

Discriminant Analysis of Red Wines from Different Aging Ways by Information Fusion of NIR and MIR Spectra

Sijia Tao¹, Jingming Li², Junhui Li¹, Jianbo Tang¹, Jinrui Mi¹, Longlian Zhao^{1*}

¹ College of Information & Electrical Engineering, China Agricultural University, Beijing, 10083, China

² College of Food Science & Nutrition Engineering, China Agricultural University, Beijing, 10083, China

Abstract. Two different levels' information fusion of near infrared (NIR) spectra and mid infrared (MIR) spectra coming from red wines aged in different ways has been investigated. A total of 96 red wines, including 44 wines aged in oak barrel, 26 wines aged in stainless steel tank with oak chips and 26 wines aged in stainless steel tank, were analyzed. The NIR transmission spectra and MIR attenuated total reflectance (ATR) spectra of samples were collected. For data level fusion, the NIR and MIR spectra were max-min normalized respectively and then merged. For feature level fusion, five PCs and db3 coefficients of 3-layer wavelet decomposition were calculated. Discriminant models of the three different aging wines were established using DPLS and Fisher method before and after fusion. All the fusion methods mentioned above have achieved better results on classification than individual spectroscopy. These results suggest that the spectral fusion of NIR and MIR is a promising technology for discriminating different aging wines.

Keywords: Information fusion, NIR, MIR, Wine, Aging way

1 Introduction

The red wine is welcomed by customers at home and abroad for its rich nutritional value. The wine aging process is the wine's mature process. It is a necessary procedure in the production of the aged wine. Different aging ways had great effects on the content of polysaccharides and wine chromaticity. So that it can affect the quality of the wine.

The research from home and abroad using spectroscopy studying wine are mainly focused on two aspects. On one hand, the application of mid infrared spectroscopy alone is widely used for classifying the varieties^{[1][2]}, vintages^[3] and geographical region of red wine. On the other hand, near infrared spectroscopy is applied for quantitative analysis which is becoming a new rapid analysis technology that it is primarily applied for contents of samples as well as wine^{[1][4]}. Although a large

¹ Corresponding author. Tel.:010-62737414, E-mail address: taosijia530@163.com
Acknowledgements: This study was supported by National Science Foundation of China (Project No: 31101289) and Chinese Universities Scientific Fund (Project No: 2009JS104).

number of studies have been published, there are very few systematic investigations exploring classification of different aging ways of red wines to our knowledge.

Information fusion is the merging of information from disparate sources, thus we can obtain more accurate and reliable description than single source achieved. The information of MIR comes from the frequency of the atomic vibration between the molecules while the information of NIR comes from the conjunction of vibration frequency and times frequency. The MIR has a lower limit of detection of 10^{-5} ~ 10^{-6} compared with NIR of 10^{-3} ~ 10^{-4} . Based on the different characteristics of NIR and MIR, the information fusion of near infrared spectra and mid infrared spectra is supposed to get better classification results.

Here, two different levels' information fusion, data level fusion and feature level fusion, are studied, and the applicability of spectra fusion technique to discriminate different aging wines is investigated. The aim of this study is to realize the rapid and reliable discrimination of wines aged in three different ways.

2. Experimental

2.1 Wine Samples

A total of 96 red wines were analyzed in this study, including 44 wines aged in oak barrel, 26 wines aged in stainless steel tank with oak chips and 26 wines aged in stainless steel tank. The wine samples were from two kinds of grape varieties (Cabernet Sauvignon and Merlot) ranged from 2003 to 2007 vintages. The red wine samples were collected from wineries in Changli and Huailai of Hebei, China.

2.2 Instruments and procedures

2.2.1 Spectra collection

NIR transmission spectra: Wine samples were scanned in transmission mode using FT-NIR spectrometer of MPA (Bruker Optics), equipped with a tungsten lamp and a cooled PbS detector. Samples were scanned in a rectangular cuvette with a 1 mm path length. The spectra were measured by summing 32 scans acquired over the wavelength range 12500 ~ 3900cm^{-1} at a resolution of 4 cm^{-1} . The spectrum of each sample was obtained in triplicate and the mean of the three measurements was used for classification. NIR spectra are shown in Fig. 1.

