

Rapid Detection of Cotton Content Based on Near Infrared Spectroscopy Technology

Li Dejun and Xing Chigang

School of Electronic and Electrical Engineering, Wuhan Textile University
ldj@wtu.edu.cn; xingchigang@hotmail.com

Abstract

The main research object in this paper is cotton-PET blended fabric. The study hopes that the content of cotton can be accurately obtained from the data of only a few near-infrared spectral point. In the research we choose spectra data within 1400nm-2000nm to establish the calibration model. At the time of modeling, we use the method which combine stepwise regression and normalization. The result indicates that the method can not only reduce the influence of fabric structural difference on quantitative analysis, but also achieve a relative accurate analysis of cotton content with the presence of a small amount of wavelength data.

Keywords: *Near-infrared; Cotton content; Normalization; Stepwise*

1. Introduction

In recent years, with the development of electronic technology and computer technology, near infrared spectroscopy technology has received increasing attention relying on its high efficiency and rapidness. By using near infrared spectroscopy technology, a non-destructive qualitative or quantitative analysis can be carried out for samples on the basis of existing calibration models [1]. For solid samples, the diffuse reflectance method can be used to measure near infrared spectra data. Near infrared spectroscopy analysis technology is becoming increasingly popular in quantitative analysis of textile fabrics [2].

Regarding the near infrared spectroscopy analysis of textile fabrics, different fiber components will result in different absorbance peaks of near infrared spectroscopy; besides, the size of absorbance peak will vary with the fiber content. Thus, during the quantitative analysis of fabrics through near infrared spectroscopy analysis technology, variables can be filtered and the calibration model can be built with a small amount of wavelength point data, hence to complete a rapid quantitative analysis of fabrics [3-4]. Depending on the characteristics of textile fabrics, diffuse reflection is often used for measurement during near infrared spectroscopy analysis. In the production of textiles, different processing and production methods will lead to different texture densities of fabrics^[5]. Such structural difference will influence the subsequent building of calibration models. Though traditional pre-processing of spectra data can eliminate the influence, it is no longer applicable in case of a small amount of wavelength data.

For cotton-PET blended fabrics of different structures, measure the near infrared spectra data within 1400nm-2000nm, adopt stepwise regression method, eliminate useless wavelength information with statistical methods and combine the fabrics of different structures through normalization based on the influence of different structures on spectra data to realize that not only the influence of fabric structure can be eliminated, but also the calibration model can be built for a quantitative analysis of cotton content in the cotton-PET blended fabric with the presence of a small amount of wavelength data.

2. Principle of Stepwise Regression

The stepwise regression method is derived from forward variable selection method and backward variable elimination method. The variables introduced through forward selection method will be always kept in the model, and the variables remaining in the model will become insignificant with the adding of other variables. The variables eliminated by backward elimination method will be always excluded, and the eliminated variables may become significant with the introduction of other variables. Thus a method of selection for variables introduced and eliminated based on the characteristics of both kinds of variables is needed which is called stepwise regression.

The stepwise regression method combines the characteristics of both forward variable selection method and backward variable elimination method [6].

The basic idea of stepwise regression is to introduce all independent variables x influencing dependent variables y into the partial regression equation one by one from small to large according to the degree of influence, and meanwhile check whether all x included in the regression equation then are significant or not; if not, eliminate current variables; when all x have significant influence on y in the regression equation, consider introducing new variables, and carry out the check according to the above method, until there is no significant variable which can be introduced in the equation and there is no insignificant variable required to be eliminated [7-8].

Assume the quantity of samples is n , the quantity of components is m , and the quantity of wavelength is p , then an absorbance matrix $X_{n \times p}$ and a component content matrix $Y_{n \times m}$ can be obtained. Additionally, specify the matrix of measurement error to be $E_{n \times m}$, then the following can be derived:

$$Y_{n \times m} = X_{n \times p} B_{p \times m} + E_{n \times m} \quad (1)$$

$$B_{p \times m} = (X_{p \times n}^T X_{n \times p})^{-1} X_{p \times n}^T Y_{n \times m} \quad (2)$$

When it is only aimed at cotton content which is equivalent to a single-component analysis, where $m=1$, the Formula (1) can be expressed as below:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_p \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_p \end{pmatrix} \quad (3)$$

At this point, put all spectra data obtained in the model of stepwise regression, and eliminate variables in the regression equation which have insignificant influence on cotton content constantly according to the judging criteria, until there is no variable that needs to be eliminated^[9,10]. Where the judging criteria are based on F inspection:

