Water-flooded zone identification of oil field is an important problem during development's middle-late period, and it is bursting to solve. Water-flooded zone identification mostly depends on log-curves which reflect physical and chemic character of stratum. It is a difficult problem of oil field geological analysis that how to carry out automatically identifying by using the information. The workload of manual identification is huge, and speed is so slow that it doesn't satisfy practical demand. The topmost identifying accuracy is between 70% and 80% because many conditions of subterranean affect Water-flooded zone identification .

During oil field geological analysis, we often use the method of making cored well for knowing Water-flooded zone condition better, geological analysis according such as character of subterranean rock-core which is from cored well. But it's very expensive of making cored well, so the number of cored well's data are limited. Our task is how to reflect geometrical peculiarity of actual curve by the received data which are from limited log-curve information, and then obtain the correct rules and make use of it for the future analysis of geology. There are a lot of log-curve identifying methods[1] and we can use many transforms for getting features of these curves[2][3]. This article uses wavelet transform to dispose primal data, it can reduce the calculational complexity and reflect geometrical peculiarity of primal data better in some aspects, for example: wave crest and inflexion.

In the reference [4], a model based on process neural network was proposed. The structure of process neural network is similar to traditional MP model, it contains three parts: adding weight、polymerization and prompting operation constitutes. It is different from traditional nerve cell. The differences lie in that weights and inputs can be changed with the time in processing nerve cell and its polymerization operation includeing polymerizing many special inputs and accumulating time process. A generic study algorithm based on grads descending was proposed in the reference [5] which presents a training problem of process neural network for us. A model based on wavelet transform and process SVM[6][7] (Wavelet-SVM) for automatically identifying log-curve stratum is proposed, it's a improvement on traditional method. Test results indicate this arithmetic has higher indentifying ability and generalization ability is fine on occasion that there are limited training data.

2. Support vector machine

Support vector machine (SVM) is a new machine learning method, whose theoretical basis is Statistical Learning Theory[6]. Particularly, more attentions are paid to researching theoretical rules and learning methods. The basic thoughts are described as follows: input vector x is mapped to a high-dimensional feature space Z by non-linear mapping which has been chosen, and then structure optimal margin hyperplane. The primary aim of SVM is to get the best answer under the information in existence, not under infinite information. At present, algorithms of SVM are:

1）C-support vector classification(C-SVC);
2）v-support vector classification (v-SVC);
3）distribution estimation (one-class SVM);
4）epsilon-support vector regression (epsilon-SVR);
5）v-support vector regression (v-SVR).

Thereinto, C-SVC and v-SVC are used for classified arithmetic; one-class SVM is used for classified estimation; epsilon-SVR and v-SVR are used for regressive arithmetic.

During mapping transformation from low-dimensional input space to high- dimensional feature space, the spatial

Manuscript received April 5, 2006.
Manuscript revised April 25 , 2006.

dimension rises rapidly (e.g. if the dimension of input space is n, through hypo-m polynomial mapping, its spatial dimension turns to $O(n^2)$ through hypo-m polynomial mapping), with the result that it's hard to calculate optimal margin plane in feature space. SVM translates this problem into input space for calculation by defining kernel function. At present, the primary kernel functions are: radial basis function (RBF function), linear function, polynomial basis function (poly function) and sigmoid function.

We consider two- classifiable problem which is a linear dividable problem as following after linear transform:

$$(x_1, y_1), (x_2, y_2), \cdots, (x_l, y_l) \in R^N \times Y, \quad Y = \{-1, +1\} \quad . \quad (1)$$

thereinto, x_i is independent and distributing the same.

Linear dividable property indicates that the experience risk of this problem is 0. According to the principle of minimizing structure risk, it's enough to minimize believable scope. The believable scope is a increasable function of VC dimension, h, so the principle of minimizing structure risk can be carried out by minimizing VC dimension.

For decreasing the repetition of classifiable plane, that using $\min |w \bullet x_i + b| = 1, i = 1, 2, \cdots, l.$ to restrict (w, b). When data $x_1, x_2, \cdots x_i$ exist in a ball with the radius r, VC dimension of $\{f = w \bullet x_i + b \ \vdots \|w\| <= A\}$ is $h \le \min\{r^2 A^2, N\}$ .In SVM, for the linear dividable problem, the problem for receiving expectation risk can boil down to:

$$\min_{w,b} \frac{1}{2} \|w\|^2 \quad s.t. y_i (w \cdot x_i + b) \ge 1, i = 1, 2, \cdots, l \qquad (2)$$

By anterior analyses, the meaning of this optimizing problem is to minimize the bound of VC dimension for minimize the VC dimension when the experience risk is zero. So we can say that SVM is a approximate actualize of the principle of structure risk minimization.

