AN IMPROVED APPROACH OF CARS FOR LONGJING TEA DETECTION BASED ON NEAR INFRARED SPECTRA

Dong Ren,* Chang Zhang,* Shun Ren,* Zhong Zhang,* Ji-hua Wang,** and An-xiang Lu**

Abstract

In this paper, the near-infrared spectroscopy is used to obtain the near-infrared spectra data of tea for the detection of West Lake Longjing tea and the general Longjing tea. Noise and other redundant information contained in the full spectrum will have a negative impact on the accuracy of the models during the data processing. Using the characteristic wavelength variables to build the models is more effective than the full spectrum. The competitive adaptive reweighted sampling (CARS) is one of the most common and effective methods for the characteristic wavelength variables selection. However, the regression coefficients of variables will change with the selected samples of the model varying randomly in CARS method. Therefore, the absolute value of the regression coefficients is not always able to fully reflect the importance of the variables. This paper introduces the variable effectiveness and proposes a wavelength selection approach called effectiveness competitive adaptive reweighted sampling (ECARS) to make up for this shortfall. This study is mainly to classify the 110 samples of West Lake Longjing tea and the general Longjing tea. The training set consists of 72 samples and the prediction set contains 38 samples. After the preprocessing of the second derivative, CARS, uninformative variable elimination, backward interval partial least squares, and ECARS algorithm proposed in this paper are used for the variables selection. Then the variable subset and the full spectrum are, respectively, used to build support vector machine (SVM) model and linear discriminant analysis model for the identification of West Lake Longjing tea and the general Longjing tea. The experiment results show that: (1) the accuracy of models that are processed by the variables selection methods is higher than those of the full spectrum models and all the other models; (2) the accuracy of the ECARS-SVM model is highest, and the accuracies of the training set and prediction set are 100% and 98.4%, respectively; (3) the ECARS algorithm proposed in this paper can efficiently reduce the number of variables, simplify the models, and improve the accuracy and stability of the models.

Key Words

NIR, spectra variable selection, ECARS, West Lake Longjing tea, identification

1. Introduction

China is the original place of tea. China is also the first country to start the cultivation and processing of tea. Tea is one of the three largest soft drinks in the world. Along with the development of society and the improvement of people's living standards, consumers' requirements of the quality of tea are also increasing. In China's current tea quality standards, the sensory quality of tea is described by qualitative, which is difficult for consumers to understand and use. Therefore, the standard is difficult to play a role in the pricing [1].

Traditionally, the primary methods for the classification and identification of tea are sensory evaluation and chemical methods. The results of sensory evaluation are easy to be disturbed by human factors and the external environment, and there is a big objectivity in the assessment results. Chemical methods are to measure the physical and chemical elements of tea. By using relatively simple data-processing methods to analyse the measurement data, then seek the inherent law and assess the quality of tea. However, these methods are time-consuming and difficult to be promoted. Therefore, it is necessary to establish a scientific, quantified, and convenient method for the tea detection.

Near-infrared spectroscopy (NIRS) analysis is a fast, low-cost, non-destructive, and reproducible detection method [2]. Research shows the mathematical methods and computer technology can effectively extract the weak information of near-infrared spectrum, NIRS has become a very promising technology. It has been used in the petrochemical, tobacco, fruits, and other areas in the application and has generated certain economic benefits [3]–[6].

* College of Computer and Information Technology, China Three Gorges University, Yichang, 443002, China; e-mail: rendong5227@163.com, {854760972, renshun_ctgu, 1335117856}@qq.com
** Beijing Research Center for Agricultural Standards and Testing, Beijing, 100097, China; e-mail: wangjh@nercita.org.cn, luax@ brcast.org.cn
Corresponding author: Shun Ren
Recommended by Prof. Simon X. Yang
(DOI: 10.2316/Journal.206.2018.1.206-5444)

