Available online at www.sciencedirect.com
SCIENCE DIRECT
Journal of Pharmaceutical and Biomedical Analysis 41 (2006) 1198–1204
JOURNAL OF PHARMACEUTICAL AND BIOMEDICAL ANALYSIS
www.elsevier.com/locate/jpba

Qualitative identification of tea categories by near infrared spectroscopy and support vector machine

Jiewen Zhao [a,*], Quansheng Chen [a,*], Xingyi Huang [a], C.H. Fang [b]

[a] *Department of Food Engineering, School of Biological and Environmental Engineering, Jiangsu University, 212013 Zhenjiang, PR China*

[b] *Laboratoire de Mecanique et Genie Civil, Universite Montpellier 2, 34095 Montpellier, France*

Received 21 November 2005; received in revised form 27 February 2006; accepted 28 February 2006
Available online 18 April 2006

Abstract

Near-infrared (NIR) spectroscopy has been successfully utilized for the rapid identification of green, black and Oolong tea. The spectral features of each tea category are reasonably differentiated in the NIR region, and the spectral differences provided enough qualitative spectral information for the identification of tea. Support vector machine (SVM) as the pattern recognition was applied to identify three tea categories in this study. The top five principal components (PCs) were extracted as the input of SVM classifiers by principal component analysis (PCA). The RBF SVM classifiers and the polynomial SVM classifiers were studied comparatively in this experiment. The best experimental results were obtained using the radial basis function (RBF) SVM classifier with $\sigma = 0.5$. The accuracies of identification were all more than 90% for three tea categories. Finally, compared with the back propagation artificial neural network (BP-ANN) approach, SVM algorithm showed its excellent generalization for identification results. The overall results show that NIR spectroscopy combined with SVM can be efficiently utilized for rapid and simple identification of the tea categories.

Keywords: NIR spectroscopy; Support vector machine; Tea; Identification

1. Introduction

Tea is one of the most popular beverages worldwide, which is of great interest due to its beneficial medicinal properties [1–6]. There are many different tea categories throughout the world, which are usually classified as green tea, black tea, Oolong tea, yellow tea, white tea and dark compressed tea. Green tea, black tea and Oolong tea are among the most popular categories across the world. Drying after roasting the leaves produce green tea, and with black tea, leaves are additionally fermented. If this fermentation is partially carried out, an intermediate kind of tea is obtained—the Oolong tea. In the fermentation, the enzymatic oxidation of tea polyphenols takes place leading to the formation of chemical compounds. Therefore, the different tea categories possess the different chemical and medical properties.

Nowadays, the identification of a tea category is performed according to various wet chemical methods such as high-performance liquid chromatography (HPLC) [7,8], gas chromatographic (GC) [9], capillary electrophoresis [10], plasma atomic emission spectrometry [11], etc. However, all of the methods mentioned above are time-consuming in identification of tea.

Near-infrared (NIR) spectroscopy has been proved to be a powerful analytical tool. It has been applied widely in the agricultural, nutritional, petrochemical, textile and pharmaceutical industries, especially the application of NIR spectroscopy for the qualitative and quantitative analysis of pharmaceutical samples has been significantly increased during the last decade [12–16]. It is known today that many studies have applied near-infrared reflectance spectroscopy to tea. Since the 1990s, attempts have been made to simultaneously predict some chemical compositions in green tea leaves using NIR spectroscopy technique [17,18]. Studies on the application of NIR spectroscopy to quantitative analysis of total antioxidant capacity in green tea are also reported recently [19,20]. However, most of these were

* Corresponding authors. Tel.: +86 511 8780308; fax: +86 511 8780201.
E-mail addresses: zhao@ujs.edu.cn (J. Zhao), q.s.chen@hotmail.com (Q. Chen).

0731-7085/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.jpba.2006.02.053

quantitative analyses, and few studies report applications of NIR spectroscopy to qualitative identification of tea categories.

As a new pattern recognition method, support vector machine (SVM) is incomparable to others in chemometrics, which has a good theoretical foundation in statistical learning theory. SVM fixes the classification decision on structural risk minimization (SRM) instead of the traditional empirical risk minimization (ERM). Therefore, the model by training avoid over-fitting problem [21]. It performs binary classification problem by finding a hyperplane with maximal margin in terms of a subset of the input data (support vectors). If the input data are not linearly separable, SVM firstly maps these data into a high-dimensional feature space to transform a linearly separable problem, and then classifies these data by hyperplane. Moreover, SVM is capable of learning in high-dimensional feature space with fewer training data. Recently, SVM has been successfully applied to NIR spectroscopy non-linear prediction model [22,23], but few studies have been reported in the qualitative analysis using NIR spectroscopy technique combined with SVM pattern recognition method. Therefore, NIR spectroscopy combined with SVM pattern recognition method was proposed to identify rapidly and simply tea categories in this study.

