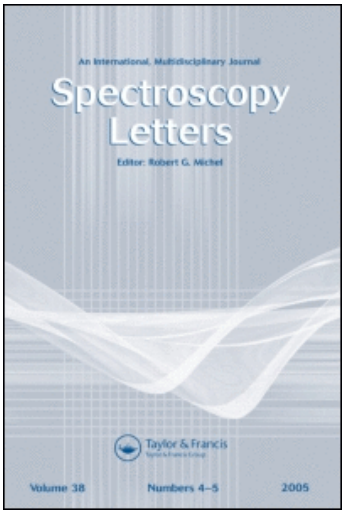

Spectroscopy Letters
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597299

Identification of Rhubarb Samples by Using NIR Spectrometry and Takagi-Sugeno Fuzzy System

Zhuoyong Zhang[a]; Yanfeng Tang[a]; Guoqiang Fan[b]
[a] Department of Chemistry, MOE Key Lab for 3-D Information Acquisition and Applications, Capital Normal University, Beijing, China [b] Tongrentang Group Company, Research Institute, Beijing, China

To cite this Article Zhang, Zhuoyong , Tang, Yanfeng and Fan, Guoqiang(2005) 'Identification of Rhubarb Samples by Using NIR Spectrometry and Takagi-Sugeno Fuzzy System', Spectroscopy Letters, 38: 4, 447 — 457
To link to this Article: DOI: 10.1081/SL-200062897
URL: http://dx.doi.org/10.1081/SL-200062897

Spectroscopy Letters, 38: 447–457, 2005

ISSN 0038-7010 print/1532-2289 online
DOI: 10.1081/SL-200062897

Taylor & Francis
Taylor & Francis Group

Identification of Rhubarb Samples by Using NIR Spectrometry and Takagi–Sugeno Fuzzy System

Zhuoyong Zhang and Yanfeng Tang
Department of Chemistry, MOE Key Lab for 3-D Information Acquisition and Applications, Capital Normal University, Beijing, China

Guoqiang Fan
Research Institute, Tongrentang Group Company, Beijing, China

Abstract: The Takagi–Sugeno fuzzy system is implemented based on several back-propagation neural networks (BP-NNs) and has been applied to identification of official and unofficial rhubarb samples based on their near-infrared spectra. Rhubarb is one of the most important Chinese medical herbs. It is of importance to identify official and unofficial rhubarb samples based on nondestructive near-infrared spectrometry for quality control in Chinese herbal products. Near-infrared diffuse reflectance spectrometry and the Takagi–Sugeno fuzzy system were used to classify 52 rhubarb samples, and the effects of the number of hidden neurons and of momentum parameters on prediction were investigated. The results obtained by using the Takagi–Sugeno fuzzy system were better than those by commonly used BP networks. With proper network training parameters, 100% correctness can be obtained by using the Takagi–Sugeno fuzzy system.

Keywords: Artificial neural network, quality control, rhubarb, Takagi–Sugeno fuzzy system

INTRODUCTION

Rhubarb is one of the most important herbal medicines in China. There are about 60 species of rhubarb in the world. Rhubarb is grown in broad areas

Received 13 March 2004, Accepted 17 January 2005
Address correspondence to Zhuoyong Zhang, Department of Chemistry, MOE Key Lab for 3-D Information Acquisition and Applications, Capital Normal University, Beijing 100037, China. E-mail: gusto2008@vip.sina.com

in Asian temperate and subtropical zones. Rhubarb has been used for thousands of years in China. The main pharmaceutical functions include purging the pathogenic fire, antibacterial, normalizing functioning of the gallbladder, liver protection, hemostasis, promoting blood circulation, and so forth.[1,2] In the Chinese Pharmacopoeia, three species of rhubarb are formally recorded. They are *Rheum palmatum* L., *R. tanguticum*, and *R. officinale Baillon*.[3] Some other species of rhubarbs that are not given in the Chinese Pharmacopoeia are known as unofficial rhubarb samples. The medicinal functions of the unofficial rhubarb were not as good as those of official rhubarb. Therefore, the unofficial rhubarb samples are not permitted to be used in traditional Chinese herbal medicinal products. However, in some cases root and rootstock of unofficial rhubarb were mixed into official rhubarb to formulate commercial products. To ensure the quality of official rhubarb products and their clinical curative effect, several techniques for identifying crude rhubarb-based drugs according to their exterior configuration, appearance characteristics by microscopy, were used,[1] but these methods can hardly identify the powdered official and unofficial rhubarb samples.

