

Research on Infrared Spectrum Food Detection Technology Based on Markov Distance Singular Point Identification Method

Ling Yang^{*a}, Ting Wu^a, Juan Zou^a, Yunmao Huang^b, Sarath Babu V^b, Li Lin^{*b}

^aSchool of Information Science and Technology, Zhongkai University of Agriculture and Engineering, Guangzhou 510225, China

^bGuangzhou Key Laboratory of Aquatic Animal Diseases and Waterfowl Breeding, Guangdong Provincial Key Laboratory of Waterfowl Healthy Breeding, College of Animal Sciences and Technology, Zhongkai University of Agriculture and Engineering, Guangzhou, Guangdong 510225, China
yang98613@163.com

Food safety is one of the important problems in the development of society. Conventional polymerase chain reaction and other detection methods were costly and complicated. The paper presented a method for the infrared spectroscopy identification for food (chicken, beef, mutton as examples). In the paper, we used the Markov distance singular point Identification method to remove the singular samples of meat, and improved the accuracy and robustness of the model. Six kinds of pretreatment methods were used to exclude spectral noise, distortion and eliminate sample production. In the process of modeling, the paper used partial least squares discriminant analysis and BP neural network. The optimal modeling method was determined by comparing decision coefficient and the root mean square error of the correction set and the detection set. The results showed that the Mahalanobis distance identification method can effectively eliminate the singular points, improve the model correlation coefficient and reduce the error. The effect of normalized pretreatment was the best. When the number of principal components was 7, both PLS-DA and BP neural network can effectively identify three kinds, and the prediction accuracy of the detection set was 100%. The correction set and detection set decision coefficient of the PLS-DA modeling method was up to 0.99. RMSEC, RMSECV and RMSEP were 0.06, 0.08 and 0.08. The model performance was superior to BP neural network modeling method. The infrared spectrum detection technique, including PLS-DA modeling method and Markov distance singular point discrimination method, solved effectively the adulteration problem of common livestock meat. Key words: Infrared spectroscopy, Meat detection, Partial Least Squares Discriminant Analysis, BP neural network

1. Introduction

Food safety is one of the important problems in the development of society today, such as meat adulteration and so on. Traditional food identification methods include bioassay methods, such as polymerase chain reaction (PCR) (Safar et al., 2014), enzyme-linked immunosorbent assay (ELISA) (Li et al., 2013) and so on. Although the detection method is very high accuracy, it has the complicated process, high cost, long detection time and other shortcomings. Infrared spectroscopy is a technique, in accordance with different absorption characteristics of the object internal atomic, molecular and other specific structures to electromagnetic wave, which analyzes qualitative and quantitative of object specific components. Because of its fast, lossless, simple and so on, it is used quickly in agricultural product testing (Yang et al., 2016), such as, tea, honey, wine, rice, olive oil, meat products (Huang et al., 2015, Kamruzzaman et al., 2015) and so on. In the paper, we carried out research on infrared spectroscopy food detection technology, based on Markov distance singular points identification and PLS-DA modeling method, and took chicken, beef and mutton as examples.

2. Materials and methods

2.1 Materials

The experiment collected respectively 60 pieces of chicken, beef and mutton, which collected from meat shops and supermarkets in 2015, 11. Put them to the refrigerator, cut them into thin slices, put them in the 45 incubator for 48 hours, grind into powder and store in a dry plastic wrap, before spectral scanning.

2.2 Instrument and Spectral Acquisition

Use Fourier Transform Infrared Spectroscopy (FTIR) to carry out the transmission spectra acquisition. Spectrum analyzer wave number range was 4000-450cm⁻¹, resolution was 0.4cm⁻¹, scanning ambient temperature was 25 °C, and humidity was 30 ± 5%. Each sample was scanned three times to gain the original spectral data. During the test, the environment was consistent and the background spectrum was measured every 2 hours.