Assuming t independent variables are used in the regression equation, the equation can be expressed as below:

$$y = b_0 + \sum_{i=1}^t b_i x_i \quad (4)$$

The relation between the sum of squares of deviations and the sum of squares of regression variances for the calibration model is listed as below:

$$S_{all}(t) = S_{reg}(t) + S_{res}(t) \quad (5)$$

The definition of each item in Equation (5) is provided as below:

$$S_{all} = \sum_{i=1}^n (y_i - \bar{y})^2;$$

$$S_{reg} = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2;$$

$$S_{res} = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Where, \bar{y} is the mean of component contents to be detected in the calibration sample set; y_i is the true value of No. i sample (chemical solution method); \hat{y}_i is the calculated content of No. i calibration sample.

After the elimination of a variable x_i , the total variance of component content y to be detected is calculated as below:

$$\Delta_i(t) = S_{reg}(t) - S_{reg}(t-1) \quad (6)$$

In this case, the statistic F_i is:

$$F_i = \frac{\Delta_i(t)}{S_{reg}(t)/(n-t-1)} \quad (7)$$

F_i follow the distribution of $F(t, n-t-1)$. In this subject, the significance level 0.95 is used as the selection threshold. If $F_i < F_{\alpha=0.95}$, eliminate the variable from the regression equation.

3. Selection of Normalized Standard Wavelength

Before the selection of wavelength, observe the near infrared spectrum chart of pure cotton fabric and pure PET fabric first:

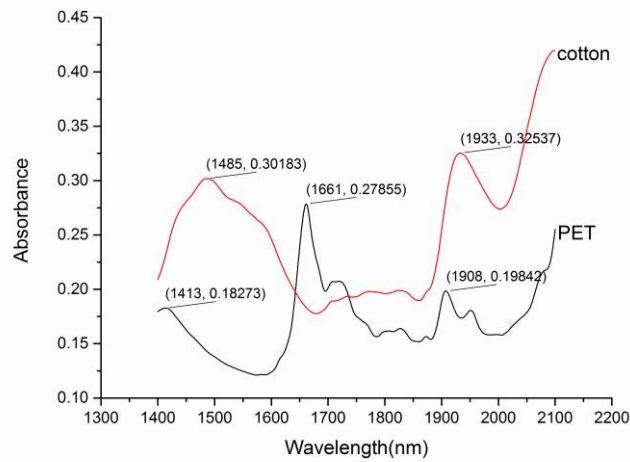


Figure 1. Absorbance Peaks of Pure Cotton and Pure PET

According to Figure 1., the variations of cotton and PET are almost the same after the wavelength exceeds 2000nm. So the wavelength within 1400nm-2000nm is selected. Within 1400-2000nm, the absorbance peaks of pure cotton are 1485nm and 1933nm, while the absorbance peaks of pure PET are 1413nm, 1661nm and 1908nm. The positions of wavelength where the absorbance peaks exist are obviously different for the two components. Also, it can be found that there is a significant linear variation near 1900nm in the spectrum chart. By reference to the literature of James Rodgers, it can also be found

that the absorbance at 1900nm presents a linear variation, too, depending on the structural difference caused by fabric production process and the gradient is basically the same.

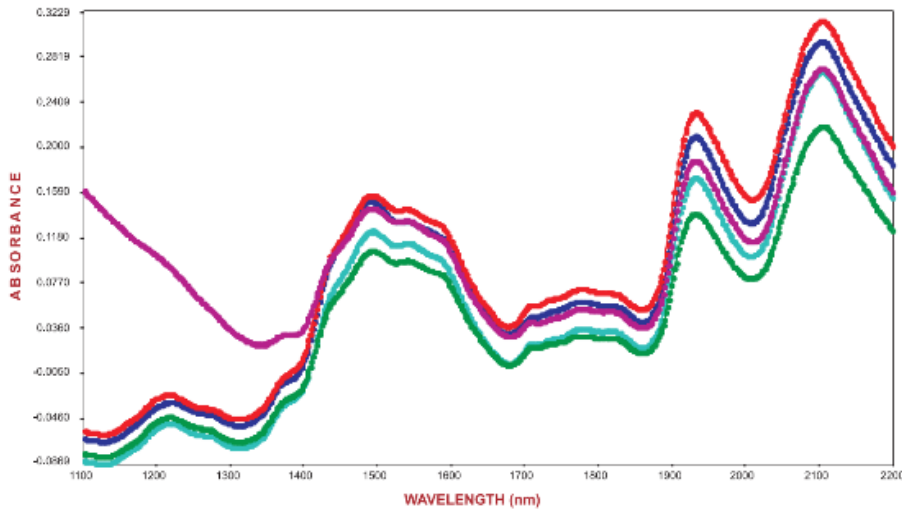


Figure 2. Spectrum Chart for Pure Cotton Fabrics with Different Structures in the Literature

