Research Journal of Applied Sciences, Engineering and Technology 5(19): 4671-4676, 2013
ISSN: 2040-7459; e-ISSN: 2040-7467

Submitted: September 26, 2012 Accepted: December 11, 2012 Published: May 10, 2013

Study of Flue-Cured Tobacco Classification Model Based on the PSO-SVM

[1]Hongmei Li, [1]Jiande Wu, [3, 4]Kuake Huang, [1]Xiaodong Wang and [2]Tingting Leng
[1]Faculty of Information Engineering and Automation,
[2]Faculty of Civil and Architectural Engineering, Kunming University of Science and Technology, Kunming, 650500, China
[3]Yunnan Agricultural University, Kunming 650000, China
[4]Qujing Branch, Yunnan Tobacco Company, Kunming 655000, China

Abstract: In this study, we study the flue-cured tobacco classification model based on the PSO-SVM. Firstly we use the Gaussian Radial Basis Function (RBF) as the kernel function of SVM and then use the Particle Swarm Optimization algorithm (PSO) to optimize the structural parameters of the SVM classifier, established the flue-cured tobacco classification model based on the PSO-SVM. Collecting a wide range of tobacco data in Qujing Yunnan Province, to train and validate the model. At last, compared with the grid parameter optimization and genetic algorithm-based parameter optimization model, the results show that the proposed model based on particle swarm optimization with high prediction accuracy and better adaptability when used in tobacco grading.

Keywords: Flue-cured tobacco, tobacco grade, particle swarm optimization algorithm, SVM

INTRODUCTION

Tobacco quality inspection and grading division are based on the national grading standards rely on the sense organs, usually identify with the Grading Factors that measure the tobacco leaf quality level and appearance features. Therefore, we need to avoid the subjectivity and vagueness in the process of tobacco leaf level division. The chemical composition of tobacco is the internal factors to determine tobacco grade quality, so can analysis the chemical composition of tobacco and other quality factors to determine tobacco grading (Yan and Zhao, 2003). Currently computer technology used in aspects of tobacco, such as cultivation, production, procurement and marketing, has great significance (Wang *et al*., 2007; Yan *et al*., 2001).

Support vector machine referred SVM, was proposed in 1995 by Vapnik and his colleagues, it is the algorithm that built based on VC dimension theory and SLT structural risk minimization (Vn, 1995). SVM combine statistical learning optimization methods with kernel function method, taking into account the minimized of training error (empirical risk) and test error (expected risk), find the most good compromise between the complexity of the model and the ability to learn that based on a limited sample information, in order to get the best ability to promote to solve the practical problems of the small sample, nonlinearity, high dimension and local minima points (Liang, 2008; Zheng, 2010). However, it is a critical issue that how to choose the penalty parameter C and the kernel function parameters G of SVM, choose different parameters has an important impact on the SVM machine learning performance (Pan and Luo, 2010; Chen and Mei, 2011; Shuaishi and Yantao, 2010). Traditional parameters choose to rely on long-term use of repeated experiments and personal experience methods. At present, some scholars have proposed a variety of parameter optimization method, such as Grid optimization algorithm, Genetic Algorithm (GA) and PSO algorithm (Li *et al.*, 2009; Gao *et al.*, 2010; Ming-Bao and Jia-Wei, 2009). Particle Swarm Optimization algorithm referred PSO, was proposed by Eberhart and Kennedy in 1995, is a population-based stochastic optimization techniques, this study use particle swarm algorithm optimization SVM parameter, the two key parameters C and G, greatly simplify the optimization process in the condition of improve SVM classifier accuracy and avoid the shortcoming that choice the optimal SVM parameters rely on experience, experimental contrast and a wide range of search (Wang and Chen, 2008; Lin, 2008). Finally, through examples show that the PSO-SVM method proposed have a high precision applied tobacco grading, make the tobacco quality evaluation method is more objective and also to provide a scientific method for the field of quality evaluation in cigarette industry.

In this study, we study the flue-cured tobacco classification model based on the PSO-SVM. Firstly we use the Gaussian Radial Basis Function (RBF) as the kernel function of SVM and then use the particle swarm optimization algorithm (PSO) to optimize the structural

Corresponding Author: Jiande Wu, Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China

parameters of the SVM classifier, established the flue-cured tobacco classification model based on the PSO-SVM. Collecting a wide range of tobacco data in Qujing Yunnan Province, to train and validate the model. At last, compared with the grid parameter optimization and genetic algorithm-based parameter optimization model, the results show that the proposed model based on particle swarm optimization with high prediction accuracy and better adaptability when used in tobacco grading.

