Near infrared rapid detection of protein, fat, starch and water in rice

Accepted 30th September, 2015

ABSTRACT

In this paper, partial least squares (PLS) method and BP neural network were employed to predict the content of protein, fat, starch and water in rice which was based on the Near-infrared (NIR) model. Kennard-Stone method was used to select the calibration and prediction set samples. Then we used the segmented wavelet de-noising algorithm to eliminate spectrum noise before screening the band which is useful to the samples in order to apply the competitive adaptive reweighted sampling (CARS) method. The results show that the PLS model of fat outperformed the BP neural network model; and the PLS models of protein, starch and water are the same as the BP neural network model. Therefore, the PLS method and BP neural network are suitable for rapid and non-destructive testing of rice protein, fat, starch and water.

Key words: Rice content, NIR spectroscopy, PLS, BP neural network, Competitive adaptive reweighted sampling (CARS)

INTRODUCTION

Protein, fat, starch and water are important indices for evaluating the quality of rice which, as we know, is one of the main food in the world (Huang et al., 2013). Traditional methods for detecting these indices are usually very tedious and time consuming. Moreover, the detection condition is not easy to control and a large amount of drug is needed during the detection process. Near infrared (NIR) spectroscopy analysis technology has many advantages, such as simple and rapid detection process, reduced material requirement, none destructive, good reproducibility, low cost, and therefore, has wide applications in food, fruit and meat inspection (Fan et al., 2009).

During the past few years, based on NIR spectroscopy analysis technology, some interesting results have been reported. In the study of Sun et al. (2008), based on NIR spectroscopy analysis technology and BP neural network, the prediction accuracy of the water content in rice was up to 96.67%. In Guo et al. (2012) and Wang and Chen (2007), NIR diffuse reflectance spectroscopy has been used to detect the contents of protein, fat, sugar and water in rice and brown rice. In a study by Lin et al. (2011), NIR spectroscopy was employed to detect the starch and amylase in Guangxi rice. As reported by Chen et al. (2010), by the utilization of CARS method together with NIR spectroscopy, the prediction correlation coefficients of protein, fat, sugar and water in milk were larger than 0.9951. In Pan et al. (2009), NIR wave band has been optimized to detect sugar by genetic algorithm. However, it should be pointed out that the modeling method in the above literatures were single, that is, only one or two methods were discussed. Moreover, it is obvious that the above results are not enough to reveal the characteristics of rice because only one or two kinds of indices were detected.

Therefore, this paper intends to determine a better method to detect the contents in rice, and compare the prediction accuracies based on PLS and BP neural network. The contents included protein, fat, starch and water.

MATERIALS AND METHODS

Experiment samples of rice are nine different kinds of 'Yangzhou Liang You’ rice (200 g for each one). Based on permutation and combination, 246 mixed samples have been obtained. Each sample had two kinds of rice and the
ratio of quality was 1:1. The total number of samples was 255.

Determination of protein, fat, starch and water in rice

The kjeldahl method of GB/T 5511-2008 titled 'Cereals and pulses-Determination of the nitrogen content and calculation of the crude content-Kjeldahl method' (2008) has been used to detect protein content in rice. Based on the fact that the conversion coefficient of nitrogen content in rice that converses into protein is 6.25, the protein content of the nine kinds of rice has been calculated. The soxhlet extraction method of GB/T 5511-2008 titled 'Inspect of grain and oilseeds-Determination of crude fat content in grain' (2008) has been employed to detect the fat content in rice. According to the calculation formula of crude fat dry basis content of rice, the fat contents of nine kinds of rice have been calculated. The direct titration method of GB/T 5009.7-2008 titled 'Determination of reducing sugar in foods' (2009) has been utilized to detect starch content in rice. The direct drying method of GB 5009.3-2010 titled 'National food safety standard Determination of moisture in foods' (Ministry of Health of the People's Republic of China, 2009) has been employed to detect the moisture content in rice. The protein, fat, starch and water contents of other 246 mixed samples were calculated by the four contents of the nine kinds of rice. Specific calculation method has two steps: first, sum the four contents of the two mixed rice respectively; second, take the average value of the above results.

NIR spectrum collection

The NIR spectra of rice samples have been collected by diffused SupNIR-2700 NIR spectrometer. The holographic digital grating InGaAs detector has been utilized. The wavelength ranged from 1000 to 1799 nm and the spectral resolution was 10 nm. The environment temperature was 15°C. There were 255 rice samples. The analytical contents were protein, fat, starch and water in rice.

A certain amount of samples were put in the sample cup. Integrating sphere and universal stage have been used to get NIR diffused spectra. The samples were scanned 6 times and these spectra were averaged. 255 rice samples were scanned.

Modeling method of NIR

Calibration and prediction set selection

The Kennard-Stone method (Zhang et al., 2012) has been employed to select calibration and prediction set. Before the experiment, the number of calibration set samples was set as 204, and the number of prediction set samples was set as 51.

Spectra pretreatment and key variables selection

In order to improve spectral signal to noise ratio, and make spectra be more representative to the information of protein, fat, starch and water in rice, segmented wavelet de-noising algorithm (Zhou et al., 2006; Zhu et al., 2003) has been used to eliminate spectrum noise and improve the quality of spectra. The parameters of segmented wavelet de-noising algorithm have been set as follows: the number of segment of spectra was \( n = 1 \); the maximum decomposition scale was \( m = 8 \); the generating function of wavelet transform was 'Daubechies'(‘db2’); the threshold determination method was SURE (TPTR was ‘heursure’); the wavelet coefficient deduction method was soft threshold (SORH was ‘s’).

