doi:10.3788/gzxb20164505.0530001

基于太赫兹时域光谱技术与PCA-BPN网络的转基因大豆鉴别

聂君扬[1,3],张文涛[1,3],熊显名[1,3],陈涛[1],占平平[1,3],涂闪[2,3]

(1 桂林电子科技大学 电子工程与自动化学院,广西 桂林 541004)

(2 广西师范大学 理学院,广西 桂林 541004)

(3 广西高校光电信息处理重点实验室,广西 桂林 541004)

摘　要:基于太赫兹波段内的光谱分析技术以及主成分特性分析与反向前馈神经网络建模,提出了一种转基因大豆鉴别方法.从光谱数据中提取累计方差贡献率达到97.582%的前8种主成分因子,并将其作为输入源导入神经网络模型,通过剔除冗余数据、降低数据维数,所建立的神经网络模型能准确识别校验集.该方法可以实现转基因大豆的快速、无损检测,在农业安全领域有广泛的应用前景.

关键词:转基因大豆;太赫兹;主成分分析;反向前馈神经网络;无损检测

中图分类号:O433.4　　**文献标识码**:A　　**文章编号**:1004-4213(2016)05-0530001-7

Recognition of Transgenic Soybeans Based on Terahertz Spectroscopy and PCA-BPN Network

NIE Jun-yang[1,3], ZHANG Wen-tao[1,3], XIONG Xian-ming[1,3], CHEN Tao[1], ZHAN Ping-ping[1,3], TU Shan[2,3]

(1 *Institute of Electrical Engineering and Automation, Guilin University of Electronic Technology, Guilin, Guangxi* 541004, *China*)

(2 *College of Physical Science and Technology, Guangxi Normal University, Guilin, Guangxi* 541004, *China*)

(3 *Guangxi Colleges and Universities Key Laboratory of Optoelectronic Information Processing, Guilin, Guangxi* 541004, *China*)

Abstract: An approach for recognition of transgenic soybeans was proposed based on spectral analysis in the terahertz (THz) range combing with Principle Component Analysis (PCA) and Back Propagation Neural (BPN) network. Eight principal component factors, whose accumulated variance reached 97.582%, were extracted from the original spectra data and then fed as inputs into the BPN network model. The utilization of the dimension-reduced data in training the network model can recognize the validation set accurately. The nondestructive testing of transgenic soybeans could be achieved by using THz spectroscopy, which could be widely applied in agricultural security areas.

Key words: Transgenic soya beans; Terahertz; Principle component analysis; Back propagation neural

Foundation item: The National Natural Science Foundation of China (No. 61565004), the Natural Science Foundation of Guangxi (Nos. 2015GXNSFBA139252, 2014GXNSFGA118003, 2013GXNSFDA019002), Guangxi Education Scientific Research Program (Nos. ZD2014057, KY2015B103), Guilin Scientific Research and Technology Development Program (Nos. 20140127-1, 20150133-3), Guangxi Scientific Research and Technology Development Program (No. 1598017-1), Guangxi Key Laboratory of Automatic Detecting Technology and Instruments (No. YQ15104).

First author: NIE Jun-yang (1990—), male, M.S. degree candidate, mainly focuses on photoelectric detection technology. Email: njyang@live.com

Responsible author: ZHANG Wen-tao (1976—), male, professor, Ph.D. degree, mainly focuses on the interactions between lasers and materials. Email: uestczrk@126.com

Corresponding author: CHEN Tao (1984—), male, Lecturer, Ph.D. degree, mainly focuses on terahertz science and technology and its applications. Email: ct63307@163.com

Received: Dec. 21, 2015; **Accepted**: Apr. 11, 2016

network; Nondestructive testing

OCIS Codes: 300.6495; 040.2235; 040.2235; 120.4290; 120.4290

0 Introduction

Terahertz radiation refers to electromagnetic waves at frequencies from 1.0 to 10 THz (or wavelength 30mm to 3μm). Located in the special transition region between electronics and photonics, THz waves occupy a very important position in the electromagnetic spectrum as well as possessing a great value in both theory and practice[1-2]. Recent theoretical researches show that the vibration and rotational energy levels of most biomolecules are in the THz range. And in consideration of the low energy of the THz radiation (about 1～10 meV), which is harmless to biomolecules, the molecular response of biological material to the THz radiation is of quite particular interest in numerous fields[3-5].

