Article

Automatic Identification of Cigarette Brand Using Near-Infrared Spectroscopy and Sparse Representation Classification Algorithm

Yang Shuangyan,[a,b] Hou Ying,[b] Yang Lingchun,[c] Zhang Jianqiang,*[a,d] Liu Weijuan,[a] Qiu Changgui,[b] Li Ming[a] and Yang Yanmei[a]

[a]*Yunnan Reascend Tobacco Technology (Group) Co., Ltd., 650000 Kunming, China*

[b]*Yunnan Comtestor Co., Ltd., 650000 Kunming, China*

[c]*Yunnan Entry-Exit Inspection and Quarantine Bureau, 650000 Kunming, China*

[d]*Faculty of Science, Kunming University of Science and Technology, 650000 Kunming, China*

A cigarette brand automatic classification method using near-infrared (NIR) spectroscopy and sparse representation classification (SRC) algorithm is put forward by the paper. Comparing with the traditional methods, it is more robust to redundancy because it uses non-negative least squares (NNLS) sparse coding instead of principal component analysis (PCA) for dimensionality reduction of the spectral data. The effectiveness of SRC algorithm is compared with PCA-linear discriminant analysis (LDA) and PCA-particle swarm optimization-support vector machine (PSO-SVM) algorithms. The results show that the classification accuracy of the proposed method is higher and is much more efficient.

Keywords: near-infrared spectroscopy, deep learning, sparse representation classification, non-negative least squares

Introduction

The applications of spectroscopy in various fields of modern science are increasing every year.[1-4] Near-infrared (NIR) spectroscopy has the advantages of being fast, accurate, easy and non-destructive and is becoming a useful tool for process analytical chemistry.[5] NIR spectroscopy is based on the absorption of electromagnetic radiation in the region from 780 to 2500 nm. Analysis of NIR spectroscopy usually involves a combination of multiple samples, each of which has a large number of correlated features. So how to reduce the complexity accompanying such large amounts of data is meaningful.[6]

Deep learning is a new area of machine learning research, which has been introduced with the objective of moving machine learning closer to one of its original goals: artificial intelligence. The term deep learning was introduced to the machine learning community by Dechter and Kleinrock in 1986,[7] and artificial neural networks by Aizenberg *et al.*[8] in 2000, in the context of Boolean threshold neurons. In 2005, Schmidhuber *et al.*[9] published a paper on learning deep partially observable Markov decision process (POMDP) through neural networks for reinforcement learning. In 2006, a publication by Hinton *et al.*[10] showed how a many-layered feed-forward neural network could be effectively pre-trained one layer at a time,[11] treating each layer in turn as an unsupervised restricted Boltzmann machine, then fine-tuning it using supervised back-propagation.[12] In recent years, deep learning architectures have been applied to fields including computer vision,[13] speech recognition,[14] natural language processing,[15] audio recognition,[16] social network filtering,[17] machine translation and bioinformatics,[18] where they produced results comparable to and in some cases superior[19] to human experts.[20] However, it has not been widely used in analytical chemistry and tobacco area yet. As one of deep learning methods, sparse representation is a parsimonious principle that a sample can be approximated by a sparse linear combination of basis vectors. It has two advantages: first, it is very robust to redundancy,[21] because it only selects few among all of the basis vectors; second, it is very robust to noise.[22] NIR spectral data is usually noisy and redundant, so it can be represented efficiently by sparse representation.

*e-mail: zhangjianqiang@reascend.com

In our paper, a novel spare representation classification method based on NIR spectroscopy and deep learning algorithm is first investigated to classify cigarette brands. In the method, the high-dimensional NIR spectral data will be investigated sparse representation for data analysis systematically. The large cigarette NIR spectral data is made to be sparse and adopt some transformation matrix to reduce dimensionality from a Bayesian viewpoint instead of principal component analysis (PCA) operation. The method has the advantages of denoising and avoiding overfitting and is more convenient in practice. In order to verify the reliability of the proposed method, the different brands of cigarettes have also been classified by the chemical index data. The two results are compared and show that the combination of NIR spectroscopy and deep learning algorithm is a promising tool for discriminating cigarettes of different brands in tobacco industry.

