Computers and Electronics in Agriculture 147 (2018) 64–69

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original papers

Discrimination of tea varieties using FTIR spectroscopy and allied Gustafson-Kessel clustering

Xiaohong Wu[a,b,*], Jin Zhu[a], Bin Wu[c], Jun Sun[a], Chunxia Dai[a,d]

[a] School of Electrical and Information Engineering, Jiangsu University, Zhenjiang 212013, China
[b] Key Laboratory of Facility Agriculture Measurement and Control Technology and Equipment of Machinery Industry, Jiangsu University, Zhenjiang 212013, China
[c] Department of Information Engineering, ChuZhou Vocational Technology College, Chuzhou 239000, China
[d] School of Food and Biological Engineering, Jiangsu University, Zhenjiang 212013, China

ARTICLE INFO

Keywords:
Tea
Fourier transform infrared reflectance
Fuzzy clustering
Classification methods

ABSTRACT

For the purpose of classifying tea varieties, allied Gustafson-Kessel (AGK) clustering was proposed to cluster the Fourier transform infrared reflectance (FTIR) spectra of tea samples. As a fuzzy clustering algorithm, AGK can not only produce fuzzy membership and typicality values but also cluster various shapes of data with the help of Gustafson-Kessel (GK) clustering. After FTIR spectra were collected by FTIR-7600 infrared spectrometer, they were preprocessed with multiple scatter correction (MSC). To reduce the dimensionality of FTIR spectra and make the classification of data easily, principal component analysis (PCA) and linear discriminant analysis (LDA) were used to process the FTIR spectra. After that, fuzzy c-means (FCM) clustering, possibilistic c-means (PCM) clustering, AGK clustering and allied fuzzy c-means (AFCM) clustering were performed to cluster data, respectively. The clustering accuracy of AGK achieved 93.9% which was the highest one than other fuzzy clustering algorithms. The results obtained in experiments showed that AGK coupled with FTIR spectroscopy could provide an effective discrimination model for classification of tea varieties successfully.

1. Introduction

As a healthy beverage (Sun et al., 2017), tea is being greeted with increasing approval in the world. Tea originated in China and the custom of drinking tea is from Sichuan, China. Drinking tea often keeps you in the pink because tea contains some wholesome compounds such as catechins, cholestenone, caffeine (Li et al., 2015; Sinija and Mishra, 2009), inositol, pantothenic acid, amino acid (Li et al., 2013) and folic acid. Besides a beverage, tea can be served as the good medicine of the bowel disease, coronary heart disease, hypertension, etc. For different varieties of tea, the effectiveness is often distinct. For example, pu'er tea can lower blood fat and prevent from atherosclerosis and coronary heart disease. Tie Guanyin, a variety of oolong tea, has anti-aging and anti-cancer effect. Zhu Yeqing can relieve heat and quell thirst with the effect of detoxication and diuresis. Moreover, the factors affecting the quality of the tea, such as color and luster, sweet smell and taste, are closely related to variety, cultivation conditions and storage conditions. The consumers want to buy real tea, not the fake in the tea market. Therefore, it becomes an urgent task to set up a rapid and effective classification of tea varieties for consumers, researchers and farmers.

Human sensory evaluation is a common method for evaluating the quality of tea (Yaroshenko et al., 2014; Zhi et al., 2013). Trained sensory panels always grade tea samples according to their appearance, aroma, soup color, taste and Securinega through their vision, taste, smell and touch. However, training skilled tea personnel spends a lot of time and money. Furthermore, for the same tea sample, sensory evaluation results are not the same to different professionals because the results are influenced by experience, gender, mental state, physical condition and other factors.

Recently, some studies have been described to classify tea varieties using chemical analysis methods, such as high performance liquid chromatography (HPLC) (Wang et al., 2014a,b), gas chromatography-mass spectrometry (GC–MS) (Ding et al., 2015; Lv et al., 2015), liquid chromatography coupled with tandem mass spectrometry (LC-MS) (Zhang et al., 2014), and inductively coupled plasma optical emission spectrometry (ICP-OES) (Szymczycha-Madeja et al., 2015). Nevertheless, chemical analysis methods mentioned above are complex and time-consuming to be used in discriminating tea varieties.