THE BASIC METHODS

Principle of SVM: Assuming exist two types of linear data samples (x_i, y_i), $i = a,\ldots, n$, $x_i \in R_d$ $y_i \in \{+1, -1\}$. Linear discriminant function is $f(x) = w \cdot x+b$ and the corresponding classification surface equation is $f(x) = w \cdot x+b$, x is the input, w is adjustable weight vector and b is the bias. Normalized the function, make the two types of samples are satisfied, that is $|f(x)| \geq 1$, if all of the samples are correctly classified must satisfy:

$$y_i[(w \cdot x_i) + b] - 1 \geq 0, i = 1, \ldots, n \tag{1}$$

Now the classification interval equal to $2/\|w\|$, while the optimal classification surface can be expressed as the following constrained optimization problem, that is to say, solving the minimum of following formula (2) in the requirements of formula (1) :

$$\phi(w) = \frac{1}{2}\|w\|^2 \tag{2}$$

Define a Lagrange function (3):

$$L(w, b, \alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{n} \alpha_i [y_i (w \cdot x_i + b) - 1] \tag{3}$$

Seeking partial differential for w, b and α_i respectively and make them equal to zero, so the original problem can be transformed into the dual problem:

$$\begin{aligned} &\max \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \\ &s.t. \quad i \geq 0, i = 1, \ldots, n \\ &\sum_{i=1}^{n} \alpha_i y_i = 0 \end{aligned} \tag{4}$$

The parameter α_i is Lagrange multiplier in equation (4) and it existence a unique solution. Solve the above problems will get the optimal classification function:

$$f(x) = \operatorname{sgn}\{\sum_{i=1}^{n} \alpha_i^* y_i (x_i \cdot x) + b^*\} \tag{5}$$

Machine symbiotic intelligent system: Human can interact with robot each other and enjoy the robot's service. Even complex problems are solved through collaboration. This shows that intelligent robotic system is typical platform for studying complex system. Therefore, Metasynthesis are suitable for guiding the robot system design.

The parameter α^*_i is the optimal solution in formula (5), wherein the nonzero value is the support vector and the parameter b^* is classification threshold.

For the linear inseparable circumstances, introducing slack variables ξ and allowing the presence of misclassified samples, so the objective function is turn into:

$$\phi(w, \xi) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \xi_i \tag{6}$$

In formula (6), C is a positive number, called penalty factor. Similarly, the above equation can be solved by quadratic programming.

The basic idea of the non-linear support vector machine is that the input variables are mapped to a high-dimensional space through nonlinear transformation and then get the optimal hyperplane in this high-dimensional space. The transformation may be more complicated, but can be found in the above dual problem, the calculate process only relates to the dot product operation $<x_i, x_j>$between training samples. According to the functional theory, as long as a kernel function $K(x_i, x_j)$ satisfies the Mercer condition, it corresponds to the dot product of a particular space. If it satisfied the condition $0 \leq \alpha_i \leq C$, $\sum_{i=1}^{n} \alpha_i y_i = 0, i = 1, \ldots, n$ the objective function can be expressed as:

$$Q(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \tag{7}$$

And the classification function becomes:

$$f(x) = \operatorname{sgn}\{\sum_{i=1}^{n} \alpha_i^* y_i K(x_i, x_j) + b^*\} \tag{8}$$

Particle swarm optimization: The particle swarm algorithm is a global optimization technique that based on swarm intelligence and the basic concept stems from the research of the flock predatory behavior, each potential solution of the optimization problem is a bird in the search space, there are called "particles".

Firstly PSO algorithm is initialized a group of particles in the solution space, each particle. Figure 1 shows the PSO algorithm flow.

Represents a potentially optimal solution of the global optimization problem, with three indicators that the position, velocity and fitness value indicates the

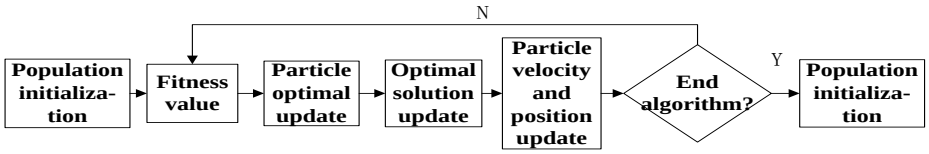

Fig. 1: The PSO algorithm flow

particle characteristics, the speed of each particle determine the direction and distance of their flying, the fitness value is calculated by the fitness function, its value on behalf of the particles merits. These particles movement in the solution space and update individual position by track the individual extremum PBest and groups extremum GBest, each updated will calculate the fitness value of the particle and by comparing the new particle's fitness value, the individual extremum and groups extremum fitness value in order to update the position of individual extremum pBest and groups extremum GBest.

