

# Edible Oil Classification Based on Molecular Spectra Analysis with Image Recognition

CAO Yu-ting<sup>1</sup>, ZHAO Zhong<sup>1\*</sup>, YUAN Hong-fu<sup>2</sup>, LI Bin<sup>1</sup>

1. College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China  
2. College of Materials Science and Engineering, Beijing University of Chemical Technology, Beijing 100029, China

**Abstract** Molecular spectra analysis combined with the chemometrics is becoming a popular method for rapid classification of edible oil. However, when the molecular spectral differences among the different types of samples are tiny, it is usually difficult to identify them with the traditional classification techniques. In this work, a method of molecular spectra analysis based on image recognition for rapid classification of edible oil is proposed. In order to accomplish recognition of different types of edible oil, the attenuated total reflectance infrared spectra of seven types of edible oil are scanned on ATR-FTIR. To enhance the spectral differences among different types of samples and visualize the identification process, the pretreated IR spectra are transformed into two-dimensional spectral image with auto correlation operation. Then, the local extrema are extracted with the method of image expansion and are used as the classification features. The back propagation (BP) neural network is chosen as the classifier to identify the extracted local extrema of the two-dimensional spectral image. Comparative experiments to identify the same samples with the proposed method, PCA-BP and KL-BP have also been done. Comparative experiment results have verified that the classification results with the proposed method (correct classification rate is 94.4%) are obviously better than those with PCA-BP (correct classification rate is 66.7%) and with KL-BP (correct classification rate is 83.3%). The proposed method has provided a new way to classify the edible oil rapidly based on molecular spectra analysis.

**Keywords** Edible oil; Molecular spectrum; Spectral visualization; Feature extraction; Image recognition  
中图分类号: O561.3 文献标识码: A DOI: 10.3964/j.issn.1000-0593(2019)02-0659-06

## Introduction

Edible oil is a necessity of daily life, and it is very important to ensure the safety of edible oil because its quality greatly affects people's health. There are many kinds of edible oil such as the peanut oil, rapeseed oil, soybean oil, corn oil, tea seed oil, sesame oil and olive oil in the market. The nutritional values and prices of different kinds of edible oil in the market are different due to their different ingredients. To avoid spurious activities in the market, it is necessary to establish effective detection methods to classify the quality of the edible

oil in the market.

There are some methods which have been reported to detect the quality of edible oil, such as the detection methods based on gas chromatography (GC)<sup>[1]</sup>, GC combined with mass spectrometry (GC-MS)<sup>[2]</sup>, thin-layer chromatography (TLC)<sup>[3]</sup> and high performance liquid chromatography (HPLC)<sup>[4]</sup>. However, detection methods mentioned above are usually time-consuming and require complex pre-treatment<sup>[4]</sup>. Spectral analysis based detection methods<sup>[5]</sup>, including near infrared spectroscopy (NIR) and Fourier transform infrared spectroscopy (FTIR) based recognition methods have been developed to analyze the edible oils with the advantages of

Received: 2017-03-21; accepted: 2017-09-15

Foundation item: Beijing Natural Science Foundation(4172044), Chaoyang District Collaborative Innovation Project (CYXC1707)

Biography: CAO Yu-ting, (1991—), female, master, College of Information Science and Technology, Beijing University of Chemical Technology  
e-mail: 294095465@qq.com \* Corresponding author e-mail: zhaozhong@mail.buct.edu.cn

fast speed and non-destructive testing. In most spectra analysis methods<sup>[6-7]</sup>, the dimension of spectral data need to be compressed to extract the classification features. Principal component analysis (PCA)<sup>[8]</sup>, Karhunen-Loeve transform (KL)<sup>[9]</sup> and cluster analysis (CA)<sup>[10]</sup> are of ten applied for data compression. PCA is probably the most widely applied method among the existing methods of data compression. PCA has been proved to be an effective data compression method. Usually, the main components of unclassified samples for edible oil identification are very similar, such as the samples of the peanut oil, rapeseed oil, soybean oil, corn oil, tea seed oil, sesame oil and olive oil. When PCA is applied to extract the classification features from the samples with similar components, the number of principal components could be very large in order to include the classification information. Then, the classification information could be hidden by other principal components. It is usually difficult to identify the hidden classification information even with the nonlinear classification method. Therefore, the key step for edible oil identification based on molecular spectra analysis is extracting the effective classification features that can enhance the difference among the samples. Transforming the one-dimensional molecular spectra data into two-dimensional molecular spectral image data is an effective way to improve the visualization of the molecular spectra data and enhance the differences among the samples. Image processing techniques<sup>[11]</sup> can be applied to extract the classification features from the two-dimensional molecular spectra image data and classify the two-dimensional molecular spectral image. In this work, a method of molecular spectra analysis based on image recognition for rapid classification of edible oil is proposed. In order to accomplish recognition of the different types of edible oil, the IR spectra of seven types of edible oil are scanned on ATR-FTIR. To enhance the spectral differences among different types of samples and visualize the identification process, the pretreated IR spectra are transformed into two-dimensional spectral image with auto correlation operation. Then, the local extrema are extracted with the method of image expansion and are used as the classification features. The back propagation (BP) neural network is chosen as the classifier to identify the extracted local extrema of the two-dimensional spectral image. Comparative experiments to identify the same samples with the proposed method, PCA-BP and KL-BP have also been done. Comparative experiment results have verified that the classification results with the proposed method (correct classification rate is 94.4%) are obviously better than those with PCA-BP (correct classification rate is 66.7%) and with KL-BP (correct classification rate is 83.3%). The proposed method has provided a new way to classify the edible oil rapidly based on molecular spectra analysis.