Transgenic crops, which have been firstly commercialized in 1996, have undergone a huge development and evolution in the last decade and its cultivated area has been continuously increased. By the year 2014, the cultivated area of transgenic crops reached 181 million hectares, which exceeded 10% of the world′s total cultivated area. Among all kinds of transgenic crops, transgenic soybeans own the biggest cultivated area, which accounts about half of the total cultivated area of the transgenic crops. America, Brazil and Argentina are the world′s main exporters of soybeans, and the cultivated scale of transgenic soybeans in these three countries is over 90%[6-7]. Even despite being directly used as food, through being the raw materials for oil manufacture and being the fodders for feeding animals, transgenic soybeans had already entered the daily life of most people in the global world indirectly[8].

Although the economic benefit and the ecological advantage of the transgenic soybeans are quite significant, no research institution can deny the possibility of the potential problems yet[9]. Out of the responsibility for the public consumers and in consideration of the agricultural security, nondestructive testing techniques for transgenic soybeans are of great significance. At present, most of the transgenic detection methods are based on protein detection and nucleic acid detection[10]. Among them, PCR[11] and ELISA[12] are the two most widely used methods. However, even including the latest advances of gene chip technology[13], such existing technologies still cannot meet the requirements of speed, sensitivity, high throughput and quantitative accuracy. It is important and urgent to establish an efficient and rapid transgenic detection method, which is also to be safe, convenient and cheap.

Since Markelz et al.[14] firstly used Terahertz Time-Domain Spectroscopy (THz-TDS) to study examine low-frequency collective vibrational modes of calf thymus DNA, bovine serum albumin and type I collagen in 2000, many researches had been reported. Gen et al.[15] proposed an Attenuated Total Reflection (ATR) method to extract drug′s THz spectrum information. Tu et al.[16] proposed a Principal Component Analysis (PCA) method to study the THz spectrum property of transgenic cotton. Yang et al.[17] studied optical property and spectroscopy of edible flavor based on THz-TDS. Chen Tao et al.[18] used fuzzy pattern recognition in the research of terahertz absorption spectra of different kinds of biomolecules.

In this article, we focus on detection of transgenic soybeans by combing with THz spectroscopy. The spectra of 3 kinds of different transgenic soya bean powders have been investigated both in time and frequency domain, from where few obvious characteristic absorption peaks was able to be observed. Through smoothing the original spectra, preliminary recognition is achieved. Principle component analysis was thus adopted to reduce the redundancy of the spectra. 8 principle component features were extracted and then fed as inputs into a BPN network model. After network training, the previous selected spectra data were then used as the validation set to verify the accuracy of the model. The experiment result is satisfactory. This method is cheap and fast comparing with many traditional methods, which can be widely used in many agriculture-related areas and provide reference for further study.

1 Experimental

1.1 Experimental apparatus

A Zomega Z-3 THz-TDS system was adopted in this experiment, with a maximum-delay time more than 1 minute, a peak of dynamic range above 70 dB, a time of data acquisition approximately equal to the 1 minute and two patterns including transmission mode and reflection mode. A Toptica Photonics AG ultrafast femtosecond laser, which generates laser pulses of 100 fs duration at a central wavelength of around 800 nm with a repetition frequency of 80 MHz, was served for the THz pump pulse and probe light. The schematic diagram of the experimental apparatus is shown in Fig. 1.

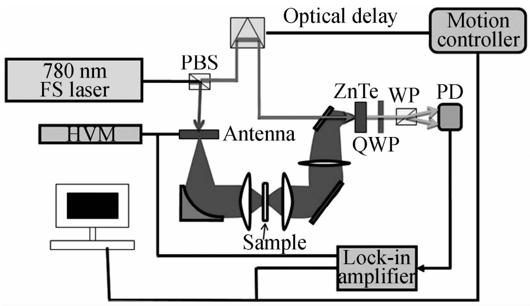


Fig. 1 Schematic diagram of the THz-TDS system

Considering the strong absorption of water to the THz wave, the experiment was processed in a particular environment, where Nitrogen would be injected until the internal relative humidity drops to 2% below. And the experiment was conducted at room temperature.