This optimizing problem, using Lagrange multiplication is equal to:

$$\max \sum_{i=1}^{l} \lambda_i - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} \lambda_i \lambda_j y_i y_j x_i \bullet x_j \quad s.t. \lambda_i \ge 0, i = 1, 2, \cdots, l, \sum_{i=1}^{l} \lambda_i y_i = 0 \qquad (3)$$

Obviously, this optimizing problem is a quadratic programming problem (QP problem), its local solutions must be the global best solutions. Feedforward neural networks can't translate classified problem into QP problem although it has tried its best. Another important meaning of this optimizing problem is it relate nothing but netscape, by reason of this, it is the basis of kernel application. By optimizing problem can gained λ_i , and then

$$w = \sum_{i=1}^{l} \lambda_i y_i x_i \qquad (4)$$

thereinto, λ_i is the solution of dual programming. The normal vector of classification hyperplane is a linear combination of swatch. It is one of the features of SVM. x_i which corresponds with λ_i , called support vector.

The final decision function has format like this:

$$f(x) = \mathrm{sgn}(\sum_{sv} a_i y_i K(x, x_i) + b) \qquad (5)$$

3. Wavelet

3.1 Basic theory of wavelet

Wavelet[7] is a function or singnal $\psi(x)$ which satisfies one of following conditions in function space

$$C_V = \int_{R^*} \frac{|\psi(w)|^2}{|w|} dw < \infty \qquad (6)$$

Thereinto, $R^* = R - \{0\}$ shows all real number except zero. Sometimes, $\psi(x)$ is called wavelet mother function, preceding condition is "permissibility condition". The function for any real number (a, b) like this

$$\psi_{(a,b)}(x) = \frac{1}{\sqrt{|a|}} \psi(\frac{x-b}{a}) \qquad (7)$$

is called continuous wavelet which is created by wavelet mother function $\psi(x)$ and depends on parameter (a, b), it's short wavelet, and parameter a must be a real number which is not a zero.

3.2 Wavelet transform

Wavelet transform[8] is a new kind of method for transform and analysis, its origin is multiresolution analysis. It expresses function $f(t)$ in $L^2(R)$ as a series of imminent expression, and each of them is the format which is received by smooth disposal and it The sampling pace of wavelet transform is accommodative with different frequency component in time-space or airspace, it's a high frequency when time is short, and it's a low frequency when time is long. In this senses, wavelet transform is called the microscope of mathematics, it can divide signals or images into many scale elements which can complect together. Using corresponding sampling pace in time-space with different frequency component can focus on any minute detail of object, and it is the better aspect than standard Fourier transform and windowed Fourier transform.

4. Results and Conclusions of Experiment

Water-flooded zone identification mostly depends log-curves which reflect physical and chemic character of stratum[9]. We choose representative five parameters (thick, deep-lateral resistivity,spontaneousness potentio,interval transit time,puny potentio,water-flooded grade) as input by expert experience and test analysis, while output is water-flooded grade.

During oil field geological analysis, we often use the method of making cored well for knowing Water-flooded zone condition better, geological analysis according such as character of subterranean rock-core which is from cored well. But it's very expensive of making cored well, so the number of cored well's data are limited. Our task is how to reflect geometrical peculiarity of actual curve by the received data which are from limited log-curve information, and then obtain the correct rules and make use of it for the future analysis of geology. There are a lot of log-curve identifying methods[1] and we can use many transforms for getting features of these curves[10][11]. This article uses wavelet transform to dispose primal data, it can reduce the calculational complexity and reflect geometrical peculiarity of primal data better in some aspects, for example: wave crest and inflexion.

We choose representative five parameters (thick, deep-lateral resistivity,spontaneousness potentio,interval transit time,puny potentio,water-flooded grade) as input by expert experience and test analysis, while output is water-flooded grade.

We take two experiments for validating the correctness of this model.