2. Materials and methods

2.1. Sample preparation

All tea samples of three categories came from different provinces in China. Tea categories, origins and the numbers of samples are shown in Table 1. Each tea category has some different brands, which come from different provinces in China.

All tea materials were already in stock within a 4 months period. Taking into consideration the heterogeneity of tea samples, major attention was paid to the sampling stage. The samples were ground before analysis. For the grinding, the whole tea leaves were put into a small electric coffee mill and ground for 10 s. After this procedure, the powders were sieved with a mesh width of 500 μm and these sieved powders were used for the subsequent analyses.

2.2. Spectra collection

The NIR spectra were collected in the reflectance mode using a *NEXUS 670 FT-IR* spectrophotometer (Nicollet, USA) with an optical fiber. Each spectrum was the average spectrum of 64 scans. The spectral used for the data analysis covered the range from 11000 cm^{-1} to 3800 cm^{-1}, and the data were measured in 1.928 cm^{-1} intervals, which resulted in 3735 variables.

Table 1
Categories, numbers and origin of tea samples

Tea categories	Samples numbers	Tea origins
Green tea	50	Zhejiang, Anhui, Jiangsu, Fujian, Henan
Black tea	50	Jiangsu, Sichuan, Anhui, Jiangxi, Yunnan
Oolong tea	50	Anhui, Fujian

The standard quartz cup was used for performing the tea spectra collection. For each tea sample, respectively, 10 ± 0.1 g of tea powder was filled into the quartz cup in the standard procedure depending upon the bulk density of materials. The corresponding amount of powder was densely packed into the quartz cup and then compressed by closing it. Each tea sample was collected three times after rotating the cup at 120°. The mean of the three spectra which were collected from the same tea sample was used in the following analysis step. The temperature was kept around 25 °C and the humidity was kept at a steady level in the laboratory.

2.3. Preprocessing methods

In this study, three data preprocessing method were applied comparatively; these were standard normal transformation (SNV), first derivative, second derivative, etc. SNV is a mathematical transformation method of the log (1/*R*) spectra used to remove slope variation and to correct for scatter effects. Compared to SNV, first and second derivatives eliminate baseline drifts and small spectral differences are enhanced. To avoid enhancing the noise, which is a consequence of derivative, spectra are first smoothed. This smoothing is done by using the Savitzky–Golay algorithm [24], which is a moving window averaging method: a window is selected where the data are fitted by a polynomial of a certain degree. The central point in the window is replaced by the value of the polynomial.

2.4. Basic principle of SVM

SVM is a new generation of learning systems based on statistical learning theory as proposed by Vapnik and Chervonenkis [25,26]. Here, a brief introduction of SVM is presented, and readers can refer to the tutorials on SVM [27–29] for details.

The basic SVM deals with two-class problems, in which the data are separated by a hyperplane defined by a number of support vectors. The SVM can be considered to create a hyperplane between two sets of data for classification. In case of two-dimensional situation, the action of the SVM can be explained easily following as Fig. 1. A series of points for two different

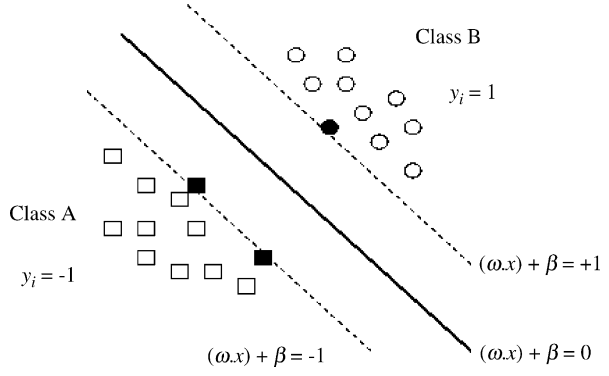

Fig. 1. Classification of data by SVM. The solid line and dashed line denote the hyperplane and margins, respectively. Squares and circles denote the negative and the positive training samples, respectively. The grey ones in margins denote the support vectors.

classes of data are shown, squares (class A) and circles (class B). The SVM tries to set an appropriate boundary so that the distance between the boundary and the nearest data point is maximal. The boundary is then placed in the middle of this margin. The nearest data points that are used to define the margins are known as support vectors (SVs, represented by grey circles and squares). Once the support vectors are selected, the rest of the feature set can be discarded, because SVs contain all the necessary information for the classifier.

