



Application of LS-SVM and Variable Selection Methods on Predicting SSC of Nanfeng Mandarin Fruit

Tong Sun, Wenli Xu, Tian Hu, Muhua Liu

► To cite this version:

Tong Sun, Wenli Xu, Tian Hu, Muhua Liu. Application of LS-SVM and Variable Selection Methods on Predicting SSC of Nanfeng Mandarin Fruit. 7th International Conference on Computer and Computing Technologies in Agriculture (CCTA), Sep 2013, Beijing, China. pp.249-262, 10.1007/978-3-642-54344-9_31 . hal-01220926

HAL Id: hal-01220926

<https://inria.hal.science/hal-01220926>

Submitted on 27 Oct 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Application of LS-SVM and variable selection methods on predicting SSC of Nanfeng mandarin fruit

Tong Sun^a, Wenli Xu^b, Tian Hu^c, Muhua Liu^d

College of Engineering, Jiangxi Agricultural University, 1225 Zhimin Street,

310045 Nanchang, P.R. China;

^asuntong980@163.com, ^bxuwenli.1990@163.com, ^c784643819@qq.com,

^dsuikelmh@sina.com

Abstract. The objective of this research was to investigate the performance of LS-SVM combined with several variable selection methods to assess soluble solids content (SSC) of Nanfeng mandarin fruit. Visible/near infrared (Vis/NIR) diffuse reflectance spectra of samples were acquired by a QualitySpec spectrometer in the wavelength range of 350~1800 nm. Four variable selection methods were conducted to select informative variables for SSC, and least squares-support vector machine (LS-SVM) with radial basis function (RBF) kernel was used develop calibration models. The results indicate that four variable selection methods are useful and effective to select informative variables, and the results of LS-SVM with these variable selection methods are comparable to the results of full-spectrum partial least squares (PLS). Genetic algorithm (GA) combined with successive projections algorithm (SPA) is the best variable selection method among these four methods. The correlation coefficients and RMSEs in LS-SVM with GA-SPA model for calibration, validation and prediction sets are 0.935, 0.560%, 0.912, 0.631% and 0.933, 0.594%, respectively.

Keywords: Vis/NIR, LS-SVM, variable selection, soluble solids content, Nanfeng mandarin fruit

1 Introduction

Soluble solids content (SSC) is one of the most important properties of fruits that match human's taste. In recent years, Visible /near infrared (Vis/NIR) spectroscopy has become a well-accepted method for SSC assessment of fruit because it works fast and non-destructively, allows no sample preparation and brings no environmental chemistry pollution. And it has been used to measure SSC in a variety of fruits such as mandarin fruit [1-2], apple [3-4], pear [5-6], kiwifruit [7-8], melon [9-10] and so on. Due to high resolution in modern spectroscopic instrument, the spectral data usually has hundreds or thousands wavelength variables, so contains substantial information. However, some of the information is useless or irrelevant to component properties, which will worsen the predictive ability of the model for component properties and should be eliminated. Recently, many variable selection methods were developed to

eliminate irrelevant information or variables, and reserve important information or variables. Ying et al. (2008) [11] used genetic algorithm (GA) method to select important variables for sugar content (SC) of apples, then developed calibration model by partial least squares (PLS). Compared to full-spectrum PLS, the root mean square error of prediction (RMSEP) in GA-PLS was decreased from 0.512% to 0.395%. Liu et al. (2009) [12] applied successive projections algorithm (SPA) to choose variables for organic acids of plum vinegar. Least squares-support vector machine (LS-SVM) was used to develop calibration models. The models developed by SPA-LS-SVM for organic acids were better than that of full-spectrum PLS. Sorol et al. (2010) [13] investigated PLS combined with several variable selection methods to assess Brix degrees of sugar cane juice, and GA-PLS obtained the best results. Wu et al. (2011) [14] reported that uninformative variable elimination (UVE) was necessary to conduct before SPA on spectra of dried laver for predicting protein content. The average improvement percentage of RPD (standard deviation/RMSEP) was 38.66%. Huang et al. (2011) [15] selected two important wavelength variables (431 nm, 976 nm) by SPA for total soluble solid contents (TSS) of mulberry fruit, and acceptable results were obtained. Sun et al. (2012) [16] applied competitive adaptive reweighted sampling (CARS) to sort important variables for SSC of navel oranges. The CARS-PLS model for SSC was better than other methods (UVE-PLS, SPA-PLS, full-spectrum PLS). Besides, there were other research works concerning variable selection in Vis/NIR application [17-19]. From above research papers, it can be found that variable selection will simplify models, and reduce predicted errors.

LS-SVM is an evolution of the standard support vector machine. It has the capability for both linear and nonlinear multivariate calibration, and can handle some latent nonlinear information existed in the spectral data [20](Suykens et al., 2002). Therefore, LS-SVM combined with several variable selection methods was proposed to assess SSC of Nanfeng mandarin fruit, with the purpose of developing a fast and accurate nonlinear model using important variables for quality assessment of Nanfeng mandarin fruit.