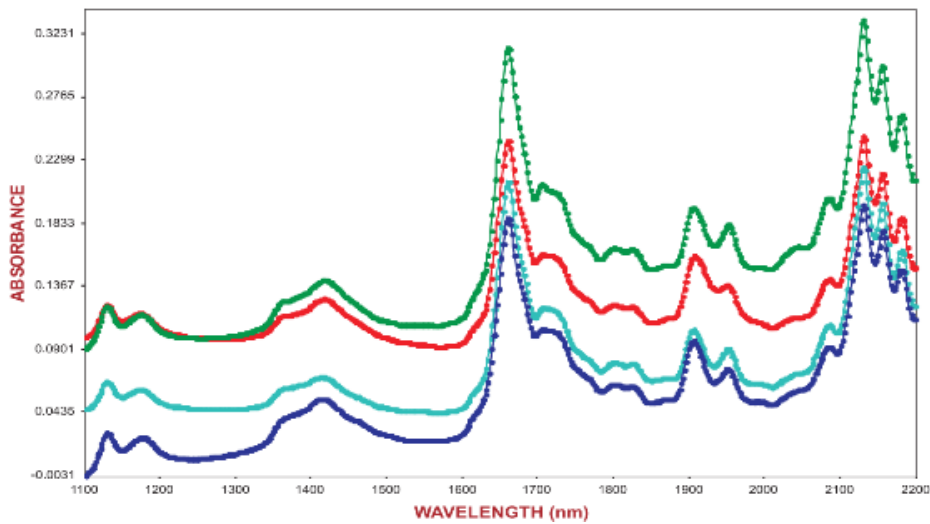


Figure 3. Spectrum Chart for Pure PET Fabrics with different Structures in the Literature

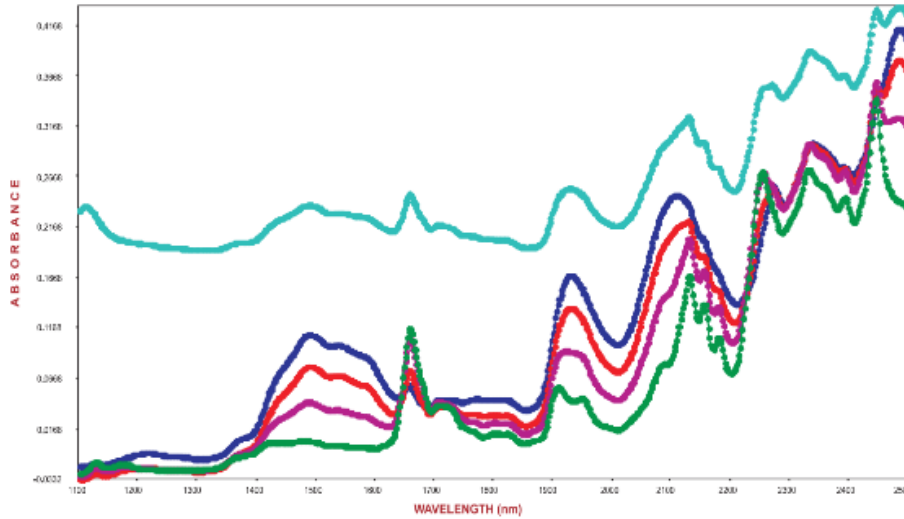


Figure 4. Spectrum Chart for Cotton-PET Blended Fabrics with Blending Ratios and Different Fabric Structures

Carry out the differential coefficient processing for the spectra data near 1900nm with first order derivatives, with the result shown as below:

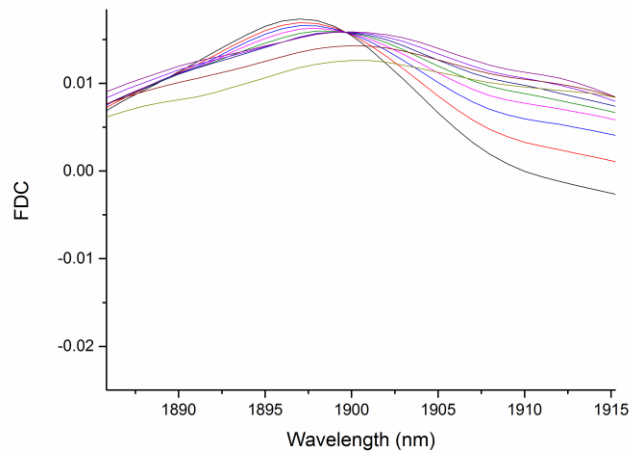


Figure 5. Spectrum Chart at 1900nm After First Differential Processing (Partial Enlarged View)

According to the chart, the result of original spectra data of samples after the first order differential coefficient indicates that the spectra data intersect in one point at about 1900nm. So, conduct a linear fitting for the spectra data within the range. It can be seen that all linear fitting coefficients are greater than 0.9, which means the data within the range not only have a certain linear variation, but also vary based on the same gradient. Hence, the data from the wavelength of 1900nm is used as the normalized standard. Through the normalization:

$$A_j = \frac{A_i}{A_{1900}} \quad (8)$$

Make the spectrum chart coincide at about 1900nm to reduce the influence of fabric structural difference.

4. Experiment Results and Analysis

Take 30 cotton-PET blended fabrics with different structures for the experiment, with 20 as the prediction set and 10 as the calibration set. All fabrics are not dyed. The complex at NIR2500 fiber optic near infrared spectroscopy experiments, spectral data acquisition way using diffuse reflectance detection, in order to ensure the measurement of diffuse reflection of incident light angle, at the same time with the specular reflection of light from the experiment using P-TIP, diffuse reflectance probe, using a series of NIR fiber FIB matching with the optical fiber. WS-1-S2 diffuse reflection standard board, HL2000 light source. Collect the data with the near infrared spectroscopy from Ocean Optics through diffuse reflection detection.



Figure 6. The Spectrometer used in the Study

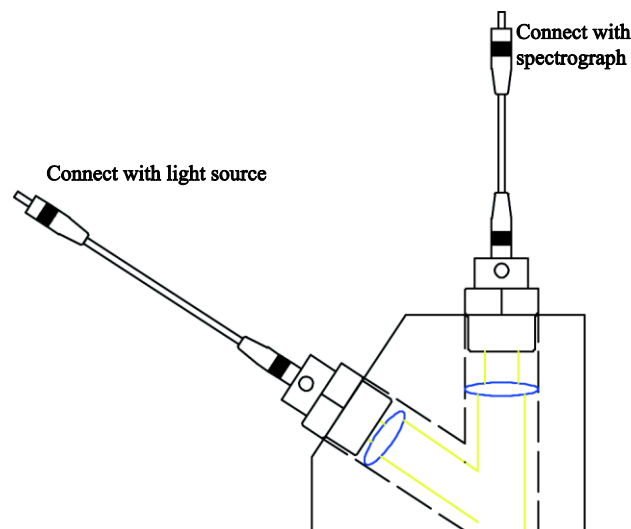


Figure 7. Connection of the Probe and the Spectrograph

The spectrum chart of collected spectra data after normalized is provided as below:

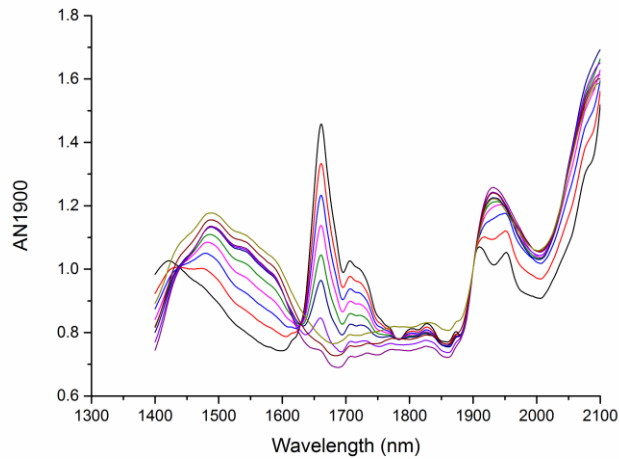


Figure 8. Spectrum Chart with 1900nm as the Standard After Normalized

First use all spectra data after normalized to carry out the stepwise regression modeling analysis, In order to evaluate model, typically use these indicators:RMSECV, RMSEP, R^2 .