2.3 Methods

2.3.1 Markov distance method to exclude singular points

Because of the pollution during the process of sample preparation or the influence of equipment and environment during the process of spectral acquisition, there might be individual singular points. They had a great influence on the robustness and precision, so they need to be eliminated. In the paper, we used Markov distance method to exclude singular points. As was shown to be:

$$D(X_i) = \sqrt{(X_i - \bar{X})^T S^{-1} (X_i - \bar{X})} \quad (1)$$

Where X_i was the i th spectral vector, \bar{X} was the average spectrum of the spectral vector, S was the covariance matrix, and $D(X_i)$ was the Markov distance from the sample vector X_i to the average spectrum. Since the 60 samples collected from the same meat were similar, a certain Mahalanobis distance threshold can be set to exclude the abnormal sample. As was shown in equation (2), T was the threshold, \bar{D} was the average spectrum, δ was the adjustment parameter, σ_D was the Markov distance standard deviation.

$$T = \bar{D} + \delta \times \sigma_D \quad (2)$$

In the paper, ensure $\delta = 3^{[6]}$ to determine the Mahalanha distance threshold. When $D(X_i)$ was greater than T , it can be considered as a singular point and excluded.

2.3.2 Partial Least Squares Discriminant Analysis

Partial Least Squares (PLS) is a common multivariate statistical method, and widely used to the establishment of spectral models now. This algorithm combines factor analysis and regression analysis, spectral matrix X and the concentration matrix Y in the spectrum are decomposed at the same time, and find its potential scalar. Partial least squares are always used for quantitative analysis. In qualitative analysis, since it has not concentration matrix Y , it is necessary to manually assign Y , and then use the partial least squares method to convert to quantitative analysis. In this study, we used the cross validation method to calculate the sum of squares of predicted residuals in the modeling process, fought the number of the best potential variables, and established a linear regression model.

2.4 Spectral analysis

Use Unscrambler X 10.3 and TQ Analyst 8.0 to pretreat original spectral information and do PLS-DA analysis. The Mahalay distance calculation and BP neural network analysis were carried out by MATLAB.

2.5 Establish PLS-DA and BP neural network models

In the paper, PLS-DA and BP neural network were used to establish the identification model of chicken, cattle and mutton, compare the advantages and disadvantages of the two methods, and find the most appropriate identification model. In the modeling process, the test samples were divided into calibration sets and detection sets. In the paper, 38,39,39 samples of chicken, cattle and sheep were randomly selected as the calibration set, and the remaining 20,20,20 samples were used as the test set, to establish the model.

3. Results and Analysis

3.1 The original spectral curve

The original spectrum of the collected chicken, beef, and lamb was shown in Figure 1. In order to judge the differences between different meat, three kinds of meat were averaged. As was shown in Figure 2, three kinds in the 2922 cm^{-1} , 1742 cm^{-1} , 1653 cm^{-1} , 1546 cm^{-1} and other bands had a strong absorption peak. That was absorption, from hydrocarbon group, carbonate, and fatty acid ester of meat. It could be seen from Figure 2 that the absorbance of the spectrum of beef was stronger. But when the three kinds of spectra were normalized, the spectral absorption bands were similar and the difference was not obviously.