Jian and Hao [7] combined partial least squares (PLS), Euclidean distance and other methods based on near infrared (NIR) to achieve the identification of special materials of four different kinds of teas. Accuracies of the training set and prediction set were 90.3% and 83.5%, respectively. The study provided a reference method to achieve identification of special varieties of teas and the source tracing of finished teas. Chen and Zhao [8] combined NIRS with support vector machine (SVM) to establish the identification model of Biluochun tea. The accuracy of the training set contained 138 tea samples was 93.48%, and the accuracy of the prediction sample set that consisted of 90 independent sample was 84.44%. It showed that SVM based on NIR was feasible to identify the authenticity of the famous tea. Zhao *et al.* [9] combined NIR and principal component – Mahalanobis distance to identify Longjing tea, Biluochun tea, Mao Feng tea, and Tieguanyin tea. The result showed that the accuracies of the training set and prediction set were 98.75% and 95%, respectively.

Near-infrared spectral region contains a large number of absorption peaks of octave frequency and co-frequency of hydrogen-containing group. Although the samples can reflect the spectral information of the test substance (*e.g.*, concentration, category), there is a problem of spectral information overlap. A lot of noises and samples background redundant information it contains affect the accuracy and stability of the model, and the information is difficult to be eliminated by preprocessing. If these data is involved in modelling, the amount of calculation is larger and the models are more complex. Besides, the accuracy is not necessarily high [10]. Therefore, the variable selection of NIR spectrum is very necessary. It is good for removing the redundant information.

There are many common variable selection methods, such as interval partial least-squares [11], backward interval partial least squares (BiPLS) [12]–[14], moving window partial least-squares [15], uninformative variable elimination (UVE) [16], successive projections algorithm [17], competitive adaptive reweighted sampling method (CARS) [18], and so on.

The regression coefficients of variables will change with the selected samples of the model varying randomly in CARS method. Therefore, the absolute value of the regression coefficients is not always able to fully reflect the importance of the variables. In this study, the variable effectiveness is introduced to make up for this shortfall.

This study proposes a variable selection method called effectiveness competitive adaptive reweighted sampling (ECARS) to process NIR spectral data of ordinary Longjing and West Lake Longjing. Then the SVM [19], [20] model and linear discriminant analysis (LDA) model are built to identify them. This method makes up for the shortcomings of traditional methods and improves the accuracy and efficiency of recognition. It provides a theoretical basis and foundation for the authenticity of identification of West Lake Longjing. There is a great significance for promoting the standardization of the tea market.

2. Principles and Methods

2.1 Definition of Variable Effectiveness

The basic form of PLS regression model can be expressed as:

$$Y = X\beta + \varepsilon \tag{1}$$

where matrix $X_{n\times p}$ is the spectral of samples, n is the number of samples, p is the number of variables, $Y_{n\times 1}$ represents the nature of the target vector, and ε is the random error vector. Randomly select m samples from n samples to establish PLS regression model by Monte Carlo sampling method, then obtain the corresponding regression coefficient vector β, model correlation coefficient R^2 and root mean square error (RMSE). After M sampling, the effectiveness e_i of the i-th variable is represented by the following formula:

$$e_i = \left| \left(1 + \frac{\overline{R^2}_i}{\overline{RMSE_i}} \right) \times \overline{\beta_i} \right| \quad i = 1, \ldots, p \tag{2}$$

where $\overline{R_i^2}, \overline{RMSE_i}$, and $\overline{\beta_i}$ are the averages of model correlation coefficient, the average of RMSE, and corresponding regression coefficient. As can be seen from the above equation, correlation coefficient is greater and RMSE is smaller; the greater the value of the variable effectiveness, the stronger its importance.

2.2 Effectiveness Competitive Adaptive Reweighted Sampling

The variable effectiveness is the measure of ECARS algorithm. The greater the effectiveness, the more possibility variables are selected. This algorithm main includes N cycles, calculate the effectiveness of each variable every cycle first, then using adaptive reweighted sampling (ARS) method to screen out variables with relatively large effectiveness to form a subset of variables. At the end of the cycle, N variables subsets are obtained. PLS models are conducted for each variables subset. According to the root mean square error of cross-validation (RMSECV) of the models, select the variables subset of the minimum RMSECV as the best variables subset. The specific steps of ECARS algorithm are as follows:

Step 1. In the first cycle, calculate the effectiveness of each variable in the original variables set;