First: select the proper kernel function

In data of oil-saturated stratum, we choose 430 samples for training and 205 for testing. Thereinto, water-flooded grade divides into strong, middle, weak and non water-flooded zone.

corresponds with different resolving power.

We input the results which is received form training samples by wavelet transform to SV, then we will receive support vectors and weight parameters. We choose rbf function as kernel function because it can receive the most support vectors, the highest accuracy and the rapid is very rapid. The result is showed in Table 1 and Table

Second: compared with neural network in the precondition of same data

Training the training 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 zone identification. When we train the same data by network., the accuracy of training samples is 96.4%, but the accuracy of measuring data is 73.4%. The result is showed in Table 3.

We can see the generalization abilityof this model is stronger than neural network, although it's accuracy of training samples is 90.5%.

5. Conclusion

This paper introduces the basic theory and proposes a Wavelet-SVM model, and then provides the experiment to support. Wavelet-SVM can avoid longer training time and weaker generalization ability, and it is a good reference for solving the pattern recognition, system identification and simulation modeling problem. We can see that Wavelet-SVM can avoid longer training time and weaker generalization ability, and it is a good reference for solving the pattern recognition, system identification and simulation modeling problem. In the future, more and more theory research and experiment should do for filtering training samples and increasing the quantitative analysis of accuracy.

Table 1: Support Vector Received by Some Kernel Function

kernel function	parameter setting	number of support vector	number of support vector of each kind			
			strong	middle	weak	no
rbf	d＝3.0,c＝1000.0,g=0.25	343	46	110	41	146
Poly	d＝3.0,c＝1000.0,g=0.25	250	43	83	37	87
sigmoid	d＝3.0,c＝1000.0,g=0.25	269	38	86	39	106

Table2: Training Speed and Accuracy Received by Some Kernel Function

kernel function	parameter setting	time（second）	the accuracy of training data (%)	the accuracy of measuring data (%)
rbf	d＝3.0,c＝1000.0,g=0.25	40	90.5	78.1
Poly	d＝3.0,c＝1000.0,g=0.25	1145	80.5	68.4
sigmoid	d＝3.0,c＝1000.0,g=0.25	254	78.3	65.2

Table 3: Comparison Between Wavelet-SVM Algorithm and Network

algorithm	Training time (second)	the accuracy of training data (%)	the accuracy of measuring data (%)
B-SVM	40	90.5	78.1
ANN	8145	96.4	73.4

References

[1] Fuhua Shang, Tiejun Zhao and Sheng Li, A Normalized Fuzzy Neural Network and Its Application, iclm2003. pp.1088-1091

[2] Nie L,Wu S G.Continuous wavelet transform and its application to resolving and quantifying the overlapped voltammetric peeks[J].Analytical Chimina Acta,2001,450:185-192

[3] Xu L， Shi J. Geometric Hermite interpolation for space curves. Computer Aided Geometric Design, 2001.18(9):817-829.

[4] Xingui He, Jiuzhen Liang. Theoretical problems of process neural network. Engineering Science, 2000，2(12)：40-44

[5] Xingui He, Jiuzhen Liang, Shaohua Xu. Research and Applications of process neural network.. Engineering Science,2001，3(4)：31-35

[6] Vapnik V.: The Nature of Statistical Learning Theory[M]. Springer-Verlag (1995)

[7] Ingrid Daubechies. Ten Lectures on Wavelets. SIAM. Philadelphia, PA. 1992

[8] Fusheng Yang. Engineering Analysis and Applications of Wavelet Transform. Beijing: Publishing House of Science. 2000

[9] Li H.Y.: Water Injection to Stratum for Well Logging' 96 International Academic meeting Survey. Well Logging Technology. Vol. 9. (1996)

[10] Nie L, Wu S.G.: Continuous wavelet transform and its application to resolving and quantifying the overlapped voltammetric peeks[J]. Analytical Chimina Acta. Vol. 450. (2001) 185-192

[11] Xu L.， Shi J.: Geometric Hermite interpolation for space curves. Computer Aided Geometric Design. Vol. 18. (2001) 817-829

Fuhua **Shang** Received the M.S. degrees in Harbin Institute of Technology, Computer Science and Technology College in 1990.Now, he stayed in Machine Translation Laboratory (MTLab).The mail .Research interests are Artificial Intelligent and Machine learning.