## 1 Methods

### 1.1 Spectra enhancement algorithm

Due to the environment factors (such as temperature, and humidity), the measured spectra may contain noise, baseline drift, wavelength drift and other measurement errors. To remove the useless information, the spectra are pretreated with multivariate scatter corrected (MSC) as follows

Step 1: Calculate the average spectra of the sample  $\bar{x}$  ( $1 \times m$ ) as

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (1)$$

where  $n$  is the number of the samples and  $m$  is the number of the wavelengths.

Step 2: Do linear regression between each sample spectra  $x_i$  and average spectra  $\bar{x}$  as

$$x_i = a_i + \bar{x}b_i \quad (2)$$

where  $a_i$ ,  $b_i$  are linear regression coefficients.

Step 3: Do multiple scattering correction for each sample spectra as

$$x_{i, MSC} = \frac{x_i - Ia_i}{b_i} \quad (3)$$

where  $I \in R^{n \times 1}$  is unit vector.

### 1.2 The procedure of spectra visualization

The infrared spectra can reflect the vibration of the molecule. When different frequencies of infrared lights are projected to the molecule in proper order, the different absorption can be obtained. Let  $Z_i$  denotes the absorbance matrix of spectra as

$$Z_i = [A_1 \ A_2 \ \cdots \ A_{i-1} \ A_i]^T \quad (4)$$

where  $[A_1 \ A_2 \ \cdots \ A_{i-1} \ A_i]$  are the infrared light absorbance vector. The one-dimensional infrared spectra are obtained by plotting the absorbance  $T\%$  versus frequency  $\lambda$ . To transform the one-dimensional spectrum into 2D spectral contour map, the absorbance matrix is multiplied by its transpose as

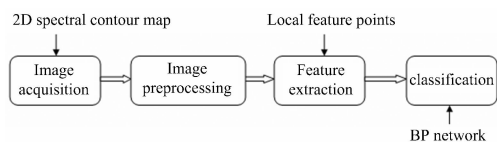
$$Z_{ii} = Z_i Z_i^T = \begin{bmatrix} A_1^2 & A_2 A_1 & \cdots & A_i A_1 \\ A_1 A_2 & A_2^2 & \cdots & A_i A_2 \\ \cdots & \cdots & \ddots & \cdots \\ A_1 A_i & A_2 A_i & \cdots & A_i^2 \end{bmatrix} \quad (5)$$

The different correlation coefficient values are represented with different color.

### 1.3 Image recognition

Image recognition is a technology to identify a variety of different patterns in the field of computer application, which uses computer to process and analyze the image. Generally, the system of image recognition mainly includes the information acquisition of image, preprocessing of image, feature extraction and classification<sup>[12]</sup>. In this work, the local extrema

of image are extracted as classification features and the BP neural network is applied as the classifier. The procedure of molecular spectra analysis based on image recognition is illustrated in Fig. 1.



**Fig. 1 The procedure of molecular spectra analysis based on image recognition**

Feature extraction is a very important step in the procedure of image recognition. The extracted features need to be stable and can represent the typical characteristics to ensure the accuracy of the classification. The features for image recognition include color feature, edge feature, shape feature and local extrema feature. The local extrema feature is usually not affected by the disturbance of external environment. In this work, the local extrema feature is chosen as the classification feature for edible oils.