Step 2. Using the exponential decay function to retain variables of the larger effectiveness, retention of variables is $r_i = ae^{-ki}$, where $a = (p/2)^{1/(N-1)}$, $k = [\ln(p/2)]/(N-1)$, p is the number of original variables;

Step 3. Using ARS to screen out variables with relatively large effectiveness from the variables subset of Step 2, the variables selected are as the i-th variables subset S_i, then, cycles plus one;

Step 4. In the i-th cycle, calculate the effectiveness of each variable in the subset of variables S_{i-1}, then repeat Step 2 and Step 3;

Step 5. If $i \leq N$, repeat Step 4, if $i = N+1$, executive Step 6;

Step 6. After N cycles, use N variables subsets $(S_1, S_2, \cdots, S_N)$ to build PLS models, calculate the RMSECV of each model;

Step 7. According to the RMSECV, select the variables subset of the minimum RMSECV as the best variables subset.

3. Materials and Preprocessing

3.1 Materials

All samples used in the experiments were produced in accordance with Longjing tea processing. Fifty-seven samples of West Lake Longjing were from Hangzhou City, as well as other 53 general Longjing samples purchased in the supermarket. All samples were divided into two portions by random sampling (RS) method [21], [22]. The training set consists of 40 West Lake Longjing samples and 32 general Longjing samples, which was used to build the identification model. The remaining samples made up the prediction set as the unknown samples to verify the identification model. Therefore, the samples set was divided into a training set of 72 samples and a prediction set contained 38 samples. To avoid the contingency of experimental results and special samples set, this experiment was repeated using RS carrying out 10 times, and the results of experiments were averaged.

3.2 Sample Treatment

Before collecting the NIR spectrum, all of the samples had been pulverized. About 20 g tea samples were weighed and then they were placed in pulverizer for about 30 s. After a sieve, accurately weighed 10 g samples for analysis the near-infrared spectrum.

3.3 Acquisition of NIRS

An infrared spectrometer manufactured by PerkinElmer Inc. (USA) was used for the experiments. The wavenumber ranged from 7,300 to 5,500 cm^{-1}. A total of 32 scans were performed, with a resolution of 2 cm^{-1} and an interval of 2 cm^{-1}. Thus, each spectrum had 901 variables. The experimental instruments also included a PC. The spectrometer software used to collect the spectral data was Spectrum Version 10.4.1,which was produced by PerkinElmer, Inc. (USA). The indoor temperature was kept at about 25°C, and the humidity remained basically unchanged (less than 60%). Each sample was collected three times in parallel, and the original spectrum of the sample was taken as the average of these three samples. To ensure the consistency of the measurement environment and manual operations, the background was scanned every 10 samples to eliminate drift. The original spectrum is as shown in Fig. 1.

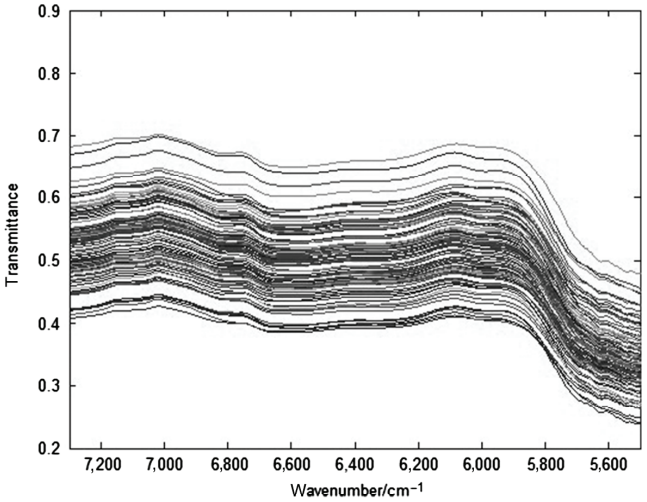

Figure 1. The original spectrum.