EXPERIMENTAL ANALYSIS AND VERIFICATION

Experimental realization using MatlabR2011a Programming, tobacco samples comes from the 9 tobacco growing areas in Qujing County Fuyuan in 2010 as the source data of predictive model, including Huize, Luliang, Luoping, Malong, Shizong and Zhanyi County and Sherwin Qilin District, there are three levels of tobacco data of a total of 221 groups, including 75 groups of C3F, 74 groups of B2F and 69 groups X2F, each group of data include seven chemical composition indicators, there are total sugar, reducing sugar, total nitrogen, nicotine, potassium, soluble chlorine, volatile alkali and petroleum ether extract and seven sensory quality indicators respectively are aroma, aroma quantity, irritating, aftertaste, concentration, strength and aroma type. The model uses 163 groups samples as the training set, the rest of the 58 groups as the test samples set, in order to make the Matlab identify tobacco level data, each grade were given a different label, the C3F level given label 1, B2F grade given label 2 and X2F grade given label.

Due to the different dimensionless among the various indicators of the tobacco chemical composition and the chemical composition and sensory quality indicators, so the data need (0, 1) normalization processing before establish model, to reduce errors caused by different dimensions. The most commonly used Radial Basis Function (RBF) as the kernel function of support vector machines. Respectively use the Grid, GA and PSO parameter optimization method to optimize the SVM penalty parameter C and kernel function parameters G.

Figure 2 is the PSO parameter optimization algorithm fitness curve, it can be seen from the figure the PSO cognitive learning factor c1 and social learning factor c2 value are c1 = 1.5, c2 = 1.7, the termination algebraic value is 100, that is the maximum iterative optimization 100 times, the population quantity pop = 20, combined with the running program results of SVM toolbox in Matlab: Bestc = 2.1312, Bestg = 0.1,the average accuracy rate is CVAccuracy = 96.875% in the sense of the CV (cross-validation).

Based on 221 samples in the experiment, use the front of 163 samples as training set to train SVM classification model and use the back of 58 samples as testing set to test the accuracy of the SVM model. In order to better shown the experimental results in the figure, training set sample label (three levels) is in randomized order, while the label of the test set is in accordance with the collation sorting of the Level.

Training set SVM predict results contrast is shown in Fig. 3, the "asterisk" on behalf of the real category, the "circle" represents the predicted category, it can be seen the training set of model accuracy is 100%.

Figure 4 is GA algorithm parameter optimization predicted results. Similarly, the "asterisk" on behalf of the real category, the "circle" represents the predicted category, it can be seen from Figure the prediction accuracy is 91.3793%, there are 5 samples have error predict and respectively the first, second, twelfth, twenty-fifth and fifty-fourth, the first, second and twelfth samples from the C3F level (Category 1) mistaken for B2F level (category 2); the twenty-fifth samples from the C3F level (Category 1) mistaken for X2F level (Category 3); the fifty-fourth samples from the X2F level (Category 3) mistaken for B2F level (category 2).

Figure 5 is Grid algorithm parameter optimization predicted results, it can be seen from Figure the prediction accuracy is 94.8276%, there are respectively the first, seventh, twenty-fifth have errorpredict, the first and seventh from the C3F level (Category 1) mistaken for B2F level (category 2); the twenty-fifth samples from the C3F level (Category 1) mistaken for X2F level (Category 3).

PSO algorithm parameter optimization predicted results shown in Fig. 6, the prediction accuracy is 96.5517%, only the first and fifty-fourth have error predict, the first from the C3F level (Category 1) mistaken for B2F level (category 2); the fifty-fourth from the X2F level (Category 3) mistaken for B2F level (category 2).

