Food Chemistry 138 (2013) 192–199

Contents lists available at SciVerse ScienceDirect

Food Chemistry

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

Analytical Methods

Rapid detecting total acid content and classifying different types of vinegar based on near infrared spectroscopy and least-squares support vector machine

Shi Ji-yong [a], Zou Xiao-bo [a,b,*], Huang Xiao-wei [a], Zhao Jie-wen [a], Li Yanxiao [a], Hao Limin [c], Zhang Jianchun [c]

[a] School of Food and Biological Engineering, Jiangsu University, 301 Xuefu Rd, 212013 Zhenjiang, Jiangsu, China
[b] Key Laboratory of Modern Agricultural Equipment and Technology, 301 Xuefu Rd, 212013 Zhenjiang, Jiangsu, China
[c] The Research Center of China Hemp Materials, Beijing, China

ARTICLE INFO

Article history:
Received 19 October 2010
Received in revised form 10 October 2012
Accepted 18 October 2012
Available online 8 November 2012

Keywords:
Near infrared spectroscopy
Least-squares support vector machine
Vinegar
Total acid content
Principle component analysis
Back propagation artificial neural network
Partial least-square

ABSTRACT

More than 3.2 million litres of vinegar is consumed every day in China. There are many types of vinegar in China. How to control the quality of vinegar is problem. Near infrared spectroscopy (NIR) transmission technique was applied to achieve this purpose. Ninety-five vinegar samples from 14 origins covering 11 provinces in China were collected. They were classified into mature vinegar, aromatic vinegar, rice vinegar, fruit vinegar, and white vinegar. Fruit vinegar and white vinegar were separated from the other traditional categories in the two-dimension principal component space of NIR after principle component analysis (PCA). Least-squares support vector machine (LS-SVM) as the pattern recognition was firstly applied to identify mature vinegar, aromatic vinegar, rice vinegar in this study. The top two principal components (PCs) were extracted as the input of LS-SVM classifiers by principal component analysis (PCA). The best experimental results were obtained using the radial basis function (RBF) LS-SVM classifier with $\sigma = 0.8$. The accuracies of identification were more than 85% for three traditional vinegar categories. Compared with the back propagation artificial neural network (BP-ANN) approach, LS-SVM algorithm showed its excellent generalisation for identification results. As total acid content (TAC) is highly connecting with the quality of vinegar, NIR was used to prediction the TAC of samples. LS-SVM was applied to building the TAC prediction model based on spectral transmission rate. Compared with partial least-square (PLS) model, LS-SVM model gave better precision and accuracy in predicting TAC. The determination coefficient for prediction (R_p) of the LS-SVM model was 0.919 and root mean square error for prediction (RMSEP) was 0.3226. This work demonstrated that near infrared spectroscopy technique coupled with LS-SVM could be used as a quality control method for vinegar.

1. Introduction

Vinegar has been made and used dating from around 3000 BC and is an important element in Asian, European, Western, and other traditional cuisines of the world (Chen, Chen, & Guo, 2010).The Codex Alimentarius (1987) Commission of FAO/WHO joint commission defined vinegar as a liquid fit for human consumption, produced from a suitable raw material of agricultural origin, containing starch, sugars or starch and sugars, by the process of double fermentation, alcoholic and acetous, that can constrains a specified amount of acetic acid (Moros, Inon, Garrigues, & de la Guardia, 2008). Many medicinal components that are good for health have been reported in vinegar, such as carbohydrates (Leeman, Ostman, & Bjorck, 2005), organic acids (acetic, formic, lactic, etc.) (Cocchi et al., 2006), alcohols and polyols (ethanol, acetoin, 2,3-butanediol and hydroxymethylfurfural), amino acids and peptides, and volatile substances (ethyl acetate). Many remedies and treatments have been ascribed to vinegar, and great effort has been focused on identifying the functional components in vinegar and their reaction mechanisms (Xu, Tao, & Ao, 2007).

