Computers and Electronics in Agriculture 112 (2015) 28–35

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Determination of tea polyphenols content by infrared spectroscopy coupled with iPLS and random frog techniques

Xiaoli Li, Chanjun Sun, Liubin Luo, Yong He *

College of Biosystems Engineering and Food Science, Zhejiang University, 866 Yuhangtang Road, Hangzhou 310058, China

ARTICLE INFO

Article history:
Received 1 April 2014
Received in revised form 31 December 2014
Accepted 3 January 2015
Available online 9 February 2015

Keywords:
Infrared spectroscopy
Tea polyphenols
Interval partial least-squares
Random frog
Wavenumber selection

ABSTRACT

The potential of infrared spectroscopy for fast determination of tea polyphenols (TP) of 14 cultivars of tea trees was investigated based on data mining technique. And the TP determination models were respectively developed for large leaf cultivars, middle leaf cultivars and all the cultivars. Interval partial least squares (iPLS) was proposed to extract and optimize feature from full-spectrum data. Regression models were respectively established based on PLS, iPLS and biPLS. Modeling results showed that the model based on the biPLS with the optimal subinterval selection (2452-dimensional wavenumbers) outperformed the other models, and the optimal regression model was obtained with high validation correlation of 0.9059, and low RMSE of 1.0277. On the basis of the optimal subinterval selection from biPLS, a further excavation of characteristic wavenumber was done by random frog. Thus, 18 optimal wavenumbers were selected for the TP determination, and the corresponding linear formula of the TP measurement was established. The results proved the feasibility of infrared spectra for measurement of the TP content of tea.

1. Introduction

Tea is widely consumed all over the world and tea polyphenols (TP) is considered to be an important source of beneficial bioactive compounds, such as antioxidants, anticarcinogenics, anti-inflammatory, antimicrobials and anti-arteriosclerotic compounds (Lu et al., 2005b). These natural chemicals exhibit excellent radical scavenging and hence antioxidative properties, and have been widely employed as food antioxidants (Graf et al., 2005).

The main components of TP are catechins and a small amount of caffeine. In the fresh leaves, the catechins content decrease with the increase of maturity (Vuong and Roach, 2014); Polyphenols in tea is a major part of water-soluble pigment, which is also the main color of tea; The main ingredient in the astringency of tea catechins is ester type; The composition and concentration of polyphenols are important material bases of tea quality and tea shade. Caffeine content varies in different leaf positions, usually the more content was appeared in the fresher position, so it can be used as a degree market of freshness; Also, it is an important material basis of bitter in the tea soup. Moreover, caffeine can excite the nerve center (Wanyika et al., 2010), has diuretic and other pharmacological function, making a very significant impact on the quality of tea.

The traditional method for measuring TP is a time-consuming chemical operation (Chen et al., 2008a). So it is an urgent requirement to develop a fast and nondestructive method in tea industry. Nowadays Infrared spectroscopy (IR) technology is widely used in varies disciplines. It is a non-destructive testing technology, with low equipment cost, fast spectral acquisition, and no chemical reagent.

Infrared spectra is generally the baseband frequency range of organic compounds, which can give very rich structure information (Zhao, 2012). Different compounds have different specific infrared spectra, the number, position, shape and intensity of band change with the variation of compounds and the states of aggregation. Therefore, infrared spectra could be used as a fingerprint identification tool to determine the existence of functional groups and compounds according to the spectra and thus the qualitative analysis of the compounds could be performed. Fourier transform mid-infrared (FT-MIR) spectroscopy was applied to determine the degree of substitution (DS) of carboxymethyl starch (CMS) by Liu et al. (2012). Abdul Rohman used Fourier transform mid infrared (FT-MIR) spectroscopy to detect and quantify the adulteration in virgin coconut oil (Rohman and Che Man, 2011). Liu used Fourier transform mid-infrared spectroscopy to detect if lotus root powder was adulterated with potato starch and/or sweet potato starch (Liu et al., 2013). Fen (2007) investigated the mid-infrared spectra of 40 cultivars of tea to develop a discriminant method for the identification of tea varieties. In this study, the inner structure correlated with the TP was tentatively explored.

* Corresponding author. Tel./fax: +86 0571 86971143.
E-mail address: yhe@zju.edu.cn (Y. He).

http://dx.doi.org/10.1016/j.compag.2015.01.005
0168-1699/

Near-infrared spectroscopy (NIRS) is a relatively new technology, which can take advantage of the full spectrum for qualitative or quantitative analysis (He et al., 2006). Li and He (2008) selected three varieties of tea to develop a method for discriminating varieties of tea plant based on their visible/near infrared reflectance (Vis/NIR) spectral characteristics. NIR technology can also be used in quantitative analysis (Chi-Leung et al., 2004). Chen et al. (2006) and Lu et al. (2005a) explored the possibility to use near infrared (NIR) spectroscopy as a rapid method to predict the content of caffeine and total polyphenols in green tea on the basis of 50 tea samples and 46 tea polyphenol samples, respectively. Bian et al. (2013) investigated the reflectance spectroscopy of total tea polyphenols and compared the accuracy of the determinations at three different levels (powder, leaf and canopy) with 48 observations. Although both qualitative and quantitative experiments on tea by near infrared spectroscopy had been done before, these experiments with large sample size were mainly used to identify the cultivars of tea, while the determinations of TP content were just based on small sample size. Few experiments were focused on the detection of TP content in various kinds of tea and the unrepresentative samples may influence the final results. In this study, 14 different cultivars of tea were used to obtain broadly representative results between the TP content and infrared spectroscopy.