In order to meet the rapid development of tea market and people's demand for higher quality of tea, it is urgent to develop a fast, safe and green detection technology for identification of tea varieties. Fourier transform infrared reflectance (FTIR) spectroscopy or near-infrared

* Corresponding author at: School of Electrical and Information Engineering, Jiangsu University, Xuefu Road No. 301, Zhenjiang 212013, China.
E-mail address: wxh_www@163.com (X. Wu).

https://doi.org/10.1016/j.compag.2018.02.014
Received 4 August 2017; Received in revised form 26 January 2018; Accepted 10 February 2018

spectroscopy (Sinija and Mishra, 2008), as a fast and nondestructive technology, can be served as a powerful analytical tool and it has been widely used to discriminate tea varieties (Cai et al., 2015; Wu et al., 2016), apple (Wu et al., 2016a; Wu et al., 2016b), meat (Kodogiannis et al., 2014), jujube (Yang et al., 2015), olive oil (Jiménez-Carvelo et al., 2017), fish species (Alamprese and Casiraghi, 2015), etc. Some researchers studied the feasibility of rapidly discriminating tea varieties using FTIR spectroscopy or near-infrared spectroscopy coupled with supervised pattern classification methods (Diniz et al., 2014; He et al., 2012, 2007), such as back propagation artificial neural network (BP-ANN) (Chen et al., 2007, 2009) and support vector machine (SVM) (Chen et al., 2007, 2009). BP-ANN, a kind of artificial neural network (ANN), is a nonlinear learning method with a few layers networks, and it can solve nonlinear classification problem. But BP-ANN easily gets trapped at local minima and does not converge to the global minimum point (Chandwani et al., 2015). Under the principle of structural risk minimization, SVM can solve local minima problem and carry out nonlinear classification with kernel trick. However, it is difficult to adjust the parameters of SVM such as the kernel parameter for the radial basis function (RBF) kernel, the most frequently used kernel function, and the soft-margin parameter C (Chang and Chou, 2015). Moreover, to search the proper parameters for optimizing SVM may require a large number of calculations (Liu et al., 2011). Therefore, it makes sense that simple and effective classification methods should be researched for identification of tea varieties.

As an unsupervised classification method, fuzzy clustering always shows better performance than traditional one. Fuzzy clustering has been widely used in digital image processing, computer vision and pattern recognition (Bezdek et al., 1999). Fuzzy c-means (FCM) clustering, a well-known fuzzy clustering, derives its origin from hard c-means (HCM) clustering algorithm (Bezdek, 1981). FCM has the probabilistic constraint interpreting memberships as degrees of sharing (Bezdek, 1981). However, because of the probabilistic constraint, FCM is sensitive to noise (Barni et al., 1996). How to deal with the noisy data is an important issue in designing fuzzy clustering models. To overcome the noise sensitivity drawback of FCM, Krishnapuram and Keller have presented the possibilistic c-means (PCM) clustering algorithm by abandoning the constraint of FCM and constructing the novel objective functions (Krishnapuram and Keller, 1993). PCM can deal with noisy data better than FCM, but it is very sensitive to initializations and sometimes generates coincident clusters. PCM attaches importance to the possibility (typicality) but neglects the important membership. To combine the benefits of FCM and PCM, an allied fuzzy c-means (AFCM) clustering has been proposed to produce memberships and possibilities simultaneously (Wu and Zhou, 2006). On the other hand, Gustafson and Kessel proposed Gustafson-Kessel (GK) clustering to cluster data containing different geometric shapes (Costel et al., 2007). However, there are few reports on applying fuzzy clustering algorithms for classification of tea varieties together with FTIR spectroscopy.

