Available online at www.sciencedirect.com
SCIENCE DIRECT®
Talanta 64 (2004) 943–948
Talanta
www.elsevier.com/locate/talanta

Quality control of the powder pharmaceutical samples of sulfaguanidine by using NIR reflectance spectrometry and temperature-constrained cascade correlation networks

Xiujun Cui [a], Zhuoyong Zhang [b,*], Yulin Ren [c], Sidong Liu [a], Peter de B. Harrington [d]

[a] *Faculty of Chemistry, Northeast Normal University, Changchun 130024, PR China*
[b] *Department of Chemistry, MOE Key Laboratory of 3D Information Acquisition and Applications, Capital Normal University, Beijing 100037, PR China*
[c] *Faculty of Chemistry, Jilin University, Changchun 130023, PR China*
[d] *Center for Intelligent Chemical Instrumentation, Clippinger Laboratories, Department of Chemistry and Biochemistry, Ohio University, Athens, OH 45701-2979, USA*

Received 2 February 2004; received in revised form 4 April 2004; accepted 14 April 2004

Abstract

Temperature-constrained cascade correlation networks (TCCCNs) were applied to the identification of the powder pharmaceutical samples of sulfaguanidine based on near infrared (NIR) diffuse reflectance spectra and their first derivative spectra. This work focused on the comparison of performances of the uni-output TCCCN (Uni-TCCCN) and multi-output (Multi-TCCCN) by near infrared diffuse reflectance spectra and their first derivative spectra of sulfaguanidine. The TCCCN models were verified with independent prediction samples by using the "cross-validation" method. The networks were used to discriminate qualified, un-qualified and counterfeit sulfaguanidines pharmaceutical powders. The results showed that single outputs network generally worked better than the multiple outputs networks, and the first derivative spectra were more suitable for the identification comparing with original diffuse reflectance spectra. With proper network parameters the pharmaceutical powders can be classified at rate of 100% in this work. Also, the effects of parameters and related problems were discussed.

Keywords: Temperature-constrained cascade correlation; Artificial neural network; Near-infrared reflectance spectra; Classification; Sulfaguanidine

1. Introduction

Near infrared (NIR) spectrometry is a simple, fast and non-destructive way for organic compound analysis. The near infrared region covers the wavelengths between 700 nm (near the red end of the visible spectrum) and 3000 nm (near the beginning of infrared stretches of organic compounds). Absorption peaks in the NIR region originate from overtones and combinations of the fundamental (MIR) bands. The near infrared spectra are used chiefly for identifying or quantifying molecules that include unique hydrogen atoms. The vibration information including H group in molecules can be obtained. The samples can be measured without or with little pre-treatment. Derivative spectra are often used to detect small peaks, to enhance small peak separation and to eliminate slop background. NIR spectrometry has broad applications including agricultural products and foods [1–6], pharmaceutical samples [7–16]. NIR spectra are seriously overlapped, so they are only useful with the aid of chemometric methods [17–22].

The most popular artificial neural network in analytical chemistry is back-propagation neural network (BNN) [23–25]. The BNN adjusts all the weights in the network simultaneously, so it may be easier to cause chaos. The BNN constructs its own topology prior to training. The main drawbacks of BNN include slower training rate and being easily trapped in local minimum. The cascade correlation network (CCN) was developed to alleviate these problems [26]. The CCN configures its own architecture as it trains. CCN starts with a minimal network (network with only one hidden neuron) and then sequentially adds hidden neurons until the output error decreases below a user-defined value.

* Corresponding author. Tel.: +86-1086755907; fax: +86-1068903040.
E-mail address: gusto2008@vip.sina.com (Z. Zhang).

0039-9140/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.talanta.2004.04.009