Experimental

Equipment

Firstly, the NIR spectrometer was preheated for one hour. Next, instrument test was carried out. If the test was passed, the samples would be scanned. The spectra in the near-infrared range of 1000-2500 nm were recorded in triplicate using a Nicolet Nexus 670 Fourier transform (FT)-NIR spectrometer with a spectral resolution of 4 cm^{-1} and 64 scans co-added. A mean spectrum was then calculated for each sample by averaging the triplicate spectra. The samples of tobacco were put into the rotating sample groove and the spectrum of a polytetrafluoroethylene sample was used as background. As the result, the NIR spectra reflects only the contribution of the tobacco of the cigarette. Examples of the different diffuse reflectance spectra of four different brands of cigarettes are shown in Figure 1.

Besides, 19 routine chemical index[23] were also detected. The contents of total sugar, reducing sugar, potassium, total plant alkali, chlorine, total nitrogen in tobacco were determined by continuous flow analytical method using Skalar SANPWS flow analyzer. The contents of glucose and fructose in tobacco were determined by high performance liquid chromatography (HPLC) using HP-5MS and DB-35MS (30 m × 0.25 mm i.d.) capillary column. The contents of malonic acid, succinic acid and malic acid in tobacco were determined by gas chromatography-mass spectrometry (GC-MS) method using Agilent 6890/5973 GC-MS. The content of aromatic components in tobacco was determined by simultaneous distillation and extraction GC-MS method
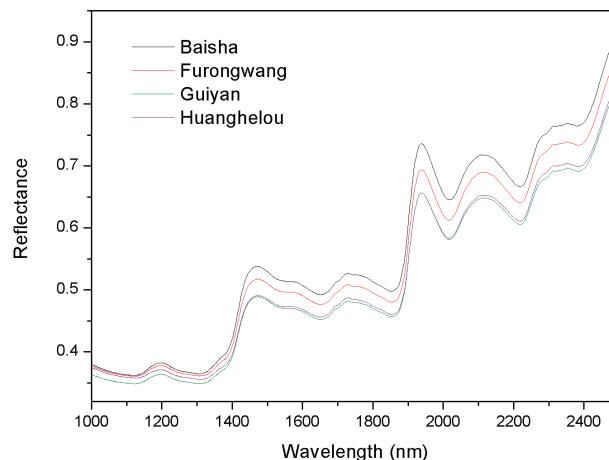

Figure 1. Examples of diffuse reflectance NIR spectra for the four different cigarette brands: Baisha, Furongwang, Guiyan and Huanghelou.

using R-215 rotary evaporator (Büchi). Examples of the main chemical index of four different brands of cigarettes are shown in Table 1.

Samples

Different brands of cigarettes differ in compositions, aroma and retail prices and as well as in the levels of potentially hazardous substances. Thus, it is important to have appropriate methods to distinguish different types of cigarettes. However, distinguishing different types of cigarettes mainly depends on human sensory responses, which are time consuming, laborious, and subjective, and may lead to unreliable results, and so it is necessary to develop alternative methods that are faster and more objective. In our research, two experimental sets were chosen. All samples of the two experimental sets were collected from 9 cigarette factories in south and north China, respectively. Data set 1 has 200 cigarette samples and it contains four different brands: Baisha, Furongwang, Guiyan and Huanghelou. Data set 2 has 240 cigarette samples and it contains five different brands: Changbaishan, Huangjinye, Taishan, Lanzhou and Jiaozi. The NIR spectra of two data sets is shown in Figure 2. The chemical index of the two data sets are two matrices with the dimensionality 200 × 19 and 240 × 19.