Considering a two-class classification problem with k labeled training samples, it is represented by $\{(x_i,y_i)|i=1,\ 2,\ \ldots,\ k\}$ where $x \in R^n$ is a n-dimensional vector and $y \in \{-1,+1\}$ is the class label. The boundary can be expressed as follows:

$$\{(x,\ y)|y = (\omega \cdot x) + \beta,\ \omega \in R^n,\ \beta \in R\} \tag{1}$$

where the vector ω defines the boundary and β is a scalar threshold. At the margins, where SVs are located; the equations for classes A and B, respectively, are as follows:

$$(\omega \cdot x) + \beta = -1 \tag{2}$$

$$(\omega \cdot x) + \beta = +1 \tag{3}$$

As SVs correspond to the extremities of the data for a given class, the following decision function can be used to classify any data point in either class A or B:

$$f(x) = \text{sign}((\omega \cdot x) + \beta) \tag{4}$$

The aim is to find a hyperplane that can be used to classify these data points between classes A and B. The optimal hyperplane separating the data can be obtained as a solution to the following optimization problem:

Minimize

$$\tau(\omega) = \min_{\omega,\beta} \left\{ \frac{1}{2}\|\omega\|^2 \right\} \tag{5}$$

subject to

$$y_i((\omega \cdot x_i) + \beta) \geq 1. \tag{6}$$

In case there is an overlap between the two classes, a slack variable $\xi_i, i = 1, \ldots, k$ can be introduced. The optimization problem changes as follows:

Minimize

$$\tau(w) = \min_{\omega,\beta,\xi_1 \ldots \xi_k} \left[\frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{k} \xi_i \right] \tag{7}$$

subject to

$$y_i(\omega \cdot x_i + \beta) - 1 + \xi_i \geq 0 \tag{8}$$

where the parameter C is a penalty coefficient, and it determines the tradeoff between minimizing the training error and minimizing model complexity. The solution of the constrained optimization problem can be obtained as follows:

$$\omega = \sum_i y_i \alpha_i x_i; \quad \alpha_i \geq 0 \tag{9}$$

where α_i is called Lagrange multipliers and x_i is a support vector obtained from training. Putting (9) in (4), the decision function is obtained as follows:

$$f(x) = \text{sign}\left\{ \sum_{i,j} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) + \beta \right\} \tag{10}$$

In cases where the linear boundary in input spaces will not be enough to separate two classes properly, it is possible to create a hyperplane that allows linear separation in the higher dimension (corresponding to curved surface in lower-dimensional input space). In SVM, this is achieved through the use of a transformation function $\Phi(x)$ that converts the data from an n-dimensional input space to ε-dimensional feature space:

$$s = \Phi(x) \tag{11}$$

where $x \in R^n$ and $s \in R^\varepsilon$. Fig. 2 shows the transformation from the input space to the feature space where the non-linear boundary has been transformed into a linear boundary in feature space. The transformation into higher-dimensional feature space is relatively computation-intensive. A kernel can be used to perform this transformation and the dot product in a single step provided the transformation can be replaced by an equivalent kernel function. This helps in reducing the computational load and at the same time retaining the effect of higher-dimensional transformation. The kernel function $K(x_i,x_j)$ is defined as follows:

$$K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j). \tag{12}$$

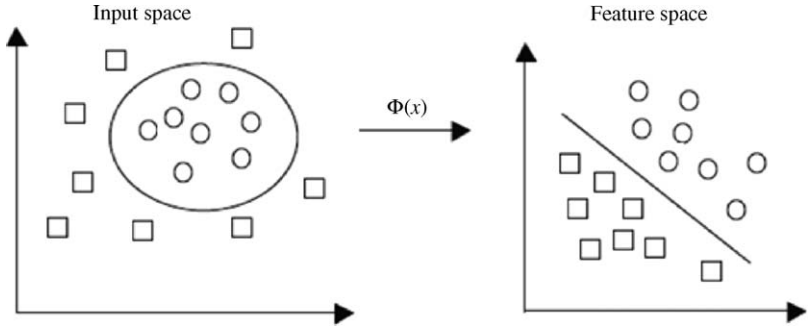

Fig. 2. Non-linear separation of input and feature space. Squares and circles denote the negative and the positive training samples, respectively. These points are non-linear separation in input space, and linear separation in feature space.