In this work, FTIR spectroscopy was used as a nondestructive technology in detecting tea samples with different varieties. A novel fuzzy clustering called allied Gustafson-Kessel (AGK) clustering was proposed by combination of AFCM and GK for classification of tea varieties. After FTIR spectra of tea samples were processed by multiple scatter correction (MSC), principal component analysis (PCA) and linear discriminant analysis (LDA), AGK was performed for classifying the tea varieties and its clustering accuracy and the computing time were compared with those of FCM, PCM and AFCM. Furthermore, we pointed out the noise sensitivity problem of GK and our proposed AGK can solve this problem. The main objectives of this research are: (1) to investigate the potential of FTIR for classification of tea varieties; (2) to propose allied Gustafson-Kessel (AGK) clustering based on AFCM and GK; (3) to compare the clustering accuracy and the computing time of FCM, PCM, AFCM and AGK in classifying FTIR spectra of tea samples.

2. Materials and methods

2.1. Sample preparation

Three varieties of tea samples: Emeishan Maofeng, high quality Leshan trimeresurus and low quality Leshan trimeresurus were prepared in this study. They came from two regions: Emeishan (Emeishan Maofeng) and Leshan (Leshan trimeresurus). The number in each variety of them is 32 and the total number is 96. After all samples were ground with a small mill, 0.5 g tea powder from each sample was evenly mixed with KBr according to ratio 1:100, and 1 g mixture was chosen to be pressed with film as one sample for FTIR experiments.

2.2. Spectral acquisition

The FTIR spectra of tea samples were collected using a FTIR-7600 infrared spectrometer (Lambda Scientific Pty Ltd, Edwardstown, Australia) with the high-sensitivity Deuterated Triglycine Sulphate (DTGS) detector. After the spectrometer was turned on and warmed up for one hour, each spectrum was acquired as the average of 32 scans over the range of 4001.569–401.1211 cm^{-1} with a sampling interval of 1.9285 cm^{-1}. The dimensionality of the spectrum is 1868. Each tea sample was made three separate spectral measurements and the average of the three spectra was prepared as the final datum for further data analysis. Because infrared spectrophotometer is sensitive to the change of the temperature and the humidity in laboratory, the spectral collection was operated at temperatures of around 25 °C, and on the relative humidity of 50–60%.

2.3. Spectra preprocessing method

The FTIR spectra of tea samples with the wave numbers from 4001.569 cm^{-1} to 401.1211 cm^{-1} were shown in Fig. 1. The spectra contain not just spectral absorption information related to chemical content of tea samples but light scatter information. Light scattering were influenced by physical factor such as particle size, shape and distribution, and there are possible differences in the light scatter information of different samples (Wang et al., 2014a,b). Under the influence of light scatter information, classification results are not satisfactory, and multiple scattering correction (MSC) is the commonly used method of resolving it effectively. Standard normal variate (SNV) is also the widely used pre-processing method in FTIR spectroscopy and it is similar to MSC in reducing the scattering effects of the spectrum (Zhao et al., 2016). In this study, MSC was utilized to preprocess the FTIR spectra and Fig. 2 illustrated the results.

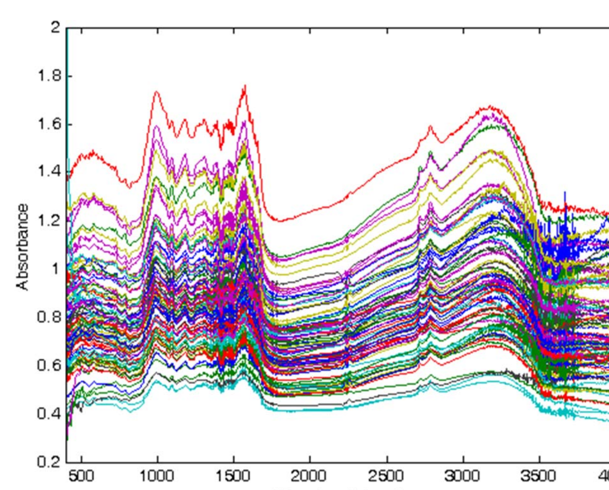

Fig. 1. The FTIR spectra of tea samples.