1.2 Sample preparation

Three kinds of different soya bean powder Certified Reference Materials (CRMs) containing different mass fractions of genetically modified 305423 soya are supplied by Institute for Reference Materials and Measurements of the European Commission's Joint Research Centre (EC-JRC-IRMM). According to the certificate, all the powders have been produced from whole beans of the transgenic soya 305423 Soya and non-modified soya, both supplied by Pioneer Hi-Bred International, Inc. (Johnston, IA, USA). The presentations of the samples are shown in Table 1.

Table 1 Samples of soya bean powders

Soya bean powder CRMs	Mass fraction of 305423 Soya		Samples
	Certified value	Uncertainty	
BF426a	0.0%	0.0%	44
BF426c	1.0%	0.1%	44
BF426d	10.0%	0.7%	44

To recognize the three different types of transgenic soya bean powders, 44 samples of each type were prepared and detected by THz-TDS in the frequency range of 0.3~1.6 THz. Both front and back of each sample were scanned 3 times, the average of every above 6 times scan spectral data of each sample is the corresponding THz spectrum. All samples were pressed into tablets with a diameter of 13.000mm and a thickness of 1.000 mm.

1.3 Mathematics manipulation

Complex refractive index $\tilde{n}(\omega)$ is commonly used to describe the macroscopic optical properties of materials, and is defined as follows

$$\tilde{n}(\omega)=n(\omega)-\mathrm{j}\kappa(\omega) \tag{1}$$

where $n(\omega)$ is the real refractive index describing materials' dispersion characteristics, $\kappa(\omega)$ is the extinction coefficient, and further more $\alpha(\omega)=2\omega\kappa(\omega)/c$ is the absorption coefficient referring to the absorption characteristics.

In this experiment, we obtained both the amplitude and phase of the reference signal $E_{\mathrm{ref}}(t)$ and the sample signal $E_{\mathrm{sam}}(t)$ by using THz-TDS. Thus the corresponding frequency spectrum $E_{\mathrm{ref}}(t)$ and $E_{\mathrm{sam}}(t)$ were acquired by Fourier transform. Without the consideration of the Fabry-Perot effect, the transfer function of the detecting THz pulse can be calculated as follows

$$H_\omega=\frac{E_{\mathrm{sam}}(\omega)}{E_{\mathrm{ref}}(\omega)}=A\exp(-\mathrm{i}\varphi(\omega))\approx \frac{4\tilde{n}}{(\tilde{n}+1)^2}\exp\left[\frac{\mathrm{i}\omega(\tilde{n}-1)\mathrm{d}}{c}\right] \tag{2}$$

where A is the amplitude ratio, $\varphi(\omega)$ is the phase difference between the sample signal and the reference signal, c is light speed in a vacuum, and d is the thickness of the sample. Consequently, the real refractive index $n(\omega)$ and absorption coefficient $\alpha(\omega)$ can be calculated as follows [19-21]

$$n(\omega)=\frac{c\varphi(\omega)}{\omega d}+1 \tag{3}$$

$$\alpha(\omega)=\frac{2}{d}\ln\left\{\frac{4n}{A[n(\omega)+1]^2}\right\} \tag{4}$$

The thickness relative error of the sample tablets is less than 5%, but nonetheless its affection to the real refractive index and the absorption coefficient cannot be ignored. To avoid the artificial error caused by thickness measuring and the uncertainty of powder particles' refractive index, we adopt Absorbance, a dimensionless relative amount that indicates material's light absorption levels, as the reference variable of the spectra. The Absorbance formula is defined as follows

$$\mathrm{Absorbance}=-\lg\left[\frac{E_{\mathrm{sam}}(\omega)^2}{E_{\mathrm{ref}}(\omega)^2}\right] \tag{5}$$

2 Theory and modeling

2.1 Principal component analysis

As a typical kind of data processing method, principal component analysis, a linear transform based on original features, is used to reduce the dimensionality of a data set consisting of a large number of interrelated variables, while retaining as much as possible of the variation present in the data set [22]. Generally, when the accumulated variance of the new variables is larger than 85.000%, these new variables can be approximated as the principal components of the original data set.