Theory of classification algorithms

Theory of linear discriminant analysis (LDA) algorithm

Linear discriminant analysis (LDA) is a well-known dimension reduction and classification method.[24] In the algorithm, the data is projected into a low dimension space so that the different classes can be well separated. If the method is used for binary classification problems, a set of

Table 1. Nineteen routine chemical index of Baisha, Furongwang, Guiyan and Huanghelou

Chemical index	Baisha	Furongwang	Guiyan	Huanghelou
Total sugar / %	22.800	21.400	21.200	20.400
Potassium / %	2.340	2.680	2.170	2.260
Reducing sugar / %	20.600	19.700	19.000	18.700
Total plant alkali / %	2.260	2.170	2.190	2.460
Chlorine / %	0.500	0.660	0.490	0.750
Total nitrogen / %	1.960	2.110	2.120	2.300
Malonic acid / (mg g^{-1})	0.740	1.030	0.850	1.350
Succinic acid / (mg g^{-1})	0.280	0.280	0.450	0.270
Malic acid / (mg g^{-1})	38.530	40.060	37.220	40.810
Glucose / (μg g^{-1})	7.620	7.540	8.590	7.270
Fructose / (μg g^{-1})	5.990	5.830	5.770	5.850
Furfural / (μg g^{-1})	1.644	1.739	1.301	1.737
Furoic acid / (μg g^{-1})	0.256	0.232	0.235	0.211
Benzaldehyde / (μg g^{-1})	0.149	0.202	0.179	0.172
6-Methyl-5-heptene-2-ketone / (μg g^{-1})	0.162	0.189	0.177	0.189
Solanone / (μg g^{-1})	9.363	9.505	9.084	10.557
β-Damascenone / (μg g^{-1})	2.148	2.531	2.239	2.452
Geranyl acetone / (μg g^{-1})	1.605	1.634	2.155	1.675
Megalanone / (μg g^{-1})	23.782	22.941	29.538	24.373
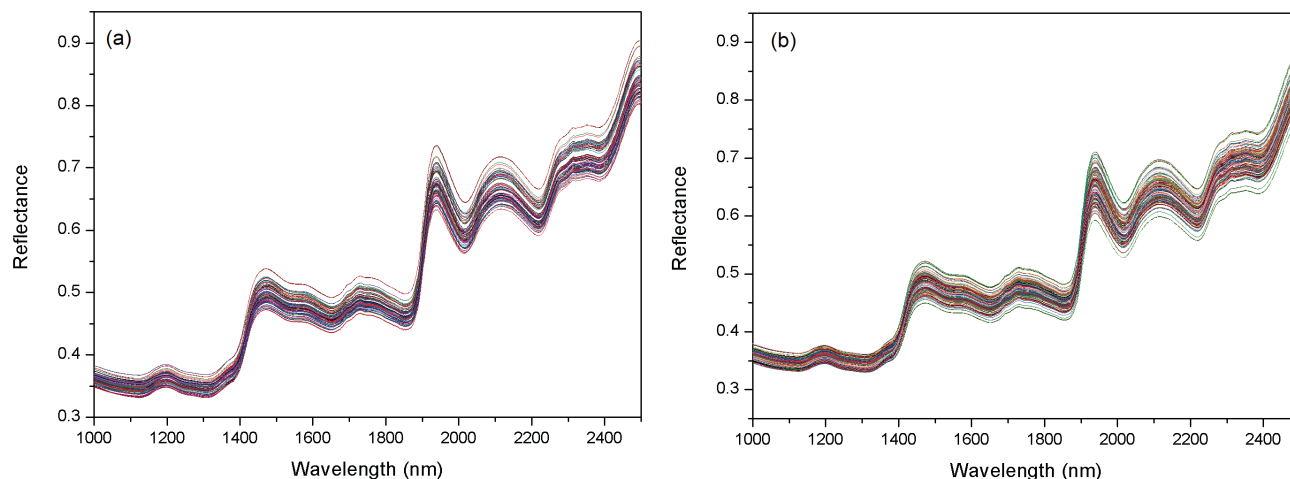
Figure 2. Two spectral data sets. (a) Original data of data set 1 with 200 samples and (b) original data of data set 2 with 240 samples.

