TELKOMNIKA Indonesian Journal of Electrical Engineering
Vol.12, No.2, February 2014, pp. 1643 ~ 1650
DOI: http://dx.doi.org/10.11591/telkomnika.v12i2.4167

The Oil Layer Recognition Based on Multi-kernel Function Relevance Vector Machines

Qianqian Zhao[1], Xinyuan Xia[2], Kewen Xia*[3], Zhaozheng Hu[4]
School of Information Engineering, Hebei University of Technology, Tianjin 300401, China
Corresponding author, e-mail: zhaoqq123_456@163.com[1], will_9898@sina.com[2], kwxia@hebut.edu.cn[3], zzhu@hebut.edu.cn[4]

Abstract

In the oil layer recognition, Relevance vector machines (RVM) have a good effect. But the single kernel function RVM has some limitations, a kind of multi-kernel function RVM based on particle swarm optimization (PSO) is proposed, which includes the model parameter estimation, model optimization on multi-kernel function RVM, PSO-based training, and recognition. The results of simulation experiment for a typical recognition dataset show that its effect is superior to that of classical RVM and PSO-based single kernel function RVM, and its actual application for oil layer recognition in well logging indicates that the recognition results are completely consistent with the conclusions of oil trial, and it has the high recognition accuracy and the good recognition effect.

Keywords*: Oil layer recognition, relevance vector machines (RVM), multi-kernel function, particle swarm optimization (PSO)*

1. Introduction

Now the technology of oil well logging has developed well. It developed from the original analog logging to the digital logging and then to the informational logging. In the informational logging, the information technology is applied to the oil well logging. Oil layer recognition [1] is an important part of petroleum exploration and development and it is also one of the main works for logging explanation. In previous oil layer recognition, the managers usually obtained the recognition results by an experience-based parameter. So for the multi-parameters, the recognition becomes inaccurate or unscientific which eventually caused serious economic losses. In recent years, some researchers used neural networks [2], fuzzy clustering analysis and Support Vector Machines (SVM) [3] for oil layer recognition. The recognition was not limited by a single parameter and began to analyze and forecast the multi-parameters which made oil layer recognition more scientific and reasonable. But neural networks, fuzzy clustering analysis and SVM have some shortcomings in the recognition, such as the structure of neural networks is not easy to determine and easy to fall into local minimum; fuzzy clustering analysis has low extraction capacity for some implicit classes that may cause the recognition results do not match with actual results; SVM lacks probability factor and its kernel function must be the continuous symmetric positive definite kernel [4-6].

Therefore, many scholars began to study the Relevance Vector Machines (RVM) [7-9]. In fact, RVM suffer from none of the other limitations of SVM outlined above. However, the kernel function parameters of the classic RVM are obtained based on experience, so some scholars began to optimize them by using evolutionary computation algorithms to get good application results, such as the method of PSO-based RVM. In actual application, the performance of the single kernel RVM still has some limitations and can't be greatly improved by optimization of the kernel function parameters.

Hence we present a kind of PSO-based multi-kernel function RVM, and use it to carry on the simulation analysis and the actual application in the oil layer recognition. The main improvements cover that the multi-kernel RVM parameters are optimized by PSO algorithm, and the RVM have the high recognition accuracy.

Received August 17, 2013; Revised September 25, 2013; Accepted October 17, 2013

2. RVM and Its Improvement

2.1. The Principle of RVM

Tipping M E proposed the Relevance Vector Machines (RVM) in 2000s [10]. Define a training sample group $\{x_i, t_i\}_{i=1}^{n}, x_i \in R^n$ is an input vector, the output definition of RVM model is as follows.

$$y(x; w) = \sum_{i=1}^{M} w_i k(x, x_i) + w = \Phi(x) w \tag{1}$$

Here, $w = [w_0, w_1, w_2, \cdots w_M]$, $\Phi(x)$ is the $M \times (M+1)$ matrix organized by the kernel function.

The likelihood of the complete dataset can be written as

$$p(t \mid w) = \prod_{i=1}^{N} N(t_i \mid y(x_i; w)) = (2\pi)^{-\frac{N}{2}} \exp\left(-\frac{1}{2} \| t - \Phi(x) w \|^2\right) \tag{2}$$

Here $t = [t_0, t_1, t_2, \cdots, t_N]^T$.