(1). RMSEP (root mean standard error of prediction)

$$RMSEP = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_{ni} - \hat{y}_{ni})^2} \quad (9)$$

Where n is the number of samples in calibration set, the I samples of near infrared determination value, for the sample as involved in the analysis of the samples the true value.

(2). RMSECV (cross validation root mean square deviation)

$$RMSECV = \sqrt{\frac{1}{n} \sum_{i=1}^n PRESS} \quad (10)$$

The 2.32 type of PRESS to predict the residual sum of squares.

$$PRESS = E_1^2 + E_2^2 + \dots + E_n^2 \quad (11)$$

Cross validation refers to, each model calibration sample concentration extract samples were predicted, and then the rest of the calibration samples to establish the model, then build model, can be out of sample forecast to generate a predictive value by the model, the predicted value and the true value is extracted samples generally there is a difference, the difference between the two is the residual. Cross validation of the prediction calculation characteristics of each sample in the sample values are not included in the model, the results compared to other authentication method is more objective. Under normal circumstances the value of RMSECV is less than RMSEP, the value of RMSEV can objectively represents the deviation degree between forecasted and real value of the sample, the smaller the value the better. The stability of the model with the RMSEC value and RMSEP value as a reference, the value is to show the better stability

(3). R^2 (The coefficient of determination)

$$R^2 = \left(1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - y_m)^2} \right) \times 100\% \quad (12)$$

The coefficient of determination value closer to 1 indicates the better fitting effect equation.

The smaller RMSECV, RMSEP are the better model is. R^2 is closer to 1 means the model is better. The result is shown in Table 1:

Table 1. Calibration Model Built with Normalized Full Wavelength

Normalization	Evaluation parameters		
	RMSECV	RMSEP	R^2
Full wavelength	1.2234	1.3215	0.9618

Based on the result of the use of normalized full wavelength, along with the cotton content peaks, select normalized spectra data with the center wavelength of 1408nm, 1933nm, 1485nm and spectral width of ± 10 nm to re-conduct the stepwise regression analysis:

Table 2. Model Built with Filtered Data

Normalization	Evaluation parameters		
	RMSECV	RMSEP	R^2
After filtering	1.5814	1.6113	0.9571

At this point, the regression model built is:

$$y = 6.0524x_1 - 4.1380x_2 - 1.0795 \quad (13)$$

Where, x_1 is the normalized data at 1495nm; x_2 is the normalized data at 1943nm. Compare the model corresponding to Formula (14) with the model built with spectra data without normalization. According to the model in Formula (14), use the near infrared light with the central wavelength of 1490nm, 1900nm and 1940nm, get the monochromatic light required for near infrared analysis with an optical filter, and detect the cotton content with a spectrometer. The detection result is provided as below:

Table 4. Prediction Result

Predicted value	Reference value	Difference
1.8%	0%	1.8%
7.0%	10%	-3.0%
22.7%	20%	2.7%
32.1%	30%	1.2%
39.0%	40%	-1.0%
46.3%	50%	-3.7%
63.4%	60%	3.4%
71.2%	70%	1.2%
82.7%	80%	2.7%
92.3%	90%	2.3%
103.6%	100%	3.6%

The cotton content can be detected accurately with the monochromatic light data through the model corresponding to Formula (14), which means the normalization with the wavelength data at 1900nm as the standard can still effectively reduce the influence of fabric structural difference when there is only a small amount of wavelength data.

5. Compared with the Traditional Pretreatment Methods

Partial least squares method in previous studies has proved to be a very effective modeling method [11-14]. Partial least squares method provides a linear multivariate regression method of multiple variables. This method is especially suitable for the two group, there are a lot of variables, and the existence of multiple correlation, at the same time for the small number of calibration samples, the partial least square method with other algorithms do not have the advantage. Traditional pretreatment methods mainly include: second differential, SNV, MSC[15].