Near-infrared (NIR) spectroscopy has many advantages over mid- and far-infrared spectroscopic methods in terms of speed, simplicity in sample preparation, cost, efficiency, and flexibility. With the development of computer technology and chemometrics, NIR techniques have been applied in broad areas including agriculture,[4,5] food industry,[6] petroleum industry,[7,8] pharmaceutical analysis,[9,10] biomedical technology,[11] and many others.

The classification of medicinal herbs by NIR spectroscopy relies on chemometric methods. Principal component analysis (PCA) and partial least squares (PLS) are commonly used for interpretation of NIR spectroscopic data.[12] Calibration of NIR spectroscopic measurements is limited by the use of linear mathematical models. Neural network calibration shares many of the advantages of other calibration techniques, including avoiding overfitting the training set and determination of optimal levels of model parameters. Neural networks are mathematical models inspired by biological models. The basic processing unit is a neuron, which receives many inputs (dendrites) and generates one output. The neuron computes a weighted sum of the inputs. The weights are determined by learning. Learning in a feed-forward network consists of adjusting the weights of a neuron based on the error of the output. The implementation in the output error is gradual, and many training iterations are required. Various artificial neural networks (ANNs) have been used for pattern recognition and data interpretation in near-infrared spectroscopy[13–15] and many other fields.[16,17] ANN has been used to classify Chinese traditional medicine and to discriminate between the medicinal materials by application of near-infrared spectroscopy.[2]

The Takagi–Sugeno fuzzy system has been used in many fields, especially in fuzzy controls.[18–20] Relatively few applications of the Takagi–Sugeno

system in chemistry have been reported. In the current paper, near-infrared spectra of rhubarb samples were measured and then compressed by wavelet transform. Takagi–Sugeno fuzzy systems were used to establish a model for identification. The effects of some training parameters on the identification of samples are discussed.

THEORETICAL BASIS

The Takagi–Sugeno model can be implemented based on a neural network–driven fuzzy reasoning system. The system consists of P + 1 neural networks. The networks NN_1-NN_P denote the functions in the conclusion parts of the P rules respectively, and NN_{P+1} is used to calculate the fitness of each rule corresponding to an input vector.

If x_1 is A_1^j, x_2 is $A_2^j, \ldots,$ x_n is A_n^j, then y is B_j, where $A_1^j, A_2^j, \ldots, A_n^j$, and B_j is fuzzy subsets. In the conclusion part, the model substituting fuzzy set B_j with a function is commonly called the Takagi–Sugeno model. If x_1 is A_1^j, x_2 is $A_2^j, \ldots,$ x_n is A_n^j, then $y = f_j(x)$, where $f_j(x)$ is a linear combination of input variables, that is

$$f_j(x) = c_j^1 x_1 + c_j^2 x_2 + \cdots + c_j^n x_n + c_j^{n+1} \tag{1}$$

The model divides input space into linear spaces because the input variables are independent of each other. To avoid an increase of fuzzy rules when the two models mentioned above are used, it is convenient to adopt the number of the following model:

$$\text{If } X \in P_j, \quad \text{thus} \quad y = f_j(X) \tag{2}$$

in which $X = (x_1, x_2, \ldots, x_n)$, and P_j is a partial space divided from input space. In the formula (2), the membership function cannot be determined as a formula independently, so we only acquire the joint membership function in the condition part using neural networks; similarly, the function in the conclusion part can be expressed by neural networks. The schematic diagram of the fuzzy system based on integration of neural networks is shown in Fig. 1.

In Fig. 1, there are, in total, $P + 1$ neural networks, in which $NN_1 \sim NN_P$ are used to represent P rules of functions $f_i(X)$, and NN_{mf} is used to calculate the fitness of each rule corresponding to input vector. The output of the fuzzy system can be calculated by the following formula

$$y = \sum_{j=1}^{P} \mu_j \cdot g_j \tag{3}$$

where g_j is the output value of NN_j, and μ_j is weights of subnet.

