Research Article

Honey Discrimination Using Visible and Near-Infrared Spectroscopy

Yun Li and Haiqing Yang

College of Information Engineering, Zhejiang University of Technology, Hangzhou 310032, China

Correspondence should be addressed to Haiqing Yang, yanghq@zjut.edu.cn

This study aims to investigate the potential of honey discrimination by visible and near-infrared (vis-NIR) spectroscopy with wavelength reduction. A total of 80 samples from four brands of honey produces were measured by a mobile fiber-type USB4000 spectrophotometer with recorded wavelength range of 380.17~939.98 nm for model calibration. Firstly, principal components analysis (PCA) was used for extracting principal components (PCs). Next, the first seven PCs, which accounted for 97% of variance of the spectra, were combined separately with support vector machine (SVM) and linear discriminate analysis (LDA) to develop PC-SVM and PC-LDA models, both of which achieved 100% discrimination accuracy. In addition, the spectra were subjected to successive wavelength reduction rates (WRRs) of 2^x, $x = 1$–9, for wavelength reduction. The PC-LDA and PC-SVM models developed for these reduced wavelengths produced almost the same performance as compared with those developed for original full wavelengths. This experiment suggests that vis-NIR spectral wavelengths can be reduced at large spacing interval, which allows easing data analysis as well as developing a simpler and cheaper sensor for honey discrimination in practice.

1. Introduction

Honey is considered as healthy and wholesome food with curative properties as it contains plenty of nutrients and plays effective antimicrobial effects against many bacteria [1]. Honey produces are quite popular in China. Generally, consumers classify honeys in terms of color, smell, and taste. Although sensory evaluation is easy to use, subjective experience is often biased and low accurate. Using chemical test methods in laboratory may be accurate but quite time- and cost-consuming, which produces unsafe guarantee to ordinary consumers and hinders the development of apiculture. Thus, it is necessary to develop a convenient and accurate way for honey discrimination.

Recently, visible and near-infrared (vis-NIR) spectroscopy has received wide attention as it is suitable for nondestructive analysis of biological and biomedical materials. For example, the vis-NIR spectroscopy can be used for discriminating tea beverages [2], coffee [3], milk power [4], and other materials [5–7]. Using this technique, Gallardo-Velázquez et al. [8] and Zhu et al. [9] qualified adulterants in some local origins of honey. Although these researches presented good

accuracy for honey discrimination, the calibration models were developed using full range of wavelengths, which would result in high complexity in computation and cause difficulty in practical application.

This study was to investigate the possibility of reducing spectral wavelengths in the process of using the vis-NIR spectroscopy to discriminate Chinese honeys with an aim of fabricating a simpler and cheaper sensor for practical use.

2. Marerials and Methods

2.1. Honey Origins and Spectrum Acquisition. Four brands of honey (linden tree: 1, locust: 2, wild chrysanthemum: 3, milk vetch: 4) were purchased from a local supermarket on December, 2011 and sealed in room temperature of $20 \sim 25°C$. Before spectral measurement, the samples were placed in a water container at $50°C$ until the soluble substances fully dissolved. Eighty samples were obtained with twenty for each brand. Spectral scanning was conducted by a mobile fiber-type USB4000 spectrophotometer (Ocean Optics, USA) in transmission mode with recorded wavelength range of

346.01 ~ 1038.08 nm at 0.3 nm sampling space. Before sample measurement, twenty reference scans were taken on a ceramic standard supplied with the spectrophotometer. Ten photometric scans were conducted and averaged on each sample.

2.2. Spectrum Transformation. Due to signal noise at both ends of the original spectra, only the wavelengths from 380 to 940 nm were retained for further investigation. Firstly, sample spectral reflectance at a wavelength of λ was calculated by

$$S_{\text{ref}}(\lambda) = \frac{(S_\lambda - D_\lambda)}{(R_\lambda - D_\lambda)} \times 100\%, \qquad (1)$$

where $S_{\text{ref}}(\lambda)$ is sample reflectance, and S_λ, D_λ, and R_λ are digital counts of sample, dark, and reference light intensities, respectively [10]. Then, the reflectance spectra were transformed into absorbance spectra using log $(1/R)$, as absorbance is directly proportional to the concentration of an absorber according to the Beer-Lambert Law.