The distribution of w is the zero-mean Gaussian prior distribution. Thus the probability prediction formula is

$$p(t_* \mid t) = \int p(t_* \mid w) p(w, \alpha \mid t) dw d\alpha \tag{3}$$

According to delta resembling function, the formula (3) can be changed into

$$\begin{aligned} p(t_* \mid t) &= \int p(t_* \mid w, \alpha) p(w \mid t, \alpha \mid t) \times p(\alpha \mid t) dw d\alpha d \\ &\approx \int p(t_* \mid w, \alpha) p(w \mid t, \alpha) \times \delta(\alpha - \alpha_{MP}) dw d\alpha \\ &= \int p(t_* \mid w, \alpha_{MP}) P(w \mid t, \alpha_{mp}) dw \end{aligned} \tag{4}$$

Adopt following formula of Maximum-likelihood calculation.

$$\begin{cases} p(\alpha \mid t) \propto p(t \mid \alpha) p(\alpha) \\ (\alpha_{MP}) = \arg \max_{\alpha} p(t \mid \alpha) \end{cases} \tag{5}$$

And it only needs to adopt the following formula to maximize $p(t \mid \alpha)$.

$$p(t \mid \alpha) = \int p(t \mid w) p(w, \alpha) dw = (2\pi)^{-\frac{N}{2}} |\Omega|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} t^T \Omega^{-1} t\right) \tag{6}$$

Here $\Omega = \Phi A^{-1} \Phi^T$. By partial differential on the formula (7) for α, then the solution of stagnation point is as follows.

$$\alpha_i^{new} = \frac{\gamma_i}{u_i^2} \tag{7}$$

Here $\gamma_i = 1 - \alpha_i \Psi_{i,i}$.

2.2. A Kind of PSO-Based Multi-Kernel Function RVM

The traditional RVM use a single kernel function to complete the mapping process. In many actual applications, using a single kernel of RVM has achieved good results. But when the features of samples contain different data structures or the samples size are relatively large,

simply using a single kernel function for all samples has limitations which can't well adapt to the diverse data types of the samples. In order to adapt to the complex situations, in RVM many basic kernel function can be selected, the commonly used kernel functions are as follows:

(1) Linear kernel function: $K(x, y) = xy$

(2) RBF kernel function: $K(x, y) = \exp(-\frac{\|x - y\|^2}{\sigma^2})$

(3) Laplace kernel function: $K(x, y) = \exp(-\sqrt{\gamma \|x - y\|^2}$

(4) Polynomial kernel function: $K(x, y) = (x(\frac{y}{\sigma^2})' + 1)^2$

(5) Sigmoid kernel function: $K(x, y) = \tanh(\rho xy - \sigma)$

The main parameter of RBF kernel function and Polynomial kernel function is σ, the main parameter of Laplace kernel function is γ, the main parameter of Sigmoid kernel function is ρ, σ. Different kernel functions have different characteristics and the same kernel function with different parameters will produce different effects on kernel function.

We make the different kernel functions together and the combined kernel function can have better characteristics, which is the basic idea of multi-kernel function. In actual applications, the RBF kernel function exhibits an excellent property which is a typical local kernel function. The Polynomial kernel function is a typical global kernel function. For both local kernel and global kernel advantages, we combine RBF kernel function with Polynomial kernel function.

The multi-kernel function is

$$K = \lambda G(x, y) + (1 - \lambda) P(x, y) \tag{8}$$

Here, $G(x, y) = \exp(-\frac{\|x - y\|^2}{\sigma_1^2})$, $P(x, y) = (x(\frac{y}{\sigma_2^2})' + 1)^2$, $\lambda, \sigma_1, \sigma_2$ are the kernel parameters to be determined. When $\lambda = 0$ or $\lambda = 1$ it is a single kernel function respectively.