2.2 Back propagation neural network algorithm

BPN network, developed by Rumelhart et al. [23], is a way of simulating human thinking and is the most representative learning model for the ANN. Based on a back propagation algorithm of a three-layer network structure (input layer, hidden layer and output layer), BPN, a nonlinear multilayer feedforward network system formed by a large number of individual neuron in a layer mode, can realize the nonlinear mapping between input data and output data, while featuring fast learning, high accuracy and strong fault-tolerance. Learning in BPN is an optimization process of an error correction learning algorithm which comes in two parts: positive signal propagation and error back-propagation. The procedure of BPN network repeatedly adjusts the weights of the connections in the network so as to minimize the measure of the difference between the actual output vector of the net and the desired output vector. The topology of a typical BPN is shown in Fig. 2.

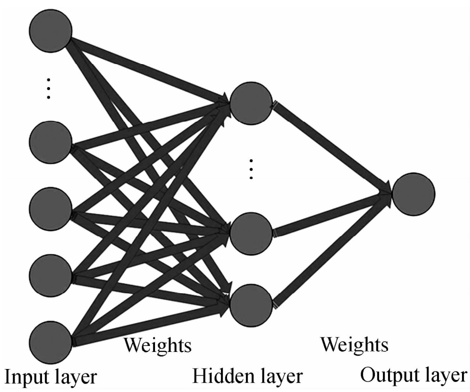


Fig. 2 An example of BPN network topology

Consider a three-layer neural network $f(\boldsymbol{x},\boldsymbol{V},\boldsymbol{W})$ with $\boldsymbol{x}$ as a vector representing inputs, $\boldsymbol{V}$ as a vector representing variable interconnection weights between the input layer and the hidden layer and $\boldsymbol{W}$ as a vector representing variable interconnection weights between the hidden layer and the output layer. In addition, $f(\boldsymbol{x})$ is the transfer function, a unipolar sigmoid function described as $f(\boldsymbol{x})=(1+\mathrm{e}^{-x})^{-1}$, which is continuous, derivative and is closer to the output forms of the biological neurons. For the hidden layer and the output layer, we can get the following equations

$$h_i = f(\sum_{i=0}^{n} v_{ij}x_i) \quad j=1,2,\cdots,m \tag{6}$$

$$y_k = f(\sum_{j=0}^{m} w_{jk}h_j) \quad k=1,2,\cdots,l \tag{7}$$

$$E=\frac{1}{2}\sum_k (z_k - y_k)^2 \tag{8}$$

which is a function of the interconnection weights w_{jk} and v_{ij}. Hence, through adjusting all the interconnection weights, we can achieve the aim of reducing the error rate E so as to get a better learning effectiveness. According to the gradient descent algorithm, the following weight adjustment formulas of BP learning algorithm are obtained[23]

$$\begin{cases} \Delta w_{jk} = -\eta\dfrac{\partial E}{\partial w_{jk}} = \eta\delta_k y_j = \eta(z_k - y_k)y_k(1-y_k)h_j \\ \Delta v_{ij} = -\eta\dfrac{\partial E}{\partial v_{ij}} = \eta\delta_j x_i = \eta(\sum_{k=1}^{l}\delta_k w_{jk})h_j(1-h_j)x_i \end{cases} \tag{10}$$

where Δv_{ij} and Δw_{jk} are the adjustment of the corresponding interconnection weights, δ_k and δ_j are the error signal of the output layer and the hidden layer respectively, and η is the learning rate.

It is easy to see from the above formulas that the formula form of each layer′s adjustment in the BPN network is the same. In addition, the error signal of the output layer is related to the difference between the desired output and the actual output of the network, which directly reflecting the output error. Error signal δ_k then transmits back to the hidden layer, and error signal δ_j is obtained, therefore we can calculate all the adjustments of the interconnection weights in the network.

3 Results and discussion

3.1 Spectrum analysis and preprocessing

We first used the THz-TDS system to collect the waveform transmitted through the sample tablets. Fig. 3 shows the THz waveforms of the 3 transgenic soya bean powders in the time domain, which is composed of the average of the spectrum data of the training set. Contrast with the reference signal, there are some amplitude attenuations and redshifts of all the three spectrum lines, both of which are quite varied from each other. Given the shortage of the raw data points (640 points only), it really requires zero-padding (to reach 32768 points) before performing a FFT so as to get smoother spectrum curves in the frequency domain.