The local extrema are extracted by the image dilation with the two dimensional data. The image dilation is an image processing method based on the dilation element operator in mathematical morphology. In the image dilation, target image  $X$  and structural elements  $S$  are looked as two sets of pixels, move structure element  $S$  from the origin to  $S_{xy}$ , if  $S_{xy}$  hit or is included in  $X$ , then form the coordinates of the dilation set as

$$X \oplus S = \{(x, y) \mid S_{xy} \uparrow X\} \quad (6)$$

where  $x$ ,  $y$  are the horizontal and vertical coordinate respectively of the structural elements moved to.

Comparing the original data and image dilation, only a relatively large value can overlap with the result after the dilation. According to this principle, the local extrema can be extracted. In this work, the classifier of back propagation (BP) neural network is designed based on the supervised learning.

## 2 Experiment

### 2.1 Materials

72 edible oil samples are provided by the National Institute of Metrology (NIM) of China. They belong to nine types of edible oil of sesame, corn, canola, blend, sunflower, peanut and olive. 54 samples are chosen as calibration set and the remaining 18 samples are chosen as the validation set. The number of calibration set and validation set for each types of samples is shown in Table 1.

### 2.2 Spectra measurement

In this work, the infrared spectra of the samples are scanned by Attenuated Total Internal Reflectance Fourier

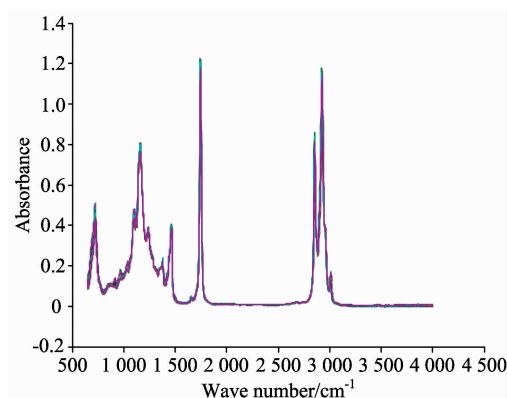
Transform Infrared (ATR-FTIR) spectrometer. The samples of edible oil and spectrometer are shown in Fig. 2. The spectra are collected from 650 to 4 000  $\text{cm}^{-1}$  with a resolution 4  $\text{cm}^{-1}$  and with 10 scans. Each sample is scanned three times and the average is used for analysis. The spectra of all samples measured on ATR instrument are shown in Fig. 3.

**Table 1 Statistics of samples**

Name	No.	Total sample number	Samples in calibration set	Samples in validation set
Sesame oil	1	15	12	3
Corn oil	2	4	3	1
Canola oil	3	8	5	3
Blend oil	4	10	8	2
Sunflower oil	5	7	5	2
Peanut oil	6	12	9	3
Olive oil	7	16	12	4
Total	7	72	54	18



**Fig. 2 ATR spectrometer and edible oil samples**



**Fig. 3 The spectra of all samples measured on ATR spectrometer**

## 3 Results and Discussion

### 3.1 Spectra visualization

2D spectra contour map is used to generate the spectra image based on autocorrelation analysis. The spectra of two samples belong to sesame oil and blend oil respectively are

shown in Fig. 4. The shapes of spectra of edible oils are almost the same in Fig. 4. The spectra from 558 to 598  $\text{cm}^{-1}$  are processed into a spectral image by the visualization method. The spectral images of two kind edible oil are shown in Fig. 5. Obviously, the spectral difference after visualization becomes apparent and it is easier to classify the visualized spectra of edible oils.

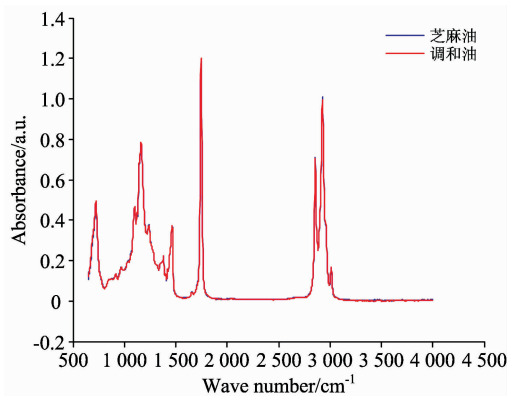


Fig. 4 The IR spectra of sesame and blend oil

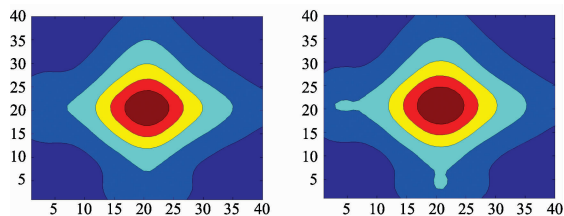


Fig. 5 The IR spectral images of sesame (a) and blend oil (b)

### 3.2 Feature extraction results

The characteristic absorption band in IR is caused by vibration of molecular groups and appears as the local extrema, which can be extracted by the image dilation in the visualized spectra image. Choose the local extrema of the spectra image and then observe the distribution of the extrema. The distribution of the first extreme is more representative, as shown in Fig. 6.