The principle of PSO:

Suppose there are N particles, each particle is a D dimension vector, $X_i = (x_{i1}, x_{i2}, \cdots, x_{id})$ and $V_i = (v_{i1}, v_{i2}, \cdots, v_{id})$ respectively stand for position and velocity of the particles. $P_b = (p_{i1}, p_{i2}, \cdots, p_{id})$ stands for the best position of the i particles and $P_g = (p_{g1}, p_{g2}, \cdots, p_{gd})$ stands for the best position of all the particles. The updating of the i velocity and positions are as follows.

$$v_i = v_{i-1} + c_1 \cdot r_1 \cdot (pbest - x_{i-1}) + c_2 \cdot r_2 \cdot (gbest - x_{i-1}) \tag{9}$$

$$x_i = x_{i-1} + v_i \tag{10}$$

Fix $V_{\max}$ the maximum velocity, and then the velocity of the particle can be changed as follows:

$$\begin{cases} v_i = V_{\max}, & if \quad v_i > V_{\max} \\ v_i = -V_{\max}, & if \quad v_i \le V_{\max} \end{cases} \tag{11}$$

In the process of PSO-multi-kernel RVM, $\lambda, \sigma_1, \sigma_2$ are the three parameters to be determined and the particles' positions vectors stand for the three parameters, the main steps are as follows:

Step1: The initialization of PSO: Initialize the iterations, the number of particles, weight parameter and other parameters; randomly initialize position $X_i(0)$ and velocity $V_i(0)$ of the particles;

Step2: Calculate the fitness of each particle; 5-fold cross-validation criteria are used to evaluate the fitness [11]: randomly divide the training samples into 5 folds; for each parameters setting, train the model five times; during every training, use 4 folds for training and the remaining fold for testing and record the fitness of every test; finally use the average fitness of 5 times test as the generalization performance. The fitness function is defined as.

$$F_{fitness} = \sum_{i=1}^{N} \frac{\sqrt{y - y_i}}{M} \tag{12}$$

Here, y is the measured value of the sample, y_i is the predictive value of the sample, and M is the number of the samples.

Step3: Change and update the particles; update the velocity and the position of each particle respectively in accordance with formula (9) and formula (10), and calculate the fitness of each particle to find the best position.

Step4: Judge the conditions: determine whether the iteration meets the termination condition, if meets, then stop and apply the optimized kernel parameters in the multi-kernel function RVM model, otherwise repeat step3.

2.3. Simulation on Classic Example

In order to test the effects of the recognition of the RVM algorithm, the dataset Pipley's synthetic is used. We choose 300 data as training sample and 1000 data as test sample. Pipley's synthetic dataset contains many mixed Gaussian noise. We use the classic RVM, the PSO-based single kernel function RVM and PSO-based multi-kernel function RVM to simulate the data recognition in computer respectively. In PSO algorithm, parameters are as follows: the number of particles is 40, the maximum number of iterations is 100, c_1, c_2 are respective taken 2.1.The obtained results are shown in Figure 1, Figure 2 and Figure 3 respectively. Here 1and 2 represent two classes of data. The results of three methods are shown in Table 1.

