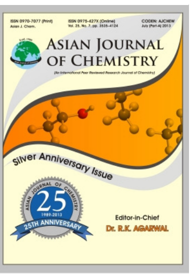
Asian Journal of Chemistry; Vol. 25, No. 7 (2013), 3668-3672

ASIAN JOURNAL OF CHEMISTRY

http://dx.doi.org/10.14233/ajchem.2013.13712



Multi-class Support Vector Machine for On-line Spectral Quality Monitoring of Tobacco Products

C. TAN[1,2,*], T. WU[1] and X. QIN[1]

[1]Department of Chemistry and Chemical Engineering, Yibin University, Yibin 644007, P.R. China
[2]Computational Physics Key Laboratory of Sichuan Province, Yibin University, Yibin 644007, P.R. China

*Corresponding author: Tel/Fax: +86 831 3551080; E-mail: chaotan1112@163.com

(*Received*: 10 March 2012; *Accepted*: 9 January 2013) AJC-12671

The combination of on-line near-infrared spectroscopy and chemometrics methods, *i.e.*, three kinds of multi-class support vector machine (SVM), namely, BSVM, one-against-one support vector machine (OAOSVM) and one-against-all support vector machine (OAASVM), was explored for monitoring the quality of tobacco products. The influence of the training set size on the performance was also investigated. A total of 165 samples from a cigarette factory were used for simulation. To compare different algorithms, three performance criteria were defined. The results revealed that as a whole, BSVM shows the best performance, especially in situations where the training set is comparatively small, while OAOSVM and OAASVM make no difference. Also, BSVM bears least computational cost since it can often build a classifier with less support vectors by only solving a single optimization. It seems that that BSVM could be a powerful tool for quality control based on high-dimensional spectral information.

Key Words: Quality control, Tobacco, Near-infrared spectroscopy, Chemometrics.

INTRODUCTION

Maintaining the best quality is essential for the survival of a tobacco factory in a globally competitive world. Cigarette is a complex mixture of components responsible for aroma and flavour and fixed compounds consisting of a large variety of substance with different characteristics. It is exactly the special composition that make cigarette of a brand unique among those of other brands[1]. The major risk for both buyers and sellers is that the product will not meet specifications and expectations when delivered. Thus, the purpose of quality control is to ensure consistency of product quality and to distinguish cigarettes of different brands. Even with extensive automation, it is still a very difficult task due to the multi-component nature. So, even today, cigarettes of different brands are mostly distinguished by human sensory responses, which are time-consuming, laborious and may also be susceptible to subjective factors and so there is an urgent need to develop alternative methods that are faster and more objective.

Generally, an ideal method used for brand classification of cigarettes should not contain a process of sample pretreatment, but can accomplish a fast data acquisition and treatment with relatively low cost[2]. Nowadays, near-infrared spectrometry has become an effective alternative to wet chemical methods in various fields such as food[3-5], pharmaceutical[6-8], medical[9-11] and petrochemical[12-14] industries, because it enable rapid and nondestructive analysis with little or no sample preparation. In applications related to near-infrared spectroscopy, a key step is to develop a prediction model by chemometrics, which allow the analytical information to be extracted from near-infrared spectra[15]. Over the past decade, chemometricians have developed many valuable algorithms intended for near-infrared applications, in which support vector machine (SVM) is an outstanding representative[16-18]. The standard support vector machine are originally designed for binary classification problem. How to effectively extend it for multi-class classification is still an on-going research issue. Currently, there are two types of approaches for multi-class support vector machine. One is by constructing and combining several binary classifiers such as one-against-one support vector machine and one-against-all support vector machine, while the other is by directly considering all classes in one single optimization formulation. Vojtech Franc has proposed to modify slightly the original optimization formulation by adding a bias term to its objective function and to transform the modified problem to a single-class problem, which is simpler than its original formulation[19]. This is so-called BSVM, which is especially appropriate for on-line applications where the simplicity is of great importance. Besides, it has been recognized that one of the very challenging works for spectroscopists is to select the most appropriate algorithm for a given task since the superiority of one algorithm over another for a task can not be generalized to another task[20].