Table 1
The Model Results of Different Preprocessors

Methods	Accuracy			
	SVM		LDA	
	Training Set	Prediction Set	Training Set	Prediction Set
RAW	**0.954**	**0.782**	**0.979**	**0.911**
MSC	0.965	0.737	0.871	0.479
SNV	0.967	0.887	0.989	0.911
SNV+DT	0.969	0.855	0.997	0.945
S-G	0.950	0.766	0.969	0.903
SW	0.951	0.776	0.979	0.908
1-Dr	0.992	0.903	1	0.947
2-Dr	**1**	**0.937**	**1**	**0.953**

3.4 Spectral Preprocessing

Different spectral processing methods have different impacts on model performance. The following methods were considered to determine which was best for all 110 samples: multiplicative scatter correction (MSC), standard normal variable (SNV) transformation, SNV added to the trend method (SNV&DT), Savitzky–Golay smoothing convolution, sliding window smoothing (SW), first-order derivative (1-Dr) and second-order derivative (2-Dr) spectra methods. Table 1 presents the results calculated by the SVM model and LDA model. As can be seen, 2-Dr produced the best performance; the accuracies of the training set and prediction set are 100% and 93.7%, respectively, of the SVM model. At the same time, the results of LDA model are 100% and 97.4%.

Figure 2 shows a spectrum that has been processed by 2-Dr. The spectral absorption peaks have increased and are more obvious and it eliminates background interference and makes them more conducive to spectral

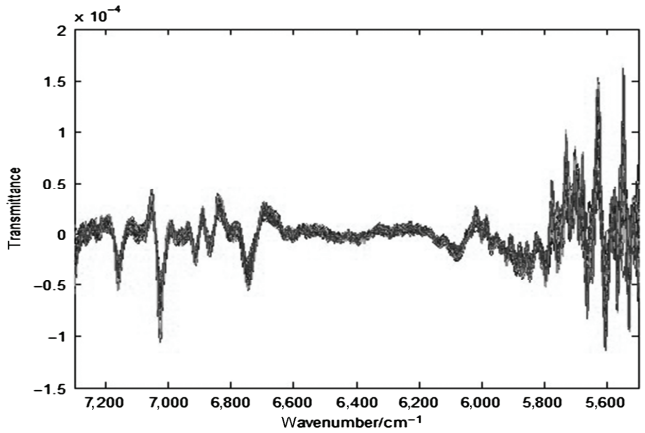

Figure 2. Spectrum that has been processed by 2-Dr.

analysis. Therefore, 2-Dr was selected as the final preprocessing method for the following comparative experiments.

4. Experiments and Results Analysis

In this paper, the training set of 72 samples was preprocessed by the second derivative firstly. Then the CARS, UVE, BiPLS, and ECARS methods were used to select the characteristic variables. The SVM model and the LDA model were built to predict the prediction set of 38 unknown samples and analyse the results of the experiments. To avoid the contingency of experimental results and special sample set, this experiment was repeated using RS carrying out 10 times, and the results of experiments were averaged.

4.1 Backward Interval Partial Least Squares

BiPLS algorithm was described in reference [9]. It was an improvement and development of the algorithm that was proposed for the characteristic variables selection by Lars Norgaard in 2000. Please refer to the specific algorithm processes as described in reference [12].

The experiments used BiPLS to select the characteristic variables. The number of interval was set from 20 to 65. The interval was 5 and there were a total of 10 data points. Ten-fold cross-validation was used to select the principal component. According to the RMSECV, the best wavelength was selected to build the classification model. The results were shown in Table 2. As can be seen, when the interval was 30, the best accuracies of the training set and the prediction set of the SVM model were 99.9% and 96.6%, respectively, and those of the LDA model were 100% and 97.9%, respectively. When the interval was 45, the best accuracies of the training set and the prediction set of the SVM model were 100% and 96.3%, respectively.

4.2 UVE Algorithm

The UVE algorithm was used for variables preferred in the full-spectrum wavelength range of the training set. Results of variable stability analysis of UVE are shown in Fig. 3. Vertical shaft solid line was the dividing line between wavelength variables and random noise variables.