Each new hidden unit receives values from network input units and previously installed hidden units. Therefore, the hidden units are connected in the cascade form. Several candidate units can be trained in parallel, and the one with largest covariance between candidate unit and its residual errors can be selected as the next hidden unit to install into the network. Once trained, the hidden units are no longer adjusted, thus, only one unit is trained at a time. This trait is somewhat unique for the CCN and eliminates the chaos of simultaneously adjusting all processing units as adjustable parameters as in the BNN training. The number of hidden units of the CCN will increase until the desired error is obtained or the network converges. So the most significant advantages of the CCN over the BNN are the auto-adjusting network architecture and faster training rate. But the CCN still does not solve the problem of overfitting. One method that overcomes overfitting is the use of computational temperature, which is a similar parameter to that used in simulated annealing [27]. The advantage afforded by a temperature constraint is the removal of an extra degree of freedom manifested in the lengths of the weight vectors in the processing units of the neural network. It is known that overfitting is related to the Euclidean length of a set of regression coefficients. The weight vectors of neural network processing units can be considered as non-linear regression coefficients. The temperature constraint optimizes the length of the weight vectors for each hidden processing unit. The temperature constraint results in softer models that tend to have greater generalizing abilities [27–29]. Sulfaguanidine is a commonly used medicine which acts as antimicrobial agents by inhibiting bacterial growth and activity. In the production of sulfaguanidine tablet, the quality control of the sulfaguanidine powder is required and have significant importance. It is desirable that a fast and non-destructive analytical method is developed in pharmaceutical tablet production. The aim of this work is to test the applicability of the temperature-constrained cascade correlation networks (TCCCNs) in quality control in pharmaceutical production. Many methods including various spectrometric and chromatographic ones have been applied in quality control of pharmaceutical industry. NIR techniques provide fast, non-destructive, and little or no sample pretreatment. The combination of NIR techniques with various chemometric methods is becoming a powerful tool in quality control. The TCCCN has been successfully used in classification examples including ion mobility spectrometry (IMS), GC–MS, etc. In the present study, TCCCN is attempted to apply to the classification of sulfaguanidine in near infrared diffuse reflectance spectra as quality control method. This method was verified with independent prediction samples.

2. Theoretical basis

The TCCCN does not connect the inputs directly to the output units. Instead, all the inputs to the output unit must pass through the fuzzy hidden units, which help to prevent overfitting. Fig. 1 is a schematic architecture of the single-output TCCCN.

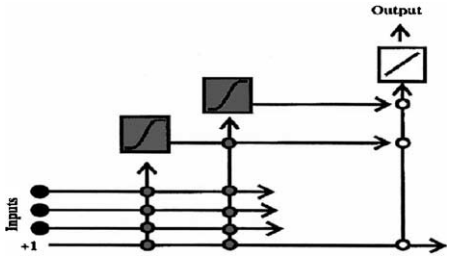

Fig. 1. Schematic architecture of TCCCN.

The number of inputs into the output units will equal to the number of hidden units. The hidden units are added sequentially. Each time a hidden unit is added, the output weight vector is recalculated by regression, and new residual errors are generated. The hidden units are trained by adjusting their weight vectors so that the pooled magnitude of the covariance between a hidden unit's output and the residual error from the output units is maximized.

The candidate units for the CCN are temperature-constrained sigmoidal functions. The hidden unit comprises a linear and a non-linear operation. The linear operation is obtained by

$$\text{net}_{ij} = \frac{\sum_{m=1}^{\gamma} w_{jm} x_{im}}{|w_j|} + b_j \tag{1}$$

in which γ is the number of input connections to unit j, w_{jm} a component of the weight vector, and x_{im} the input activation coming from the mth neuron in the preceding layer for the ith object. For the temperature parameter to be meaningful, the weight vector must be constrained to a constant length. The weight vector is normalized to unit Euclidean length. The vector length of the jth hidden unit weight vector is defined as $|w_j|$. The bias value for the jth hidden unit is designated as b_j.

The non-linear operation is given by

$$o_{ij} = f(\text{net}_{ij}) = (1 + \mathrm{e}^{-net_{ij}/t_j})^{-1} \tag{2}$$

in which the results (net_{ij}) are input to the transfer function, the outputs of the jth hidden unit is o_{ij} and t_j is computational temperature. t_j is adjusted so that it maximizes the magnitude of the first derivative of the covariance between the output and the residual error with respect to temperature. The temperature constrained transfer function is applied to hidden layer. The weight and bias parameters are adjusted so that the magnitude of pooled covariance between a hidden unit output and the residual error from the output units is maximized. The covariance magnitude ($|C_j|$) of the output from candidate unit j and the residual error from output k is obtained from

$$|C_j| = \sum_{k=1}^{p} \left| \sum_{i=1}^{n} (o_{ij} - \bar{o}_j)(e_{ik} - \bar{e}_k) \right| \tag{3}$$

Table 1
The component content of sulfagunidine (%)

