Journal of Applied Science and Engineering Innovation, Vol.2 No.6, 2015, pp.205-208
ISSN (Print): 2331-9062
ISSN (Online): 2331-9070

# The Auto-Identifying Log-curve Formation Based on Wavelet SVM

Wang Jie
*Department of Electrical Information Engineering, Northeast Petroleum University, Daqing, 163318, China*

**Abstract:** 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.

**Keywords** Log-curve; Process SVM; Wavelet transform

## INTRODUCTION

Nowadays, many oil fields take the method of water flooding to get the oil. The identity of the situation of water out behavior is an important problem that is urgent to solve in the middle and later period of the oilfield development [Yan *et al.*, 2014]. The identity of the water-flooded zone is mainly based on the log-curves that reflect the formation's physical and chemical properties. How to classify automatically according to this information is a problem in the analysis of the oilfield geology [Yan *et al.*, 2014]. Artificial perception calls for a lot of work, and it's too slow to meet the actual need. The highest accuracy of the recognition at the present is between 70 to 80 percent, for the water flooded layer recognition is influenced by a lot of conditions underground [Wang *et al.*, 2012].

The literatures put forward the model of identification method [Wang *et al.*, 2010]. The process neuron has a similar formation with the traditional MP model, which is constituted by weight, aggregation and excitation operation [Peng *et al.*, 2013]. The difference between the traditional neuron and the process neuron is that the later one calls for time-varying inputting and weighting, its aggregation operation contents multi-input aggregation as well as the accumulation of the time process [Xu *et al.*, 2011]. For the training of the process neuron net, the literature has given us a general learning algorithm based on gradient descent algorithm. In this text, we improved the traditional algorithm, and put forward a model of identifying process SVM log-curve [Shang *et al.*, 2006]. 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.

## PROCESS SUPPORT VECTOR MACHINE MODEL

The process support vector machine is made up of wavelet transform, Kernel Function transform and maximum liberal classification. The output (decision rule) is resolution analysis. Its basic thought

$$Z = \mathrm{sgn}(\sum_{i=1}^{m} a_i z_i K(y_i, y) + b) \tag{1}$$

In this formula,(x1(t),x2(t),x3(t),…,xn(t))are the input vectors,(y1(t) ,y2(t),x3(t),…,ym(t)) are vectors we get after wavelet transform. K(yi,y) is the Kernel Function. yi(i=1,2,⋯,k)in the formula is the support vector after training. y is the input vector y=(y1(t),y2(t),x3(t),…,yn(t)).Weight Wi=aizi,.b is the constant. The structure is shown in the Fig.1.

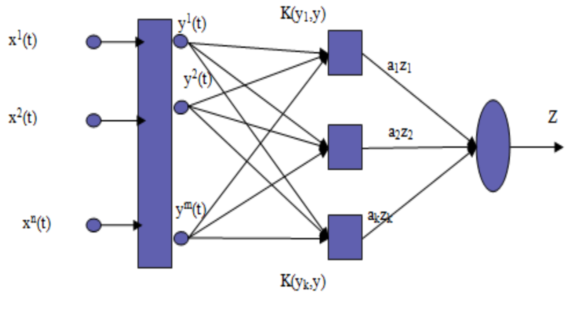


Figure1. Process Support Vector Machine Model

In the work of oil field geological analysis, people get a group of oil reservoir information from different depths after analyzing. Each information of a reservoir is a group of log-curve information, i.e. a group of time functions in different lengths. In order to standardize the information, we must standardized interpolate the information.

The fitted time function must reflect the character of the origin data; this character should present the space geometric characteristics. So it is not just a simple question of numerical approach.

**Basic wavelet theory:** Wavelet is a function or signal $\psi(x)$ which meets the following condition in the function space $L^2(R)$

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

Here, $R^* = R - \{0\}$ stands for all the nonvanishing real numbers. Sometimes, $\psi(x)$ is also known as wavelet generating functions, with the forementioned condition called admissible condition. For any real number pair $(a,b)$, in which the parameter $a$ must be a nonvanishing real number, we call the functions of the following form

$$\psi_{(a,b)}(x) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{x \quad b}{a}\right) \qquad (3)$$

The parameter $(a,b)$ -dependent continuous wavelet functions, which is created by the wavelet generating functions $\psi(x)$.It is also called xiaobo for short in Chinese.