Table 2
The Model Results of Different Intervals

Interval	Accuracy			
	BiPLS + SVM		BiPLS + LDA	
	Training Set	Prediction Set	Training Set	Prediction Set
20	0.999	0.953	1	0.974
25	0.997	0.950	1	0.971
30	**0.997**	**0.966**	**1**	**0.979**
35	0.999	0.948	1	0.971
40	1	0.948	1	0.974
45	**1**	**0.963**	0.999	0.966
50	1	0.955	1	0.966
55	0.999	0.955	1	0.968
60	0.999	0.950	1	0.974

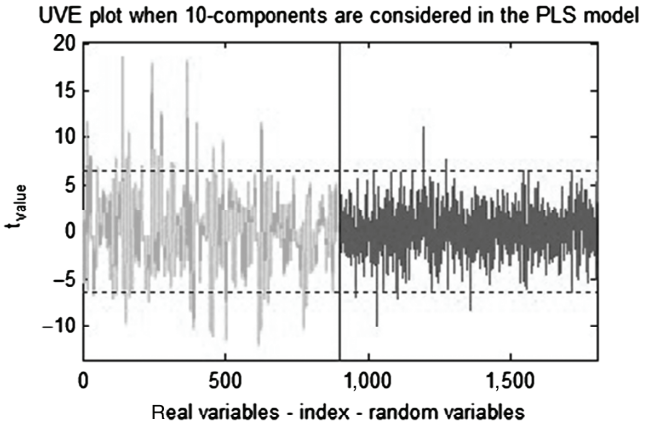

Figure 3. UVE algorithm variable stability schematic.

Wavelength variables were on the left and random noise variables were on the right. Two horizontal dotted lines represented the upper and lower threshold value of the variable stability. Variables whose stabilities were between the two threshold were useless, and they needed to be removed. The variables whose stabilities were outside the two thresholds were useful information, and would be retained.

After UVE wavelength variables preferring, the cross-validation was used to determine the optimal principal components of classification models. The maximum principal component is set to 10. Then the experiments were conducted on 10 different samples set. The result showed that the accuracy of training set of SVM classification model was 99.7%, the prediction set was 97.3%, while the LDA classification model was 99.9% and 97.8%.

4.3 CARS and ECARS

Table 3 showed the results of the models after CARS and ECARS processing. The number of sampling was from 110

Table 3
The Model Results of Different N

N	Accuracy							
	CARS + SVM		CARS + LDA		ECARS + SVM		ECARS + LDA	
	Training Set	Prediction Set	Training Set	Prediction Set	Training Set	Prediction Set	Training Set	Prediction Set
110	0.999	0.954	**1**	**0.968**	1	0.981	1	0.980
120	0.999	0.956	1	0.963	0.999	0.981	1	0.980
130	0.999	0.954	1	0.966	1	0.981	1	0.979
140	0.999	0.956	1	0.961	**1**	**0.984**	1	0.981
150	1	0.957	1	0.962	1	0.980	1	0.981
160	0.999	0.951	1	0.963	1	0.979	**1**	**0.983**
170	0.999	0.957	1	0.962	1	0.983	1	0.977
180	1	0.957	1	0.962	1	0.982	1	0.980
190	**1**	**0.958**	1	0.965	**1**	**0.984**	1	0.980
200	0.999	0.956	1	0.960	1	0.980	1	0.977

to 200 at intervals of 10, a total of 10 groups. As could be seen, when N was 190, the result of the model that CARS combined with SVM was best. The accuracies of the training set and the prediction set were 100% and 95.8%, respectively. The accuracies of the model that EACRS combined with SVM were 100% and 98.4%, when N was 140 or 190. Additionally, the best result of the model that CARS combined with LDA were 100% and 96.8%, while the model that ECARS combined with LDA were100% and 98.3%.

5. Result Analysis and Conclusion

5.1 Comparison of Spectra Variable Selection Models and the Full Spectrum Models

Table 4 showed the best performance of the models with four variables selection methods and the full spectrum combined with the SVM and LDA. Compared with the full-spectra models, it could be seen, four variable selection methods selected the characteristic variables very well and the prediction accuracies of four predictive effect detection models were higher. The results showed that the spectra variables selection could effectively simplify the models and improve prediction accuracy.