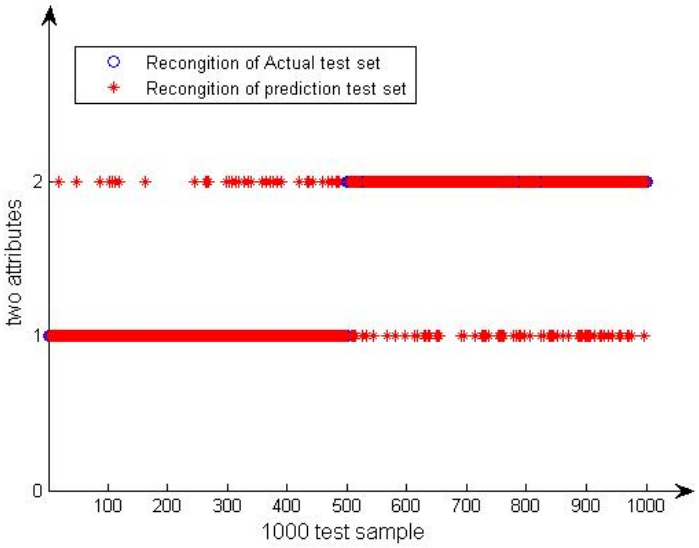

Figure 1. The Pipley's Recognition Results with Classic RVM

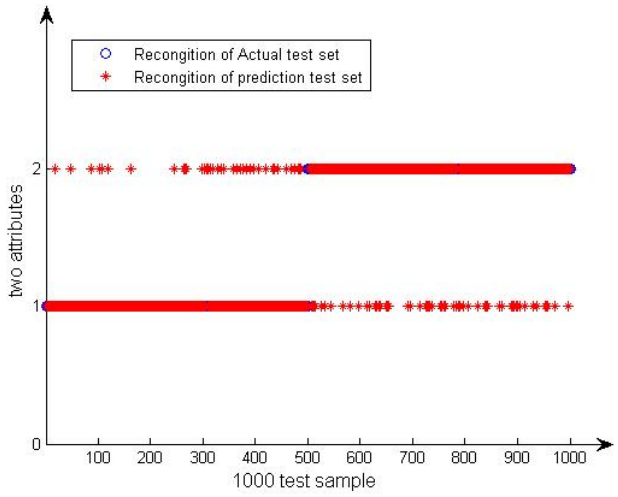

Figure 2. The Pipley's Recognition Results with PSO-based Single Kernel Function RVM

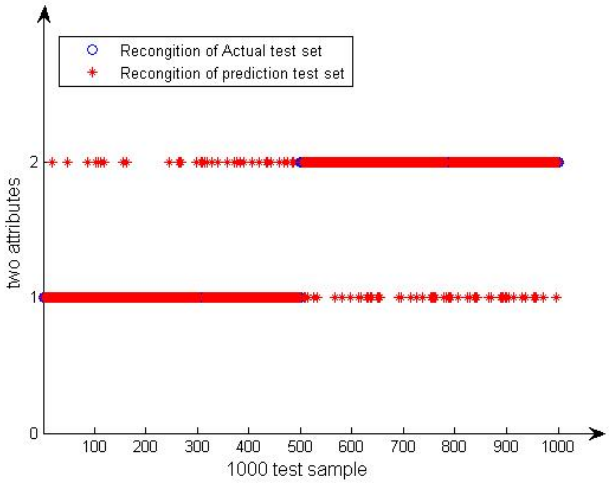

Figure 3. The Pipley's Recognition Results with PSO-based Multi-kernel Function RVM

We use recognition accuracy and false positive rate to describe the recognition results, defined as follows:

Recognition accuracy = number of correctly samples / total number of samples;

False positive rate = number of incorrectly samples / total number of samples.

Table 1. Comparison of Three Methods

Algorithm	Recognition accuracy (%)	False positive rate (%)
Classic RVM	88.8	11.2
PSO-based single kernel function RVM	90	10
PSO-based multi-kernel function RVM	90.5	9.5

Table 1 shows that for the same dataset, compared to the classical RVM, using PSO-based single kernel function RVM and PSO-based multi-kernel function RVM can improve the recognition accuracy. In particular, using PSO-based multi-kernel function RVM can get the higher recognition accuracy and lower false positive rate. And the effect of PSO-based multi-kernel function RVM is superior to that of the classic RVM and PSO-based single kernel function RVM.

3. Actual application

We firstly use PSO-based multi-kernel function RVM for a typical oil well in Xinjiang oil field. The model of oil layer recognition is shown in Figure 4. Oil recognition model includes the following steps: sample selection and preprocessing, attributes discretization and generalization, attribute reduction, building PSO-based multi-kernel function RVM, recognition and output.

Step1: Sample selection and preprocessing: we try to select comprehensive data of a various features and ensure the selected data without overlap. In addition, we should normalize the sample data, the formula is as following. Here, $x_{\min}$ and $x_{\max}$ are the minimum and maximum values of a property of the sample points respectively;

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{13}$$

Step2: Attribute discretization and generalization: the rough set method is used;

Step3: Attribute reduction: we reduction the logging attributes based on the degree of attribute dependency [12]. By calculating the degree of decision attribute dependence on logging attributes, we get the result of attribute reduction;

Step4: Build PSO-multi-kernel function RVM model: build the multi-kernel RVM and use PSO to optimize the multi-kernel function parameters;

Step5: Recognition and output: input the sample to the build model and get the recognition results.