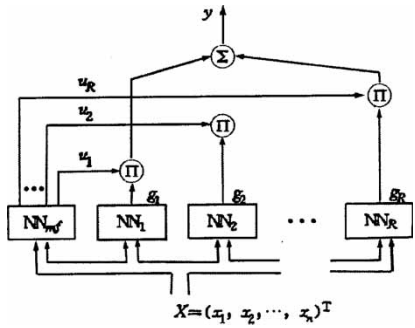

Figure 1. Schematic description of Takagi–Sugeno fuzzy system.

The steps of establishing and training the fuzzy system are as follows:

Step 1. Collect train samples for the fuzzy system.

Step 2. Use K-means method to cluster input vectors, and each cluster corresponds to a rule. Because the samples are clustered into P teams, there are P fuzzy rules for the fuzzy reasoning system.

Step 3. Train the network NN_{P+1}, which has n inputs and P outputs in its structure. The training samples can be constructed using following method: If the training sample X_i is clustered into Sth team, we have:

$$\omega_j^i = \begin{cases} on & j = S \\ off & j \neq S \quad j = 1, 2, \cdots, P \end{cases} \tag{4}$$

$$W_i = \left(\omega_1^i, \omega_2^i, \cdots, \omega_{S-1}^i, \omega_S^i, \omega_{S+1}^i, \cdots, \omega_P^i\right)^T \tag{5}$$

where *on*/*off* correspond to 1 or 0, respectively. The training samples $(X_i,\ W_i)$ are constructed for the network NN_{P+1}. After training, the final fitness vector $U = (\mu_1, \mu_2, \ldots, \mu_P)$ is used for each network NN_j $(j = 1, 2, \ldots, P)$.

Step 4. Train the networks $NN_1 \sim NN_P$. Assume that NN_S corresponds to the Sth rule, thus, all of the samples in the S-th cluster is the training data for the network NN_S. Then, back-propagation algorithm is applied to train the networks $NN_1 \sim NN_P$.

EXPERIMENTAL

Instrumentation and Data Acquisition

A FOSS 6500 NIR scanning monochromator (Foss NIR Systems Inc., Laurel, MD, USA) with halide lamp and PbS detector was used for this work. The

52 rhubarb samples used were different species and were collected from different regions. The rhubarb samples were classified as official and unofficial samples according to the Chinese Pharmacopoeia. Among these samples, 25 were official rhubarb samples (nos. 1–25) and 27 were unofficial rhubarb samples (nos. 26–52). The diameter of the sample cell was 38 mm; the thickness of the sample cell was 10 mm. The sample was shaken after each measurement, and 50 measurements were averaged to obtain the final NIR spectrum of each sample. The measurement wavelength ranged from 1100 nm to 2498 nm in 2-nm intervals. Therefore, 700 data points were obtained for each sample measurement. The spectra were then compressed by using the wavelet transform described in the next section. The spectral data were saved as ASCII code and then processed with another Pentium personal computer.

Data Processing and Computation

The measured near-infrared spectra were processed with a second-order derivation to eliminate sloping backgrounds. To speed up the training process, the NIR spectra with 700 variables were compressed by using wavelet transform to 44 variables. Wavelet transform can effectively compress the NIR spectra and keep the spectral features. The 44 variables of the NIR spectra obtained by wavelet transform were used as input to the networks. The Takagi–Sugeno fuzzy system was used to establish the classification model for identification of official and unofficial rhubarb samples. The prediction set was selected by using the "leave-one-out" cross-validation method.[21] This method leaves one sample for prediction at a time, and the other $n - 1$ samples are used for the training set. For the next test, another sample is left as the prediction sample that is different from the previous one, and also the training set includes the remaining $n - 1$ samples. This procedure is repeated until every sample has been selected as prediction sample for one time. The output values were compared with expected output values, and the network performance was evaluated based on the comparison.

The internal functions of Matlab 5.0 (Mathworks, Inc., Natick, MA, USA) were used for the wavelet transform computation. The Matlab Neural Network toolbox was used for writing the calculation program for Takagi–Sugeno in this lab. The Takagi–Sugeno fuzzy system was implemented based on three BP networks, in which NN_{mf} is used to give suitability of classification rules, while NN_1 and NN_2 were used to represent functions of classification rules. The trainbpx.m for fast back-propagation function and log sigmoid transfer function were used.