Nowadays, the popular kernel functions are the radial basic function (RBF), polynomial, sigmoid kernel function, etc., as follow:

$$\text{RBF kernel function}: \ K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right); \quad (13)$$

$$\text{polynomial kernel function}: \ K(x_i, x_j) = (1 + x_i \cdot x_j)^{\sigma}; \quad (14)$$

$$\text{sigmoid kernel function}: K(x_i, x_j) = \tanh(\sigma(x_i \cdot x_j) + \upsilon). \quad (15)$$

Finally, the basic form of SVM is accordingly obtained after substituting (12) in the decision function (10) as follows:

$$Y = \text{sign}\left\{\sum_{i,j} \alpha_i \alpha_j y_i y_j K(x_i, x_j) + \beta\right\}$$

2.5. Software

All SVM algorithms were implemented with Matlab V6.5 (Mathworks, USA) under Windows XP. The implementation of Multi-class SVM [30] algorithm was used for the classification of tea in all experiments. For the spectral acquisition, ***OMNIC 5.2a*** (***NEXUS 670 FT-IR*** Systems) was used.

3. Results and discussion

3.1. Spectra investigation

Fig. 3(a) shows the mean spectra of each class for the original data. The spectra of each class after first derivative preprocessing are presented in Fig. 3(b). As seen from Fig. 3(a) and (b), there are water absorption bands around 5155 cm^{-1} and 7000 cm^{-1} corresponding to O–H stretching + O–H deformation. These were excluded during analysis along with some regions exhibiting a high noise level (e.g. 11000 − 9000 cm^{-1}; Fig. 3(b)).

Also seen from Fig. 3(b), the most intensive band in the spectrum belonged to the vibration of the second overtone of the carbonyl group (5352 cm^{-1}), followed by the C–H stretch and C–H deformation vibration (7212 cm^{-1}), the $–CH_2$ (5742 cm^{-1}), and the $–CH_3$ overtone (5808 cm^{-1}). The vibration of the carbonyl group, the –C–H and $–CH_2$ vibrations are caused by ingredients such as polyphenols, alkaloids, protein, volatile and non-volatile acid and some aroma compounds.

In general, the water content in the dry tea leaves is up to 4–6% (w/w), therefore, the effect of water must be considered. To keep away from the water absorption band, the spectral regions between 5300 cm^{-1} and 6500 cm^{-1} were selected, because there is a great deal of information from organic substances in this NIR spectroscopy region according to the spectra investigation.

3.2. Principal components analysis (PCA)

All NIR spectral data from three tea categories were used for the PCA. Although PCA itself cannot be used as a classification tool, this behavior may indicate the data trend in visualizing dimension spaces. For visualizing the data trends and the discriminating efficiency of the three NIR spectral preprocessing methods, the scatter plots of data using the top three principal components (PCs) issued from PCA were obtained as showed in Fig. 4(a–c). As can be seen, SNV preprocessing method is superior to the other preprocessing methods. SNV preprocessing method removed physical spectral information (due to particle size), so that PCA was performed based on mainly chemical spectral information. On the other hand, SNV decreased the within-class variance. Fig. 5(a) and (b) show the spectra of Oolong tea without and with SNV preprocessing method, respectively. In the raw spectra, a small offset can be observed in some spectral regions such as the lower wavenumbers regions. These phenomena typically occur in powdered materials due to multiplicative effects of scatter and particle size. They are often corrected by SNV preprocessing method over the entire spectral range. Therefore, we selected the SNV preprocessing method in this study.