After nonlinear transformation, we come to consider the following linear classification problems can be divided into two.

$$\begin{aligned} &(x_1, y_1), (x_2, y_2), \cdots, (x_l, y_l) \in R^N \times Y, \\ &Y = \{\ \ 1, +1\} \end{aligned} \qquad (4)$$

In the formula, the xi is independent and identically distributed.

The property of linear separability reveals that such classification problem bears no empirical risk, according to the theory of Structural Risk Minimization, all we have to do is to minimize the confidence interval. As the confidence interval is the increasing function of the VC dimension h, the Structural Risk Minimization reflects in minimizing the VC dimension h. In order to reduce the repeat of the classification plane, we bind the(w,b)as When the data points x1,x2…,xl w situate in the globe with the radius of r, in the formula of ,the VC dimension h⩽ min{r2A2,N}.in SVM, under the situation of linearly separable, the problem of calculating the(w,b) with the minimal expected risk can be attributed as the following:

$$\begin{aligned} &\min_{w,b} \frac{1}{2} \| w \|^2 \\ &s.t. y_i (w x_i + b) \geqslant 1, i = 1,2,\cdots,l \end{aligned} \qquad (5)$$

From the former analysis, we know that this optimization problem means to minimize the bound of the VC dimension when there is no expected risk, thereby minimizing the VC dimension. So that we say SVM is the approximate realization of the structural risk minimization theory.

Aiming to the former optimization problem, we can use the Lagrange multiplier method, which is equivalent to

$$\begin{aligned} &\max \sum_{i=1}^{l} \lambda_i \ \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} \lambda_i \lambda_j y_i y_j x_i \bullet x_j \\ &s.t. \lambda_i \geqslant 0, i = 1,2,\cdots,l, \sum_{i=1}^{l} \lambda_i y_i = 0 \end{aligned} \qquad (6)$$

Obviously, this optimization problem is a convex optimization problem, so that its local solution must be a global optimal solution. Transforming the classification problem into a convex optimization problem has never been achieved, though the feed forward neural network has made much effort in many ways. The other important meaning of this optimization problem is that it is only related to the inner product, and it lays the foundation of the application of the kernel trick. From the optimization problem, we can get the λi, thus comes

$$w = \sum_{i=1}^{l} \lambda_i y_i x_i \qquad (7)$$

In the formula, λi is the solution towards the dual programming problem given by the former optimization problem. It is one of the most important features of SVM that the vector of the classification hyperplane is the linear combination of the sample points. The data point xi that is corresponding λi is called support vector.

The final decision function is as following

$$f(x) = \operatorname{sgn}\left(\sum_{sv} a_i y_i K(x, x_i) + b\right) \qquad (8)$$

**Description of the Algorithm:** The support vector machine (SVM) we choose is the C-SVC, in the following algorithm, we separately use three kind of kernel functions, RBF function, poly function and sigmoid function, and compare the experiment results.

The algorithm is as the following:

Step 1: Reflect the log-curve to the vector space by wavelet transforming;

Step 2: Input the vector we get to the SVM training model;

Step 3: Output the support vector and the relative parameters;

Step 4: Found the SVM curve identification model based on wavelet;

Step 5: Input the log-curve that is to be analyzed;

Step 6: Output the recognition effect.

## RESULTS AND DISCUSSION

The recognition of the water-flooded zone is largely based on the log-curves that can reflect the physical and chemical properties. After the relative analysis and statistics, according to the experience of the field experts, the writer chose spontaneous potential (SP), High resolution acoustic transit time (AC), High resolution deep lateral resistivity Rlld and the difference between micro potential and micro gradient, Rmn-Rmg, as the logging feature parameters for the recognition of the water-flooded zone's water flooded grade, and the output is the water flooded grade.

From the limited reservoir data of the core holes, we chose 450 representative water-out reservoir sample to form a training set, and 225 reservoir sample to form a test suite. According to the determination method of the pattern classes number, the water flooded grade of the reservoir can be divided into 4 situations, strong water flooding, secondary water flooding, weak water flooding and not flooded.

We deal with the 450 training samples by wavelet transform, then input the results to SVM to train. After training, we get the corresponding support vectors and weight parameters, thus getting the model as shown in Fig1.In the experiment, when we use RBF function as the kernel function; we get the most support vectors, the highest classification accuracy in a fast running speed. So we choose RBF function as the kernel function, the experiment results are shown in Table1 and Table2.

