Infrared Physics & Technology 77 (2016) 65–72

Contents lists available at ScienceDirect

Infrared Physics & Technology

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

Regular article

Variables selection for quantitative determination of cotton content in textile blends by near infrared spectroscopy

Xu-dong Sun [a,b], Ming-xing Zhou [a], Yi-ze Sun [a,*]

[a] College of Mechanical Engineering, Donghua University, Shanghai 201620, China
[b] College of Mechanical and Electrical Engineering, East China Jiaotong University, Nanchang 330013, China

HIGHLIGHTS

- Variables selection was developed for analysis of textile blends by NIR spectroscopy.
- MCUVE-SPA could separate irrelevant variables and improve model robustness.
- Correlation coefficient and RMSEP of PLS model was 0.988 and of 2.100%.

ARTICLE INFO

Article history:
Received 2 August 2015
Revised 17 May 2016
Accepted 17 May 2016
Available online 18 May 2016

Keywords:
Near infrared spectroscopy
Monte Carlo uninformative variables elimination
Successive projections algorithm
Cotton
Blend fabrics

ABSTRACT

Investigations were initiated to develop near infrared (NIR) techniques coupled with variables selection method to rapidly measure cotton content in blend fabrics of cotton and polyester. Multiplicative scatter correction (MSC), smooth, first derivative (1Der), second derivative (2Der) and their combination were employed to preprocess the spectra. Monte Carlo uninformative variables elimination (MCUVE), successive projections algorithm (SPA), and genetic algorithm (GA) were performed comparatively to choose characteristic variables associated with cotton content distributions. One hundred and thirty-five and fifty-nine samples were used to calibrate models and assess the performance of the models, respectively. Through comparing the performance of partial least squares (PLS) regression models with new samples, the optimal model of cotton content was obtained with spectral pretreatment method of 2 Der-Smooth-MSC and variables selection method of MCUVE-SPA-PLS. The correlation coefficient of prediction (r_p) and root mean square errors of prediction (RMSEP) were 0.988% and 2.100%, respectively. The results suggest that NIR technique combining with variables selection method of MCUVE-SPA has significant potential to quantitatively analyze cotton content in blend fabrics of cotton and polyester; moreover, it could indicate the related spectral contributions.

1. Introduction

Determining the composition of textile blend is a significant textile testing process for textile products. The textile products must be labeled their fiber composition on the label. The textile products with incomplete or incorrect information of fiber composition will affect their sale directly. Chemical dissolution method is widely used in the textile industry for accurate analysis of fiber composition [1]. Taking two-component blend fabrics as an example, some related reagent is utilized for dissolving of one component, leaving the other as an insoluble fraction. Then blending rate is calculated through precise weighing of the insoluble fraction. However, this method is time-consuming, extensive use of auxiliary chemicals and costly. Besides, a large number of laboratory technicians have to contact with the chemicals directly, which may cause health hazards. Therefore, a rapid and high-efficiency analytical method is essentially required.

Near infrared (NIR) spectroscopy with the fast, accurate and nondestructive advantage has been applied in many fields [2,3]. NIR spectroscopy has shown promising ability for chemical content analysis in textile industry [4–10]. However, accurate determining of the composition of textile blend is difficult when samples have wide and diverse origins [11]. Actually, in textile manufacturing, many parameters such as yarn, fiber, color, finish material, or supplier are impossible to control. And coloration methods, fabric parameters and surface effects may dramatically impact the NIR spectral response among the samples [12–14]. This raises a number of problems when developing a robustness regression model

* Corresponding author at: College of Mechanical Engineering, Donghua University, No. 2999 Renminbei Road, Songjiang District, Shanghai 201620, China.
E-mail address: sunyz58@sina.com (Y.-z. Sun).

http://dx.doi.org/10.1016/j.infrared.2016.05.020
1350-4495/

for blend composition from NIR spectra. What is more, the modern spectroscopy instrumentations usually possess high resolution, with hundreds or thousands of spectral variables including collinearity, redundancies, and noise [15]. Therefore, many variables selection methods such as genetic algorithm (GA), successive projections algorithm (SPA) and Monte Carlo uninformative variables elimination (MCUVE) have been proposed to separate the irrelevant variables and improve model robustness [16–18]. GA has stochastic nature, results are realization dependent and variables selection may not be reproducible. In this work, a hybrid variables selection algorithm, MCUVE by partial least square (PLS) with SPA, was applied for NIR spectral variables selection. Both MCUVE and SPA are deterministic algorithms, which do not employ stochastic operations. So far, there are few publications focusing on variables selection for measuring cotton content in blend fabrics of cotton and polyester by NIR spectroscopy.

In this work, the feasibility of measuring cotton content in blend fabrics of cotton and polyester was investigated by NIR spectroscopy. The specific efforts had been attempted to a hybrid variables selection method of MCUVE and SPA for selecting informative variables in the wavenumber range of 3999.732–7096.921 cm^{-1}. And the establishment of the relationships between NIR spectra and cotton content of textile product based on PLS regression had been successfully developed.