Samples	Sulfagunidine			Starch			Mg stearate		
	Maximum	Minimum	Average	Maximum	Minimum	Average	Maximum	Minimum	Average
Qualified samples (27)	96.22	87.98	92.1	11.47	3.23	5.69	0.68	0.34	0.51
Un-qualified samples (8)	74.50	58.56	65.87	40.91	25.02	33.56	0.87	0.36	0.57
Counterfeit samples (5)	0	0	0	99.65	99.36	99.51	0.64	0.37	0.49

for which the covariance is calculated with respect to the n observations in the training set. The absolute values of the covariances are added for the p output units. The averages are obtained for the n objects in the training set for the hidden unit output (o_j) and error (e_j). The denominator of $n-1$ is omitted from the calculation, because it is constant through the entire training procedure. The weight vectors are adjusted so that $|C_j|$ is maximized.

As shown in Fig. 1, the output unit is linear, and the hidden units are temperature-constrained sigmoidal functions. The weight vector is adjusted so that the magnitude of the covariance between the residual output error and the output of the hidden unit is maximized. The weight vector will be constrained to unit vector length. Once the hidden unit is adjusted, it will be held constant. The output unit is readjusted by regression with the two hidden units' output and bias value as inputs. This procedure continues until a specified residual error is reached. Each time a hidden unit is added and trained, the output unit weight vectors are recalculated by regression of the target values onto the column space defined by the hidden unit outputs.

The temperature-constrained cascade correlation networks with uni-output (denoted as Uni-TCCCN) and with multi-output (Multi-TCCCN) were used for the classification. The multi-output network is similar to the one given in Fig. 1 except for multi-output units are used.

3. Experimental

3.1. Instrument and working conditions

A Shimadzu NIR Spectrophotometer (model ISR-3100) with diffuse reflectance accessory was used. Slit height is 12 nm. Scan range is 1300–2500 nm. Each sample was scanned two times and the average was used. Data were collected in every 1 nm. So 1201 data points were collected for each spectrum. The temperature constrained cascade correlation neural network software was written in C++ and compiled with Borland C++ 5.0. The program run on an Intel Pentium II 450 MHz processor equipped with 128 MB of RAM, which was operated under Windows 2000.

3.2. Reagents and samples

The pharmaceutical powder samples of sulfaguanidine comprises sulfagunidine, magnesium stearate, and starch, which were prepared according to the China Pharmacopoeia. Twenty seven qualified samples (nos. 1–27), eight un-qualified samples (the amount of sulfagunidine were lower than the requirement, nos. 28–35) and five counterfeit samples (do not contain any sulfagunidine at all, nos. 36–40) were used. The amount of sulfagunidine in each classes of samples were given in Table 1.

3.3. Discrimination method

Two data sets were used for each neural network model. The training set was used to build the model, and the testing set was used to evaluate the model. Binary coding of the target values was used for the classification. For example, output vectors [1 0 0], [0 1 0], and [0 0 1] were used to distinguish qualified, un-qualified and counterfeit samples, respectively.

A threshold value of 0.5 was used to classify the network outputs. An output value that was greater than or equal to 0.5 indicated class assignment. The criterion for class assignment is more robust than the 'winner-take-all' criterion and allows for multiple classifications of the same object.

The cross-validation method was used for assessment of the networks.

4. Results and discussions

Near infrared spectrometry has been found more and more applications in various fields, mainly including pharmaceutical, food, textile industries. The main advantage of measurements based on NIR technique is non-destructive analysis and none or little pretreatment for sample preparation. However, because the NIR spectra are composed of many overtones and combination bands, the NIR spectra are usually severely overlapped. These make the NIR spectra difficult to interpret. Therefore, chemometric methods have to be used in data interpretation and calibration in almost all the NIR applications. The near infrared diffuse reflectance spectra and their derivative spectra of sulfagunidine, magnesium stearate, starch were given in Fig. 2(A) and (B), respectively. There were some differences of dampness and brightness between samples, however, the effects of dampness and brightness of samples in calibration set were minimized by randomly mixing the samples. It can be seen that the spectra of the three components were seriously overlapped. The overlapped spectra deteriorate the separation