The objective of this research was to investigate the performance of LS-SVM combined with several variable selection methods to assess SSC of Nanfeng mandarin fruit. Calibration models developed by LS-SVM with different variable selection methods were compared to identify an optimal method for SSC assessment of Nanfeng mandarin fruit.

2 Experiments and Methods

2.1 Samples

A total of 280 Nanfeng mandarin fruits were purchased from a local fruit market and stored for 24 hours at about 20 °C and 60 % relative humidity (RH) to equilibrate before Vis/NIR spectra measurement. Samples were separated randomly into three sets: a calibration set (n=140), a validation set (n=70), and a prediction set (n=70), with the ratio of 2:1:1. In order to ensure the adaptability of the model, the calibration set contained the samples with the minimum and maximum values of SSC. Spectra

collection and SSC determination of Nanfeng mandarin fruits were completed at the same day for same samples.

For SSC measurement, the pulp of Nanfeng mandarin fruit was macerated using a manual fruit squeezer after removing pericarp. And the filtered juice were used for SSC measurement with a hand-held refractometer (Atago Co. Ltd., Tokyo, Japan)

2.2 Spectra Acquisition

Diffuse reflectance spectra of Nanfeng mandarin fruits were acquired in the Vis/NIR system which consisted of a QualitySpec spectrometer (Analytical Spectral Devices, Inc., USA), light source, an optic fiber and a computer. The QualitySpec spectrometer was equipped with a Si detector (350~1000 nm) and an InGaAs detector (1000~1800 nm). The light source was tungsten quartz halogen lamp (4W/5.5V), and the color temperature was about 2901 K.

In this research, diffuse reflectance spectra of samples were acquired at two positions, and each position was acquired once. The average spectrum of each sample was used for further analysis. The positions of samples were put as the stalk-stem of fruit was perpendicular to horizontal plane, and one position was stalk up, another was stem up. A polytetrafluoroethylene (PTFE) board was used to obtain reference spectrum prior to sample spectra acquisition. The integration time and scan number for reference and samples were both 136 ms and 2, respectively. Spectrometer parameters setting, spectra data obtaining and storing were carried out via the Indic software v4.0 (Analytical Spectral Devices, INC., USA). The measurement was expressed as $\log(1/R)$. R was percent reflectance of sample spectra.

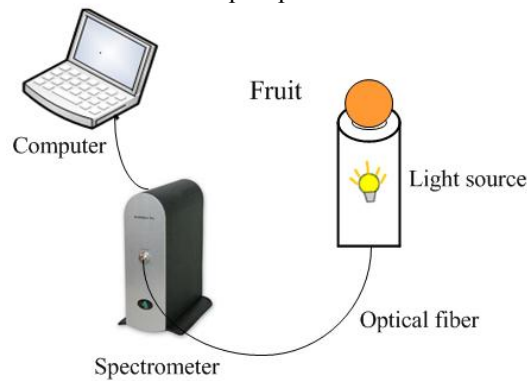


Fig. 1. Schematic diagram of Vis/NIR system

2.3 Variable Selection

2.3.1 UVE

Uninformative variable elimination (UVE) is a novel variable selection method based on stability analysis of regression coefficient [21]. The objective of UVE is to eliminate the variables that have no more useful information for modeling than noise.

After reducing the uninformative variables, it can avoid a model to be over-fitting and improve model predictive ability. In the UVE algorithm, a noise matrix is appended to instrumental response data for estimating the uninformative variables, and a PLS regression coefficient matrix $b=[b_1, \dots, b_p]$ is calculated from new matrix through leave-one-out cross validation. Then the stability value S of each variable can be calculated using equation (1). The large S value indicates the variable is more reliable and important for modeling. The variable whose S value is less than cut-off threshold will be realized as uninformative variable and eliminated. In this study, the cut-off threshold was set as 99% of maximal absolute S value of noise variables, and 1231 random variables (noise matrix) were added to spectral data.

$$s_j = \frac{\text{mean}(b_j)}{\text{std}(b_j)} \quad (1)$$

Where s_j is the reliability value of variable j , $\text{mean}(b_j)$ and $\text{std}(b_j)$ are the mean and standard deviation values of PLS regression coefficients of variable j .

2.3.2 GA

Genetic algorithm (GA) is an optimization method based on the principles of genetics and natural selection, and it has been successfully applied in the Vis/NIR spectroscopy for variable selection. The basic steps of GA algorithm are described as follows: first, a population of chromosomes is generated; second, PLS analysis is conducted and the fitness of each chromosome is calculated; third, chromosomes are chosen according to the fitness value; fourth, the chosen chromosomes are paired and reproduced by cross-over and mutation; then, step 2 to step 4 are repeated until the fitness reaches an acceptable value. The detailed principle of GA can be found in the literature [22-23]. In this study, the parameters of GA were set as follows: population size (30 chromosomes), probability of mutation (1%), probability of cross-over (50%), and number of runs (100).