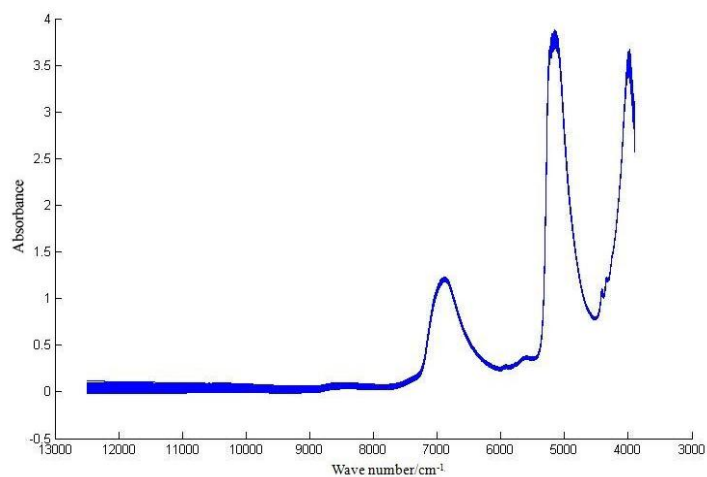


Fig. 1 Near infrared spectra of red wine samples

MIR ATR spectra: The MIR spectra of 96 wine samples were obtained from 550 to 4000 cm^{-1} on TENSOR 27 (Bruker Optics) FT-MIR instrument with an ATR attachment. The experiment used air as reference in room temperature. For both background and samples, the spectra were measured by summing 32 scans acquired at a resolution of 4 cm^{-1} . MIR spectra of the wine samples are shown in Fig. 2.

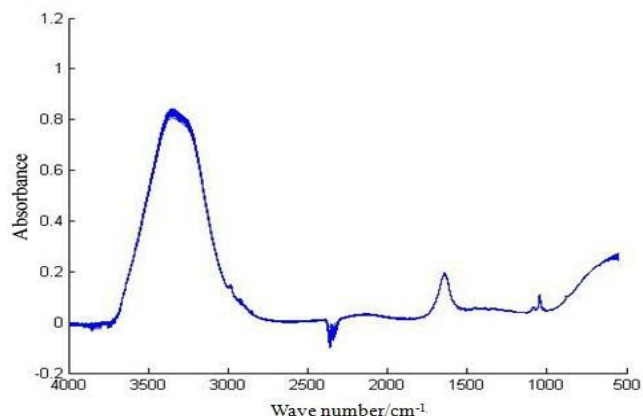


Fig. 2 Mid infrared spectra of red wine samples

2.2.2 Chemometric techniques for classification analysis

The spectra before and after fusion were used to develop discrimination models for different wine aging ways using two pattern recognition methods, discriminant partial least squares (DPLS) and Fisher discriminant analysis (FDA). DPLS is an application of Partial Least Squares (PLS) [5] for discrimination analysis. Different from PLS, the output variable of DPLS was expressed as a binary variable, while the PLS was concentration value. Generally a critical threshold was pre-set, by comparing the predictive value

with the threshold, the samples' categories were determined. FDA has been widely used in the domain of pattern recognition. The Fisher discriminant criterion was to find a linear projection that the classes of the samples are separated with maximal ratio of between-class and within-class variability^[6].

3. Results and discussion

3.1. Data level fusion

For data level fusion, the NIR and MIR spectra are preprocessed (min-max normalized) respectively. These two sources of information are merged then.

To develop the discriminant models, 96 wine samples were divided into two sets randomly, 2/3 training set and 1/3 for validation sets. Both sets were independent; the training set was used to build discriminant models and the validation set was only used for testing the models. Then DPLS and FDA modeling methods were used to develop classification models of different aging wines before and after fusion respectively. To test the stability of the models, wine samples were divided into training and validation set for 10 times randomly. Table 1 and table 2 show the mean recognition accuracy of cross-validation (training set) and prediction.

Table 1. Classification results for 3 different aged wine samples by DPLS

	Recognition accuracy (%) of training set	Recognition accuracy (%) of validation set
NIR	90.60	86.02
MIR	97.81	96.25
After fusion	99.06	96.59

Table 2. Classification results for 3 different aged wine samples by FDA

	Recognition accuracy (%) of training set	Recognition accuracy (%) of validation set
NIR	93.68	93.66
MIR	96.00	97.34
After fusion	98.01	97.98

From table 1 and table 2 we can see, both recognition accuracy of training set and recognition accuracy of validation set are improved. For DPLS model, the recognition accuracy of training set was improved to 99.06% while the recognition accuracy of validation set was advanced to 96.59%. For FDA model, the recognition accuracy of training set and validation set were 98.01% and 97.98%, respectively.

3.2 Feature level fusion

Two kinds of feature extraction methods, PCA and wavelet transform, were used to achieve the goal of feature level fusion.