RESULTS AND DISCUSSION

The chemical composition of rhubarb sample is very complicated. The main components in rhubarb include aloe-emodin, rhein, emodin, sennoside A,

sennoside B, sennoside E, sennoside F, and many other minor components. Because of the complexity of overlapped overtones and combinational bands in the near-infrared region, it is not possible to establish direct relationships between chemical composition and NIR spectra of rhubarb samples like those used in UV-Vis spectrometry.

In the current work, the fingerprints (overall features) of official and unofficial rhubarb samples were used for the classification. It is impossible to recognize the differences of spectral features between official and unofficial rhubarb samples, because the spectral differences between official and unofficial rhubarb samples are so small. Correlation coefficients of NIR spectra between some official and unofficial rhubarb samples have been calculated to evaluate the similarity of the NIR spectra. The correlation coefficients between official and unofficial rhubarb samples with conventional NIR spectra are close to 0.999. However, when second-order derivative NIR spectra were used, the correlation coefficients were lowered down to 0.96–0.97 for most samples. Therefore, to enhance the resolution of spectra and to eliminate the sloping backgrounds of the NIR spectra, second-order derivative NIR spectra were used for the classification study.

Principal component analysis (PCA) and partial least squares (PLS) are commonly used methods for sample identification and qualification. However, for such a difficult classification problem, the conventional statistical methods would not be work well. Thus, various artificial neural networks have been tried to solve this problem in our laboratory.

Unfortunately, only 52 rhubarb samples were used for the current study, because the rhubarb samples, especially the official rhubarb samples, are quite difficult to collect. This number of samples is sufficient to establish a reliable classification model, although more samples are helpful to get a better model. In the optimization process, five samples were selected at random from each class of samples as a test set, and the remaining samples are used as a training set. The SEC and SEP were calculated to evaluate the model performance.

Effect of Number of Neurons in Hidden Layer

The complexity of the network architecture depends on the number of hidden units. In most situations, there is no way to determine the best number of hidden units without training several networks and estimating the generalization error of each. If too few hidden units are used, the high training error and high generalization error will be obtained due to underfitting and high statistical bias. However, if too many hidden units are used, low training error can be obtained, but high generalization error will result due to overfitting and high variance.

The Takagi–Sugeno fuzzy system is implemented based on three BP networks, NN_{mf}, NN_1, and NN_2. Therefore, it seems necessary to investigate the effect of hidden neurons and to determine the optimal number of

hidden neurons. In the current work, SEP (standard error of prediction) was used to evaluate the network performance. The effect of hidden neurons on SEPs is shown in Fig. 2.

It can be seen that different numbers of hidden neurons have different effects on the SEPs. Therefore, it is possible to adjust the number of hidden neurons to get better performance. From Fig. 2a it can be seen that when 7 hidden neurons of NN_{mf} were used, the minimal SEP, 1.757, was obtained. When 8 hidden neurons of NN_1 were used, the minimal SEP, 1.833, was obtained, and for 3 hidden neurons of NN_2, 1.863 was obtained.

Effect of Momentum Parameter

Learning rate and momentum are two important parameters in network training. In a standard back-propagation network, too low a learning rate makes the network learn very slowly, and too high a learning rate makes the weights and objective function diverge, so there is no learning at all. Momentum is also an important parameter in training computation formula. A momentum parameter is used to avoid the training network being trapped in a local minimum. In the internal function of MATLAB, trainbpx.m