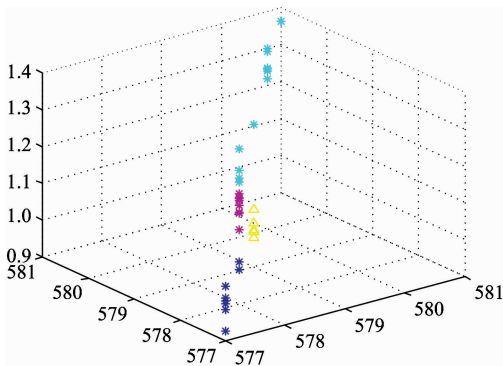


Fig. 6 The distribution of the first extreme (four kinds of edible oil)

### 3.3 Classification results

In this work, BP neural network with supervised learning is applied to classify the local extrema in the spectra image of edible oil, and the minimum training rate of the BP neural network is 0.1, and the error is 0.01, and the number of iterations is 800, and the training process is shown in Fig. 7, and the BP classification model is established.

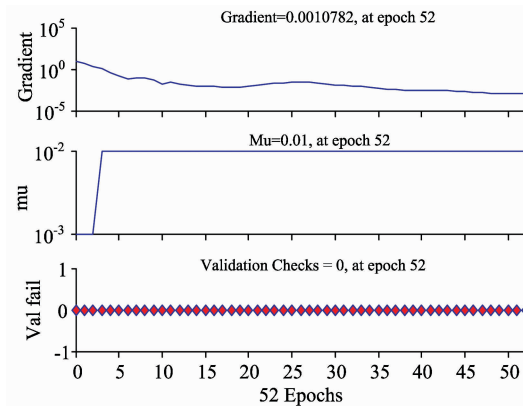


Fig. 7 The training state of BP neural network

18 validation samples are predicted by the established BP classification model, and the results are shown in Fig. 8 and the accuracy are shown in Table 2.

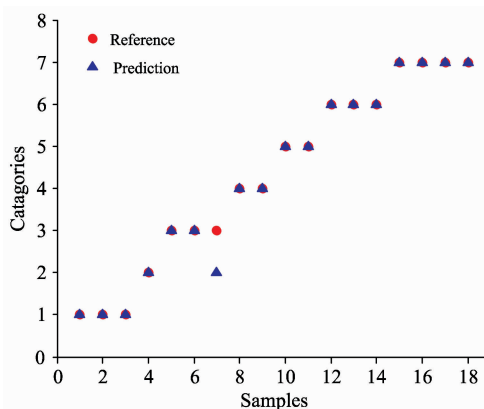


Fig. 8 The prediction results of with the trained BP classification model

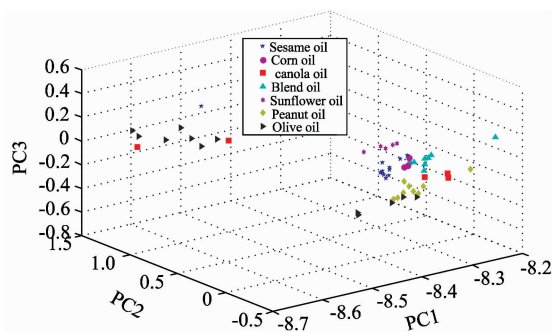
Table 2 Classification results based on molecular spectra image recognition

Method	Sample number	Correct classification	Error classification	Accuracy /%
Spectra image recognition	18	17	1	94.4

### 3.4 Comparative experiments

PCA is applied to analyze the calibration set of edible oil. The principle component number is chosen as 3 and the first three principal components score graphic of IR spectra is shown in Fig. 9. From Fig. 9, it can be seen that the projec-

tions of the IR spectra overlap on the principal component subspaces, therefore it is difficult to classify them directly.



**Fig. 9 Spatial distribution of IR spectra PCA for different edible oils**

Let the IR spectra data compressed by PCA and KL decomposition as the input of the neural networks to establish the classifiers of PCA-BP and KL-BP. The classification results are shown in Table 3. Since IR spectra projections of edible oils overlap on the principal component subspaces, and the nonlinear classifiers such as the neural networks are

needed to classify them.