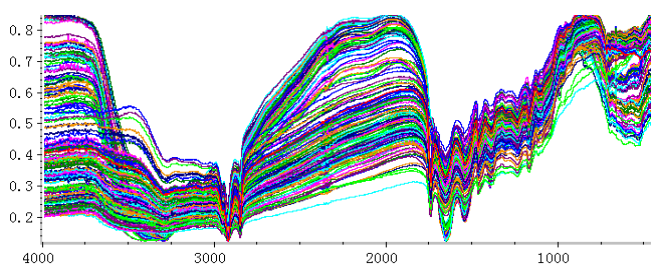


Figure 1: The original spectrum of chicken, beef, lamb

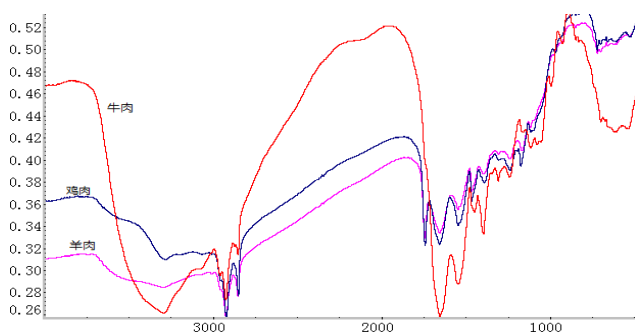
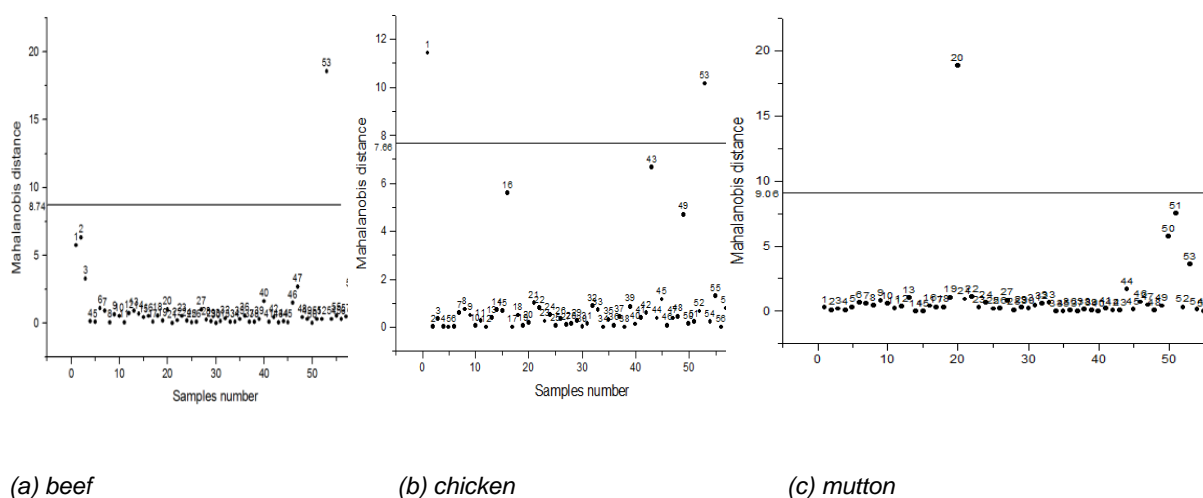


Figure 2: The average spectrums

3.2 Markov distance exclusion singular points

Respectively, use the formula (1) (2) for the Mahalanobis distance between samples and the average spectrum, and the adjustment parameter δ was 3, singular points was shown in Figure 3:



(a) beef

(b) chicken

(c) mutton

Figure 3: The singular points recognition of chicken, cattle, sheep samples

In the beef sample, the Mahalanobis distance of 53th sample was 18.55, far higher than 8.74(the set threshold), so it was excluded. In the chicken samples, the Markov distances of the first and the 53th samples were 11.44 and 10.14, respectively, far higher than 7.66 (the set threshold), so it was excluded. Similarly, the 20th sample in the mutton sample was excluded. The performance of the model before and after the removal of the singular points was shown in Table 1. It could be seen that the decision coefficient of the model calibration set and the detection set, mean square error and the cross-validation variance of the corrected set were improved. The singular points had a great influence on the model, and the model performance was obviously optimized after the singular point excluded.

Table 1: The performance of the model before and after the removal of the singular point

Sample	R^2_{cal}	RMSEC	R^2_{val}	RMSECV
before thremoval	0.96	0.17	0.93	0.21
after removal	0.98	0.09	0.99	0.11

3.3 Spectral pretreatment

In the original spectrum acquisition process, in order to exclude affection caused by the equipment, environment, sample production and other factors, the original spectral analysis was carried out by six preprocessing methods, including multiple scattering correction (MSC), standard normal transform (SNV), first order derivation, second order derivation, Savitzky-Golay smoothing, normalization. As is shown in Table 2:

Table 2: Effects of six pretreatment methods on the identification model

Pretreatment methods	R^2_{cal}	RMSEC	R^2_{val}	RMSECV	prediction accuracy of detection set /%
original spectrum	0.98	0.09	0.99	0.11	91.7
Second derivative	0.94	0.03	0.99	0.2	3
First order derivative	0.97	0.03	0.99	0.16	33
Normalization	0.99	0.06	0.99	0.08	100
Savitzky-Golay	0.98	0.09	0.99	0.11	90
MSC	0.97	0.12	0.98	0.14	98.3
SNV	0.97	0.11	0.98	0.13	98.3

The first-order derivation and second-order derivation amplified the characteristic absorption band and the noise, which reduced performance and accuracy. The normalization was able to exclude thickness and transmittance of different sample, and it could effectively eliminate the difference between the same kinds, so the effect was best. It could be seen that the decision co-efficiency of correction set and the detection set was both 0.99, RMSEC and RMSECV was 0.06 and 0.08. The model performance was excellent. The accuracy of 20 samples was 100%.

3.4 The analysis of results

In the establishment of PLS-DA model, cattle, chicken, sheep were given -1, 0, 1 in the paper. After the model established, the difference between the predicted value and the reference value was used to determine the model accuracy. As was shown in Figure 4:

It could be seen that after the normalized pretreatment method, the predicted values of beef, chicken and mutton located at true value -1, 0, 1. RMSEC and RMSECV were 0.06 and 0.08, respectively. The decision coefficient was 0.99. The model performance was excellent. Use it to predict the 20 samples, as was shown in Figure 5: In the paper, a reference rang, ± 0.5 , was taken to determine the accuracy of the model prediction. It was observed that the predictive value of beef was between -1.09 and -0.92, the chicken was between -0.05 and 0.18, and the mutton was between 0.70 and 1.08. Three kinds of meat predictions were in the

corresponding reference range, and the prediction accuracy was 100%. At the same time, three kinds of meat identification models were established by BP neural network modeling method. Compare the performance of the PLS-DA model with the BP neural network. As was shown in Table 3:

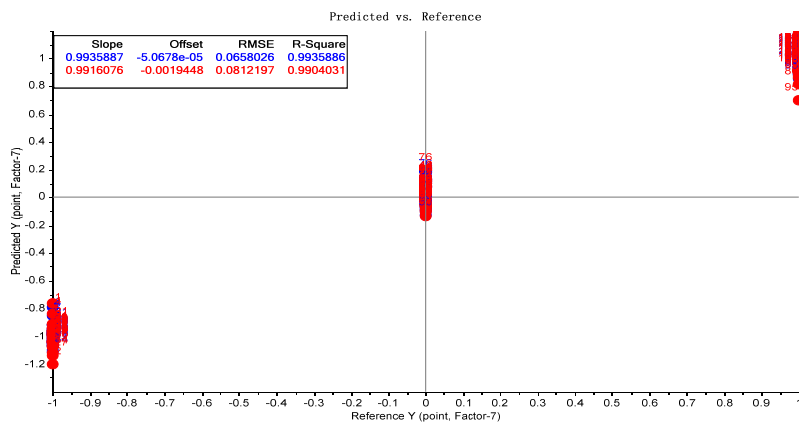


Figure 4: The distribution of PLS-DA model correction set predictive value and true value

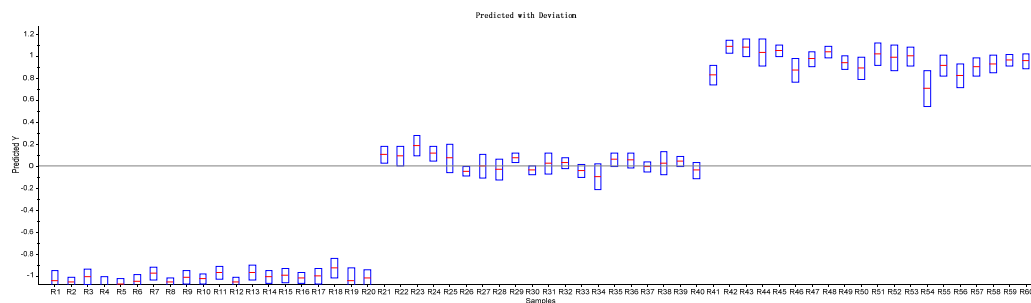


Figure 5: The distribution of PLS-DA model detection set predictive value and true value