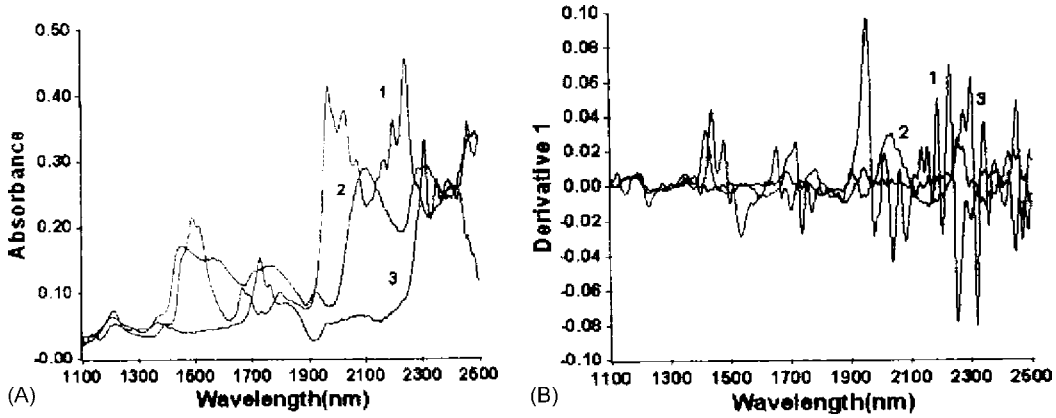

Fig. 2. (A) Near-infrared diffuse reflectance spectra and (B) first derivative spectra (1) sulfagunidine (2) Mg stearate, and (3) starch.

of characteristic peaks. Besides, the original spectra have slop background. The slop background also deteriorates final sample classification. To enhance the separation of the spectra and to eliminate the slop background, the first derivative spectra were used for sample classification. It will be seen in the classification results section of this paper, that the classification results are quite satisfactory, these are partly due to the enhancement of peak separation and the elimination of the slop background.

4.1. Effect of pre-defined relative errors

In network training, a pre-defined relative error should be given so that the training process can terminate when output error reaches or lower than this value. If the calculated relative error is larger than the pre-defined relative error, the networks add automatically a hidden unit, and calculate again. The training error is defined as the relative root mean square error of calibration (RRMSEC) and is calculated by

$$\mathrm{RRMSEC} = \sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{p}(\hat{y}_{ij} - y_{ij})^2}{\sum_{i=1}^{n}\sum_{j=1}^{p}(y_{ij} - \bar{y}_{ij})^2}} \quad (4)$$

in which p is the number of classes or network outputs, n the number of training spectra, and the target value for the ith object and jth class is y_{ij}, $\hat{y}_{ij}$ is the corresponding network estimate of the target value, and $\bar{y}_j$ is the average target value for the pth output.

Prediction performance is evaluated by using root mean square error of prediction (RMSEP) defined as

$$\mathrm{RMSEP} = \sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{p}(\hat{y}_{ij} - y_{ij})^2}{np}} \quad (5)$$

The relative root-mean-square error of prediction (RRMSEP) is similar to the RRMSEC given in Eq. (4) except only the test set data is used [30].

For the Multi-TCCCN, each output node represents one specific class, and the network is trained so that a specific node produces the largest output among all nodes to show the presence of that specific class. Classification results can be obtained simultaneously by using a Multi-TCCCN. In the classification temperature-constrained sigmoid function was used in hidden unit, and the output unit was sigmoid function. The RMSEPs of the pharmaceutical sample are given in Fig. 3. With the decrease of pre-defined relative errors, the RMSEPs decreased. It can also be seen that the standard errors for derivative spectra are lower than those for original spectra.

Uni-TCCCN was also used for comparison. In uni-output network there is only one output unit in the network, so only one class can be predicted at a time. The whole Uni-TCCCN is composed of several uni-output networks, therefore, multi-class classification can also be obtained by using the Uni-TCCCN system. The RMSEPs of the pharmaceutical sample with sulfagunidine are also shown in Fig. 3. The prediction standard errors seem to be less affected by the pre-defined training errors when the values are at 0.6 or above.

4.2. Effect of the number of candidate units

In TCCCN the hidden units were added automatically when needed. In the training process, several units can be

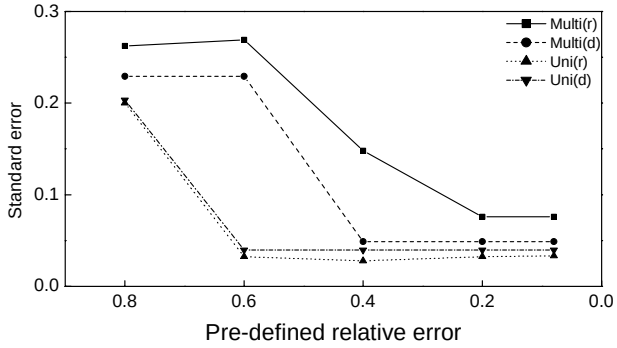

Fig. 3. Changes in prediction errors with different pre-defined relative errors (r refers to diffuse reflectance spectra, d refers to first derivative spectra) (Multi stands for using Multi-TCCCN, Uni stands for using Uni-TCCCN).