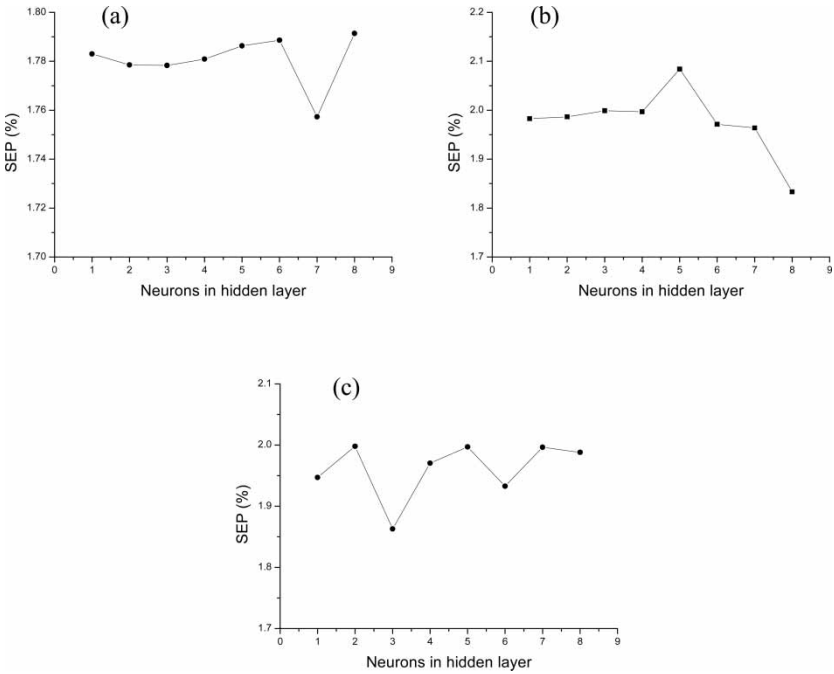

Figure 2. Effect of hidden neurons on prediction. (a) NN_{mf} subnet, (b) NN_1 subnet, and (c) NN_2 subnet.

adjusts the learning rate automatically during the training process. Therefore, the effects of momentum parameter were investigated in the current study. The effects of momentum parameters on SEPs are shown in Fig. 3.

It can be seen that different momentum parameters have various effects on prediction errors. In our case, a momentum of 0.9 seems optimal for the three networks, NN_{mf}, NN_1, and NN_2. If the momentum used is too large, the network will not converge.

There is no significant difference between the SEPs for different momentum parameters. For example, SEP = 1.89% for momentum = 0.2 and SEP = 1.86% for momentum = 0.9 are not significantly different. In this case, a relatively larger momentum parameter was chosen to avoid the possibility of a model being trapped in a local minimum in the training process.

Identification of Official and Unofficial Rhubarb Samples

The identification results obtained by using the Takagi–Sugeno fuzzy system for official and unofficial rhubarb samples are shown in Fig. 4. It should be

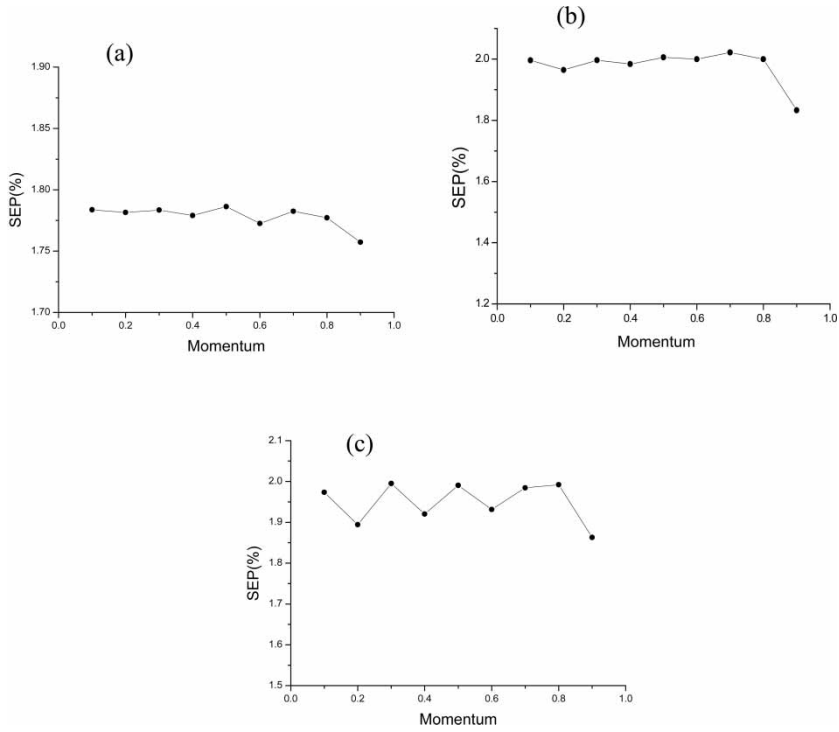

Figure 3. Effect of momentum on prediction. (a) NN_{mf} subnet, (b) NN_1 subnet, (c) NN_2 subnet.