3.2.1 Principal components analysis (PCA)

PCA^[7] was a widely used feature extraction method for reduction the number of variables showing co-linearity. In this study, five PCs were extracted from NIR spectra and MIR spectra respectively, then these PCs were merged as the input of the chemometric methods in order to get the fused classification discrimination through modeling at the feature level. FDA method was applied to build the classification models of three different aging wines using the NIR spectra, MIR spectra and the merged spectra. Table 3 shows the classification result of the 96 red wine samples.

Table 3. Classification results for 3 different aged wine samples by FDA

	Recognition accuracy (%) of training set	Recognition accuracy(%) of validation set
NIR	90.63	87.50
MIR	94.38	91.87
After fusion	95.95	98.67

As it can be seen from table 3, the recognition accuracy of training set and validation set was 95.95% and 98.67 respectively. When compared with the results of NIR and MIR spectra, it can be easily found the accuracies were improved after spectra information fusion

3.2.2 Wavelet transform

Wavelet transform^[8] is now being adopted for a vast number of applications. Wavelets are mathematical functions that cut up data into different frequency components, and then study each component with a resolution matched to its scale. These coefficients can then be compressed more easily because the information is statistically concentrated in just a few coefficients.

The db3 wavelet was chose for 3-layer wavelet decomposition. The approximate coefficients of the 3rd layer were used to information fusion. Then DPLS and FDA modeling methods were used to develop classification models of different aging wines before and after fusion respectively.

we can obtain 99.52% as the recognition accuracy of training set and 95.96% as the recognition accuracy of validation set by taking an average of ten times as a result when use the DPLS model. While for the FDA method, the recognition accuracy of training set is 97.61% and the recognition accuracy of validation set is 98.65%. It can be found the accuracy was improved after information fusion.

4. Conclusions

In this paper, two different levels' information fusion of NIR and MIR spectra of three different aging wines were studied. For classifying three different aged wines, discriminant models are developed to analysis fused spectra with FDA and DPLS methods. The recognition accuracies were all increased to a certain extent after spectra fusion. For data level fusion, the recognition accuracies of validation set were

improved to 96.59% and 97.98% by DPLS and FDA respectively. For feature level fusion, the recognition accuracies of validation set were improved to 98.67% using DPLS when PCA was used while the recognition accuracies of validation set were improved to 95.96% and 98.65% respectively by DPLS and FDA when we used wavelet transform. All the results above showed that information fusion of NIR and MIR spectra combined with the pattern recognition methods (DPLS and FDA) had the capability to extract feature of red wines and play a good role in discriminating wines from different aging systems.

However, some work still can be done for better results: some other pattern recognition methods could be tried, and the level of information fusion could be extended to the decision level^[9]. It should be surely useful to extend the study in this direction.

References

1. Sylvie Roussel, Véronique Bellon-Maurel, Jean-Michel Roger and Pierre Grenier. Fusion of aroma, FT-IR and UV sensor data based on the Bayesian inference. Application to the discrimination of white grape varieties. *Chemometrics and Intelligent Laboratory Systems*, 2003, 65(2):209-219
2. Picque D, Cattenoz T, Corrieu G, et al. Discrimination of red wines according to their geographical origin and vintage year by the use of mid-infrared spectroscopy. *sciences des aliments*, 2005, 25(3):207
3. Andrea Edelmann, Josef Diewok, Kurt Christian Schuster and Bernhard Lendl. Rapid method for the discrimination of red wine cultivars based on mid-infrared spectroscopy of phenolic wine extracts. *Journal of Agricultural and Food Chemistry*. 2001, 49(3):1138-1145
4. Maria D Guillen, Nerea Cabo. Infrared Spectroscopy in the study of edible oils and fats. *Journal of the Science of Food and Agriculture*, 1997, 75:1-11
5. Rao, Xiang, Zhou, Wang, Xie, & Xu, 2009; Xie, Ying, Ying, Tian, Niu, & Fu, 2007
6. Garcia-Allende, Conde, Mirapeix, Cobo, & Lopez-Higuera, 2008; Markiewicz, Matthews, Declerck, & Herholz, 2009; Zhu, Zang, Cao, Yan, He, Jiang, Sui. & Wang, 2008
7. B.G.M. Vandeginsten, D.L. Massart, S. Buydens, S. De Jong, P.J. Lewi, J. Smeyers-Verbeke, *Handbook of Chemometrics and Quality Metrics: Part B*, Elsevier, Amsterdam, 1998
8. Mallat S, Hwang WL. Singularity Detection and Processing with Wavelet. *IEEE Trans. on Information Theory*, 1992, 38(2) : 618-643
9. Terry Windeatt, Reza Ghaderi. Binary labeling and decision-level fusion. *Information Fusion* 2 (2001) 103-112