n samples which belongs to two classes C_1 with n_1 samples and C_2 with n_2 samples. If each sample is described by q variables, the data forms a matrix $\mathbf{X} = (\mathbf{X_{ij}})$, i = 1, …, n; j = 1, …, q. We denote by μ_k the mean of class C_k and by μ the mean of all the samples:

$$\mu_k = \frac{1}{n_k} \sum_{x_i \in C_k} x_i \qquad (1)$$

$$\mu = \frac{1}{n} \sum_k n_k \mu_k \qquad (2)$$

Then the between-class scatter matrix $\mathbf{S_B}$ and the within-class scatter matrix $\mathbf{S_W}$ can be defined as:

$$\mathbf{S_B} = \sum_k n_k (\mu_k - \mu)(\mu_k - \mu)^t \qquad (3)$$

$$\mathbf{S_W} = \sum_k \sum_{x_i \in C_k} (x_i - \mu_k)(x_i - \mu_k)^t \qquad (4)$$

LDA seeks a linear combination of the initial variables on which the means of the two classes are well separated,

measured relatively to the sum of the variances of the data assigned to each class. For the purpose, LDA determines a vector ω such that $\omega^t S_B \omega$ is maximized while $\omega^t S_W \omega$ is minimized. This double objective is realized by the vector ω_{opt} that maximizes the criterion:

$$J(\omega) = \frac{\omega^t S_B \omega}{\omega^t S_W \omega} \qquad (5)$$

It can be proved that the solution ω_{opt} is the eigenvector associated with the sole eigenvalue of $S_W^{-1} S_B$ if S_W^{-1} exists. Once ω_{opt} is determined, LDA provides a classifier.

Theory of particle swarm optimization (PSO)-support vector machine (SVM) algorithm

SVM is developed by Vapnik.[25] It is based on some 'beautifully simple ideas'[26] and provides a clear demonstration of what learning from examples is all about. Details about SVM classifiers can be found in He *et al.*[27] In computer science, PSO is a computational method that optimizes a problem by iteratively trying to improve a candidate solution with regard to a given measure of quality. Here PSO algorithm uses particles moving in an m-dimensional space to search solutions of an optimization problem with m variables. In our approach, PSO is used to search for the optimal particle. Each particle represents a candidate solution. SVM classifier is built for each candidate solution to evaluate its performance. Velocity and position of particles can be updated by:

$$v_{ij}^{(t+1)} = \omega v_{ij}^t + c_1 rand_1 (pbest_{ij}^t - x_{ij}^t) + \qquad (6)$$
$$c_2 rand_2 (gbest_{ij}^t - x_{ij}^t) x_{ij}^{t+1} = x_{ij}^t + v_{ij}^{t+1}$$

where t is evolutionary generation, v_{ij} and x_{ij} stand for the velocity and position of particle i on dimension j, respectively, ω is the inertia weight and it is used to balance the global exploration and local exploitation, rand represents the random function, c_1 is the personal learning factor and c_2 is the social learning factor. In fact, the aim of PSO-SVM algorithm is to optimize the accuracy of SVM classifier by randomly generating the parameters and estimate the best value for regularization of kernel parameters for SVM model.

Theory of non-negative least squares (NNLS) spare representation classification algorithm

Spare representation (SR) is a principle that a signal can be approximated by a sparse linear combination of dictionary atoms.[21] The SR model can be formulated as:

$$(b|A, x, k) = x_1 a_1 + \ldots + x_k a_k + \varepsilon = Ax + \varepsilon \qquad (7)$$

where $A = [a_1, \ldots, a_k]$ is called dictionary, a_1 is a dictionary atom, x is a sparse coefficient vector, and ε is an error term. A, x and k are the model parameters. SR involves sparse coding and dictionary learning. The sparse codes can be obtained by many regularization methods and constrains. If we pool all training instances in a dictionary, and then learn the non-negative coefficient vectors of a new instance, which is formulated as a one-sided model:

$$\min_x \frac{1}{2} \|b - Ax\|_2^2 st x \geq 0 \qquad (8)$$

The model is called NNLS sparse coding.[22] NNLS sparse coding has two advantages: first, the non-negative coefficient vector is more easily interpretable than coefficient vector of mixed signs, under some circumstances. Second, NNLS sparse coding is a non-parametric model which is more convenient in practice. As the result, the NNLS sparse coding sparse coding algorithm is chosen in the following part. The main idea of spare representation classification is to represent a given test sample as a sparse linear combination of all training samples, then classifies the test sample by evaluating which class leads to the minimum residual. The spare representation classification algorithmic procedures can be summarized as follows: (*i*) input: a matrix of training samples $A = [A_1, A_2, \ldots, A_k] \in \Re^{m \times n}$ for k classes, a test sample $y \in \Re^{m \times n}$, (and an optional error tolerance $\varepsilon > 0$); (*ii*) normalize the columns of A to have unit $l^2 - norm$; (*iii*) learn the sparse coefficient matrix X, of the new instances by solving equation 8; and (*iv*) use a sparse interpreter to predict the class labels of new instances, such as nearest neighbor, K-nearest neighbor, or nearest subspace rule.

Therefore, the main idea of the algorithm can be concluded as: first, training instances are collected in a dictionary. Then, a new instance is regressed by NNLS sparse coding. Thus, its corresponding sparse coefficient vector is obtained. Next, the regression residual of this instance to each class is computed, and finally this instance is assigned to the class with the minimum residual.

Results and Discussion

Three different multivariate data analysis techniques will be used to solve the problem, including LDA, PSO-SVM and sparse representation classification (SRC) algorithms and the results are shown in the following part. Twenty five samples of each class (the total is 100) are chosen as the training set and the other 100 samples are chosen as the testing set for data set 1. For data set 2, 24 samples of each class (the total is 120) are chosen as the training set and the other 120 samples are chosen as the testing set. Here accuracy and elapsed execution time are used to measure

the classification performance. For LDA and PSO-SVM algorithms, the spectral data is a high-dimensional data. If it is directly used in the two classification algorithms, it will lead to high computational-complexity. As the result, principal component analysis is used for outlier detection and dimensionality reduction of NIR spectral data. Tables 2 and 3 are the first four and five principal components and total variance contribution rates of the principal component analysis results of data sets 1 and 2, respectively. It can be seen from the two tables that the first four and five principal components can be a good description of the two original spectral data sets. After PCA operation, the dimensionality of the two data sets have been reduced from 200 × 1550 to 200 × 4 and 240 × 1550 to 240 × 5.

We employed two-fold cross-validation to partition a data set into training sets and test sets. It means half of the samples are chosen randomly as the training set and the other half are chosen randomly as the test set for both data sets 1 and 2. All the classifiers ran on the same training and test splits for fair comparison. We defined the accuracy of a given classifier as the ratio of the number of correctly predicted test samples to the total number of test samples. For each data set, two-fold cross-validation was 10 times, and the average classification accuracy over the 10 runs were computed. The correct classification number and accuracy of each class and the average accuracy of all classifiers on the two data sets are compared in Tables 4 and 5, respectively. It can be seen from Tables 4 and 5

that the correct classification number and accuracy of spare representation classification algorithm based on NNLS sparse coding is comparable with that of other two algorithms. This is because PCA-LDA and PCA-PSO-SVM algorithms use the minimum Euclidean distance of the feature space between the training samples and test samples to classify. It will lead to ineffective classification results. However, the spare representation classification algorithm can capture the essential feature of the data by means of using the redundancy characteristic of the dictionary and has the strong robustness. Besides, as there is no pre-processing operation on the two data sets and the non-negative least squares spare coding representation classification algorithm is robust to noise, it has better classification results than that of the other algorithms. This convinces us that sparse coding spare representation classification classifiers can be very effective for classifying high-dimensional spectroscopy data.