2.3.3 SPA

Successive projections algorithm (SPA) is a novel forward variable selection algorithm for solving collinearity problems. In the SPA analysis, it has two phases. The first phase is to conduct projections on instrumental response data to generate subsets of variables. The second phase is to evaluate candidate subsets of variables obtained in the first phase by MLR according to root mean squared error of cross validation (RMSECV). And the best subsets of variables are obtained with the smallest RMSECV. The detailed principle of SPA can be found in the literature [24-25]. In this study, the parameters of minimum and maximum number of variables used in SPA were 1 and 30, respectively.

2.3.4 CARS

Competitive adaptive reweighted sampling (CARS) is a novel strategy for variable selection based on 'survival of the fittest' of Darwin's Evolution Theory. The absolute values of regression coefficients in PLS regression are used to evaluating the importance of each variable. In the CARS analysis, exponentially decreasing function based enforced wavelength selection and adaptive reweighted sampling (ARS) based

competitive wavelength selection are adopted to select the important variables. The detailed principle of CARS can be found in the literature [26]. In this study, the number of Monte Carlo sampling was set as 50, and 10-fold PLS cross validation was used to develop calibration model.

In the variable selection, UVE or GA was used to select wavelength variables for SSC of Nanfeng mandarin fruit first, then SPA or CARS was conducted to select wavelength variables based on the results of UVE or GA. So four combined variable selection methods (UVE-SPA, UVE-CARS, GA-SPA, GA-CARS) were used in this study, and conducted in the Matlab 7.6.0 (The MathWorks, Inc. USA).

2.4 LS-SVM

After variable selection, LS-SVM was used to develop calibration models for SSC of Nanfeng mandarin fruit. In order to evaluate the performance of LS-SVM with four variable selection methods, the results of LS-SVM were compared with the results of full-spectrum PLS.

In LS-SVM regression, radial basis function (RBF) kernel was used as the kernel function, simplex technique and leave one out cross-validation were used to find the optimal parameter values including regularization parameter γ and RBF kernel function parameter σ^2 . LS-SVM was conducted in the Matlab 7.6.0 (The MathWorks, Inc. USA).

3 Results and Discussion

3.1 Quality Attribute Distribution

The mean and standard deviation (S.D.) of SSC in all Nanfeng mandarin fruit were 14.45% and 1.59, respectively. Details of SSC statistic results for calibration, validation and prediction sets were presented in Table 1.

Table 1. Statistical results of soluble solids content of Nanfeng mandarin fruit in calibration, validation and prediction sets

| <i>Parameter</i> | <i>Data set</i> | <i>Samples</i> | <i>Maximum</i> | <i>Minimum</i> | <i>Mean</i> | <i>S.D.</i> |
|------------------|-----------------|----------------|----------------|----------------|-------------|-------------|
| <i>r</i> | | | | | | |
| SSC (%) | Calibration | 140 | 18.6 | 10.4 | 14.43 | 1.58 |
| | Validation | 70 | 18.0 | 10.6 | 14.26 | 1.53 |
| | Prediction | 70 | 18.6 | 10.6 | 14.51 | 1.64 |

3.2 Spectral Analysis

Fig. 2 shows the Vis/NIR diffuse reflectance spectra of all Nanfeng mandarin fruit. From Fig. 2, it can be seen that there are obvious peaks and valleys in the wavelength range of 1000 nm~1800 nm, and spectra seem flat in the wavelength range of 350

nm~1000 nm. This is due to weak absorption in the wavelength range of 350 nm~1000 nm, and this wavelength range also contains useful information such as a third overtone stretch of CH and second and third overtones of OH (Shao et al., 2011). In the wavelength range of 1000 nm~1800 nm, it develops peaks at about 1210 nm and 1465 nm, and develops valleys at 1100 nm, 1287 nm and 1676 nm. Besides, there are some noises at the two ends of spectra which lead to low signal/noise ratio, and the wavelength ranges of 350 nm~550 nm and 1780 nm~1800 nm are removed. So only wavelength range of 550 nm~1780 nm was used for further analysis, and there were 1231 wavelength variables in this wavelength range.

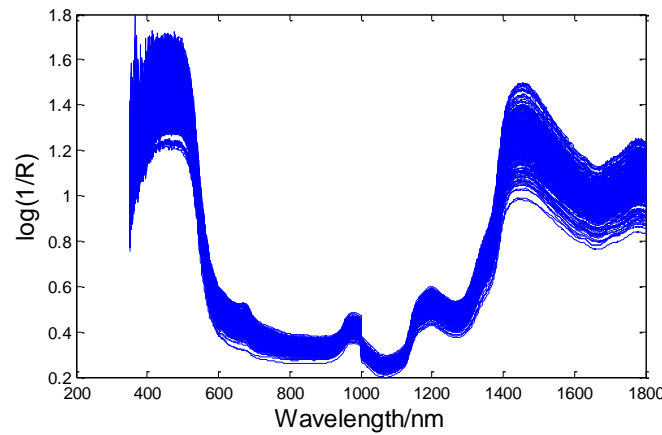


Fig. 2. Vis/NIR diffuse reflectance spectra of all Nanfeng mandarin fruit

3.3 Variable Selection

3.3.1 UVE-SPA

For UVE-SPA variable selection method, UVE was applied to select wavelength variables first, then SPA was conducted on the selected wavelength variables by UVE. Fig. 3 shows the stability results of each variable in the PLS model for SSC by UVE. In Fig. 3, the solid vertical line is the split line of wavelength variables and random variables, wavelength variables are at the left while random variables are at the right. Two dashed horizontal lines mean the upper and lower cutoffs for selecting wavelength variables. The variables beyond the cutoffs are realized as informative variables and reserved, while the variables between the cutoffs are realized as uninformative variables and eliminated. After UVE analysis, 151 variables were selected from 1231 variables for SSC.