2.3. Principal Components Analysis (PCA). The PCA is a well-known method for revealing the hidden structure within large data set [11]. By using an orthogonal transformation, which could calculate the eigenvectors of the covariance matrix of the original inputs, PCA converts a set of observations of possibly correlated variables (wavelengths) into a set of values of uncorrelated variables called principal components (PCs). Using the PCA, data can be projected from a high-dimension space to a lower-dimension space. In this study, the first few PCs were used by the following two calibration methods, that is, SVM and LDA, for developing discrimination models [3].

2.4. Support Vector Machine (SVM). The SVM is a relatively new computational learning method based on statistical learning theory. SVM maps input vectors to a higher dimensional space by constructing a maximal separating hyperplane, which can produce good performance in ill-posed situations with limited amount of training data. With appropriate choice of kernel functions, SVM is capable of modeling highly nonlinear classification boundaries [12]. However, SVM may cause some problems when dealing with large number of input variables [13]. In the study, we used PCA to transform the hundreds of variables to several principal components, which was then used for developing SVM models.

2.5. Linear Discrimination Analysis (LDA). The LDA has been proved to be a promising approach of feature extraction. The basic idea of LDA is to find a linear transform in a way that the ratio of the between-class scatter and the within-class scatter is maximized. Namely, samples can be projected to a new space with smallest within-class distance and largest between-class distance [14, 15]. Although LDA usually gives a good discrimination performance, it suffers from some deficiencies if variables are highly correlated or class boundaries are complex or nonlinear. In the former
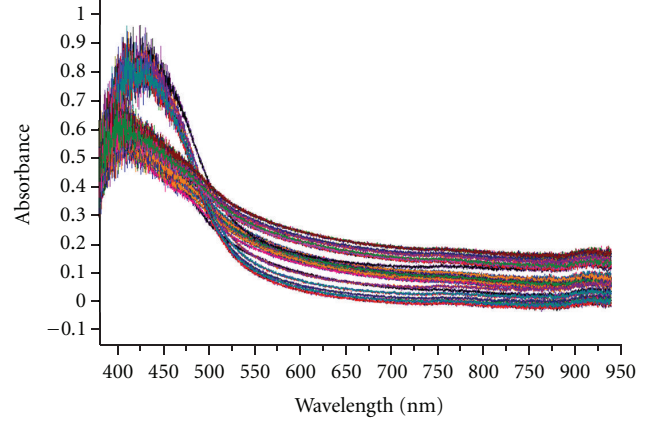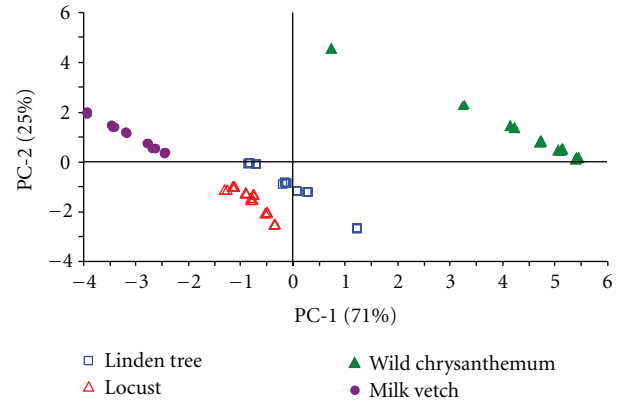

FIGURE 1: Spectra of honey samples.

FIGURE 2: PC1 × PC2 scores scatter plot of the four origins of honey.

case, LDA overfits the data, and in the latter case, LDA underfits the data [16]. To avoid such deficiencies, variables are often transformed by correlation-reducing methods such as PCA.