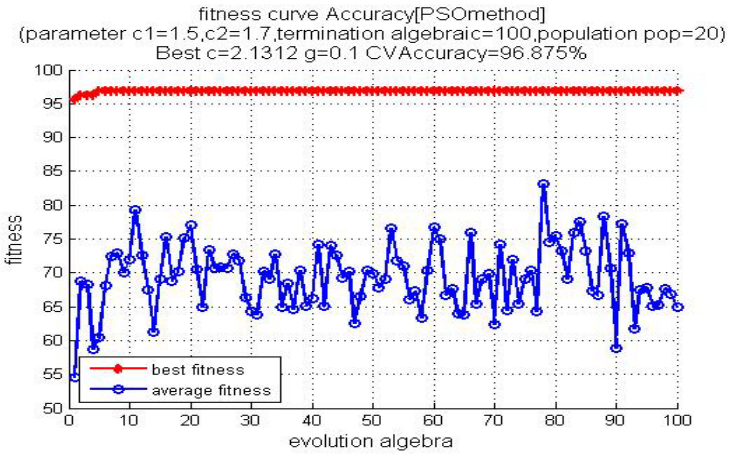

Fig. 2: The PSO fitness curve

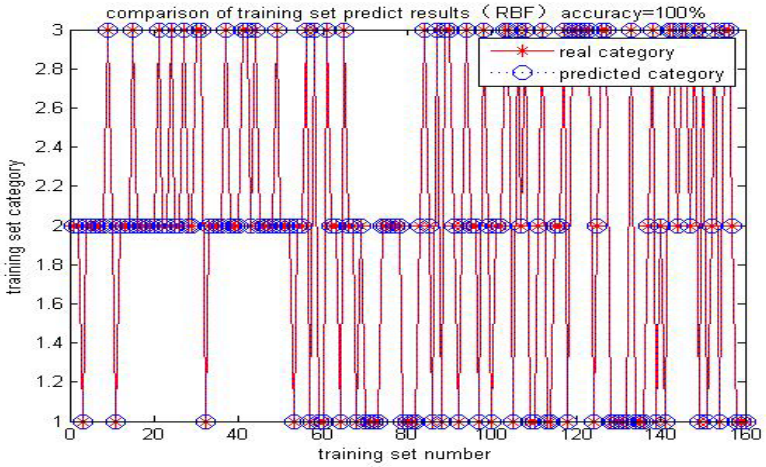

Fig. 3: Comparison of training set predict results

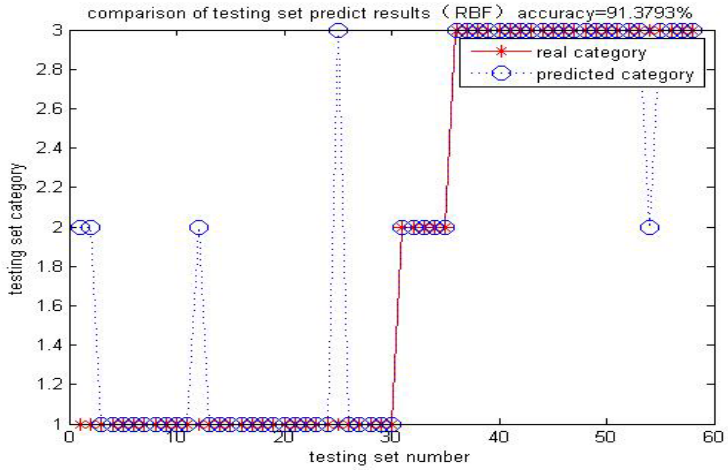

Fig. 4: Predict results of GA parameter optimization

The predict the results of three parameter optimization method as shown in Table 1, it can clearly see that the PSO method have the highest predict accuracy, From the classification accuracy angle, PSO-