Data mining and feature optimization are necessary to determine the inherent structure of spectral data and to segregate noise in the data. To measure the concentration of vitamin C in aqueous solution, Suhandy et al. (2011) used interval partial least squares (iPLS) regression to develop calibration model, and the prediction performance was improved, which revealed the superiority of iPLS in model calibration. Backward interval partial least squares (biPLS) regression was adopted to yield estimation of rice for overcoming problems caused by model overfitting (Takayama et al., 2012). Chen and Zhao (2014) determined the water content in biodiesel using random frog to select the characteristic wavelengths and obtained high accuracy with correlation coefficient of 0.965.

Considering the good performances of the models mentioned above, we first compared the capacities of the three algorithms including partial least squares (PLS) regression, interval partial least squares (iPLS) regression and backward interval partial least squares (biPLS) regression method for establishment of TP determination models. On this basis, we further combined random frog to obtain the optimal characteristic wavenumbers. All the results were evaluated by correlation coefficient (R) and root-mean-square error (RMSE).

The objectives of this study were: (1) to investigate the potential of infrared spectroscopy for detection of the TP of the 14 cultivars tea trees; (2) to evaluate the performances of the wavenumber selection algorithms including PLS, iPLS, biPLS and random frog.

2. Materials and methods

2.1. Samples

In this work, tea samples were taken from Tea Plantation (Germplasm Resources) of Zhejiang University (TPGRZU) (120.191E, 30.261N), China. The physiological age of these tea plants were about 35 years. To increase the diversity and generalization of the sample, 160 samples from 14 cultivars of tea trees at three different locations were obtained. In detail, the heights of these tea trees were around 130 cm^{-1}. As the root position on the ground was marked 0 cm^{-1} and the top part of the tree was marked 130 cm^{-1}, 3 locations were set as the up location with the height between 90 and 130 cm^{-1}, the middle location with the height between 40 and 90 cm^{-1} and the down location with the height between 0 and 40 cm^{-1}. Furthermore, only the first unfolded leaf in the up location, the second unfolded leaf in the middle location and the third unfolded leaf in the down location were respectively picked. And 6 leaves from the same type of location were taken as a sample and each cultivar consisted of about 12 samples from 3 locations. The size, shape and color of the leaves from the same location were similar.

2.2. Preparation and chemical experiment

The leaves were wiped slightly on the surface after taken, to prevent the impact of dust on experiment. Then, samples were heated at 103 °C ± 2 °C by electrothermal constant-temperature dry box (DHG-9070A, Shanghai, China) to constant weight following the national standard GB/T 8304-2002. Afterwards each of the samples was grinded by micronizer (Tissuelyser-48, Shanghai, China), under the same condition of 60 Hz, 60 s. Then samples were filtered through 60 mesh sieve, weighed 0.2 g for the chemical experiment, and 0.01 g for the spectroscopy experiment. The content of the TP was determined following the Chinese national standard GB/T 8313-2008. And the detailed information about the 14 cultivars is presented in Table 1.

2.3. Mid-IR spectroscopy experiment

The MIR spectra were collected using a Jasco FT/IR-4100 spectrometer (Japan) with a TGS detector and transmission with a ZnO crystal sampling accessory. The spectrometer was operated in the range of 7800–350 cm^{-1}, resolution of 4 cm and auto scans for each spectrum. The Signal to Noise Ratio (SNR) was 22000:1. A high-strength ceramic light source was placed inside and the 450 Michelson interferometer could be adjusted automatically. Since moisture in the environment had a great impact on the detection, drier was placed inside the detecting cavity and the temperature was kept at about 25 °C during the whole experiment.

Before the experiment, KBr (SCR, Shanghai, China) was dried at 105 °C for about 5 h in electrothermal constant-temperature dry box (DHG-9070A, Shanghai, China) and then cooled to room temperature in the dryer. Successively, the sample powder was mixed sufficiently with KBr using agate mortar. The mass ratio between them was 1:49, which consulted the previous researches (Zhu et al., 2014). In this study, the sample and KBr were weighed 0.01 g and 0.49 g, respectively. After that, the mixed powder filling in the sample holder was extruded by tablet machine. Successively, the spectra were taken after the sample tablet placed in the sample cell.

2.4. Samples partition

To evaluate the TP determination model, the whole samples were divided into calibration sample set and prediction sample set. Firstly, the total 160 samples were sorted in descending sort according to their respective values of the TP content. Subsequently, the median of every three was selected for the prediction set. Calibration and prediction set containing 106 and 54 samples respectively were obtained and the range of the TP content values in the prediction set was within the one in the calibration set. The detailed information was shown in Table 2.

2.5. Establishment of the TP model

Partial least squares (PLS) is one of the most widely used multivariate calibration methods in infrared spectroscopy analysis. It makes the variance of the main components calculated maximization, meanwhile the greatest degree of correlation between main components and chemical score is obtained. The results were evaluated by the RMSE and the correlation coefficient (r). A good model