In the present work, the combination of on-line near-infrared spectroscopy with chemometrics method (three kinds of multi-class support vector machine, namely, BSVM, one-against-one support vector machine and one-against-all support vector machine was explored for monitoring the quality of tobacco product. The influence of the training set size on the performance was also investigated. A total of 165 samples from a cigarette factory were used for simulation. To compare different algorithms, three performance criteria based on the correctly classified rate were defined. The results revealed that as a whole, BSVM shows the best performance, especially in situations where the training set is comparatively small, while one-against-one support vector machine and one-against-all support vector machine make no difference. Also, BSVM bears least computational cost since it can often build a classifier with less support vectors by only solving a single optimization. It seems that that BSVM could be a powerful tool for quality control based on high-dimensional spectral information.

EXPERIMENTAL

Support vector machine: Support vector machine[21-23], well researched in statistical learning theory, have been actively investigated in pattern classification and regression. Support vector machine map an input sample/pattern to a high dimensional feature space and try to find an optimal hyperplane that minimizes the recognition error for the training data using a special non-linear transformation function.

The standard support vector machine is designed for binary classification. The multi-class support vector machine is still an ongoing research issue. The existing methods can roughly be divided between two different approaches *i.e.*, the single machine approach, which attempts to train a multi-class support vector machine by solving a single optimization problem and the divide and conquer approach, which decomposes the multi-class problem into several binary sub-problems and builds a standard support vector machine for each. The most popular decomposing strategy is the one-against-all, which consists of building one support vector machine per class, trained to distinguish the samples in a single class from the samples in all remaining classes. Another popular strategy is the one-against-one, which builds one support vector machine for each pair of classes; *i.e.*, for the k-class problem, a total of k(k-1)/2 binary support vector machine are first trained and then, a fusion method such as majority voting is used to combine the multiple support vector machine outputs. In this study, both one-against-one and one-against-all approaches are used and termed as one-against-all support vector machine and one-against-one support vector machine, respectively.

BSVM belongs to so-called single machine approach, which deals with a multi-class classification by solving only an optimization problem. Detailed information can be found in the literature[24].

Sample set partitioning: For a given data set, in general, the selection of a representative training set upon which training the classifiers is performed is of great importance further, a test set is necessary in order to evaluate the performance of such classifiers. Strictly speaking, the evaluation is valid only if the test set has the same distribution as the training set. For this purpose, the classical Kennard-Stone (KS) algorithm[25,26], which sequentially selects a sample to maximize the minimal Euclidean distances between already selected samples and the remaining samples, is first used to rank all samples of each class, afterwards, an alternate re-sampling is applied to select one sample of every three samples in order to constitute the test set, the remaining samples constitute the training set. As a result, the training set and the test set have about two-third and one-third of samples, respectively, *i.e.*, a 2/1 division of training/test samples.

Performance criteria: In order to verify and compare different classifiers, three criteria based on the correctly classified rate were adopted. The correctly classified rate (CCR) was defined as follows:

$$\mathrm{CCR} = \frac{\sum_{i=1}^{K} \text{correctly classified samples in class i}}{\text{total number of samples}} \tag{1}$$

where k is the total number of class. Because modeling was repeated m times for each training set size, a criterion average correctly classified rate" is defined as follows:

$$\text{Average CCR} = \frac{1}{m}\sum_{i=1}^{m} \mathrm{CCR}_i \tag{2}$$

To measure the stability of classifiers, another two criteria, *i.e.*, 95 percentile of correctly classified rate and standard deviation of correctly classified rate are used. Actually, the average correctly classified rate describes the average behaviour of an algorithm, while 95 percentile of correctly classified rate describes the extremely bad behaviour of an algorithm with 5 % chance. For example, if 95 percentile correctly classified rate is 98 %, it means that the algorithm has a chance of 95 % to produce a classifier with as large as 98 % correctly classified rate. The standard deviation of correctly classified rate is the standard deviation of correctly classified rate, which can indicate the diversity of an algorithm, *i.e.*, the influence of the training set on the correctly classified rate.

Sampling and spectra collection: All samples were taken from a cigarette factory in west China. The near-infrared spectrum was on-line recorded in the diffuse reflectance model using the Matrix-E system (Matrix-E, Bruker, German), which was suspended exactly over the conveyer belt where shredded tobacco was passing. The sampling module contains 4 near-infrared light sources to illuminate the sample. Light from the sources is focused on to the conveyor belt through a window. The distance from the window to the conveyor belt is about 20 cm and the measured spot size is approximately 2.5 cm in diameter. Each final near-infrared spectrum is the average spectrum of 64 scans over the range 12000-4000 cm^{-1}, with a resolution of 8 cm^{-1}. A total of 165 spectra corresponding to three brands were obtained, among which 38, 103 and 24 spectra belonged to Jiaozi (A), Wuniu (B) and Xiongshi (C), respectively. In order to perform the later classification calculation, each spectrum was assigned a label from 1 to 3 according to its brand. Each spectrum combined with its label represents a sample.