3. Results and Discussion

3.1. Absorbance Spectra. Figure 1 shows absorbance spectra of the four brands of honey with wavelengths ranging from 480 to 940 nm. The spectra present peaks at the band of 400 ~ 450 nm and low absorbance in the NIR range above 760 nm. All the spectra show their similarity in spectral shape and absorbance. Thus, it is necessary to apply appropriate multivariate analysis methods to build calibration models for honey discrimination.

3.2. PC-SVM and PC-LDA Models for Full Wavelengths. PCA was used to reduce the spectral dimensionality to a small number of principal components, which enabled samples clustering in terms of floral origins. Figure 2 shows the scores scatter plot of the 1st and 2nd PCs. Obviously, the samples can be clearly distinguished in terms of their origins. The first two PCs explain as high as 96% of variance of the spectra (71% for PC-1 and 25% for PC-2). To balance model

TABLE 1: Discrimination accuracy of PC-SVM and PC-LDA models under different wavelength reduction rates (WRRs).

Number of wavelengths	WRRs	PC-SVM (%)	PCA-LDA (%)
2847 (full)	1	100.0	100.0
1424	2	100.0	100.0
712	4	100.0	100.0
356	8	100.0	100.0
178	16	100.0	98.7
89	32	100.0	100.0
44	64	98.7	100.0
22	128	100.0	100.0
11	256	97.5	100.0
6	512	97.5	98.7

complexity and discrimination accuracy, we chose seven PCs, which accounted for 97% of spectral variance, to develop PC-SVM and PC-LDA models. Table 1 shows their excellent performance with 100% accuracy for honey discrimination.

3.3. PC-SVM and PC-LDA Models for Reduced Wavelengths. We reduced the wavelengths by increasing spectral spacing interval. Subjected to successive wavelength reduction rates (WRRs) of 2^x, $x = 1$–9, the number of wavelengths was reduced by the times of 2, 4, 8, 16, 32, 64, 128, 256, and 512, respectively. Based on these reduced wavelengths, a series of PC-SVM and PC-LDA models were established. The results of model performance are also listed in Table 1.

Surprisingly, these new models responded insensitively to various WRRs. For example, the PC-SVM and PC-LDA models developed for 6 wavelengths (WRR = 512) achieved discrimination accuracy of 97.5% and 98.7%, which was almost as those for full 2847 wavelengths. This indicates that the six wavelengths at interval of 153.6 nm (512 points at 0.3 nm spacing) are enough to develop accurate models (PC-SVM or PC-LDA) for honey discrimination. Previous reports also revealed the feasibility of reducing wavelengths by WRRs. Yang et al. [7] developed soil nitrogen and carbon contents prediction models using vis-NIR wavelengths subjected to the WRRs of 2, 5, 10, 20, 50, 100, 200, and 500. They found that the models calibrated for 21 wavelengths at interval of 100 nm achieved almost the same as those for full 2100 wavelengths.

4. Conclusions

In this study, the vis-NIR spectroscopy was used for honey discrimination. Two classification models (PC-SVM and PC-LDA) were developed for various numbers of wavelengths ranging from 380 to 940 nm. The results indicated that the vis-NIR spectra can be reduced without compromise of model performance for honey discrimination. PC-SVM and PC-LDA achieved perfect performance with discrimination accuracy of 97% for full wavelengths as well as for those reduced at large spectral interval, which may allow developing a simpler and cheaper sensor for honey discrimination in practice.

Abbreviations

LDA: Linear discriminate analysis
NIR: Near infrared spectroscopy
PCA: Principal components analysis
PCs: Principal components
SVM: Support vector machine
Vis-NIR: Visible and near infrared
WRRs: Wavelength reduction rates.

Acknowledgment

References

[1] B. Kilicoglu, K. Kismet, S. S. Kilicoglu et al., "Effects of honey as a scolicidal agent on the hepatobiliary system," *World Journal of Gastroenterology*, vol. 14, no. 13, pp. 2085–2088, 2008.

[2] Q. Chen, J. Zhao, M. Liu, and J. Cai, "Nondestructive identification of tea (*Camellia sinensis* L.) varieties using FT-NIR spectroscopy and pattern recognition," *Czech Journal of Food Sciences*, vol. 26, no. 5, pp. 360–367, 2008.