Table1. Conditions of supporting vectors obtained by several kernel functions

| Kernel Function | Parameter Setting | Quantity of supporting vector | Quantity of supporting vector in each classification | | | |
|---|---|---|---|---|---|---|
| | | | Strong water flooding | Secondary water flooding | Weak water flooding | not flooded |
| 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 |

Table 2. Conditions of training speed and accuracy obtained by several kernel functions

| Kernel Function | Parameter Setting | Time(s) | Training Sample Accuracy (%) | Test Sample Accuracy (%) |
|---|---|---|---|---|
| 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 |

When we back to judge the training sample with the studied SVM actuator, the correct recognition rate is 90.5%; when judging the 305 samples in the test suite, the correct recognition rate is 78.1%. It is a fairly good result in terms of flooded layer's automatic recognition. When the same data are used in the neural network, the sample accuracy will come to 96.4%; but the accuracy of the test suite is only 73.4%.The experiment results are shown in Table3.

Table3. Comparison between B-SVM Algorithm and Process neural network

| Algorithm | Training time(s) | Training Sample Accuracy (%) | Test Sample Accuracy (%) |
|---|---|---|---|
| B-SVM | 40 | 90.5 | 78.1 |
| Process neural network | 8145 | 96.4 | 73.4 |

## CONCLUSION

We can know from the experiment results that although the accuracy of the training sample is only 90.5%,it has a strong ability of generalization. So the process support vector machine presented in this paper overcame the problem of long neural network training time and weak generalization ability. Furthermore, it has a very good reference value in solving the problems of pattern identification of time varying system, system identification and simulation modeling.

## REFERENCES

Lu, Y., Li, W., Tian, W., & Zhou, K. (2014). Fuzzy Entropy Thresholding Method Using Adaptive Genetic Algorithm. Journal of Applied Science and Engineering Innovation, Vol. 1(1), pp.1-6.

Peng, Y., X. Zhang, Z. Han and J. Jiao, 2013. Dynamic fault diagnosis in chemical process based on SVM-HMM. Proceedings of the IEEE International Conference on Mechatronics and Automation, August 4-7, 2013, Takamatsu, Japan, pp: 1687-1691.

Vapnik, V., 1995. The Nature of Statistical Learning Theory. 2nd Edn., Springer, New York, USA., ISBN-13: 9780387945590, Pages: 188.

Wang, A., M. Sha, L. Liu and F. Zhao, 2010. Fault diagnosis of TE process based on ensemble improved binary-tree SVM. Proceedings of the IEEE 5th International Conference on Bio-Inspired Computing: Theories and Applications, September 23-26, 2010, Changsha, China, pp: 908-912.

Wang, Y., Y. Zhang, J. Wang and Y. Chen, 2012. Research of flotation dosing control method based on SVM and image processing technology. Proceedings of the 31st Chinese Control Conference, July 25-27, 2012, Hefei, China, pp: 7067-7070.

Xu, J. and S. Hu, 2010. Nonlinear process monitoring and fault diagnosis based on KPCA and MKL-SVM. Proceedings of the International Conference on Artificial Intelligence and Computational Intelligence, Volume 1, October 23-24, 2010, Sanya, China, pp: 233-237.

Xu, J.J., L.N. Sha, Y. Zhang, D.F. Zhang, G.C. Liu, A.H. Xu and H.Y. Li, 2011. A new algorithm for minimum spanning tree. Power Syst. Prot. Control, 39: 107-112.

Yan, L.M.,Cui, J., Y.Q. Huang, Y.B. Xie, J.J. Xu, L.M. Yan and Y.S. Zhu, 2014. Power system state estimation of quadrature Kalman filter based on PMU/SCADA measurements. Electr. Mach. Control, 18: 79-84.

Yan, L., Y. Xie, J. Xu, C. Xue, H. Zhao and L. Bai, 2013. Improved forward and backward substitution in calculation of power distribution network with distributed generation. J. Xi'an Jiaotong Univ., 47: 117-123.

Yan, L., Y. Zhu, J. Xu, W. Ren, Q. Wang and Z. Sun, 2014. Transmission lines modeling method based on fractional order calculus theory. Trans. China Electrotech. Soc., 29: 260-268.