The first use of the whole range of the spectrum data for regression analysis. In the establishment of the regression model, the data pretreatment, pretreatment effect observation and analysis of the way of regression. The experiment collected 701 wavelength data, for all 701 wavelengths data using two differential treatment, standard normal transformation, multiple scattering correction, and then use the PLS method (model, most of the data using the method of comparison to facilitate) establish a regression model, using PLS for data processing build the calibration model, need to choose the factor number according to the correction of the model, the general is the factor number, RMSEP and RMSEV value is smaller, the number of the number of not less than the variable factor. In order to facilitate comparison, unified use of 5 factors, the calibration models were established, and the effect of detection model, the experimental results shown in table 4:

Table 4. The Evaluation Parameters of Pretreatment Methods and the Corresponding Model

Pre-processing method	Evaluation parameters		
	RMSECV	RMSEP	R2
Non-pretreatment	3.4724	3.7351	0.9221
second differential	3.1137	3.4252	0.9489
SNV	2.5013	2.6814	0.9516
MSC	2.3216	2.6211	0.9565

As can be seen from the table, is the best regression model with multiplicative scatter correction, followed by the standard normal transformation, two differential. No regression model and compared with pretreatment, and the processed data can improve the accuracy of the model in a certain extent.

Table 5. Result of Comparison with Models Built Through other Methods

Modeling method	Evaluation parameters		
	RMSECV	RMSEP	R ²
After filtering	1.5814	1.6113	0.9571
Stepwise regression	1.9247	1.9469	0.9502
Without pre-processing	3.4724	3.7351	0.9421
Second differential	3.1137	3.4252	0.9489
SNV	2.5013	2.6814	0.9516
MSC	2.3216	2.6211	0.9565

It can be found that though the prediction accuracy with models built by using filtered normalized data is lower than the prediction accuracy with models built by using spectra data which are all normalized, the quantity of variables used becomes less; meanwhile, compared to the result of using stepwise regression analysis alone, the fitting effect of the

model is slightly improved, but the RMSECV and RMSEP values decrease significantly.

Come along with multiplicative scatter correction of standard normal transformation, these two kinds of processing in the previous studies have shown that the effective processing method of diffuse reflection detection method. But despite the correction model is established using the MSC and SNV of the two methods are shown to not use preprocessing and model using two differential pretreatment of the fitting effect on the increase, but the RMSECV and RMSEP value was not significantly reduced.

6. Conclusions

Carrying out quantitative analysis for cotton-PET blended fabrics through the method combining stepwise regression and normalization can effectively reduce the influence of fabric structural difference on spectra data. Meanwhile, since the standard wavelength used in normalization is the data of a single wavelength, the influence of fabric structural difference can be reduced effectively when a few wavelength points are used for data collection. It also indicates that using near infrared spectroscopy analysis technology can detect the cotton content in the fabric rapidly.

References

- [1] V. Bellon, J. L. Vigneau and F. Svila, "Food Control", vol. 5, no. 94, (1994), pp. 21
- [2] S. K. Dey, "Textile Research Journal", vol. 73, no. 12, (2003), pp.1062.
- [3] J. Rodgers and K. Beck, "Textile Research Journal", vol. 79, no. 8, (2009), pp. 675.
- [4] J. E. Rodgers, "AATCC Review", vol. 2, no. 6, (2002), pp. 27.
- [5] M. Blanco, J. Coello and H. Iturriaga, "Analyst", vol. 8, no. 8, (1994), pp. 1779.
- [6] S. Bruun, B. Stenberg and T. A. Breland, "Soil Biology and Biochemistry", vol. 37, no. 12, (2005), pp. 2283.
- [7] P. Celio, "Journal of the Brazilian Chemical Society", vol. 14, no. 2, (2003), pp. 198.
- [8] W. C. Tincher and A. Luk, "Textile Chemist&Colorist", (1985).
- [9] X. Qian, "Information and Control", vol. 5, no. 63, (1986).
- [10] X. Qi and L. Zhang, Journal of Beijing Agricultural College, 3, 49 (1999)
- [11] J. Huang, H. Yuan, C.Song and X. Li, "Spectroscopy and Spectral Analysis", vol. 1, no. 60, (2013).
- [12] L. Han, Xingbing Li and Y. Tan, "Computer Technology and Development", vol. 5, no. 87, (2006).
- [13] N. Jia, L. Liu, S. Gao and Y. Wu, Academic Journal of Second Military Medical University, vol. 10, no. 967, (2000).
- [14] T. Sun, X. Geng and M. Liu, Spectroscopy and Spectral Analysis, 12, 3257 (2014)
- [15] R. Gao, S. Fan, Y. Yan and L. Zhao, "Spectroscopy and Spectral Analysis", vol. 12, (2004), pp.1563.

Authors



Li Dejun, Director of academic affairs of Wuhan Textile University, Professor. The main research directions for solar power, signal processing, image processing.



Xing Chigang, Graduate Student of Wuhan Textile University. The major is Electronic circuit and system.