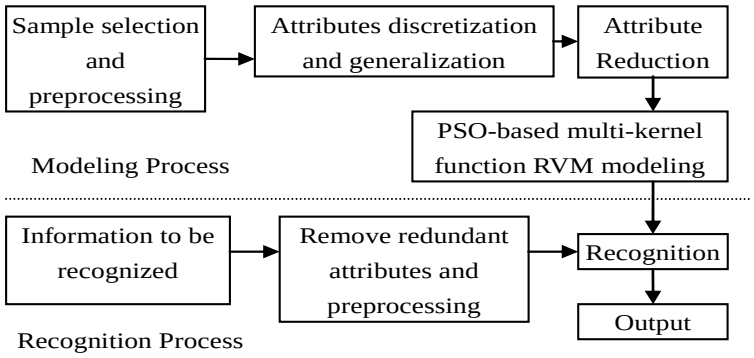

Figure 4. The Model of Oil Layer Recognition Based on PSO-multi-kernel RVM

We select an actual oil well to test. There are 11 logging attributes in the oil well sample, ie. {AC, CALI, GR, NG, RA2, RA4, RI, RM, RT, RXO, SP}. There are 3200 test samples in the oil well section 900m-1300m. The training sample is drawn between the 1220m-1290m depth of the oil well. The 281 sample points contain 59 oil layer points and 222 dry layer points. So the decision attribute is {dry layer, oil layer}. Firstly we identify the maximum and minimum of the sample points and use formula (13) to normalize them. Then through the discretization of continuous attributes and the generalization of decision attribute, we reduction the logging attributes based on the degree of attribute dependency. By calculating the degree of decision attribute dependence on logging attributes, we get the reduction results {AC, NG, RI, SP}.Then we respectively use the classic RVM, PSO-based single kernel function RVM and PSO-based multi-kernel function RVM to train the sample after reduction. The training samples recognition results show in Figure 5, Figure 6 and Figure 7 respectively. The recognition results of three algorithms are shown in Table 2. Here, Class1 is representative the dry layer and Class2 is the oil layer. In the PSO-based multi-kernel function RVM, we need to limit the range of $\lambda, \sigma_1, \sigma_{,2}$, $0 \le \lambda \le 1$, $0.3 \le \sigma_1, \sigma_{,2} \le 1$.

Table 2．Classification Contrast of Three Methods

Algorithm	Recognition accuracy (%)	False positive rate (%)
Classic RVM	94.3	5.8
PSO-based single kernel function RVM	95.5	4.5
PSO-based multi-kernel function RVM	96.83	3.17

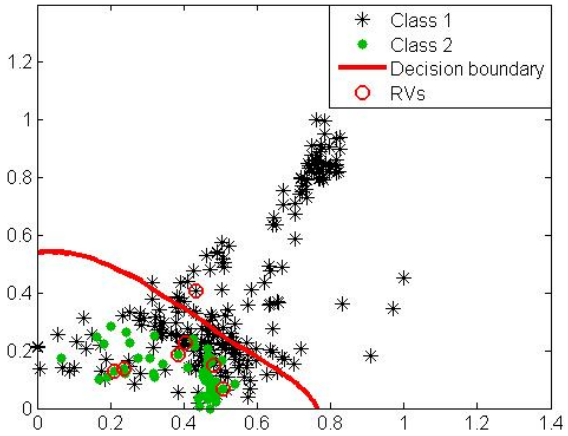

Figure 5. Recognition Results with Classic RVM for Training Sample

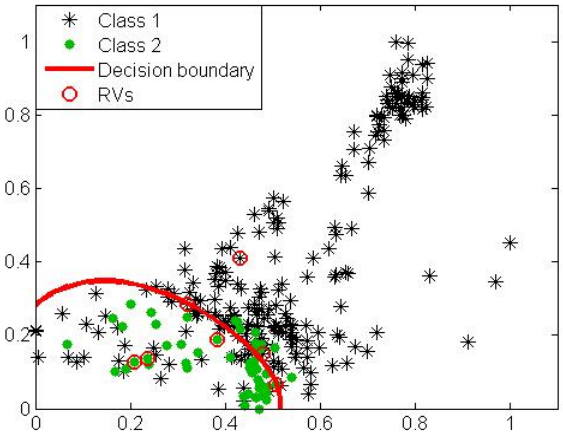

Figure 6. Recognition Results with PSO-based Single Kernel Function RVM for Training Sample