Fig. 4 shows the root mean square error (RMSE) plot and selected variables by SPA for SSC of Nanfeng mandarin fruit based on the UVE analysis. From Fig. 4(a), it can be seen that the RMSE curve drops very quickly as the number of selected variables is increasing from seven to eight, ten to eleven and fifteen to eighteen. After that, the RMSE curve is level off with further increasing of the number of selected variables. Finally, twenty wavelength variables were chosen for SSC according to an F-test

criterion with $\alpha=0.25$. Fig. 4(b) shows the distribution of the selected wavelength variables. From Fig. 4(b), it can be seen that most of the selected wavelength variables are near the peaks and valleys of spectrum.

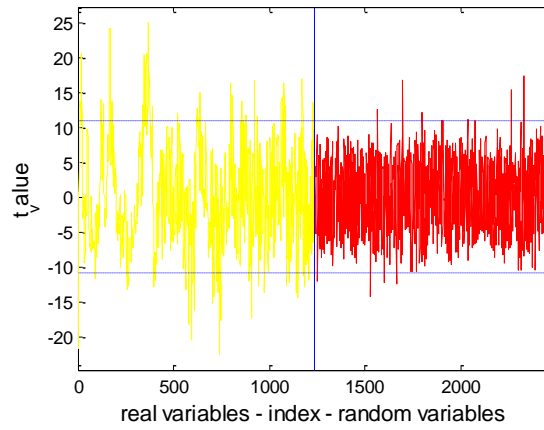
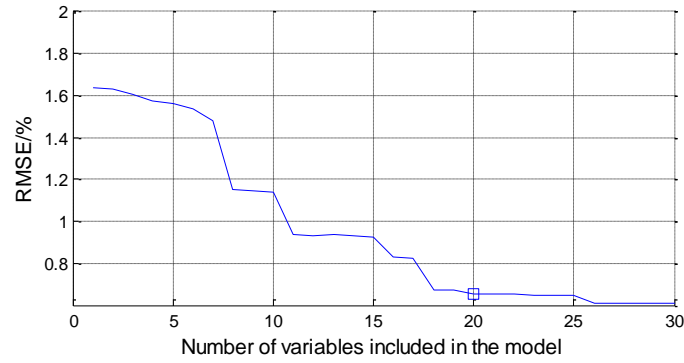
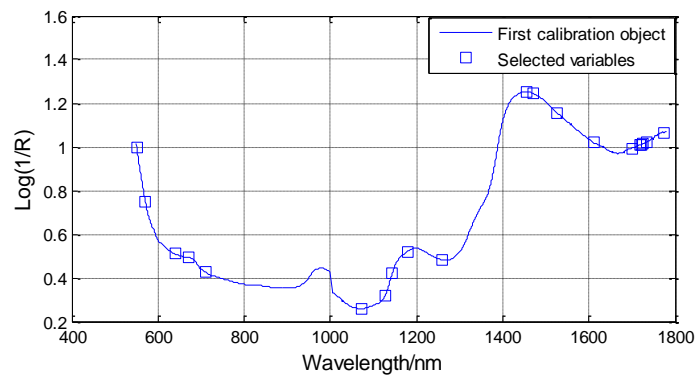


Fig. 3. Stability of each variable in the PLS model for SSC by UVE



(a)



(b)

Fig. 4. RMSE plot and selected variables by SPA based on UVE results: (a) RMSE plot; (b) Selected variables (shown in \square markers).

3.3.2 GA-SPA

Fig. 5 shows the results of GA variable selection for SSC of Nanfeng mandarin fruit. In Fig. 5, the solid horizontal line indicates the frequency cutoff for variable selection. The variables which frequency values are larger than frequency cutoff are realized as informative variables and reserved, and other variables are realized as uninformative variables and eliminated. After GA analysis, 137 variables were selected.

Fig. 6 shows the RMSE plot and selected variables by SPA based on GA results. From Fig. 6(a), it can be seen that the RMSE curve drops quickly as the number of selected variables is increasing from three to five, seven to eight, and fifteen to sixteen. The RMSE curve drops slowly as the number of selected variables is increasing from nine to fifteen, seventeen to twenty one. After that, the RMSE curve is level off with further increasing of the number of selected variables. And twenty one variables were selected at last. Fig. 6(b) shows the distribution of the selected variables for SSC by GA-SPA.