2. Materials and methods

2.1. Sample preparation

A total of 194 samples were collected at Shanghai Textile Research Institute in Shanghai, China. The textile product was composed of cotton and polyester. Meanwhile, pure cotton and polyester were employed for investigating the NIR absorbance characterization. These textile products, extracted with petroleum ether previously, were cut into small pieces weight a little more than 1 g. For increasing signal to noise ratio of NIR spectra, all of these fabric pieces were split into fibers by laboratory technicians. Then they were employed to scan spectra and analyze cotton content.

2.2. Spectra collection

Spectra of the samples were scanned on a Tensor37 Fourier transform-NIR (FT-NIR) spectrometer (Bruker Optik GmbH, Ettlingen, Germany) fitted with the Petri dish accessory. NIR spectrum was recorded in the region of 3999.732–12492.895 cm^{-1} with the wavenumber interval of 4 cm^{-1} after 32 scans. The spectrum was collected one time when the Petri dish rotated 120°. Thus, three spectra were collected for each sample, and then they were averaged to be a mean spectrum, which was used in subsequent analysis, for providing high signal to noise ratio. Moreover, during the spectra measurements, the temperature was kept at approximately 20 °C, and a steady humidity level was maintained in the laboratory, because FT-NIR spectrometer is sensitive to changes in external environmental condition.

2.3. Chemical and spectral analysis

According to ISO 1833-11:2006, 700 mL of sulphuric acid with the concentration of 1.84 g/mL was added to 350 mL of water until cool to room temperature. Then, it was diluted to 1 L with water as solvent for cotton when dissolving. Then, 80 mL of ammonia liquor with the concentration of 0.88 g/mL was diluted to 1 L with water, and it was put to use as neutralization solution. Distilled water was used in the process of rinsing. All these solutions were also supplied by Shanghai Textile Research Institute.

The sample was put into the triangle flask, and 200 mL of sulphuric acid was added. The flask was covered with the glass stopper and placed one hour at the temperature of 50 ± 5 °C. The flask was shaken one time at the interval of ten minutes. After dissolution, the fibers were rinsed with neutralization solution and distilled water, and then suction-filtered by sand core crucible. Next, the remained textile fiber of polyester was dried in thermostatic drying oven (HOC-9030A, Hengqian technology Ltd, Shanghai) for more than four hours (no more than sixteen hours) maintaining the temperature of 105 ± 3 °C. After drying, test fibers were cooled in drying vessel at ambient temperature and dry atmosphere for half an hour and weighed with accuracy of 0.1 mg by BSA124S-CW, Sartorius, in the end.

Spectra collection and chemical analysis were performed on the same day. All samples were divided into calibration and prediction sets according to normal (Gaussian) distribution around the mean value, 135 samples were used for developing calibration models, and remains were used for prediction. The samples with high or low chemical values were put in the calibration set for ensuring the adaptability of calibration models. The MCUVE, SPA and GA were implemented with Matlab 2010a software package (MathWorks, USA). The PLS regression was performed with the statistical program of Unscrambler v8.0 software package (CAMO AS, Trondheim, Norway).

2.4. Background: MCUVE, SPA, GA and PLS regression

The PLS regression is commonly applied in quantitative analysis to correlate spectroscopic data (X) with related chemical data (Y) [19]. PLS is a bilinear modeling method in which the original independent information (X) is projected into a small number of components (also known as latent factors) to simplify the relationship between X and Y for prediction. The predictions Y are computed by Eq. (1). For NIR analysis, variables selection methods can make the model simpler and improve the predictive ability [20].

$$Y = \beta X + b \tag{1}$$

where β is a $(p \times 1)$ vector of regression coefficients, and b is the model offset.

In MCUVE method, regression coefficient matrix $\beta = [\beta_1, \ldots, \beta_p]$ is calculated through Monte Carlo method rather than leave-one-out validation. Because each coefficient β_j represents the contribution of the corresponding variable to the established model, the reliability of each variable j can be quantitatively measured by Eq. (2).

$$s_j = \mathrm{mean}(\beta_j)/\mathrm{std}(\beta_j) \quad j = 1 \cdots p \tag{2}$$

where $\mathrm{mean}(\beta_j)$ and $\mathrm{std}(\beta_j)$ are the mean and standard deviation of the regression coefficients of variable j. The stability of uninformative variables should be less than a threshold. In order to estimate a suitable threshold, an artificial random variable matrix N $(n \times p)$ with very small amplitude (e.g. 10^{-11}) is added to the NIR data to compute their stability. The stability of uninformative variable is less than that of random variables, and these variables should be eliminated. The cutoff threshold is usually defined by Eq. (3).

$$\mathrm{cutoff} = k \times \max(abs(s_{\mathrm{noise}})) \tag{3}$$

where k is an arbitrary value, e.g. 0.7 in this case. MCUVE can eliminate the uninformative variables for modeling than noise and so enhancing the impact of validation on modeling and increasing the probability of selection the best model [21].

The GA is a global search and optimization methods based upon the principles of natural evolution and selection developed by