China has more than 5000 years history of producing vinegar. Every year over 26 million hectolitres of vinegar is produced in China. Chinese people treat the vinegar as favourite condiments, health products and even medicines. More than 3.2 million litres of vinegar is consumed every day in China. There are many types of vinegar in China. How to control the quality of vinegar is problem.

Different types of vinegar have the same main components, water and acid. Vinegar contains many nutrients, which contains more acids, such as acetic acid, lactic acid, pyruvic acid, formic acid, malic acid, citric acid, oxaloacetate, and succinate. These amount nutrients are connecting with the total acidity content (TAC). That is, TAC is highly connecting with the quality of vinegar. The United

* Corresponding author at: School of Food and Biological Engineering, Jiangsu University, 301 Xuefu Rd, 212013 Zhenjiang, Jiangsu, China. Tel.: +86 511 88780174; fax: +86 511 88780201.
E-mail address: zou_xiaobo@ujs.edu.cn (Z. Xiao-bo).

0308-8146/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.foodchem.2012.10.060

States Food and Drug Administration (FDA) require that any product called "vinegar" contain at least 4% acidity. The Codex standard proposed a minimum of 6% for wine vinegar and 5% for others because the percent of acetic acid present in the product varies according to what they are made from (Moros et al., 2008).

Nowadays, the quantitative analysis of vinegar acidity is performed according to various wet chemical method such as high performance liquid chromatography (HPLC) (García-Parrillaa, Camachob, Herediaa, & Troncoso, 1994; Romero, Muñoz, Alvarez, & Ibáñez, 1993), gas chromatographic (GC) (Cocchi et al., 2006), capillary electrophoresis (De Vero et al., 2006; Plessi, Bertelli, & Miglietta, 2006; Tang & Wu, 2005) and chemical titration, etc. However, all of the methods mentioned above are time-consuming in the quantitative analysis of vinegar total acid. Near infrared spectroscopy is a fast, accurate, easy and non-destructive technique that can be as a replacement of time-consuming chemical method.

Recently, NIR spectroscopy has been applied for the discrimination of aging of vinegar during storage (Casale, Abajo, Sáiz, Pizarro, & Forina, 2006), and prediction of chemical constituents such as organic acids during storage and aging (Sáiz-Abajo, González-Sáiz, & Pizarro, 2006), reducing sugars (Fu, Yan, Chen, & Li, 2005), total procyanidins (García-Parrilla, Heredia, Troncoso, & González, 1997), soluble solids content and pH of rice vinegar (Liu, He, & Wang, 2008a) and discrimination of fruit vinegar varieties (Liu, He, & Wang, 2008b). These researchers demonstrate that NIR could be a good method for quantity and quality control of vinegar.

Support vector machine (SVM) is a promising method proposed by Vapnik. Because of its attractive advantages and excellent performances, SVM has been widely applied in many fields (Burges, 1998; Comak, Arslan, & Turkoglu, 2007; Guo, Li, & Chan, 2001; Vapnik, 1998; Xiaobo, Jiewen, Povey, Holmes, & Hanpin, 2010). SVM is processed based on the statistical learning theory. The structural risk minimisation principle (SRM) is embodied instead of traditional empirical risk minimisation principle (ERM) which is employed by conventional neural network to avoid overfitting and multidimensional problem. SVM can be applied in both binary classification tool and regression tasks. Least-squares support vector machine (LS-SVM) is an optimised algorithm based on standard SVM by Suykens et al. (2002). LS-SVM has the capability for linear and nonlinear multivariate calibration and solves the multivariate calibration problems in a relatively fast way (Suykens & Vanderwalle, 1999; Xiaobo et al., 2010). It uses a linear set of equations instead of a quadratic programming (QP) problem to obtain the support vectors (SVs). It adopts least-squares linear system as the loss function and is applied in the pattern recognition and nonlinear evaluation. It is capable of learning in high-dimensional feature space with fewer training data. Several papers have been published in the spectral analysis (Borin, Ferrao, Mello, Maretto, & Poppi, 2006; Chauchard, Cogdill, Roussel, Roger, & Bellon-Maurel, 2004).