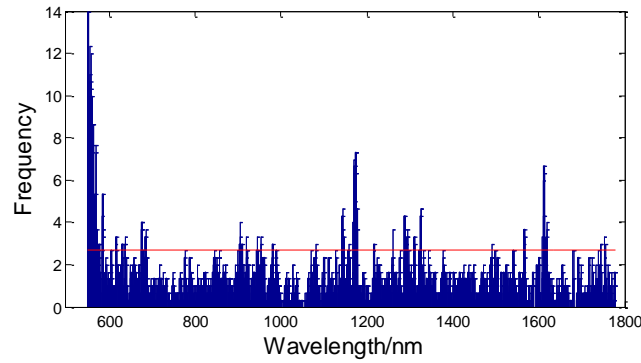
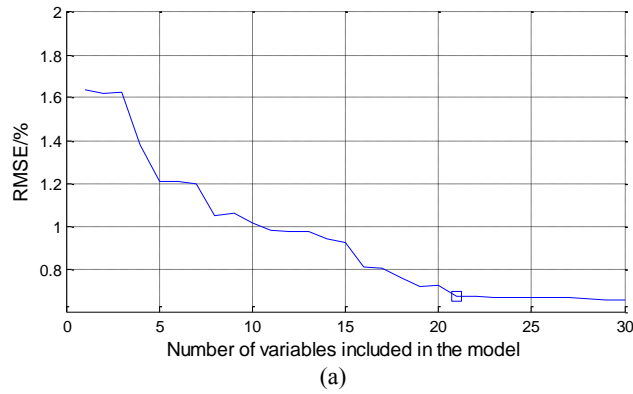


Fig. 5. Results of GA variable selection for SSC of Nanfeng mandarin fruit



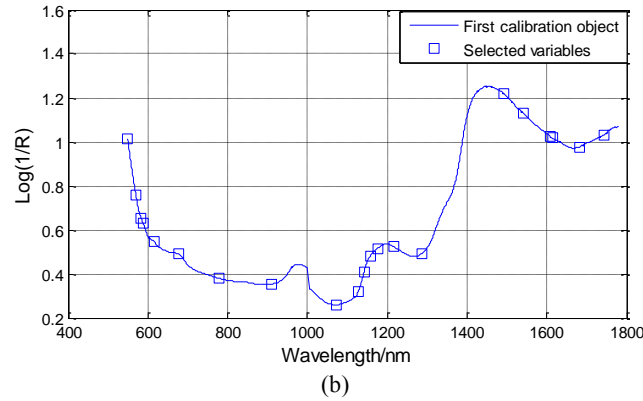


Fig. 6. RMSE plot and selected variables by SPA based on GA results: (a) RMSE plot; (b) Selected variables (shown in \square markers).

3.3.3 UVE-CARS

In the UVE-CARS analysis, the same results of UVE in UVE-SPA were used, and CARS was conducted on the results of UVE. Fig. 7 shows the results CARS variable selection for SSC of Nanfeng mandarin fruit. From Fig. 7(a), it can be seen that the number of sampled variables decreases as the number of sampling runs increases, and the declining trend is from quick to slow. Fig. 7(b) shows the change of RMSECV with increasing of sampling runs. The RMSECV decreases slowly as the number of sampling runs is increasing from one to fifteen. After that, the RMSECV increases quickly with further increasing of sampling runs, this may indicate that some useful variables are eliminated during this process. Fig. 7(c) shows the change of regression coefficients of variables with increasing of sampling runs. The position of mark “*” means the minimum RMSECV, and the number of sampling runs is fifteen. Finally, twenty nine variables were selected. Fig. 8 shows the distribution of the selected wavelength variables for SSC by UVE-CARS.

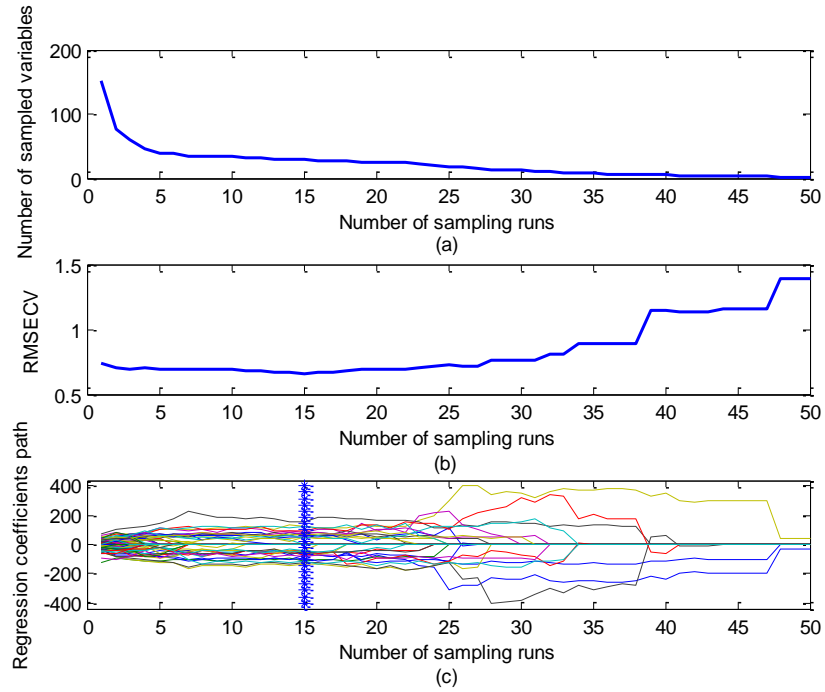


Fig. 7. Results of CARS variable selection based on UVE results: (a) Number of sampled variables; (b) RMSECV; (c) Regression coefficients path.