5.2 Comparison of ECARS and CARS

As the results showed in Table 3, the best results of the CARS-SVM models were 100% and 95.8%, while the ECARS-SVM were 100% and 98.4%. At the same time, the accuracies of the CARS-LDA models were 100% and 96.8%, and that of the ECARS-LDA models were 100% and 98.3%. The experimental results showed that whether combined with SVM or LDA for modelling, the proposed ECARS method could effectively address the disadvantages of CARS. This showed that ECARS the proposed algorithm in this paper had better performance.

Table 4
The Best Results of Different Models

Models	Accuracy			
	SVM		LDA	
	Training Set	Prediction Set	Training Set	Prediction Set
Full spectrum	1	0.937	1	0.953
BiPLS	0.997	0.966	1	0.979
	1	0.963		
UVE	0.997	0.973	0.999	0.978
CARS	1	0.958	1	0.968
ECARS	**1**	**0.984**	**1**	**0.983**

5.3 Comparison of ECARS and the Other Spectra Variable Selection Methods

From the results shown in Table 4, it can be seen that all the variables selection methods could improve the accuracy effectively. When compared with BiPLS, UVE, and CARS, the proposed ECARS behaves better. The prediction accuracies of the ECARS-SVM and ECARS-LDA were 98.4% and 98.3%, respectively. They are all the highest results in the all spectra variables selection models. Furthermore, it is shown that the use of the proposed ECARS can improve

the performance of models with high prediction accuracy and stability.

5.4 The Performance of the ECARS-SVM Model for the Detection of West Lake Longjing Tea

In this paper, we compare different preprocessing algorithms to determine the second derivative as the spectra preprocessing methods. ECARS algorithm is proposed for variables selection. The ECARS-SVM model has the best prediction accuracy. The accuracies of the training set and prediction set were, respectively, 100% and 98.4% for the detection of West Lake Longjing tea and general Longjing tea. Therefore, it can be applied for online and real-time detection of West Lake Longjing tea and can also be effectively used to select variables and applied to NIR data analysis in the identification of tea.

Acknowledgement

This work was supported in part by the Natural Science Foundations of China (Grant no. 41371349), The National Key Research and Development Projects of China (Grant no. 2016YFD0800902), and The Major Technological Innovation Project in Hubei Province of China (Grant no. 2017ABA157).

References

[1] Q-B. Liu and Z-F. Yang, The situation of Chinese tea market and the brand exploration, *Newsletter of Sericulture and Tea*, *6*, 2013, 22–23.

[2] P. Williams and K. Norris, *Near-infrared technology in the agricultural and food industries,* 2nded. (St., Paul, MN, USA: The American of Cereal Chemists, Inc., 2001).

[3] H. Ping, J-H. Wang, and G-X. Ren, Prediction of the total starch and amylose content in barley using near-infrared reflectance spectroscopy, *Intelligent Automation & Soft Computing*, *19*(3), 2013, 231–237.

[4] H-F. Yuan, W-Z. Lu, Near infrared spectrometry technology is rapidly entering into petrochemical industry, *Petroleum Processing and Petrochemicals*, *9*, 1998, 49–52.

[5] S-Z. Lei and H-G. Yao, Applications of near infrared spectrum technique for non-destructive measurement of fruit quality, *Chinese Journal of Spectroscopy Laboratory*, *4*, 2009, 775–779.

[6] H-H. Li and W. Li, Development and application of near infrared spectroscopic in tobacco fields, *Journal of Anhui Agricultural Sciences*, *29*, 2014, 10318–10321.

[7] Z. Jian and C. Hao, Study on identification and traceability of tea material cultivar by combined analysis of multi-partial least squares models based on near infrared spectroscopy, *Spectroscopy and Spectral Analysis*, *10*, 2010, 2650–2653.

[8] Q-S. Chen and J-W. Zhao, Identification of authenticity of tea with near infrared spectroscopy based on support vector machine, *Acta Optica Sinica*, *6*, 2006, 933–937.

[9] J-W. Zhao, Q-S. Chen, H-D. Zhang, and M-H. Liu, Study on the identification of tea using near infrared reflectance spectroscopy, *Spectroscopy and Spectral Analysis*, *9*, 2006, 1601–1604.