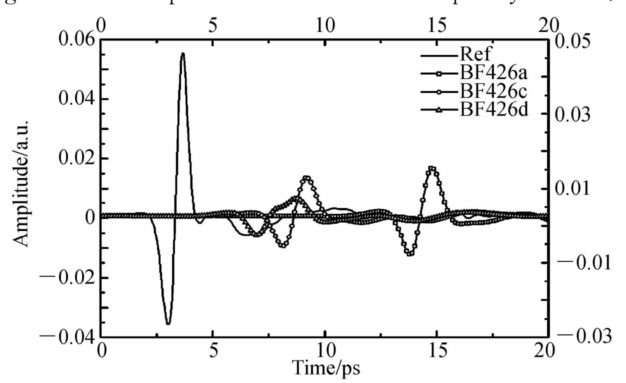


Fig. 3 Time-domain THz pulses of the reference and the samples

As Fig. 4 illustrates, the absorption cut-off frequency of the samples is different, in which BF426d′s is higher than 1.5 THz while the other two′s are about 1.0 THz, and there are no obvious characteristic absorption peaks in the THz range examined except one obvious peak at around 0.2 THz both in BF426a and BF426c. In addition we can find that the curve of the BF426a and BF426c are of a similar pattern while BF426d′s curve is absolutely different. Furthermore, the spectral curve of BF426d is more close to the curve of reference signal. As seen in Table 1, BF426a is fully produced from conventional, non-modified soya seed and BF426c and BF425d contains 1% and 10% mass fraction of genetically modified 305423 soya, respectively. Thus we can conclude that is the similar composition of BF426a and BF426c generates the similar absorption peaks at around 0.2 THz. And compared with the non-modified soya seed powders, genetically modified 305423 soya seed powders may enhance the transmittance of THz wave in the range of 0.0～1.5 THz.

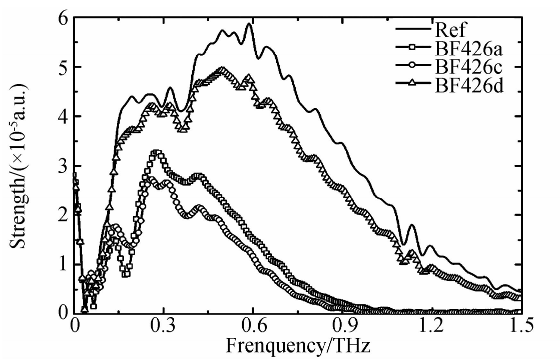


Fig. 4 Average frequency spectra of the reference and the samples

Then we employed a preprocessing method to deal with the spectra, and Fig. 5 is the absorbance spectra of the samples obtained by utilizing smoothing based on a multiplicative scatter correction and by calculating with Eq. (5). The spectral curves have been upwards shifted for ease of contrastive analysis. As shown in Fig. 5, the absorbance spectra are much smoother than the original spectra, and the peak positions as well as the peak values are quite obvious. Moreover, the oscillations caused by the differences of the particle size, the changes of the optical path and scatter are basically eliminated. To get a more accurate identification effect, pattern recognition methods are to be employed in the following part.