Table 3: The comparison between PLS-DA and BP neural network model performance

The modeling method	R^2_{cal}	RMSEC	R^2_{val}	RMSECV	RMSEP	prediction accuracy of detection set /%
PLS-DA	0.99	0.06	0.99	0.08	0.08	100
BP neural network	0.96	0.17	0.95	0.19	0.26	99

The accuracy rate of detection set prediction of PLS-DA method had reached 100%, and BP neural network had reached 99%. The predicted root mean square error of them was above 0.26. They had good performance. While comparing the advantages and disadvantages of the two methods, decision coefficient of the calibration set and the detection set of PLS-DA method of had reached 0.99, higher than BP neural network. The mean square error of the correction set of the PLS-DA method, the mean square error of the prediction set, and the cross validation root mean square error were both below 0.08, lower than BP neural network. Therefore, PLS-DA method was the best model, to detect chicken, beef and mutton.

4. Discussion

Comparing PLS-DA and BP neural networks, the RMSEC and RMSEP of BP neural network had reached 0.17 and 0.26, far higher than PLSR 0.06 and 0.08. BP neural network had a certain over-fitting phenomenon. The BP neural network used neuron weight iterative algorithm, filtered out the noise effects, and built a better nonlinear response. But the training model of BP neural network used the back propagation mechanism, so its initial value was set by random setting, which was easy to fall into the problem of local optimal and gradient diffusion. So it needs artificial adjustment parameters. So in future, solve the local optimal problem of BP neural network by introducing the depth learning algorithm and training the optimal eigenvalue.

5. Conclusion

In the paper, use infrared spectroscopy technology to identify adulterated meat of chicken, beef and mutton on the market, apply the Markov distance to eliminate singular point, and eliminate the thickness of the test sample by the normalized pretreatment method. In the paper, three kinds of meat detection models were established by PLS-DA and BP neural network modeling methods. By comparing, the R^2_{cal} and R^2_{val} of the PLS-DA method were 0.99, RMSEC was 0.06, RMSECV and RMSEP were both 0.08, and the prediction accuracy was 100%. The model performance is better than BP neural network. Therefore, based on PLS-DA method, the infrared spectrum detection technology can accurately identify different meat, and it has certain practical value.

Acknowledgments

This study was jointly supported by National Natural Science Fund (61501531); Fund for Science and technology from Guangdong Province (2015A020209173); Fund from Guangzhou Science and Technology Bureau (201704020030) and "Innovation and Strong Universities" special funds (KA170500G) from the Department of Education of Guangdong Province.

Reference

- Huang Y., Andueza D., de Oliveira L., Zawadzki F., Prache S., 2015, Comparison of visible and near infrared reflectance spectroscopy on fat to authenticate dietary history of lambs, *Animal*, 9, 1912-1920, DOI: 10.1017/s1751731115001172
- Kamruzzaman M., Makino Y., Oshita S., Liu S., 2015, Assessment of Visible Near-Infrared Hyperspectral Imaging as a Tool for Detection of Horsemeat Adulteration in Minced Beef. *Food and bioprocess technology*, 8, 1054-1062, DOI: 10.1007/s11947-015-1470-7
- Li T., Yin Y., Wang H., 2013, Quick Identification of Five Species of Meat by PCR Assay, *Food Science*, 34, 249-252, DOI: 10.7506/spkx1002-6630-201308054
- Safar M., Janejo Y., Arman K., 2014, A highly sensitive and specific tetraplex PCR assay for soybean, poultry, horse and pork species identification in sausages: Development and validation, *Meat Science*, 98, 296-300, DOI: 10.1016/j.meatsci.2014.06.006
- Yang L., Wu T., Cai X C., 2016, Application and research progress of spectroscopy in meat detection, *Journal of Food Safety and Quality, Guangdong Agricultural Sciences*, 43, 162-168, DOI: 10.16768/j.issn.1004-874X.2016.05.031