NIR spectra are usually made up of a large number of data, and the number of wavelength is larger than the number of samples. Therefore, the collinearity of spectra is very serious. Variable selection can make the model simplify, and improve the prediction accuracy of the model. The competitive adaptive reweighted sampling (CARS) method (Li et al., 2009) imitates the principle of ‘survival of the fittest’ which comes from Darwin’s evolutionary theory. Based on the adaptive reweighted sampling technology, the wavelength with big absolute value of regression coefficient in PLS model is selected, and the wavelength with small weight is removed. The cross-validation (CV) is used to select the lowest value set of Root Mean Square Error of Cross Validation (RMSECV). Therefore, the key variables can be selected. Specific algorithm can be seen in Zhang et al. (2012).

Modeling

The PLS and BP neural network have been employed to build the model, and the accuracies of these two methods were compared. All the procedures of this paper were ran by MATLAB R2010a (Zhang and Yang, 2010). The 5 fold cross-validation has been utilized in PLS, the number of Monte Carlo simulation sampling was 2500, and the frequency in sampling of Monte Carlo was 200. The three layered neural network was used in BP neural network. The number of hidden layer unit was 10. The transfer functions of hidden layer and output layer were logsig and purelin, respectively. Conjugate gradient momentum algorithm (trainscg), the BP elastic algorithm (trainelr) and Levenberg-Marquardt algorithm (trainlm) were used.

RESULTS AND ANALYSES

The original NIR spectra of rice samples are shown in Figure 1. As in the figure, it is obvious that different samples have different absorbance. This is possibly due to the fact that different samples have different component
Table 1. The parameters of calibration sets and prediction sets of rice.

<table>
<thead>
<tr>
<th>Component</th>
<th>Calibration set</th>
<th>Prediction set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Range/%</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>6.849~7.889</td>
<td>6.849~7.764</td>
</tr>
<tr>
<td>Fat</td>
<td>0.995~1.272</td>
<td>0.995~1.236</td>
</tr>
<tr>
<td>Starch</td>
<td>69.438~86.979</td>
<td>69.438~83.770</td>
</tr>
</tbody>
</table>

Results of calibration and prediction set

The calibration and prediction set have been selected by Kennard-Stone method. Results are shown in Table 1. Table 1 shows that protein, fat, starch and water contents of prediction set are included in the calibration set. Therefore, the model which is built by calibration set can predict the content of prediction set.

Spectra pretreatment and key variables selection

After spectra were treated by segmented wavelet denoising, CARS method has been employed to select key variables of protein, fat, starch and water, respectively. The key variables are mainly concentrated between 1001~1400 nm. Selection results of fat are shown in Figure 2. The tendency of key variables selection is shown in Figure 2a. The number of variables declined from fast to slowly with the increase of running times, and there are exponential relation between running times and the number of reserved variables. The tendency of RMSECV got by 5 fold cross-validation is shown in Figure 2b. The tendency of RMSECV declined from 1 to 18 times. It indicates that variables which have nothing to do with the character of samples have been removed. After 18 times, RMSECV began to increase. It indicates that variables that are related to the character of samples may be removed. The ‘*’ is the lowest point of RMSECV in Figure 2c. The all lines in Figure 2c indicate the variation tendency of regression coefficients.

Modeling and evaluation

Modeling

The results of models which were built by PLS and BP
neural network are shown in Table 2. The results of BP neural network have been obtained by conjugate gradient momentum algorithm. The R2 which was got by BP elastic algorithm and Levenberg-Marquardt algorithm was less than 0.6. So these two methods were abandoned. From Table 2, it is shown that PLS and BP neural network could get a high R2(>0.9065) and a low RMSEC(<0.6355). It indicates that the real value and prediction value have a good linear relationship. RSD is less than 1.22%, so the models have a good stability and can be used in actual detection.

### The verification of the model

The components of prediction set have been predicted by the model. The accuracy and stability have been evaluated. The results are shown in Table 3.

### CONCLUSION

Models which are built by PLS and BP neural network can predict contents of protein, fat, starch and water in rice. The R2 of fat prediction set which is got by PLS and BP neural network are 0.8420 and 0.7642, respectively. So the accuracy of the model which is built by PLS is better than BP neural network. The R2 of protein, starch and water prediction set which is got by PLS are 0.7291, 0.9015 and 0.7983, respectively; RMSEC are 0.1047, 1.2085 and 0.1050, respectively; RPD are 3.6914, 10.1524 and 4.9578, respectively; RSD are 1.44, 1.59 and 0.75, respectively. The
R2 of protein, starch and water prediction set which is got by BP neural network are 0.7322, 0.8990 and 0.8866, respectively; RMSEC are 0.1331, 1.3032 and 0.1082, respectively; RPD are 3.744, 9.039 and 8.8199, respectively; RSD are 1.83, 1.72 and 0.77, respectively. The results show that the PLS model of fat outperforms the BP neural network model; and the PLS models of protein, starch and water are the same as the BP neural network model. Therefore, PLS and BP neural network are both suitable for fast and nondestructive examination of protein, fat, starch and water in rice.

### REFERENCES