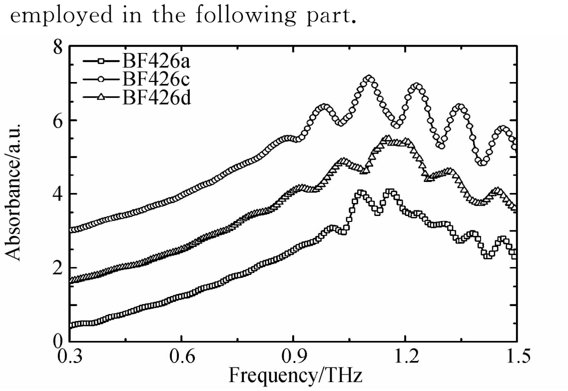


Fig. 5 Smoothing absorbance spectra of the samples

3.2 PCA analysis

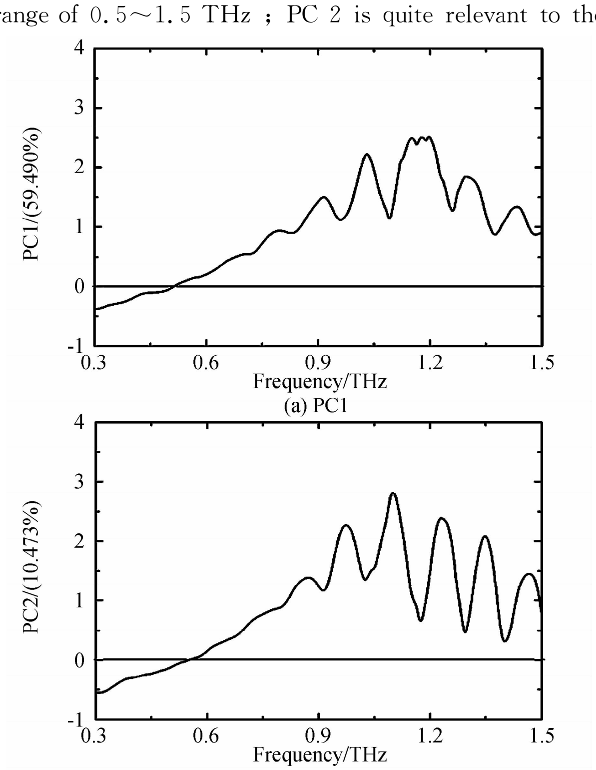
To reduce the redundancy of the spectra and promote the learning efficiency of the model, we used a PCA method to extract features of the raw spectra [24]. As seen in Table 2, the accumulated variance of the top 3 PCs reaches 90.252%, which means that the top 3 PCs can approximately substitute the spectral information. The scattered scores plot shows that, THz waves in the frequency range of 0.3～1.5THz have a certain clustering effect on different kinds of transgenic soybeans, which yet is still a long way from accurate identification. By virtue of the strong nonlinear mapping ability, generalization ability and fault tolerance, BPN has been widely applied in many biologically relevant estimation and inversion research area. Therefore, we proposed a BPN model and utilized PCA as a parameter extraction method for the inputs of the model.

Table 2 Total variance of the spectra

Component	Initial eigenvalues Total	Variance/%	Cumulative/%
PC1	1166.781	72.969	72.969
PC2	163.955	10.254	83.223
PC3	112.395	7.029	90.252
PC4	52.308	3.271	93.523
PC5	28.120	1.759	95.282
PC6	15.543	0.972	96.254
PC7	12.312	0.770	97.024
PC8	8.920	0.558	97.582
PC9	6.913	0.432	98.014
PC10	4.591	0.287	98.301

3.3 BPN network modeling

Fig. 6 illustrates the load intensity diagrams of the top 3 PCs, through where the correlation between the principle components and the frequency variables in the frequency range of 0.3～1.5 THz can be observed. As the figure shows, PC1 is quite relevant to the frequency range of 0.5～1.5 THz ; PC 2 is quite relevant to the



(a) PC1

(b) PC2

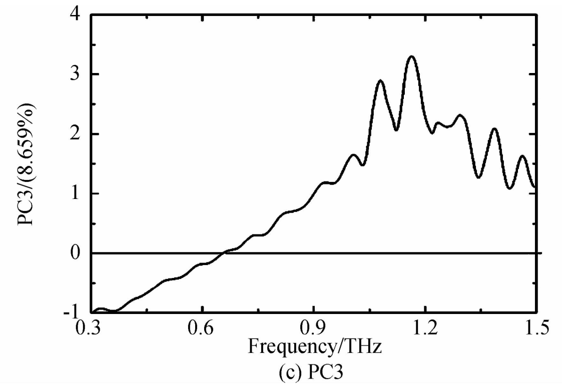


Fig. 6 Loading plot

frequency range of 0.6～1.5 THz; PC3 is quite relevant to the frequency range of 0.7～1.5 THz. As section 4.2 mentioned, the 3 PCs can approximately substitute the primal variables. Thus, we can conclude that the frequency range of 0.5～1.5 THz is the sensitive characteristic frequency band to the transgenic soybeans.

As shown in Table 2, the accumulated variance of the top 8 PCs reaches 97.582%, which are also located in the frequency range of 0.6～1.5 THz. Therefore, we take the 8 PCs as the input of the BPN network model. Accordingly, the dimensionality of the spectra dropped from 1599 to 8 without losing much original information.