Software and calculations: The Matrix-E system was controlled by Bruker Optics OPUS software package. All calculations were performed in Matlab 7.0 and Windows Xp, based on Pentium IV with 256 RAM. All the support vector machine algorithms were implemented on the Statistical Pattern recognition toolbox (http://cmp.felk.cvut.cz/cmp/software).

RESULTS AND DISCUSSION

Preliminary analysis and data preparation: To provide an overview of the data distribution, principal component analysis is employed. Fig. 1 gives the score plots of the first three/two PCs extracted by principal component analysis. As it can be seen that in both PC1-PC2 and PC1-PC2-PC3 plots, even if the near-infrared spectra seem to contain some valuable information for distinguishing different brands of cigarettes, they still shows considerably overlapped. Thus, the classification task is somewhat difficult and there is a need to seek an appropriate algorithm for building a powerful classifier.

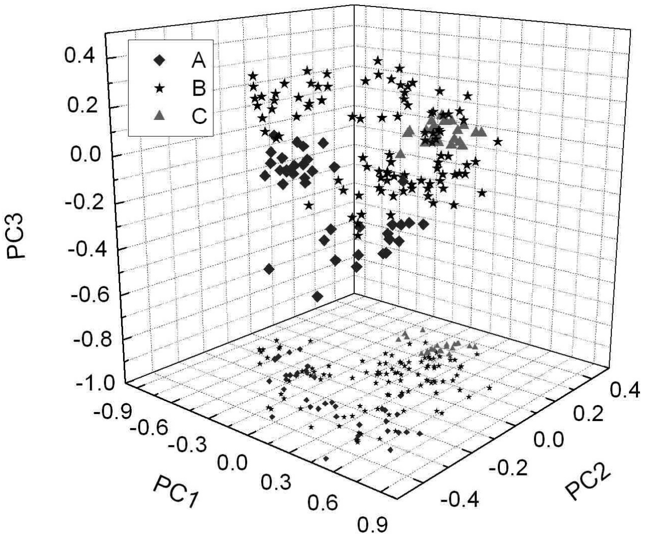


Fig. 1. Score plots of the first three/two components extracted by principal component analysis

With the sample set partitioning scheme described above, a total of 165 samples were first broken into a training set and a test set. Due to the similar information distribution existed in both the training set and the test set, it is reasonable and reliable to use only the test set for validation purpose. The training set contains 110 samples with 25, 69 and 16 samples belonging to A, B and C, respectively, while the test set contains 55 samples with 13, 34 and 8 samples belonging to A, B and C, respectively. As acquiring a sample is expensive, it is profitable to probe into the influence of the training set size on the performance of each algorithm so as to find the smallest training set size that can produce a satisfactory classifier. To achieve this, the original training set were further divided into a series of training subsets with increasing sizes at an increment of 5 (for simplicity, also called training set instead of training subset). Fig. 2 depicts the composition of samples corresponding to different training set size and the ellipse designates the case that the training set size equals the test set size (*i.e.* the most similar composition). Compared to the fixed test set, clearly, some of the training sets are larger while the others are smaller, which make it possible to analyze the effect of the training set size. It must be mentioned that, for each training set size, random sampling was carried out 100 times to create 100 training sets, on which 100 classifiers were built to compute the three performance criteria, *i.e.*, average correctly classified rate, 95 percentile of correctly classified rate and standard deviation of correctly classified rate.

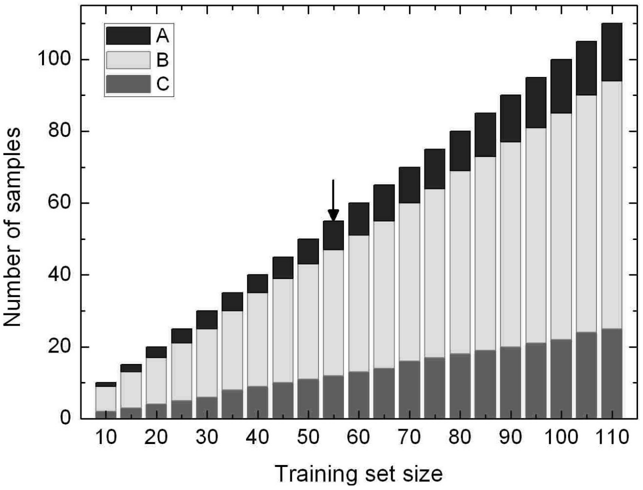


Fig. 2. Composition of samples corresponding to different training set size (the ellipse designates the case that the training set size equals to the test set size