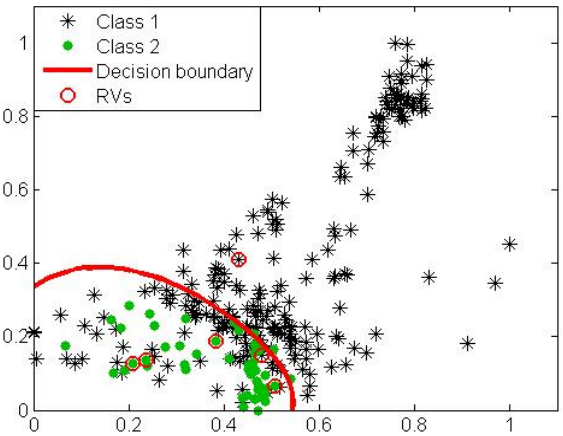

Figure 7. Recognition Results with PSO-based Multi-kernel Function RVM for Training Sample

Table 2 shows that in the actual application of oil layer recognition, compared to the classical RVM, using PSO-based single kernel function RVM and PSO-based multi-kernel function RVM can reduce the false positive rate and increase the recognition accuracy. In particular, using PSO-based multi-kernel function RVM can get the higher recognition accuracy. The recognition accuracy is over 96%. The recognition results are completely consistent with the conclusions of oil trial. It obviously, not only meets the logging industry standards and the actual logging requirements, and its application results improved significantly.

4. Conclusion

To overcome the limitations of single-kernel function RVM, based on the idea of multi-kernel function, we propose a kind of improved multi-kernel RVM method based on the PSO algorithm. The data simulation and actual application show that using PSO algorithm to optimize the multi-kernel parameters of RVM is feasible and the effect of PSO-based multi-kernel function RVM is superior to that of the classic RVM and PSO-based single kernel function RVM. PSO-based multi-kernel function RVM has significantly effect and higher recognition accuracy which has good prospects.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (No.512 08168), Tianjin Natural Science Foundation (No.11JCYBJC00900; No.13JCYBJC37700), Hebei Province Natural Science Foundation (F2013202254; F2013202102) and Hebei Province Foundation for Returned Scholars (No.C2012003038).

References

[1] Kewen Xia, Zhe Zhang, Yatong Zhou, Suicheng Si. Development on Intelligent Data Mining System in Oil Logging. *Journal of Computational Information Systems*. 2013, 9(3):1843-1850
[2] Yuhai Bao, Jianbin Ren. *Wetland Landscape Classification Based on the BP Neural Network in DaLinor Lake Area*. Procedia Environmental Sciences, 2011,10: 2360-2366
[3] Fu JH, Lee S. A multi-class SVM classification system based on learning methods from indistinguishable Chinese official documents. *Expert Systems with Applications*. 2012,39(3): 3127-3134
[4] Sankar Mahadevan, Sirish L. Shah. Fault detection and diagnosis in process data using one-class support vector machines. *Journal of Process Control*. 2009, 19(10):1627-1639.
[5] Weiguo Zhao, Yanfeng Gao. RVM based on PSO for Groundwater Level Forecasting. *Journal of Computers*. 2012, 5(7):1073-1079.
[6] Bilgin G, Erturk S, Yildirim T. Segmentation of hyper spectral images via subtractive clustering and cluster validation using one-class support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*. 2011, 49 (8): 2936-2944.
[7] Andrew R.J. Clark, Richard M. Everson. Multi-objective learning of Relevance Vector Machine classifiers with multi-resolution kernels. *Pattern Recognition*. 2012, 45(9): 3535-3543
[8] Clodoaldo A.M. Lima, André L.V. Coelho, Sandro Chagas. Automatic EEG signal classification for epilepsy diagnosis with Relevance Vector Machines. *Expert Systems with Applications*. 2009, 36(6): 10054-10059.
[9] Xiaodong Wang, Meiying Ye, C.J. Duanmu. Classification of data from electronic nose using relevance vector machines. *Sensors and Actuators B: Chemical.* 2009, 140(1): 143-148.
[10] Vapnik V N. The Nature of Statistical Learning Theory. New York: Springer-Verlag, 1995:11-13.
[11] Juan Diego Rodrı´guez, Aritz Pe´ rez, Jose Antonio Lozano. Sensitivity Analysis of k-Fold Cross Validation in Prediction Error Estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2010, 32(3), 569-575.
[12] Li Min, Deng Shao bo, Feng Shengzhong, Fan Jianping. Attribute reduction based on approximation dependency degree. *Journal of Computers*. 2013, 8(4): 920-928.