As shown in Fig. 4(a), there was clear cluster trend these data in the three-dimension (3D) principal component space represented by the top three principal components (i.e. PCs 1–3) vectors. Such good classification in this 3D space could be explained by the chemical background of tea and PCA methods. The different tea categories can exhibit considerable differences in their botanical, genetic, agronomical, characteristics, however, more connected with the different tea process. The differences detected in chemical compositions of the different tea categories can be reasonably differentiated in the NIR spectroscopy region. Therefore, in the NIR spectroscopy region, these spectral differences provided enough information

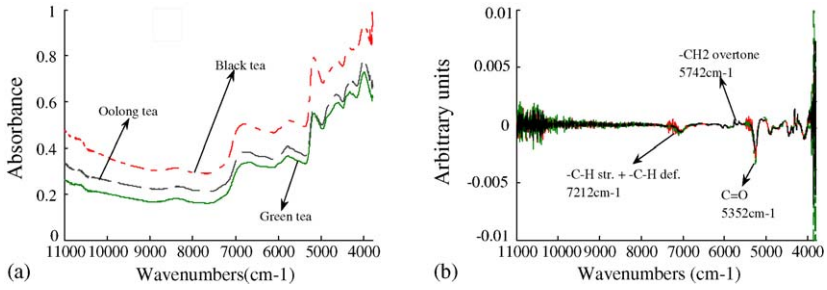

Fig. 3. Mean spectra for the three categories of tea obtained from (a) raw data and (b) first derivative date.

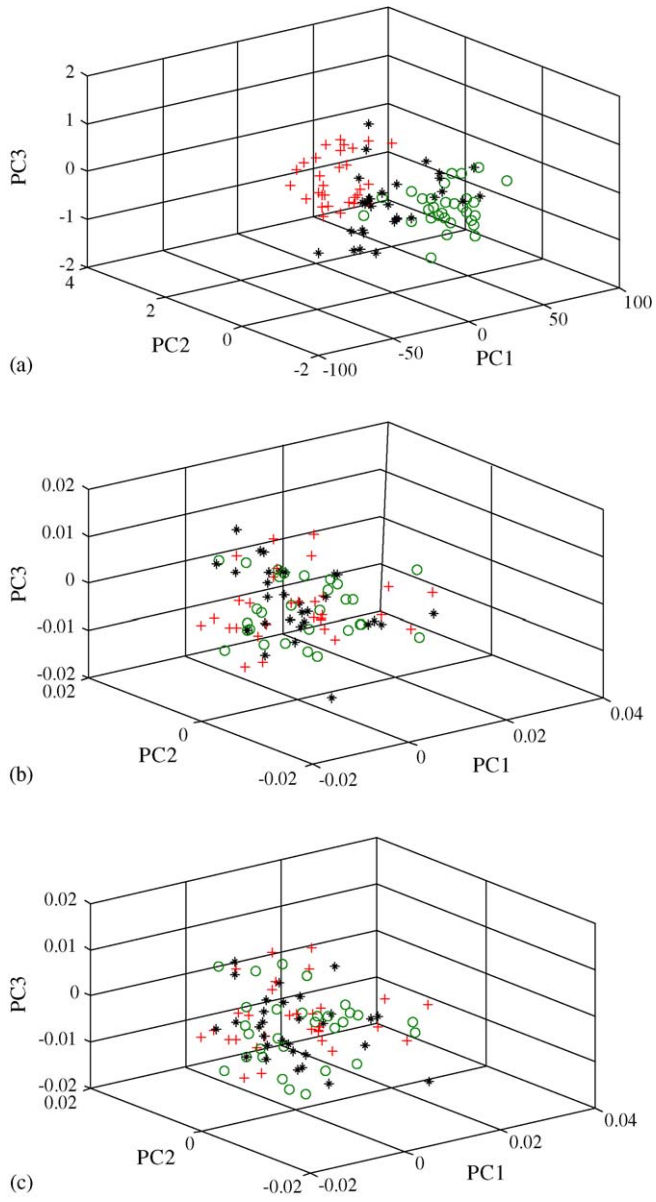

Fig. 4. Score cluster plot with top three principal components (PCs) for training set samples of green tea (◯), black tea (+) and Oolong tea (*). Score cluster obtained from (a) SNV data, (b) first derivative data and (c) second derivative data.

for further qualitative analysis. In additional, through PCA, the accumulated variance contribution rate was up to 94.7% for the top three PCs, in other words, the 3D space represented by PCs 1–3 vectors can explain 94.7% chemical composition information in the NIR spectroscopy region. Thus, the 3D space can almost express fully the information that all data are distributed in an ultra-dimensional space.