[10] U. Mall, C. Wöhler, A. Grumpe, R. Bugiolacchi, and M. Bhatt, Characterization of lunar soils through spectral features extraction in the NIR. *Advances in Space Research*, *54*(10), 2014, 2029–2040.

[11] L. Norgaard, A. Saudland, J. Wagner, J.P. Nielsen, L. Munck, and S.B. Engelsen, Interval Partial least squares regression (iPLS): A comparative chemometric study with an example from near-infrared spectroscopy, *Applied Spectroscopy*, *54*(3), 2000, 413–419.

[12] X.-B. Zhou, J.-W. Zhao, and Y.-X. Huang, Using forward and backward interval partial least squares to establish the near infrared spectroscopy model of apple sugar content, *Chinese Mechanical Engineering Society Annual Conference and China Academy of Engineering Mechanical and Vehicle Engineering Radical Annual Meeting Proceedings*, Jiangsu, 2006.

[13] F. Qu and D. Ren, The characteristic spectral selection method based on forward and backward interval partial least squares, *Spectroscopy and Spectral Analysis*, *36*(2), 2016, 593–598.

[14] D. Ren, F. Qu, and K. Lv, A gradient descent boosting spectrum modeling method based on back interval partial least squares, *Neurocomputing*, *171*, 2016, 1038–1046.

[15] J-H. Jiang, R.J. Berry, and H.W. Siesler Wavelength interval selection in multicomponent spectral analysis by moving window partial least squares regression with applications to mid-infrared spectroscopic data, *Analytical Chemistry*, *74*(14), 2002, 3555.

[16] H. Li, Y. Liang, and D. Cao, Model-population analysis and its applications in chemical and biological modeling, *Trends in Analytical Chemistry*, *38*(9), 2012, 154–162.

[17] F. Qu, D. Ren, and J. Wang, An ensemble successive project algorithm for liquor detection using near infrared sensor, *Sensors*, *16*(1), 2015, 294–300.

[18] H-D. Li, Y-Z. Liang, and Q-S. Xu, Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration, *Analytica Chimica Acta, 648*(1), 2009,77–84.

[19] N-Y. Deng and Y-J. Tian, *The new method of data mining–support vector machine* (Beijing: Science Press, 2004, 164–219).

[20] Z-Q. Bian and X-G. Zhang, *Pattern recognition* (Beijing: Tsinghua University Press, 2002, 284–304).

[21] Y-F. Li, J-Q. Yuan, and Y-F. Xue, The selection of neural network training samples in the fermentation process, *Control and Instruments in Chemical Industry*, *6*, 2004, 21–24.

[22] G. Nokas and E. Dermatas, Continuous speech recognition in noise using a spectrum-entropy beam-former, *International Journal of Research in Computer Applications & Robotics*, *22*(2), 2007, 103–111.

Biographies

Dong Ren received his Ph.D. degree in Jilin University. He is a vice professor in the College of Computer and Information Technology at China Three Gorges University, Yichang, China. His research interests include artificial intelligence, pattern recognition, and 3S technology.

Chang Zhang received her B.S. degree in Qingdao University of Technology. She is now a master's candidate of the Institute of Collaborative Innovation Center for Key Technology of Smart Irrigation District in Hubei, China Three Gorges University. Her main research interests are spectroscopy analysis and pattern recognition.

Shun Ren corresponding author of this paper, received his Ph.D. degree in Jilin University. He is currently working in the College of Computer and Information Technology at China Three Gorges University, Yichang, China. His research interests include Internet of things and wireless sensor network.

Ji-hua Wang received his Ph.D. degree in Beijing Agricultural University. He is a researcher in Beijing Research Center for Agricultural Standards and Testing at Beijing Academy of Agriculture and Forestry, Beijing, China. His research interests include agricultural information technology.

Zhong Zhang received his B.S. degree in China Three Gorges University. He is now a master's candidate of the Institute of Collaborative Innovation Center for Key Technology of Smart Irrigation District in Hubei, China Three Gorges University. His main research interests include spectroscopy analysis and pattern recognition.

An-xiang Lu is an associate professor. His research directions include farmland environment monitoring.