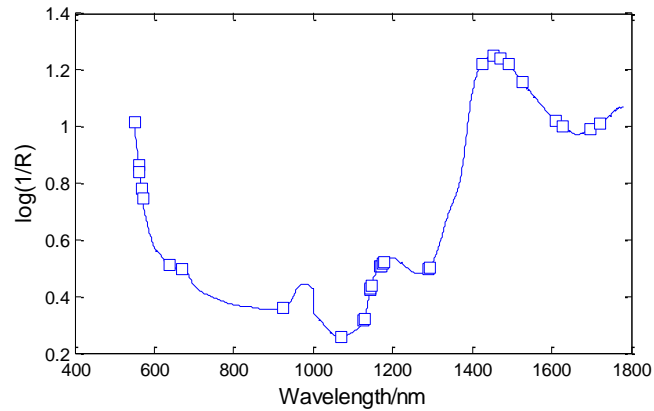


Fig. 8. Distribution of the selected wavelength variables for SSC by UVE-CARS

3.3.4 GA-CARS

In the GA-CARS analysis, the same GA results in GA-SPA were used, and CARS was conducted on the results of GA. Fig. 9 shows the results of CARS variable selection based on GA results. From Fig. 9, it can be seen that the position of mark “*” is at twenty five sampling runs, and this indicates that the minimum RMSECV is obtained when the number of sampling runs is twenty five. According to the results of

Fig. 9(a), seventeen variables were selected for SSC. Fig. 10 shows the distribution of the selected wavelength variables for SSC by GA-CARS.

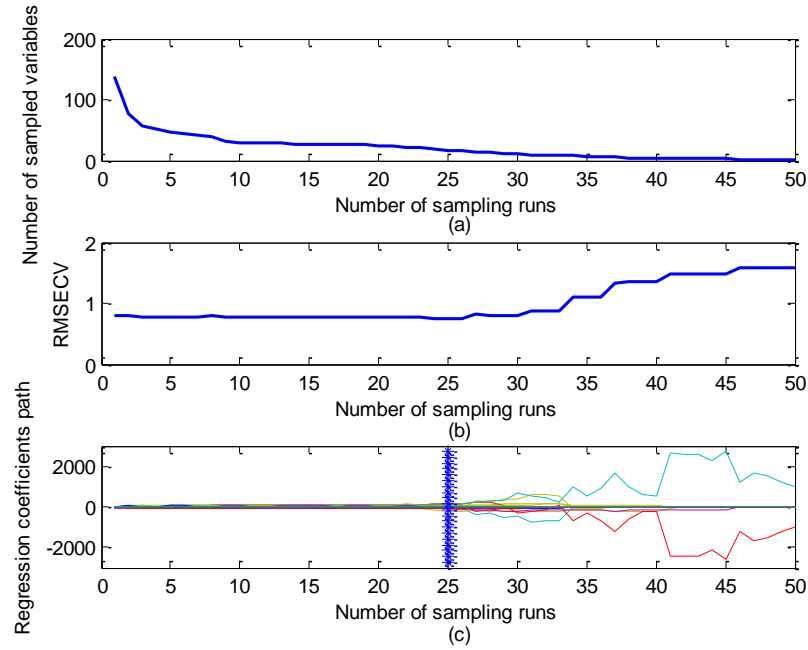


Fig. 9. Results of CARS variable selection based on GA results: (a) Number of sampled variables; (b) RMSECV; (c) Regression coefficients path.

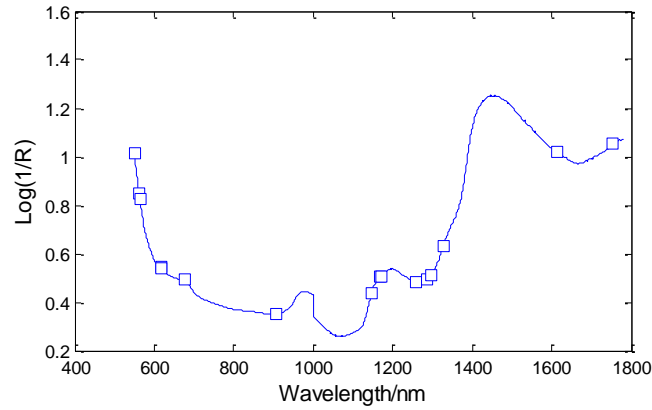


Fig. 10. Distribution of the selected wavelength variables for SSC by GA-CARS

3.4 LS-SVM

After variable selection, LS-SVM with RBF kernel was used to develop calibration models for SSC of Nanfeng mandarin fruit. Table 2 shows the results of LS-SVM and PLS regression for SSC of Nanfeng mandarin fruit. From table 2, it can be seen that

only about twenty variables are used in the LS-SVM (2% of the total number of variables), and 1231 variables are used in the PLS, while the results in LS-SVM models are comparable to the results of full-spectrum PLS model. This indicates that these four combined variable selection methods are useful and effective to eliminate uninformative variables and reserve informative variables.