Comparison of the three multi-class support vector machine: For a fair comparison, three multi-class support vector machine, *i.e.*, BSVM, one-against-all support vector machine and one-against-one support vector machine, used the same RFB kernel function, with the kernel parameter being optimized in the range of [0.1 0.2, 0.4, 0.8, 1.6, 3.2], while, the regularization constant C was fixed at 10 since it often has relatively small influence on the classifiers.

With the increase of the training set size, Figs. 3-5 give the comparison of the values of average correctly classified rate, 95 percentile of correctly classified rate and standard deviation of correctly classified rate of different classifiers based on 100 runs/classifiers, respectively. As shown in Fig. 3, when the training set size is smaller than 55 (*i.e.* equals the test set size), the Average correctly classified rate curve related to each of the five types of classifiers climbs fast. Once the training set size is larger than 55, all the curves become relatively flat, suggesting that the average correctly classified rate can only be improved slightly by increasing training samples. It seems to be difficult to construct an acceptable classifier on a training set with too small size. Basically, the curve corresponding to BSVM is above the curves corresponding to one-against-one support vector machine, one-against-all support vector machine, which means that in almost all cases, BSVM classifiers perform best. For a given training set size, BSVM can always build a classifier with higher accuracy. In other words, to build a classifier with expected accuracy, BSVM need the minimum number of training samples and therefore can save the cost of collecting samples. It can also be seen in Fig. 3 that, once the training set size is larger 95, each of the three support vector machine classifiers can achieve an average correctly classified rate value of 100 %. Besides, one-against-one support vector machine and one-against-all support vector machine make no significant difference.

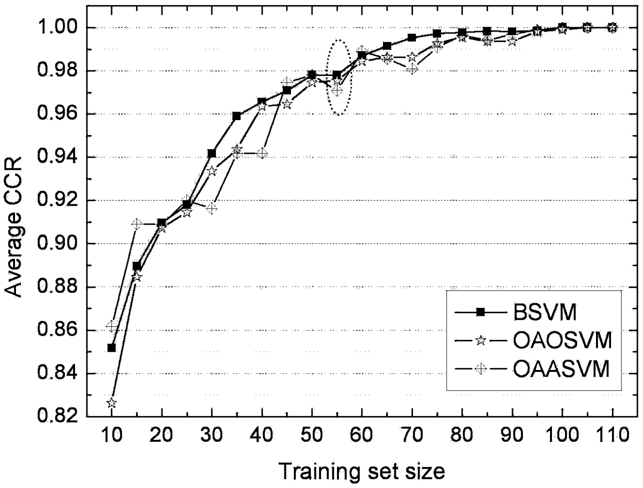


Fig. 3. Comparison of the Average CCR values of different classifiers with the changes of the training set size based on 100 runs

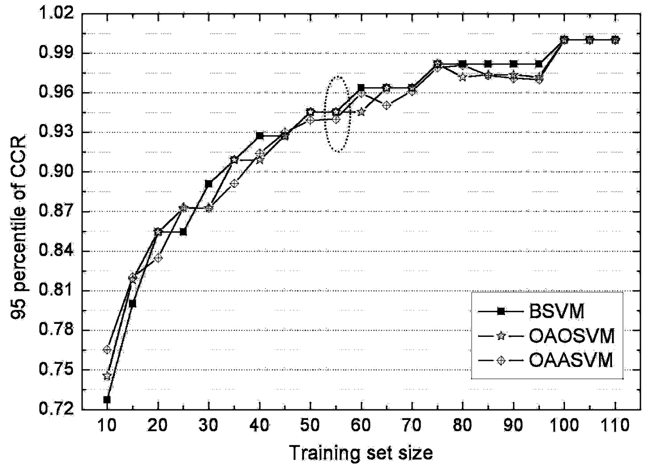


Fig. 4. Comparison of the 95 percentile of CCR values of different classifiers with the changes of the training set size based on 100 runs