The aim of this paper is to investigate the potential of LS-SVM and near infrared spectroscopy technique for non-destructive measurement of Chinese vinegar. The LS-SVM model will be used for predication TAC and classification of Chinese vinegar based on NIR spectra.

2. Materials and methods

2.1. Sample preparation

Ninety-five vinegar samples of different origins as shown in Fig. 1 were collected from the industry and several supermarkets in China (the red point in Fig. 1 means the origin of vinegar). They were from 14 origins covering 11 province in China, could cover an important range of available types of vinegar in China in the market. All of these vinegars were commonly used in Chinese people's daily life. Before the experiment, the vinegar samples were stored in the laboratory at a constant temperature of 25 ± 1 °C for more than 48 h to have an equalisation room temperature. The samples were all original vinegar liquid without dilution.

2.2. Total acid content measurement

Reference analyses were in accordance with the Official Methods of Analysis for Chinese vinegar (GB/T 5009.41-2003) (GB means national standard; 5009.41 was the codes of the official methods; and 2003 was the year in which the methods were revised.). Absorbed 5.0 ml sample and placed in 100 ml volumetric flask, add distilled water to the mark, mix and absorb 20.0 ml then placed in 200 ml beaker, add distilled water to 60 ml, starting magnetic stirrer, titration with standard sodium hydroxide solution [c (NaOH) = 0.05 mol/L] titrated to pH = 8.2 which shows in pH meter and recorded consumption of standard sodium hydroxide solution (0.05 mol/L) of ml, then total acid content can be calculated. Three groups of each sample measured in parallel, 3 groups of determinations of the maximum relative error is less than 3%.

2.3. Spectra collection and preprocessing

The NIR spectra were collected in the transmission mode using an Antaris II type Fourier transform near infrared spectroscopy (Thermo Fisher, USA) with a standard glass colorimetric ware in the range from 10,000 to 4000 cm^{-1}. The spectra used for the data analysis were the average spectra of sixteen times scans, and the data were measured in 1.928 cm^{-1} intervals, which results in 3112 variables. The temperature and humidity were kept a steady level in lab (temperature was 25 ± °C and humidity was 60%).

Raw vinegars near infrared transmission spectra shown in Fig. 2, in which can be seen in the range of 4000–5300 cm^{-1} and 6500–7098 cm^{-1} spectral response value is almost 0. Therefore, these two ranges spectra will be removed in the following processing to improve the modeling efficiency. Selected spectral as the final model range are 5301–6498 cm^{-1} and 7100–10,000 cm^{-1}, results in 2126 variables. Multiplicative scatter correction (MSC) was used to correct for additive and multiplicative effects in the vinegar spectra.

2.4. Basic principle of least square support vector machine (LS-SVM)

The SVM is a supervised learning technique which is developed by Vladimir Vapnik and co-workers at AT&T Bell Laboratories in 1998 (Vapnik, 1998). It is applicable to both classification and regression. The SVMs are based on the principle of Structural Risk Minimisation. The SVM uses structural risk minimisation, rather than a non-convex, unconstrained minimisation problem, as in standard neural network training technique using empirical risk minimisation. Empirical risk minimises the misclassification error on the training set, whereas structural risk minimises the probability of misclassifying a previously unseen data point drawn randomly from a fixed but unknown probability distribution.

Assume that the training data with k number of samples is represented by $\{x_i, y_i\}$, $i = 1, 2, \ldots, k$, where $x \in R^n$ is an n-dimensional vector with $y \in \{-1, +1\}$ is the class label and linear classifier:

$$y(x) = sign[w^T x + b] \quad (1)$$

where w is the vector, b is a real constant.

When the data of the two classes are separable one can say:

$$\begin{Bmatrix} w^T x_k + b \geqslant +1, & if \quad y_k = +1 \\ w^T x_k + b \leqslant -1, & if \quad y_k = -1 \end{Bmatrix} \quad (2)$$