For four combined variable selection methods, LS-SVM with GA-SPA obtains the best results, the correlation coefficients and RMSEs in the calibration, validation and prediction sets are 0.935, 0.560%, 0.912, 0.631% and 0.933, 0.594%, respectively. The worst results are obtained by LS-SVM with GA-CARS, and the correlation coefficients and RMSEs in the calibration, validation and prediction sets are 0.914, 0.641%, 0.836, 0.752% and 0.901, 0.716%, respectively. From above results, it can be concluded that GA-SPA is the best variable selection method among these four methods. This may due to that GA is suitable for eliminating uninformative variables and SPA is effective for selecting sensitive variables, so GA combined with SPA obtains the best results. For GA-CARS, the principles of GA and CARS are similar, this may be the reason why GA-CARS obtains the worst results. In order to evaluate the performance of LS-SVM models, the RPD (standard deviation/RMSEP) which has been defined as a measure of model performance [28] is introduced. The RPD values in the four LS-SVM models (UVE-SPA, GA-SPA, UVE-CARS, GA-CARS) are 2.67, 2.76, 2.44 and 2.29, respectively.

Table 2. Results of LS-SVM and PLS regression for SSC of Nanfeng mandarin fruit

| Method | Number of variable | γ | σ^2 | Calibration | | Validation | | Prediction | |
|----------|--------------------------|----------|------------|-------------|-------|------------|-------|------------|-------|
| | | | | r | RMSEC | r | RMSE | r | RMSE |
| | | | | | | V | | P | |
| UVE-SPA | 20 | 4.0e | 1.0e | 0.91 | 0.621 | 0.89 | 0.653 | 0.92 | 0.615 |
| | | 6 | 4 | 9 | | 3 | | 9 | |
| GA-SPA | 21 | 1.6e | 5.0e | 0.93 | 0.560 | 0.91 | 0.631 | 0.93 | 0.594 |
| | | 10 | 7 | 5 | | 2 | | 3 | |
| UVE-CARS | 29 | 6.8e | 4.3e | 0.91 | 0.651 | 0.86 | 0.699 | 0.91 | 0.671 |
| | | 10 | 7 | 0 | | 2 | | 2 | |
| GA-CARS | 17 | 4.0e | 9.0e | 0.91 | 0.641 | 0.83 | 0.752 | 0.90 | 0.716 |
| | | 4 | 2 | 4 | | 6 | | 2 | |
| PLS | 1231 | --- | --- | 0.91 | 0.619 | 0.89 | 0.649 | 0.90 | 0.710 |
| | | | | 9 | | 7 | | 1 | |

5 Conclusions

The results demonstrate that four combined variable selection methods are useful and effective to select informative variables, and the results of LS-SVM with these

combined variable selection methods are comparable to the results of full-spectrum PLS. GA-SPA is the best variable selection method among these four combined variable selection methods. The correlation coefficients and RMSEs in LS-SVM with GA-SPA model for calibration, validation and prediction sets are 0.935, 0.560%, 0.912, 0.631% and 0.933, 0.594%, respectively.

Acknowledgment

The authors gratefully acknowledge the financial support provided by the National Nature Science Foundation of China (No. 30972052), New Century Excellent Talents in Support of Ministry of Education Project (No. NCET090168) and Technology Foundation for Selected Overseas Chinese (2012).

References

1. Gómez A.H., He Y., Pereira A.G. Non-destructive measurement of acidity, soluble solids and firmness of satsuma mandarin using VIS/NIR-spectroscopy techniques [J]. *Journal of Food Engineering*, 2006, 77: 313-319.
2. Liu Yande, Sun Xudong, Zhang Hailiang, et al. Nondestructive measurement of internal quality of Nanfeng mandarin fruit by charge coupled device near infrared spectroscopy [J]. *Computers and Electronics in Agriculture*, 2010, 71S: S10-S14.
3. Fan Guoqiang, Zha Jianwen, Du Ran, et al. Determination of soluble solids and firmness of apples by Vis/NIR transmittance [J]. *Journal of Food Engineering*, 2009, 93(4): 416-420.
4. Bertone E., Venturello A., Leardi R., et al. Prediction of the optimum harvest time of 'Scarlet' apples using DR-UV-Vis and NIR spectroscopy [J]. *Postharvest Biology and Technology*, 2012, 69: 15-23.
5. Sun Tong, Lin HongJian, Xu Huirong, et al. Effect of fruit moving speed on predicting soluble solids content of 'Cuiguan' pears (*Pomaceae pyrifolia* Nakai cv. Cuiguan) using PLS and LS-SVM regression [J]. *Postharvest Biology and Technology*, 2009, 51(1): 86-90.
6. Xu Huirong, Qi Bing, Sun Tong, et al. Variable selection in visible and near-infrared spectra: Application to on-line determination of sugar content in pears [J]. *Journal of Food Engineering*, 2012, 109(1): 142-147.
7. McGlone V.A., Clark C.J., Jordan R.B. Comparing density and VNIR methods for predicting quality parameters of yellow-fleshed kiwifruit (*Actinidia chinensis*) [J]. *Postharvest Biology and Technology*, 2007, 46(1): 1-9.
8. Moghimi A., Aghkhani M.H., Sazgarnia A., et al. Vis/NIR spectroscopy and chemometrics for the prediction of soluble solids content and acidity (pH) of kiwifruit [J]. *Biosystems Engineering*, 2010, 106(3): 295-302.
9. Flores K., Sanchez M.T., Perez-Marin D.C., et al. Prediction of total soluble solid content in intact and cut melons and watermelons using near infrared spectroscopy [J]. *Journal of Near Infrared Spectroscopy*, 2008, 16(2): 91-98.
10. Tian Haiqing, Wang Chunguang, Zhang Haijun, et al. Measurement of soluble solids content in melon by transmittance spectroscopy [J]. *Sensor Letters*, 2012, 10(1-2): 570-573.
11. Ying Yibin, Liu Yande. Nondestructive measurement of internal quality in pear using genetic algorithms and FT-NIR spectroscopy [J]. *Journal of Food Engineering*, 2008, 84: 206-213.

