Engineering in Agriculture, Environment and Food 8 (2015) 195–199

Contents lists available at ScienceDirect

Engineering in Agriculture, Environment and Food

journal homepage: http://www.sciencedirect.com/eaef

Research paper

Coffee roasting degrees prediction in terms of weight loss with selected wavebands based on near-infrared spectroscopy

Jiajia Shan*, Tetsuhito Suzuki, Yuichi Ogawa, Naoshi Kondo

Graduate School of Agriculture, Kyoto University, Kitashirakawa-Oiwakecho, Sakyo-ku, Kyoto 606-8502, Japan

ARTICLE INFO

Article history:
Available online 9 October 2015

Keywords:
NIR spectroscopy
Roasted coffee
Weight loss
MLR
PLS regression

ABSTRACT

Waveband selection was investigated to improve a coffee roasting degrees model in terms of weight loss. Near infrared diffuse reflectance spectra were obtained from ground roasted coffee samples. Characteristic wavebands were determined by regression coefficients of optimal principle components (PCs) of a partial least square (PLS) regression model. After that, a multiple linear regression (MLR) model was developed based on the selected wavebands. For comparison, PLS models based on different pre-processing spectra were also established. The model results of PLS and MLR method were compared. The promising result of the MLR model (Rcv^2 of 0.983 and RMSECV of 0.612) indicated that waveband selection could improve regression model performance and reduce complexity, compared to the PLS regression model.

1. Introduction

Coffee is the second most important traded raw material in international trade, and a popular beverage that has demanding customer quality expectations. One of the most important determinants of coffee cupping quality is directly related to the coffee roasting process. During roasting, a series of chemical and physical reactions lead to formation of components responsible for sensory qualities that define cup of coffee quality (Redgwell et al., 2002; Oliveira et al., 2005). Thus, how coffee is roasted will significantly affect parameters associated with customer perception of cupping quality. This roasting process can be defined by roasting degrees.

At present, the criteria commonly used to monitor roasting degrees of coffee beans in the industry are color and cracking, while the final product is evaluated in terms of flavor and roasted coffee weight loss. Color, flavor and cracking parameters are inaccurate and depend on the subjective judgment of roasters. Thus, if weight loss could be monitored real-time during the roasting process, it could provide a more objective indicator of the roasting degrees.

Near infrared (NIR) spectroscopy has been widely used in the food chemistry field as a rapid and non-destructive technology. Moreover it has been used to investigate coffee quality, including coffee variety differentiation, dry matter content determination, and coffee sensory prediction (Ribeiro et al., 2011; Esteban-Diez et al., 2004; Huck et al. 2005). Few studies though have explored its use for measuring the roasting degrees.

The one such reported study that we could find in the literature used NIR as an analytical tool to predict roasting degrees in terms of moisture, density and weight loss (Alessandrini et al., 2008). They were able to develop successful regression models based on NIR spectra over a broad wavelength range. However, such models will contain both informative and uninformative variables, which can lead to low predictive robustness. Moreover, such broad spectra typically incorporate a number of spectra variables that can be highly correlated with each other, and lead to the over-fitting of the regression model. Thus, a more sophisticated statistical method is needed to improve the weight loss prediction models' performance.

Variable selection is one such statistical method that can be used for improving the models' performance and robustness. In this method many irrelevant, noisy and unreliable variables are removed. Moreover, the variable selection method can provide a more easily understood model, since all the variables that do not positively contribute to the model are eliminated. An additional benefit, particularly for on-line industrial applications, would be a reduction in sampling time, since the full spectral range would not be needed to be scanned (Andersen and Bro, 2010; Arakawa et al., 2011; Xu et al., 2012).

Several wavelength interval selection approaches have been proposed, such as moving window partial least square (MWPLS)

* Corresponding author.
E-mail address: sjjkyoto@gmail.com (J. Shan).

http://dx.doi.org/10.1016/j.eaef.2015.10.003
1881-8366/

and interval partial least square (iPLS). Individual wavelength selection has also studied using a genetic algorithm (GA) (Arakawa et al. 2011; Xu et al., 2012). However, the GA approach may add complexity to the computation when the number of variables increases. Principle components (PCs) were proposed in this study. PCs are a set of linearly uncorrelated variables calculated from the PLS procedure. PCs account for as much of the variability in the raw data as possible. The regression coefficient of each variable from the PCs can help us identify those wavelengths that are significant to the regression model.

The objective of this study is to improve the performance of a coffee weight loss prediction model by selecting characteristic wavebands from raw NIR spectra. A novel method was proposed to select characteristic wavebands. Regression models developed by the PLS regression method over a full wavelength range and a MLR method with selected wavebands were established separately. A full cross validation method was used to evaluate the model's performance. The accuracy and predictive ability of the MLR model was compared to the PLS regression model.