Investigated from Fig. 4(a), green tea class is first close to Oolong tea class in the 3D space, and next to black tea class. Such phenomena can be explained by the inner chemical compositions from tea. Although they come from the leaves of the plant Camellia sinensis, however, the process that the leaves undergo to make the final dry tea is different. The leaves for black tea are fully fermented; however, those for green tea are lightly steamed before being dried. Therefore, these chemical compositions differ in their chemical structure. For example, green tea leaves contain more of the simple flavonoids called catechins, while the oxidization that the leaves undergo to make black tea converts these simple flavonoids to the more complex chemical compositions called theaflavin and thearubigins. Oolong tea is a partially fermented leaf, with the flavonoids profile midway between green and black tea, therefore, its quality also represents a middle trait between green and black tea.

3.3. Support vector machine classification

In this study, green tea, black tea and Oolong tea would be classified through SVM, therefore, it is a problem of multi-class classification. SVM is usually solved by a decomposing and reconstruction procedure when two-class decision machines are implied. In the standard decomposing scheme of a multi-classification problem into dichotomies [30], SVM is trained over all the training patterns with the **1-v-r** (one versus the rest) method. SVM assign label +1 to the samples in the *i*th class, and label −1 to all the other samples.

In this experiment, three tea categories in all 150 tea samples were tried. In order to come to a 3/2 division of training/test data, 90 data (i.e. 30 green tea samples, 30 black tea samples and 30 Oolong tea samples) were selected in the training set, and the remaining 60 data (i.e. 20 green tea samples, 20 black tea samples and 20 Oolong tea samples) were selected in test set.

Firstly, the top five principal components (i.e. PCs 1–5) vectors were extracted by PCA. These vectors were input to the

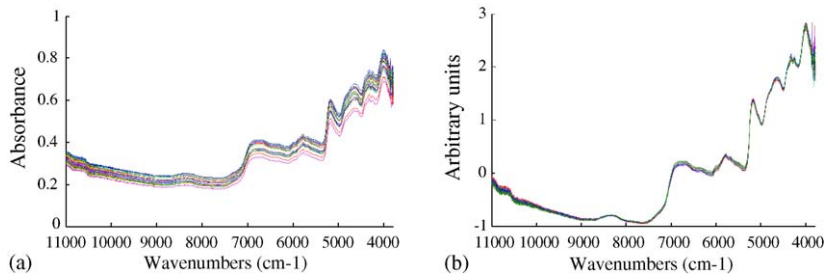

Fig. 5. Spectra for Oolong tea obtained from (a) raw data and (b) SNV data.

Table 2
The identification results with RBF SVM classifiers for training

Tea categories	Sample numbers	Classification results of RBF SVM classifiers in different parameters (σ)							
		0.2 (%)	0.5 (%)	0.8 (%)	1.0 (%)	1.2 (%)	1.5 (%)	2.0 (%)	2.5 (%)
Green tea	30	83.33	96.67	96.67	96.67	96.67	96.67	96.67	96.67
Black tea	30	96.7	100	96.67	96.67	96.67	93.3	90	86.67
Oolong tea	30	70	93.33	86.7	86.7	86.67	80	70	70

Table 3
The identification results with polynomial classifiers for training

Tea categories	Sample numbers	Identification results of polynomial classifiers in different the parameter (σ)			
		1 (%)	2 (%)	3 (%)	4 (%)
Green tea	30	80	93.33	93.33	96.67
Black tea	30	83.33	100	100	100
Oolong tea	30	66.67	93.33	93.33	90

SVM classifiers as latent variables. Through PCA, the accumulated variance contribution rate was up to 99.2% for the top five PCs, in other words, PCs 1–5 could load 99.2% of the whole spectral information; moreover, much repetitious spectral information was also removed. In fact, the experimental results were proved to the best when PCs 1–5 vectors were used.

Next, to obtain a good performance, the type of kernel function and some parameters in SVM have to be chosen carefully. We only focus on the polynomial kernel function and the RBF kernel function in this paper, because they were applied widely and their theories systems are also more mature than other kernel functions [26,31,32].

These parameters include:

(1) The kernel function parameter σ in Eqs. (13)–(15): an inappropriate choice of parameter σ may eliminate any biased performance of the SVMs, therefore, the parameter σ was focused on as an important point investigated.
(2) The regularization parameter C in Eq. (14): it determines the tradeoff between minimizing the training error and minimizing model complexity. As a penalty parameter, its value is often determined by experiment. The parameter C is set as a default value in case its effect on the classification result is few. In fact, we attempted some different C values in experiment; its effect on classification result is actually very few. In additional, optimizing simultaneously the two (C,σ) is more time-consuming for SVM approaches. Therefore, it was set as the default value 100 in this paper, and only the parameter (σ) was optimized in the experiment.