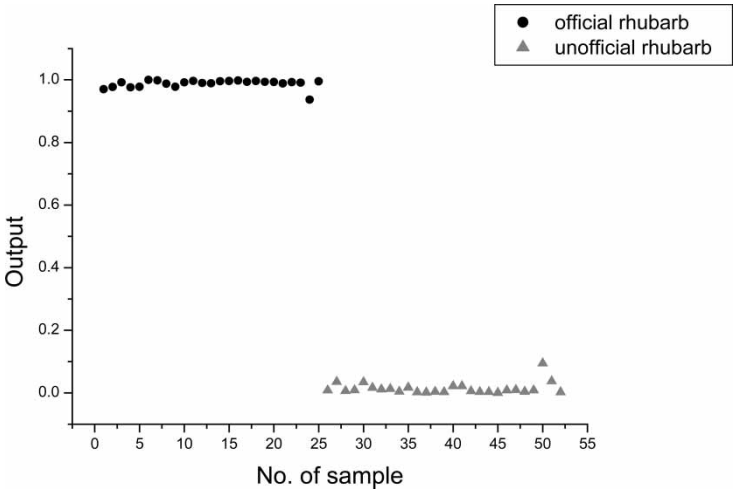

Figure 4. Prediction results of rhubarb samples by using Takagi–Sugeno fuzzy system.

noticed that the classification results shown in Fig. 4 were obtained with independent validation samples. The validation samples were selected by using the leave-one-out method. The expected output value of the official rhubarb was set to 1.0 and that of the unofficial rhubarb to 0.0. With optimized parameters, 100% of the rhubarb samples were identified correctly.

To show the effectiveness of the Takagi–Sugeno system, a back-propagation neural network (BP-NN) model was used for a comparison study. The BP-NN was constructed with 44 input units, 5 hidden units, and 1 output unit. The same official and unofficial rhubarb samples were used, and the same NIR spectra with wavelet transform compressed were used for the BP-NN training and prediction tests. Fifty rhubarb samples from among a total of 52 samples were identified correctly, so the correct identification was 96.15% with the BP-NN model.

CONCLUSIONS

Near-infrared spectroscopy and the Takagi–Sugeno system have been used for identification of official and unofficial rhubarb samples. The powdered rhubarb samples can be analyzed directly and without any complicated pretreatment. The NIR spectra were wavelet compressed from 700 to 44 variables, and then the compressed spectra were used to establish a classification model. Independent samples were used to test the model, and 100% of rhubarb samples were classified correctly, which was slightly better than those

by BP networks. This method can be used for quality control of rhubarb samples and products. Further work involving an increase in the number of samples in the database and provision of an independent validation of the models is under consideration.

ACKNOWLEDGMENTS

This work was supported by the Science and Technology Program, Beijing Municipal Education Commission (KM200310028105), and the Beijing Stars of Science and Technology Program (952871500).