[3] G. Lv and H. Yang, "Discrimination of different brands of Nescafé coffee using vis-NIR spectroscopy and comparative study," *Advances in Biochemical Engineering*, vol. 1-2, pp. 163–166, 2011.

[4] R. M. Balabin and S. V. Smirnov, "Melamine detection by mid- and near-infrared (MIR/NIR) spectroscopy: a quick and sensitive method for dairy products analysis including liquid milk, infant formula, and milk powder," *Talanta*, vol. 85, no. 1, pp. 562–568, 2011.

[5] N. Sinelli, L. Cerretani, V. D. Egidio, A. Bendini, and E. Casiraghi, "Application of near (NIR) infrared and mid (MIR) infrared spectroscopy as a rapid tool to classify extra virgin olive oil on the basis of fruity attribute intensity," *Food Research International*, vol. 43, no. 1, pp. 369–375, 2010.

[6] H. Yang, B. Kuang, and A. M. Mouazen, "In situ determination of growing stages and harvest time of tomato (*Lycopersicon esculentum*) fruits using fiber-optic visible-near-infrared (vis-NIR) spectroscopy," *Applied Spectroscopy*, vol. 65, no. 8, pp. 931–938, 2011.

[7] H. Yang, B. Kuang, and A. M. Mouazen, "Quantitative analysis of soil nitrogen and carbon at a farm scale using visible and near infrared spectroscopy coupled with wavelength reduction," *European Journal of Soil Science*, vol. 63, no. 3, pp. 410–420, 2012.

[8] T. Gallardo-Velázquez, G. Osorio-Revilla, M. Z. D. Loa, and Y. Rivera-Espinoza, "Application of FTIR-HATR spectroscopy and multivariate analysis to the quantification of adulterants in Mexican honeys," *Food Research International*, vol. 42, no. 3, pp. 313–318, 2009.

[9] X. Zhu, S. Li, Y. Shan et al., "Detection of adulterants such as sweeteners materials in honey using near-infrared spectroscopy and chemometrics," *Journal of Food Engineering*, vol. 101, no. 1, pp. 92–97, 2010.

[10] Y. Yang, P. Nie, H. Yang, and Y. He, "Rapid recognition method of nectar plant based on visible-near infrared spectroscopy," *Transactions of the Chinese Society of Agricultural Engineering*, vol. 26, no. 3, pp. 238–242, 2010 (Chinese).

[11] J. P. Reed, D. Devlin, S. R. R. Esteves, R. Dinsdale, and A. J. Guwy, "Performance parameter prediction for sewage sludge

digesters using reflectance FT-NIR spectroscopy," *Water Research*, vol. 45, no. 8, pp. 2463–2472, 2011.

[12] R. M. Balabin, R. Z. Safieva, and E. I. Lomakina, "Near-infrared (NIR) spectroscopy for motor oil classification: from discriminant analysis to support vector machines," *Microchemical Journal*, vol. 98, no. 1, pp. 121–128, 2011.

[13] R. Noori, A. R. Karbassi, A. Moghaddamnia et al., "Assessment of input variables determination on the SVM model performance using PCA, Gamma test, and forward selection techniques for monthly stream flow prediction," *Journal of Hydrology*, vol. 401, no. 3-4, pp. 177–189, 2011.

[14] H. Kim, B. L. Drake, and H. Park, "Multiclass classifiers based on dimension reduction with generalized LDA," *Pattern Recognition*, vol. 40, no. 11, pp. 2939–2945, 2007.

[15] R. M. Balabin, R. Z. Safieva, and E. I. Lomakina, "Gasoline classification using near infrared (NIR) spectroscopy data: comparison of multivariate techniques," *Analytica Chimica Acta*, vol. 671, no. 1-2, pp. 27–35, 2010.

[16] H. C. Kim, D. Kim, and S. Y. Bang, "Extensions of LDA by PCA mixture model and class-wise features," *Pattern Recognition*, vol. 36, no. 5, pp. 1095–1105, 2003.