To establish the BPN network model, 30 samples of each type of the transgenic soya bean powders were randomly selected to build a training set, and the rest of the 14 samples of each type were then used to build a validation set. After extensive practice, the optimum structure of the BPN model is obtained, where 8 nodes at the input layer, 6 nodes at the hidden layer and 3 nodes at the output layer. The relative deviation of the training set is 8.735×10^{-5}.

Table 3 BPN results of the validation set

BF426a/(Truth: 0)		BF426c/(Truth: 1)		BF426d/(Truth: 2)	
	Prediction	#	Prediction	#	Prediction
1	0.009 36	15	1.009 62	29	1.988 56
2	0.009 56	16	1.008 56	30	1.988 77
3	0.010 62	17	1.009 63	31	1.988 25
4	0.008 90	18	1.009 55	32	1.989 02
5	0.009 23	19	1.009 87	33	1.988 65
6	0.009 54	20	1.006 32	34	1.988 66
7	0.009 65	21	1.007 52	35	2.009 53
8	0.009 62	22	0.995 86	36	1.987 65
9	0.009 06	23	1.009 90	37	1.988 39
10	0.010 75	24	1.008 03	38	1.988 54
11	0.009 25	25	0.996 57	39	1.988 73
12	0.009 02	26	1.009 36	40	2.009 87
13	0.009 33	27	1.008 69	41	1.988 25
14	0.009 41	28	1.009 07	42	1.988 30
Accuracy	100%	Accuracy	100%	Accuracy	100%
Total Accuracy			100%		

Table 3 shows the BPN results of the validation set. The recognition accuracy of the 42 unknown transgenic soya bean powders reaches 100%. It can be concluded that combining THz-TDS with BPN network can achieve reliable recognition effect on different kinds of transgenic soybeans.

3.4 Discussion

The major ingredients of soybeans are protein, oligosaccharide, soybean oil and nucleic acid, which all are kinds of biomacromolecules. Previous studies found that proteins, carbohydrates and nucleic acids showed particular spectra characteristics in the THz range[5, 18]. Although it is considered that transgenic soybeans are substantially equivalent to the non-modified soybeans in terms of nutritional components, the biomolecules within the soybean cells are actually not the same owing to the unique foreign gene fragments, which provides the theoretical foundations for realizing the recognition of different transgenic soybeans using THz-TDS.

For the limited resolution and bandwidth of the apparatus we employed in this research, it is hard to recognize the certain kind of the sample just judging from the characteristic absorption peaks in the original spectra. And the limitation of the resolution and bandwidth of THz-TDS system is also the focus of current researches[1, 4]. Nonetheless, the spectra still reflect the information on physics and chemistry of the biomolecule structure. Based on it, we adopted pattern recognition methods and successfully identified the samples. Although the recognition accuracy of the BPN model reached 100%, the variety of the samples was sort of limited. More different kinds of transgenic soybeans will be added in the future study.

4 Conclusion

In this paper, a recognition model is established for the transgenic soybeans by combining THz spectroscopy with pattern recognition methods. Firstly, PCA is utilized to reduce the redundancy of the original spectra data obtained by the THz-TDS in the frequency range of 0.3～1.5THz and to extract the features of the data to fed as inputs into the model so as to promote the learning efficiency. Then a three-layer BPN network model is built based on the training set form by the selected principle component features. The recognition model can recognize the validation set accurately, which demonstrates the feasibility of nondestructive testing of transgenic crops based on THz spectroscopy.

Reference

[1] AMENABAR I, LOPEZ F, MENDIKUTE A. In introductory review to thz non-destructive testing of composite mater[J].

Journal of Infrared, Millimeter, and Terahertz Waves, 2013, **34**(2): 152-169.

[2] WANG W B, GUO B S. Control of the propagation distance of surface plasmons wave in terahertz spectrum [J]. *Acta Pohtonica Sinica*, 2016, **45**(2): 0224001.

[3] SIEGEL P H. Terahertz technology in biology and medicine [J]. *IEEE Transactions on Microwave Theory and Techniques*, 2004, **52**(10): 2438-2447.