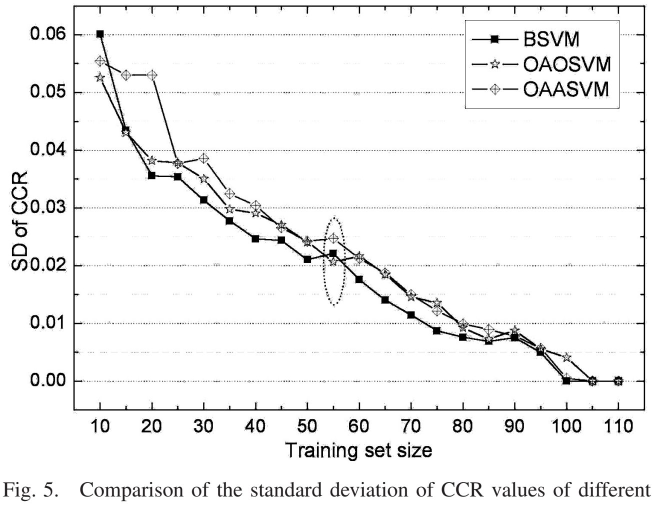


Fig. 5. Comparison of the standard deviation of CCR values of different classifiers with the changes of the training set size based on 100 runs

In Fig. 4, as a whole, all curves present as the similar trends as shown in Fig. 3. Herein, it can be noted that the performance of BSVM classifiers can always be improved steadily while each of other kinds of classifiers appears a few unusual instances. For example, the one-against-all support vector machine classifier on 65 samples, the one-against-one support vector machine classifier on 80 samples achieve a local minimum of 95 percentile of correctly classified rate, indicating that increasing training samples can not guarantee stable performance improvement. Therefore, it can be concluded that BSVM is the most robust algorithm. In practice, since one has to take the cost of collecting samples into account, the combination of average of correctly classified rate and 95 percentile of correctly classified rate provide some information on how to make a trade-off between average accuracy and robustness. It is clear in Fig. 5 that the standard deviation of correctly classified rate curve corresponding to BSVM is at the bottom, indicating that the diversity of BSVM algorithm, *i.e.*, the influence of the training set composition on the correctly classified rate, is least and thereby confirming the robustness of BSVM from another perspective. Fig. 6 depicts the mean and standard deviation of support vectors for three kinds of support vector machine classifiers trained on different training set sizes. As shown in Fig. 6, on the average, with the increase of training set size, the ratio of support vectors in the training set declines gradually and the standard deviation of support vectors rises instead, indicating that when the training set become larger, support vector machine has more choices of utilizing different subsets to construct classifiers with similar performance. For a fixed training set size, BSVM classifier often contains least support vectors. Also, it is based on solving a single optimization problem in modeling. Thus, compared to one-against-one support vector machine and one-against-all support vector machine, BSVM bears less computational burden in both constructing and applying a classifier.

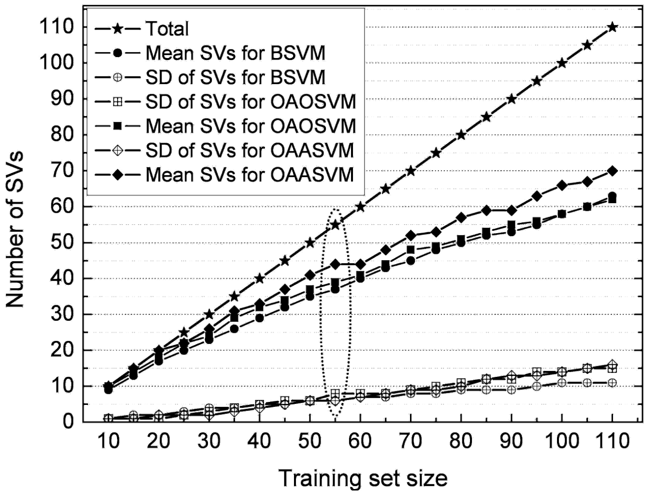


Fig. 6. Mean and standard deviation (SD) of the support vectors (SVs) for three kinds of SVM classifiers trained on different training set sizes

Conclusion

The assessment of cigarette authenticity is important and need fast classification technique. This study explored the feasibility of on-line near-infrared spectroscopy combined with three kinds of multi-class support vector machine algorithms for discriminating cigarettes of different brands. The influence of the training set size on the performance of each algorithm was also explored. In comparison to one-against-one support vector machine and one-against-all support vector machine, BSVM shows the best overall performance in almost all cases, suggesting that the combination of on-line near-infrared spectroscopy and BSVM can serve as a promising tool of discriminating cigarettes of different brands in the process quality control of tobacco industry. These properties, together with a clear theoretical background, make BSVM a good candidate to be applied to quality control systems using spectral data