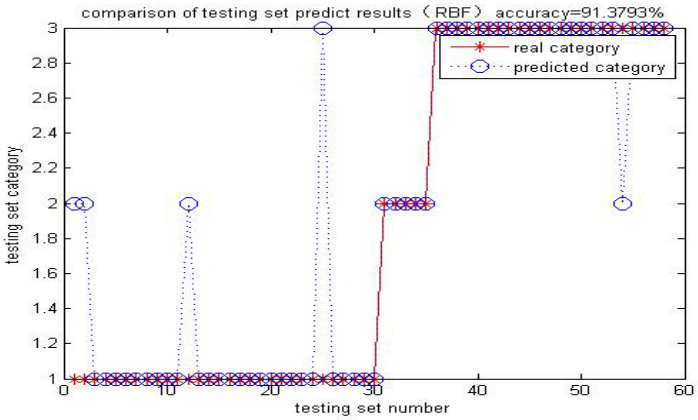

Fig. 5: Predict results of Grid parameter optimization

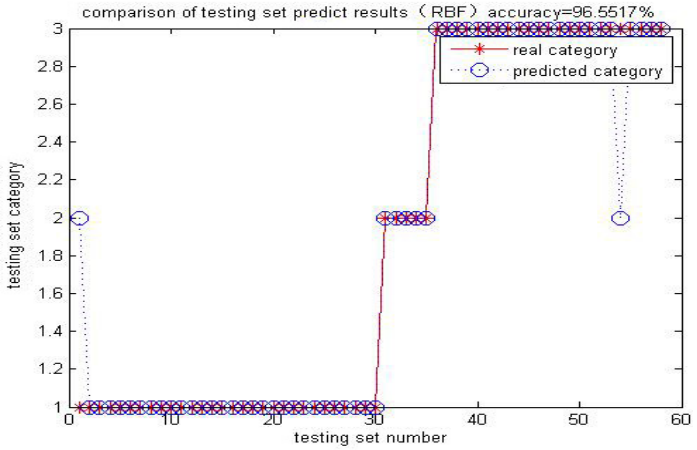

Fig. 6: Predict results of PSO parameter optimization

Table 1: Predict results of three parameter optimization method

Parameter optimization method	Number of samples	Error classification samples	Classification accuracy rate (%)
Ga	58	5	91.3793
Grid	58	3	94.8276
Pso	58	2	96.5517

SVM model is more applicable to the tobacco classification.

CONCLUSION

This studies establish the tobacco classification model by tobacco chemical composition and sensory quality and combined with SVM, using the Gaussian radial machine function as SVM kernel function and the global search characteristics of PSO to optimize the parameters of SVM, overcome the blindness of SVM parameters select. Form three levels of tobacco samples establish the modeling and compared the modeling results with the optimization method of GA, Grid and PSO, the results show that the proposed model based on particle group parameter optimization has higher predict accuracy and better adaptability. This method can better identify the different tobacco grades, the classifier model have a short training time and the classification with a high accuracy, this way can ensure the objectivity of the tobacco leaves classification, at the same time, it can provide technical support for further analysis and research of tobacco.

ACKNOWLEDGMENT

This study is supported by the foundation items: Project supported by Natural Science Foundation of China (Grant No. 51169007), Science and Research Program of Yunnan province (No. 2010DH004 and No.2011CI017 and No.2011DA005).

REFERENCES

Chen, W. and Y. Mei, 2011. Predict large forging billet internal cavity forging critical reduction by PSO and SVM. Comput. Eng. Appl., 27(47): 243-245.

Gao, J., J.Y. Peng and Z. Li, 2010. Application of improved PSO-SVM approach in image classification. Proceeding of Symposium on Photonics and Optoelectronic (SOPO), Chengdu, pp: 1-4.

Liang, Y., 2008. The Expand and Application Research of SVM Classifier. Hunan University, pp: 17-28.

Li, J., Y. Xiang and Y. Lu, 2009. Summary of particle clustering algorithm. Appl. Res. Comput., 12(26): 4423-4427.

Lin, H., 2008. Application of data mining technology in cigarette recipe and optimization. Ocean University of China, pp: 44-54.

Ming-Bao, L. and Z. Jia-Wei, 2009. SVM optimized scheme based PSO in application of engineering industry process. Proceedings of the 8th International Conference on Machine Learning and Cybernetics. Baoding, 3: 1246-1251.

Pan, L. and Y. Luo, 2010. Parameters selection of support vector machine using an improved PSO algorithm. Proceeding of the 2nd International Conference on Intelligent Human-Machine Systems and Cybernetics, Nanjing, Jiangsu, 2: 196-199.

Shuaishi, L. and T. Yantao, 2010. Facial expression recognition approach based on least squares support vector machine with improved particle swarm optimization algorithm. Proceedings of the IEEE International Conference on Robotics and Biomimetics, Tianjin, pp: 399 -404.

Vn, V., 1995. The Nature of Statistical Learning Theory. Springer-Verlag, New York.

Wang, H. and S. Chen, 2008. Soft measurement of oxygen content in flue gas based on least squares support vector machine and PS0 algorithm. Thermal power generation, 37(3): 35-38.

Wang, Q., C. Ying-Wu and M. Li, 2007. Application of one class support vector machine in tobacco selection. J. Comput. Appl., 2(27): 482-485.

Yan, K. and X. Zhao, 2003. Classification of Tobacco Leaves. China Agriculture Press, Beijing, pp: 51-62.

Yan, R., L. Han and J. Chen, 2001. Application of computer technology in the field of tobacco detection and classification. Tob. Sci. Technol., 3(13): 13-15.

Zheng, H., 2010. The Support Vector Machine Method Investigate. Northwestern University, pp: 10-16.