In the experiment, a range of parameters (σ) for RBF and the polynomial SVM classifiers were selected: $\sigma \in \{0.2, 0.5, 0.8, 1.0, 1.2, 1.5, 2.0, 2.5\}$ for the RBF SVM classifiers; $\sigma \in \{1, 2, 3, 4\}$ for the polynomial SVM classifiers.

Table 2 shows the identification results of the RBF SVM with these different parameters (σ). The RBF kernel with $\sigma = 0.5$ resulted in the best results. The best identification accuracies are 96.67%, 100% and 93.33% for green tea, black tea and Oolong tea, respectively.

Table 3 shows the identification results of the polynomial SVM with these different parameters (σ). The polynomial SVM classifier with parameter $\sigma = 2$ achieves the best results, and they are 93.33%, 100% and 93.33%. In fact, when the parameter σ is more than 2, the effect on the identification results is very few; however, the effect on the run time for the SVM algorithm is very severe. With increasing in the value of the polynomial degree, run time increases sharply. In additional, the polynomial SVM classifier with the parameter $\sigma = 1$ is a linear SVM classifier, which performs worse than any other classifiers with only 80%, 83.33% and 66.67%.

According to Tables 2 and 3, the best parameters were selected in the test experiments, which followed as: $\sigma = 0.5$ for RBF SVM classifier and $\sigma = 2$ for polynomial SVM classifier. Under these best parameters, the identification results are listed in Table 4. Seen from the whole, the RBF SVM classifier is a little better than the polynomial SVM classifier. The best prediction results are 95%, 100% and 90%, respectively.

To highlight the good performance with the generalization in the SVM algorithm, we attempted to compare the results by SVM with by back propagation artificial neural network (BP-ANN) approaches in this study. Just as the SVM approach, PCs 1–5 vectors were also input to the BP-ANN in the experiment. Table 5 shows the different identification results by SVM and

Table 4
The identification results with polynomial classifiers for test

Tea categories	Sample numbers	Identification results in the two classifiers	
		RBF SVM classifier (%)	polynomial SVM classifier (%)
Green tea	20	95	95
Black tea	20	100	100
Oolong tea	20	90	85

Table 5
Comparison between the SVM and BP-ANN approaches for train and test set

	Identification results in training set		Identification results in test set	
	SVM (%)	BP-ANN (%)	SVM (%)	BP-ANN (%)
Green tea	96.67	96.67	95	75
Black tea	100	100	100	100
Oolong tea	93.33	96.67	90	80

BP-ANN approaches in training and test set. Seen from the Table 5, the identification results by SVM are almost the same as by BP-ANN approach in training set; however, the results by SVM are obviously better than by BP-ANN approach in test set. The reasons that come to such phenomena might be explained by their basic theories of algorithm.

Traditional neural network approaches including BP-ANN are based on the empirical risk minimization (ERM) principle. They suffer difficulties with generalization, producing models that can over-fit the data. The 'best' mode by training often results in worse predictive result, in other words, the generalization of the model is worse [32].

The foundation of SVM embodies the structural risk minimization (SRM) principle, which has shown to be superior to the ERM principle. SRM minimizes an upper bound on the expected risk, as opposed to ERM that minimizes the error on the training data [21,32]. Therefore, SVM embodies the excellent generalization in its theory, which results in the better results than BP-ANN approach in the experiment.

4. Conclusion

It can be concluded that the NIR spectroscopy technique based on support vector machine has high potential to identify the tea category in a nondestructive way. Three tea categories (i.e. green tea, black tea and Oolong tea) were applied in the experiment. The best results were obtained using the RBF SVM classifier with $\sigma = 0.5$. They are up to 96.67%, 100% and 93.33% in training set; 95%, 100% and 90% in test set. Finally, compared with the BP-ANN approach, SVM algorithm shows its excellent generalization for the identification results.

Acknowledgements

This work has been financially supported by the National High Technology Research and Development Program of China (The "863" Project, No. 2002AA248051) and the National Natural Science Foundation of China (No. 30370813).