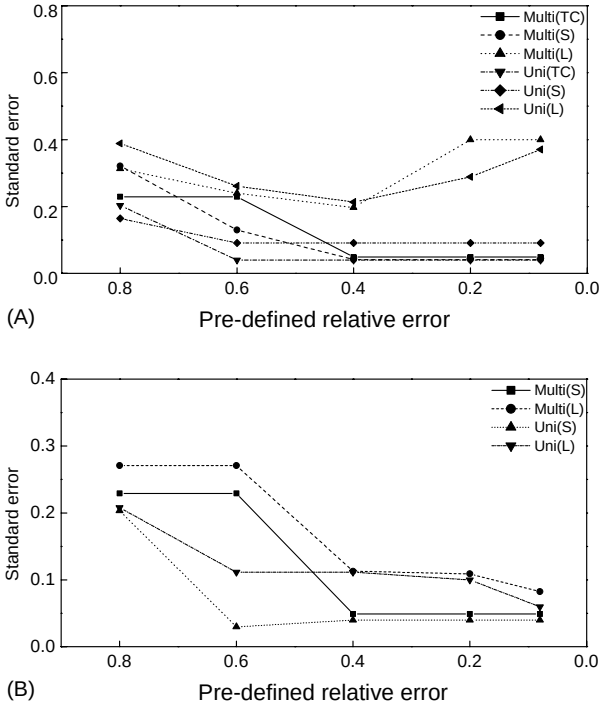

Fig. 4. Error obtained with different pre-defined relative errors in different input function and output function (Multi stands for using Multi-TCCCN, Uni stands for using Uni-TCCCN) (TC refers to temperature-constrained sigmoidal, S refers to sigmoidal, L refers to linear).

trained in parallel. These units are called "candidate units". The one with largest covariance between candidate unit and its residual error was selected to the network. To investigate the effect of number of candidate units, network training with various numbers of candidate units (from 3 to 7) were tested. Results show that the effects of candidate units number are similar for the Multi-TCCCN and Uni-TCCCN, but somewhat different for original and derivative spectra. When the number of candidate units is larger, the prediction errors seem to increase. Three candidate units were used in this work.

4.3. Effect of transfer functions

The types of transfer functions play an important role in network performance. In this section, the effect of different types of transfer functions are investigated. The effects of sigmoidal, temperature-constrained sigmoidal, and linear functions in hidden units on standard error of prediction with Uni-TCCCN and derivative reflectance spectra are shown in Fig. 4(A), respectively. It can be seen that lower prediction errors were obtained when temperature-constrained sigmoidal function was used. However, the network with linear hidden units gives relatively high prediction errors.

The effect of different types of transfer functions in output units on prediction was also investigated. The results are shown in Fig. 4(B). Networks with sigmoidal function give better predictions. Therefore, temperature-constrained sigmoidal function for hidden units and normal sigmoidal function for output units were used in this work.

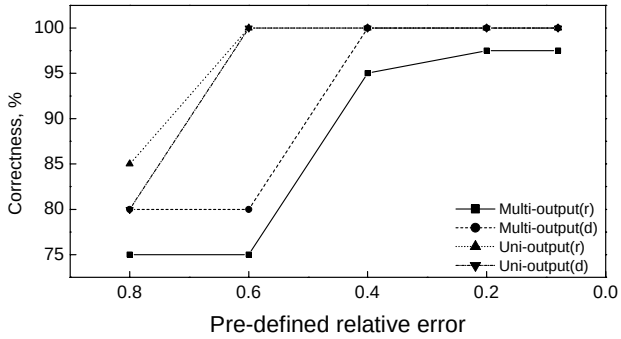

Fig. 5. Effect of different pre-defined relative error on classification accuracy.