ACKNOWLEDGEMENTS

This work was supported by the Sichuan Youth Science and Technology Foundation 09ZQ026-066), the Key Research Foundation Yibin University (2011Z22), the Innovative Research and Teaching Team Program of Yibin University Cx201104), and the Scientific Research Fund of Sichuan Provincial Education Department of China (12ZA201)

REFERENCES

1. C. Tan, M.L. Li and X. Qin, *Anal. Bioanal. Chem*., **389**, 667 (2007).
2. M.J.C. Pontes, S.R.B. Santos, M.C.U. Araújo, L.F. Almeida, R.A.C. Lima, E.N. Gaião and U.T.C.P. Souto, *Food Res. Int*., **39**, 182 (2006).
3. M. Cocchi, C. Durante, G. Foca, A. Marchetti, L. Tassi and A. Ulrici, *Talanta*, **68**, 1505 (2006).
4. H.Y. Cen and Y. He, *Trends Food Sci. Tech.*, **18**, 72 (2007).
5. F. Liu, Y. He, L. Wang and H.M. Pan, *J. Food. Eng.*, **83**, 430 (2007).
6. J. Luypaert, D.L. Massart and Y. Van der Heyden, *Talanta*, **72**, 865 (2007).
7. I. González-Martín, J.M. Hernández-Hierro and N. Barros-Ferreiro, *Anal. Bioanal. Chem.*, **386**, 1553 (2006).
8. K. Awa, T. Okumura, H. Shinzawa, M. Otsuka and Y. Ozaki, *Anal. Chim. Acta*, **691**, 81 (2008).
9. V.R. Kondepati, M. Keese, R. Mueller, B.C. Manegold and J. Backhaus, *Vib. Spectrosc.*, **44**, 236 (2007).
10. K.Z. Liu, M.H. Shi, A. Man, T.C. Dembinski and R.A. Shaw, *Vib. Spectrosc.*, **38**, 203 (2005).
11. N. Kang, S. Kasemsumran, Y.-A. Woo, H.-J. Kim and Y. Ozaki, *Chemom. Intell. Lab. Syst.*, **82**, 90 (2006).
12. R.M. Balabin and R.Z. Safieva, *Fuel*, **87**, 1096 (2008).
13. R.M. Balabin, R.Z. Safieva and E.I. Lomakina*, Chemom. Intell. Lab. Syst.*, **93**, 58 (2008).
14. M.J. Kim, Y.H. Lee and C.H. Han, *Comput. Chem. Eng.*, **24**, 513 (2000).
15. C. Tan and M.L. Li, *Anal. Sci.*, **23**, 201 (2007).
16. U. Thissen, M. Pepers, B. Üstün, W.J. Melssen and L.M.C. Buydens, *Chemom. Intell. Lab. Syst.*, **73**, 169 (2004).
17. Y.K. Li, X.G. Shao and W.S. Cai, *Talanta*, **71**, 217 (2007).
18. T.T. Zou, Y. Dou, H. Mi, J.Y. Zou and Y.L. Ren, *Anal. Biochem.*, **355**, 1 (2006).
19. V. Franc and V. Hlavac, 16th International Conference on Pattern Recognition, vol. 2, pp. 236-239 (2002).
20. J. Huang, D. Brennan, L. Sattler, J. Alderman, B. Lane and C.O' Mathuna, *Chemom. Intell. Lab. Syst.*, **62**, 25 (2002).
21. V. Vapnik, Statistical Learning Theory, John Wiley & Sons, New York (1998).
22. C. Cortes and V. Vapnik, *Mach. Learn.*, **20**, 273 (1995).
23. A.I. Belousov, S.A. Verzakov and J. von Frese, *J. Chemometr.*, **16**, 482 (2002).
24. C.W. Hsu and C.J. Lin, *IEEE Trans. Neural Networks*, **13**, 415 (2002).
25. R.W. Kennard and L.A. Stone, *Technometrics*, **11**, 137 (1969).
26. K.R. Kanduc, J. Zupan and N. Majcen, *Chemom. Intell. Lab. Syst.*, **65**, 221 (2003).