References

[1] C.S. Yang, P. Maliakal, X. Meng, Annu. Rev. Pharmacol. Toxicol. 42 (2002) 25–54.
[2] K. Nakachi, S. Matsuyama, S. Miyake, M. Suganuma, K. Imai, Biofactors 13 (2000) 49–54.
[3] V.W. Setiawan, Z.F. Zhang, G.P. Yu, Q.Y. Lu, et al., Int. J. Cancer 92 (2001) 600–604.
[4] K. Shibata, M. Moriyama, T. Fukushima, A. Kaetsu, M. Miyazaki, H. Une, J. Epidemiol. 10 (2000) 310–316.
[5] L. Jian, L.P. Xie, A.H. Lee, C.W. Binns, Int. J. Cancer 108 (2004) 130–135.
[6] H. Fujiki, M. Suganuma, S. Okabe, E. Sueoka, N. Sueoka, N. Fujimoto, Y. Goto, S. Matsuyama, K. Imai, K. Nakachi, Mutat. Res. 480–481 (2001) 299–304.
[7] P. Valera, F. Pablo, A.G. Gonzalez, Talanta 43 (1996) 415–419.
[8] Y.G. Zuo, H. Chen, Y.W. Deng, Talanta 57 (2002) 307–316.
[9] N. Togari, A. Kobayashi, T. Aishima, Food Res. Int. 28 (1995) 495–502.
[10] H. Horie, T. Mukai, K. Kohata, J. Chromatogr. A 758 (1997) 332–335.
[11] M. Angeles Herrador, A. Gustavo Gonzalez, Talanta 53 (2001) 1249–1257.
[12] K. Kachrimanis, V. Karamyan, S. Malamataris, Int. J. Pharm. 250 (2003) 13–23.
[13] Y.A. Woo, H.J. Kim, K.R. Ze, H. Chung, J. Pharm. Biomed. Anal. 36 (2005) 955–959.
[14] Y. Dou, Y. Sun, Y.Q. Ren, P. Ju, Y.L. Ren, J. Pharm. Biomed. Anal. 37 (2005) 543–549.
[15] R. De Maesschalck, T.V. Kerkhof, J. Pharm. Biomed. Anal. 37 (2005) 109–114.
[16] Y.A. Woo, H.R. Lim, H.J. Kim, H. Chung, J. Pharm. Biomed. Anal. 33 (2003) 1049–1057.
[17] M.N. Hall, A. Robertson, C.N.G. Scotter, Food Chem. 27 (1988) 61–75.
[18] H. Schulz, U.H. Engelhardt, A. Wengent, H.H. Drews, S. Lapczynski, J. Agric. Food Chem. 475 (1999) 5064–5067.
[19] J. Luypaert, M.H. Zhang, D.L. Massart, Anal. Chem. Acta 487 (2003) 303–312.
[20] M.H. Zhang, J. Luypaert, Q.S. Xu, D.L. Massart, Talanta 62 (2004) 25–35.
[21] H.C. Kim, S. Pang, H.M. Je, D. Kim, S.Y. Bang, Pattern Recogni. 36 (2003) 2757–2767.
[22] F. Chauchard, R.P. Cogdill, S. Roussel, J.M. Roger, V. Bellon-Maurel, Chemom. Intell. Lab. Syst. 71 (2004) 141–150.
[23] U. Thissena, M. Pepersb, B.U. Stuna, W.J. Melssena, L.M.C. Buydensa, Chemom. Intell. Lab. Syst. 73 (2004) 169–179.
[24] A. Savitzky, M.J.E. Golay, Anal. Chem. 36 (1964) 1627–1639.
[25] V.N. Vapnik, A.Y. Chervonenkis, Theory Prob. Appl. 17 (1971) 264–280.
[26] V.N. Vapnik, The Nature of Statistical Learning Theory, Springer-Verlag, New York, 1995.
[27] C.J.C. Burges, Data Mining Knowl. Discov. 2 (1998) 955–974.
[28] B. Scholkopf, IEEE Intell. Syst. 13 (1998) 18–19.
[29] S.R. Gunn, Support Vector Machines for Classification and Regression, Technical report: Image Speech and Intelligent Systems Research Group, (Paper available on http://www.isis.ecs.soton.ac.uk/resources/svminfo/).
[30] C. Angulo, X. Parra, A. Català, Neurocomputing 55 (2003) 57–77.
[31] S.R. Gunn, M. Brown, K.M. Bossley, Intell. Data Anal. 1208 (1997) 313–323.
[32] B. Samanta, Mech. Syst. Signal Process. 18 (2004) 625–644.