4.4. Classification results

Based on the investigation of the effect of network parameters and optimization, compromised working parameters were chosen. The classification results are shown in Fig. 5. It can be seen that Uni-TCCCN works better than the Multi-TCCCN. For example, with Uni-TCCCN, all the samples can be correctly classified by using both original diffuse reflectance spectra and their derivative spectra even though the pre-defined relative error of 0.6 was set. The classification can only reaches up to 95% when original reflectance sopectra and Multi-TCCCN were used.

5. Conclusion

Near infrared spectrometry is a non-destructive analytical method, so it is very suitable for quality control in production process. However, spectra of different components are seriously overlapped in this NIR region. NIR spectra can be useful if only proper data processing methods are used and combined with chemometric methods. The use of derivative NIR spectra enhanced the separation of spectra. Temperature-constrained cascade correlation network (TCCCN) has been approved to be a good approach for the classification of qualified, un-qualified, and counterfeit pharmaceutical powder samples of sulfaguanidine. Results show that the TCCCN with uni-output performs better than the TCCCN with multi-output.

References

[1] K.B. Whetsel, Appl. Spectrasc. Rev. 2 (1968) 1.
[2] G.S. Birth, Zachariah. Trans. Am. Soc. Agric. Eng. 16 (1973) 371.
[3] C. Watson, Anal. Chem. 77 (49) (1999) 835A.
[4] P.C. Williams, S.G. Stevenson, G.N. Irvine, Cereal Chem. 55 (1978) 263.
[5] G.S.J. Birth, Food Sci. 44 (1979) 949.

[6] W.F. McClure, A. Hamid, Am. Lab. 12 (1980) 57.
[7] P. Corti, E. Dreassi, S. Lonardi, Il Farmaco 48 (1993) 3.
[8] B.F. MacDonald, K.A. Prebble, J. Pharm. Biomed. Anal. 11 (11–12) (1993) 1077.
[9] K.M. Morisseau, C.T. Rhodes, Drug Dev. Ind. Pharm. 21 (9) (1995) 1071.
[10] L.J. Galante, M.A. Brinkley, Anal. Chem. 62 (23) (1990) 2514.
[11] E.W. Ciurczak, Appl. Spectrosc. Rev. 23 (1–2) (1987) 147.
[12] J.K. Drennen, R.A. Lodder, J. Pharm. Sci. 79 (7) (1990) 622.
[13] P. Merckle, K.A. Kovar, J. Pharm. Biomed. Anal. 17 (3) (1998) 365.
[14] R.A. Lodder, G.M. Hieftie, Appl. Spectrosc. 42 (4) (1988) 556.
[15] M.A. Dempster, J.A. Jone, J. Pharm. Biomed. Anal. 11 (11–12) (1993) 1087.
[16] J.D. Kirsch, J.K. Drennen, Appl. Spectrosc. Rev. 30 (3) (1995) 139.
[17] K. Krämer, S. Ebel, Anal. Chim. Acta 420 (2000) 155–161.
[18] D.A. Burns, E.M. Ciurczak (Eds.), Handbook of Near-Infrared Analysis, Marcel Dekker, New York, 1992.
[19] E. Cleve., E. Bach, E. Schollmeyer, Anal. Chim. Acta 420 (2000) 163–167.
[20] Paolo Carlini, Riccardo Massantini, Fabio Mencarelli, J. Agric. Food Chem. 48 (2000) 5236–5242.
[21] Martin Andersson, Anal. Chem. 72 (2000) 2099–2108.
[22] Cristoph Klawun, Charles L. Wilkins, J. Chem. Inf. Comput. Sci. 36 (1996) 69–81.
[23] P.A. Jansson, Anal. Chem. 63 (1991) 357A–362A.
[24] J.G. Zupan, J. Anal. Chim. Acta 248 (1991) 1–30.
[25] B. Wythoff, J. Chemom. Intell. Lab. Syst. 18 (1993) 115–155.
[26] S.E. Fahlman, C. Lebiere, in: D.S. Touretzky (Ed.), Advances in Neural Information Processing System, vol. 2, Morgen Kaufmann Publishers, 1990, p. 524.
[27] Peter de B. Harrington, Anal. Chem. 70 (1998) 1297–1306.
[28] W. Chuanhao, Anal. Chim. Acta 408 (2000) 1–12.
[29] C. Cai, P.B. Harrington, Anal. Chem. 71 (1999) 4134–4141.
[30] Peter de B. Harrington, Anal. Chem. 72 (2000) 5004–5013.