REFERENCES

1. Jiao, D. H.; Du, S. J. eds. *Studies on Rhubarbs*; Shanghai Science and Technology Press: Shanghai, 2000; pp. 169–239.
2. Xiang, L.; Fan, G.; Li, J.; Kang, H.; Yan, Y.; Zheng, J. The application of an artificial neural network in identification of medicinal rhubarbs by near-infrared spectroscopy. *Phytochem. Analysis* **2002**, *13*, 272–276.
3. National Pharmacopoeia Committee. *Pharmacopoeia of Peoples' Republic of China*; (Part I, 2nd Ed.); Chemical Industry Press: Beijing, 2000; pp. 18–19.
4. Shenk, J. S.; Workman, J. J., Jr.; Westerhaus, M. O. In *Handbook of Near-Infrared Analysis*, 2nd Ed.; Burns, D. A., Ciurczak, E. W., Eds.; Marcel Dekker, Inc.: New York, 2001; pp. 419–474.
5. Shenk, J. S. NIRS analysis of natural agriculture products, in near infrared spectroscopy: Bridging the gap between data analysis and NIR applications, *Proceedings of the 5th International NIRS Conference*, Haugesund-Norway, June 20–24, 1992; Hildrun, K., Isaksson, T., Naes, T., Tandberg, A., Eds.; Ellis Horwood Ltd.: New York, 1992, 235–240.
6. Thyholt, K.; Isaksson, T. Differentiation of frozen and unfrozen beef using near infrared spectrometry. *J. Sci. Food Agric.* **1997**, *73* (4), 525–532.
7. Wang, Y. B.; Liu, W.; Yuan, H. F.; Lu, W. Z. Application of artificial neural network in near infrared analysis and dark-colored oil analysis. *Pet. Process. Petrochem.* **2002**, *33* (7), 62–67.
8. Buchanan, B. NIR analysis of petrochemicals. In *Handbook of Near-Infrared Analysis*, 2nd Ed.; Burns, D. A., Ciurczak, E. W., Eds.; Marcel Dekker, Inc.: New York, 2001, pp. 647–658.
9. Ciurczak, E. W.; Drennen, J. Near-infrared spectroscopy in pharmaceutical applications. In *Handbook of Near-Infrared Analysis*, 2nd Ed.; Burns, D. A., Ciurczak, E. W., Eds.; Marcel Dekker, Inc.: New York, 2001; pp. 609–631.
10. Blanco, M.; Coello, J.; Iturriaga, H.; Maspoch, S.; de la Pezuela, C. Near-infrared spectroscopy in the pharmaceutical industry. *Analyst* **1998**, *123*, 135R–150R.
11. Ciurczak, E. W. Biomedical applications of near-infrared spectroscopy. In *Handbook of Near-Infrared Analysis*, 2nd Ed.; Burns, D. A., Ciurczak, E. W., Eds.; Marcel Dekker, Inc.: New York, 2001; pp. 633–646.
12. Ren, G.; Chen, F. Simultaneous quantification of ginsenosides in american ginseng (Panax quinquefolium) root powder by visible/near-infrared reflectance spectroscopy. *J. Agric. Food Chem.* **1999**, *47*, 2771–2775.

13. Westerhaus, M. O.; Reeves III, J. B. NIR calibrations of agriculture products with neural networks, in near infrared spectroscopy: Bridging the gap between data analysis and NIR applications, *Proceedings of the 5th International NIRS Conference*, Haugesund-Norway, June 20–24, 1992; Hildrun, K., Isaksson, T., Naes, T., Tandberg, A., Eds.; Ellis Horwood Ltd.: New York, 1992, 79–84.
14. Kvaal, K.; Næs, T.; Isaksson, T.; Ellekjær, M. R. A study of the performance of artificial neural networks in NIR spectroscopy, in near infrared spectroscopy: Bridging the gap between data analysis and NIR applications, Hildrum, K. I., Ed.; Ellis Horwood Ltd.: New York, 1992, pp. 97–101.
15. Duponchel, L.; Ruckebusch, C.; Huvenne, J. P.; Legrand, P. Standardization of near-IR spectrometers using artificial neural networks. *J. Mol. Struc.* **1999**, *480*, 551–556.
16. Zhang, Z. Y.; Urbas, A.; Harrington, P. B.; Voorhees, K. J.; Rees, J. Temperature-constrained cascade correlation network and its application to bacteria identification. *Chem. J. Chin. Universities* **2002**, *23* (4), 570–572.
17. Lavine, B. K.; Workman, J., Jr. *Chemometrics. Anal. Chem.* **2002**, *74* (12), 2763–2770.
18. Xu, J. B. J. A calculative inference method for fuzzy knowledge base by means of fuzzy frame graph. *J. Shanghai Eng. Technol. Univ.* **1999**, *15* (2), 83–87.
19. Wang, L. X. Stable adaptive control of nonlinear systems. *IEEE Trans. Fuzzy Systems* **1993**, *1* (2), 146–155.
20. Wang, S. T. Neural fuzzy systems and applications. Beihang University Press: Beijing, 1998; Vol. 48–56, pp. 176–186.
21. Rogers, W. H.; Wagner, T. J. A fine sample distribution-free performance bound for local discrimination rule. *Ann. Statistics* **1978**, *6* (3), 506–514.