**Table 3 Classification results of PCA, PCA-BP and KL-BP**

Methods	Sample number	Correct classification	Error classification	Accuracy /%
PCA	18	11	8	61.1
PCA-BP	18	12	6	66.7
KL-BP	18	15	3	83.3

## 4 Conclusions

In this work, a method of molecular spectra analysis based on image recognition for rapid classification of edible oil is proposed. The proposed method can not only increase the degree of visualization of the molecular spectra, but also enhance the differences between the spectra. The comparative experiment results have verified that the proposed method had better classification of edible oils than that of traditional methods such as PCA, PCA-BP and KL-BP. The proposed method has provided a new way to classify the edible oil rapidly based on molecular spectra analysis.

## References

- [ 1 ] Wei Ming, Cao Xinzhi, Liao Chenghua. Food Science, 2003, 24(12): 103.
- [ 2 ] Li Lin, Sun Qiuju, Xin Shigang, et al. Applied Mechanics & Materials, 2013, 395-396: 355.
- [ 3 ] Chang-Mo L I. China Oils & Fats, 2007.
- [ 4 ] Yoshimi Kitada, Yasuyuki Ueda, Masatoshi Yamamoto, et al. Journal of Liquid Chromatography & Related Technologies, 1985, 8(1): 47.
- [ 5 ] Xu L, Yu X, Lei L, et al. Food Chemistry, 2016, 202: 229.
- [ 6 ] Candolfi A, Maesschalck R D, Massart D L, et al. Journal of Pharmaceutical & Biomedical Analysis, 1999, 19(6): 923.
- [ 7 ] Chen Y, Thosar S S, Forbess R A, et al. Drug Development & Industrial Pharmacy, 2001, 27(7): 623.
- [ 8 ] Conde O M, Amado M, García-Allende P B, et al. Proceedings of SPIE - The International Society for Optical Engineering, 2007, 6565: 65650M.
- [ 9 ] Liu Jian, Wu Fei, Yao Lei, et al. Journal of Electronics and Information Technology, 2010, 1188.
- [10] Kaufman L, Rousseeuw P J. Finding Groups in Data. An Introduction to Cluster Analysis. Wiley, 1990.
- [11] Sun Q S, Zeng S G, Liu Y, et al. Pattern Recognition, 2005, 38(12): 2437.
- [12] Jordanov I, Georgieva A. Feed Forward Neural Networks for Automated Classification. Cognitive Informatics (ICCI), 2010 9th IEEE International Conference on IEEE, 2010. 578.
- [13] Ma Xiao, Yuan H F. Spectroscopy and Spectral Analysis, 2015, 35(7): 1879.

# 基于分子光谱图像识别的食用油快速分类研究

曹玉婷<sup>1</sup>, 赵 众<sup>1\*</sup>, 袁洪福<sup>2</sup>, 李 彬<sup>1</sup>

1. 北京化工大学信息科学与技术学院, 北京 100029

2. 北京化工大学材料科学与技术学院, 北京 100029

**摘 要** 分子光谱分析技术结合化学计量学已成为一种非常活跃的食用油鉴别方法。然而, 当不同类型的样本之间的光谱差异极其微小时, 利用传统的分类技术也很难将其分开。为了完成相似品种食用油的快速识别和分类, 收集了包括芝麻油、玉米油、油菜籽油、调和油、葵花油、花生油、橄榄油七种食用油的衰减全反射红外光谱, 在此基础上, 采用图像识别的方法对七种食用油进行快速分类。在所提出的图像识别方法中, 首先, 将通过多元散射校正预处理后的红外光谱吸光度矩阵进行自相关运算, 利用等高线原理根据吸光度强度值的不同生成光谱图像, 以扩大的光谱差异并提高光谱可视化。然后, 根据图像膨胀的原理找到光谱图像的局部特征点, 将其作为图像特征。最后, 使用 BP 神经网络对特征点进行训练和分类预测。为了对比所提出的方法, PCA-BP 和 KL-BP 的方法被用于与图像识别的方法进行比较, 实验结果表明, 图像识别方法的正确识别率为 94.4%, 高于 PCA-BP 的 66.7% 和 KL-BP 的 83.3%。所提方法为实现食用油的快速识别和检测提供了一条新的有效途径。

**关键词** 食用油; 分子光谱; 光谱可视化; 特征提取; 图像识别

(收稿日期: 2017-03-21, 修订日期: 2017-09-15)

\* 通讯联系人