[4] FISCHER B M, HELM H, JEPSEN P U. Chemical recognition with broadband thz spectroscopy[J]. *Proceedings of the IEEE*, 2007, **95**(8): 1592-1604.

[5] PICKWELL E, COLE B E, FITZGERALD A J, *et al*. Simulation of terahertz pulse propagation in biological systems [J]. *Applied Physics Letters*, 2004, **84**(12): 2190.

[6] JAMES C. 2014 ISAAA report on global status of biotech (GM) Crops[R]. ISAAA, 2015.

[7] GOODMAN R E. Biosafety: evaluation and regulation of genetically modified (GM) crops in the United States[J]. *Journal of Huazhong Agricultural University*, 2014.

[8] CHRISPEELS M J. Yes indeed, most Americans do eat GMOs every day! [J]. *Journal of Integrative Plant Biology*, 2014, **56**(1): 4-6.

[9] DEFRANCESCO L. How safe does transgenic food need to be? [J]. *Nature Biotechnology*. 2013, **31**(9): 794-802.

[10] AHMED F E. Detection of genetically modified organisms in foods[J]. *Trends in Biotechnology*, 2002, **20**(5): 215-23.

[11] JAMES D, SCHMIDT A, WALL E, *et al*. Reliable detection and identification of genetically modified maize, soybean, and canola by multiplex pcr analysis[J]. *Journal of Agricultural and Food Chemistry*, 2003, **51**(20): 5829-5834.

[12] MARGARIT E, REGGIARDO M I, VALLEJOS R H, *et al*. Detection of BT transgenic maize in foodstuffs[J]. *Food Research International*, 2006, **39**(2): 250-255.

[13] CUZIN M. DNA chips: a new tool for genetic analysis and diagnostics[J]. *Transfusion Clinique et Biologique*, 2001, **8**(3): 291-296.

[14] MARKELZ A G, ROITBERG A, HEILWEIL E J. Pulsed terahertz spectroscopy of DNA, bovine serum albumin and collagen between 0.1 and 2.0 THz[J]. *Chemical Physics Letters*, 2000, **320**(1-2): 42-48.

[15] TAKEBE G, KAWADA Y, AKIYAMA K, *et al*. Evaluation of drug crystallinity in aqueous suspension using terahertz time-domain attenuated total reflection spectroscopy [J]. *Journal of Pharmaceutical Sciences*, 2013, **102**(11): 4065-4071.

[16] TU S, ZHANG W T. XIONG X M, *et al*. Principal component analysis for transgenic cotton seeds based on terahertz time domain spectroscopy system [J]. *Acta Pohtonica Sinica*, 2015, **44**(1): 0430001.

[17] YANG C, TIAN L, ZHAO K. Spectroscopic studies on the edible flavoring in terahertz range [J]. *Acta Pohtonica Sinica*, 2012, **41**(5): 627-630.

[18] CHEN T, LI Z, MO W. Identification of biomolecules by terahertz spectroscopy and fuzzy pattern recognition [J]. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2013, **106**: 48-53.

[19] DORNEY T D, BARANIUK R G, MITTLEMAN D M. Material parameter estimation with terahertz time-domain spectroscopy [J]. *Journal of the Optical Society of America. A*, 2001, **18**(7): 1562-1571.

[20] DUVILLARET L, GARET F, COUTAZ J L. Highly precise determination of optical constants and sample thickness in terahertz time-domain spectroscopy[J]. *Applied Optics*, 1999, **38**(2): 409-415.

[21] DUVILLARET L, GARET F E E, COUTAZ A J. A reliable method for extraction of material parameters in terahertz time-domain spectroscopy[J]. *Journal of Selected Topics in Quantum Electronics*, 1996, **2**(3): 739-746.

[22] ESBENSEN K, GELADI P. Principal component analysis [J]. *Chemometrics and Intelligent Laboratory Systems*, 1987, **2**(1-3): 37-52.

[23] RUMELHART D E, HINTON G E, WILLIAMS R J. Learning representations by back-propagating errors [C]. MIT Press, 1988: 533-536.

[24] ZHANG Y, PENG X H, CHEN Y, *et al*. A first principle study of terahertz (THz) spectra of acephate[J]. *Chemical Physics Letters*, 2008, **452**(1-3): 59-66.