12. Liu Fei, He Yong. Application of successive projections algorithm for variable selection to determine organic acids of plum vinegar [J]. Food Chemistry, 2009, 115: 1430-1436.
13. Sorol N., Arancibia E., Bortolato S.A., et al. Visible/near infrared-partial least-squares analysis of Brix in sugar cane juice: a test field for variable selection methods [J]. Chemometrics and Intelligent Laboratory Systems, 2010, 102: 100-109.
14. Wu Di, Chen Xiaojing, Zhu Xiangou, et al. Uninformative variable elimination for improvement of successive projections algorithm on spectral multivariable selection with different calibration algorithms for the rapid and non-destructive determination of protein content in dried laver [J]. Analytical Methods, 2011, 3: 1790-1796.
15. Huang Lingxia, Wu Di, Jin Hangfeng, et al. Internal quality determination of fruit with bumpy surface using visible and near infrared spectroscopy and chemometrics: A case study with mulberry fruit [J]. Biosystems Engineering, 2011, 109: 377-384.
16. Sun Tong, Xu Wenli, Lin Jinlong, et al. Determination of Soluble Solids Content in Navel Oranges by Vis/NIR Semi-transmission Spectra Combined with CARS Method [J]. Spectroscopy and Spectral Analysis, 2012, 32(12): 3229-3233.
17. Chen Bin, Meng Xianglong, Wang Hao. Application of successive projections algorithm in optimising near infrared spectroscopic calibration model [J]. Journal of Instrumental Analysis, 2007, 26(1): 66-69.
18. Zou Xiaobo, Zhao Jiewen, Huang Xingyi, et al. Use of FT-NIR spectrometry in non-invasive measurements of soluble solid contents (SSC) of 'Fuji' apple based on different PLS models [J]. Chemometrics and Intelligent Laboratory Systems, 2007, 87: 43-51.
19. Arakawa M., Yamashita Y., Funatsu K. Genetic algorithm-based wavelength selection method for spectral calibration [J]. Journal of Chemometrics, 2011, 25: 10-19.
20. Suykens J.A.K., Van Gestel T., De Brabanter J., et al. Least Squares Support Vector Machines [M]. Singapore: World Scientific, 2002.
21. Centner V., Massart D.L., deNoord O.E., et al. Elimination of uninformative variables for multivariate calibration [J]. Analytical chemistry, 1996, 68(21): 3851-3858.
22. Leardi R. Application of genetic algorithm-PLS for feature selection in spectral data sets [J]. Journal of Chemometrics, 2000, 14: 643-655.
23. Leardi R., Gonzalez A.L. Genetic algorithms applied to feature selection in PLS regression: how and when to use them [J]. Chemometrics and Intelligent Laboratory Systems, 1998, 41: 195-207.
24. Araújo M.C.U., Saldanha T.C.B., Galvão R.K.H., et al. The successive projections algorithm for variable selection in spectroscopic multicomponent analysis [J]. Chemometrics and Intelligent Laboratory Systems, 2001, 57(2): 65-73.
25. Galvão R.K.H., Araújo M.C.U., Frago W.D., et al. A variable elimination method to improve the parsimony of MLR models using the successive projections algorithm [J]. Chemometrics and Intelligent Laboratory Systems, 2008, 92(1): 83-91.
26. Li Hongdong, Liang Yizeng, Xu Qingsong, et al. Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration [J]. Analytica Chimica Acta, 2009, 648(1): 77-84.
27. Shao Yongni, Bao Yidan, He Yong. Visible/near-infrared spectra for linear and nonlinear calibrations: a case to predict soluble solids contents and pH value in peach [J]. Food and Bioprocess Technology, 2011, 4(8): 1376-1383.
28. Williams P.C., Sobering D.C. Comparison of commercial near infrared transmittance and reflectance instruments for analysis of whole grains and seeds [J]. Journal of Near Infrared Spectroscopy, 1993, 1: 25-32.