2. Materials and methods

2.1. Sample preparation and chemicals

Arabica green coffee samples originating from Antigua Island were purchased from a supermarket. In order to develop a universal prediction model, a wide range of roasting degrees was obtained. Green coffee beans samples were roasted over the entire temperature range of a home coffee roaster (CBR-101A, Genesis Co., USA), from 60 to 250 °C at 5 °C intervals for 10 min. The resulting roasted coffee beans ranged from a light roast to a dark roast color. Totally, 39 roasted coffee bean samples with different roasting temperatures were obtained. The weight of each coffee sample was measured using electronic scales (AUW220D, Shimadzu Co., Japan), before and after roasting, and weight loss was calculated from this. To reduce the effect of uneven particle distribution, all roasted coffee samples were ground by an electric coffee grinder (DBM-8, Cinair Co., USA) and uniform particle size (600 μm) of ground roasted coffee was sieved and used for NIR measurement. All experiments were conducted carefully in closed conditions to eliminate moisture loss during grinding.

2.2. Recording of NIR spectroscopy

The NIR diffuse reflectance spectra were recorded on a UV–VIS/NIR spectrophotometer (V-670, JASCO Co., Japan), equipped with an integrating sphere detector (HISN-729, JASCO Co., Japan) connected to the spectrophotometer by optical fiber. In this setup, light was dispersed by a grating, and then routed by optical fiber to the integrating sphere, before arriving at the coffee sample. Diffuse reflectance light from the coffee sample was integrated and then detected. An integrating sphere is an optical component consisting of a hollow spherical cavity with its interior covered with a diffuse white reflective coating. As a consequence, it has a uniform scattering or diffuse effect on the light. The NIR instrument was controlled by a compatible computer, and Spectra Manager (V-670, JASCO Co., Japan) was used to acquire the data. The size of the light spot in the instrument was 4 cm in diameter. Care was taken to ensure the same amount of sample (around 2.45 g) was used to fill the sample holder (a 4 cm by 4 cm square) and a flat sample surface maintained. Each spectrum was measured from 1000 to 2200 nm, with a bandwidth of 8.0 nm, and a scan speed of 400 nm/min. All experiments were conducted at room temperature.

2.3. Data processing

A PLS regression model is a linear combination of PCs. PCs are composite variables, i.e. linearly uncorrelated functions of original variables calculated from orthogonal linear transformation, estimated to contain the main structural information of data in relationship with the Y-variable, in decreasing order. In this study, the PLS regression model was firstly developed based on NIR spectra over the whole wavelength range: 1000 nm–2200 nm. Different preprocessing methods (such as, standard normal variate (SVN), 1st derivate, and center and scale (C&S)) were performed on raw NIR spectra before PLS regression model development. Compared to different preprocessing methods, the best PLS regression model was determined.

For MLR model development, the variables selection was based on the framework of PCs. The optimal PCs were obtained from PLS regression model developed on the raw spectral data. Regression coefficient (weighted loading) of each waveband was taken from the PCs, and the regression coefficients were used to identify the significant wavebands. The MLR model was finally developed based on the selected wavebands.

A full cross validation method was used to evaluate prediction model's performance. The full cross validation method has been demonstrated to be a very useful tool to estimate the predictive ability of a regression model for future samples. Considering the limited number of samples, no attempt was made to perform additional external validation. The parameters used to evaluate the model's capability included: determination coefficient of calibration and cross validation (Rc^2 and Rcv^2), and root mean squared error of cross validation (RMSECV). Data preprocessing treatments and regression models development were carried out with Unscramble v9.8 (CAMO Software AS. OSLO, Norway).

3. Results and discussion

3.1. Observations on NIR spectra

NIR diffuse reflectance spectra of differently roasted coffee samples are shown in Fig. 1.

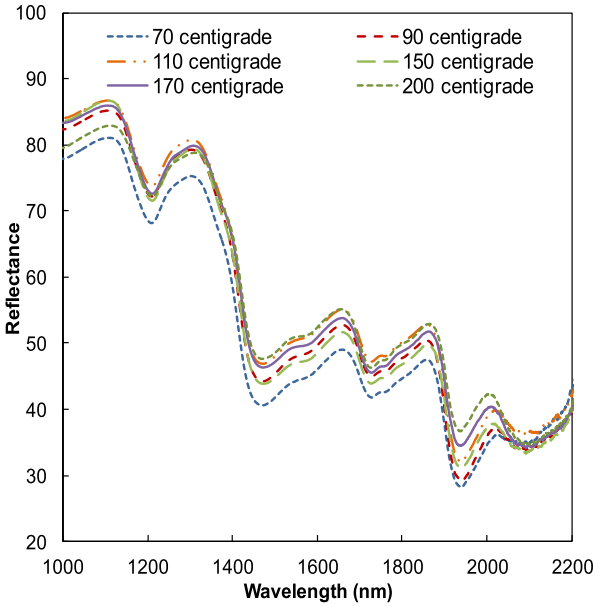

Fig. 1. NIR diffuse reflectance spectra of differently roasted coffee samples.