In order to verify the accuracy of the method using NIR spectral data, different brands of cigarettes were also classified by using chemical index data. The classifiers and the cross-validation method were the same with the above method. The results for data sets 1 and 2 are shown in Tables 6 and 7, respectively. The results show that the classification accuracy is roughly the same with Tables 4 and 5. However, the detection of the chemical index of the cigarette is much more expensive and time-consuming than NIR method. Considering the above factor and results, the

Table 2. First four principal components and total variance contribution rates of data set 1

Principal component	1	2	3	4
Total variance contribution	82.847	99.206	99.558	99.865

Table 3. First five principal components and total variance contribution rates of data set 2

Principal component	1	2	3	4	5
Total variance contribution	84.158	99.321	99.706	99.856	99.924

Table 4. Classification results by PCA-LDA, PCA-PSO-SVM and SRC algorithms of data set 1 using NIR spectral data

Brand (test No.)	Classification accuracy No., percentage / %		
	PCA-LDA	PCA-PSO-SVM	SRC
Baisha (239)	191, 79.91	197, 82.43	228, 95.40
Furongwang (257)	188, 73.15	195, 75.87	207, 80.54
Guiyan (261)	211, 80.84	218, 83.52	221, 84.67
Huanghelou (243)	201, 82.72	225, 92.60	227, 93.42
Average accuracy	791, 79.10	835, 83.50	883, 88.30

PCA-LDA: principal component analysis-linear discriminant analysis; PCA-PSO-SVM: principal component analysis-particle swarm optimization-support vector machine; SRC: sparse representation classification.

Table 5. Classification results by PCA-LDA, PCA-PSO-SVM and SRC algorithms of data set 2 using NIR spectral data

Brand (test No.)	Classification accuracy No., percentage / %		
	PCA-LDA	PCA-PSO-SVM	SRC
Changbaishan (248)	212, 85.48	224, 90.30	237, 95.56
Huangjinye (242)	187, 77.27	192, 79.33	220, 90.91
Taishan (237)	213, 89.87	198, 83.54	224, 94.51
Lanzhou (235)	191, 81.27	223, 94.89	228, 97.02
Jiaozi (238)	211, 88.66	214, 89.92	211, 88.66
Average accuracy	1014, 84.50	1051, 87.58	1120, 93.33

PCA-LDA: principal component analysis-linear discriminant analysis; PCA-PSO-SVM: principal component analysis-particle swarm optimization-support vector machine; SRC: sparse representation classification.

Table 6. Classification results by LDA, PSO-SVM and SRC algorithms of data set 1 using chemical index data

Brand (test No.)	Classification accuracy No., percentage / %		
	LDA	PSO-SVM	SRC
Baisha (256)	207, 80.85	209, 81.64	238, 92.97
Furongwang (248)	192, 77.42	192, 77.42	219, 88.31
Guiyan (239)	205, 85.77	210, 87.87	211, 88.28
Huanghelou (257)	197, 76.65	232, 90.27	235, 91.44
Average accuracy	801, 80.10	843, 84.30	903, 90.30

LDA: linear discriminant analysis; PSO-SVM: particle swarm optimization-support vector machine; SRC: sparse representation classification.

Table 7. Classification results by LDA, PSO-SVM and SRC algorithms of data set 2 using chemical index data

Brand (test No.)	Classification accuracy No., percentage / %		
	LDA	PSO-SVM	SRC
Changbaishan (239)	212, 88.70	227, 94.98	224, 93.72
Huangjinye (245)	191, 77.96	188, 77.96	231, 94.28
Taishan (247)	203, 82.19	207, 83.81	230, 93.12
Lanzhou (232)	188, 81.03	209, 90.09	222, 95.69
Jiaozi (237)	213, 89.87	221, 93.25	219, 92.41
Average accuracy	1007, 83.92	1052, 87.67	1126, 93.83

LDA: linear discriminant analysis; PSO-SVM: particle swarm optimization-support vector machine; SRC: sparse representation classification.

NIR technology combination with SRC algorithm could be an effective tool for discriminating different brands of cigarettes.

The averaged elapsed execution time in seconds of each method is also recorded as a measure of performance. All experiments are performed on an Intel machine (Core i5-4590s, 3.00 GHz, central processing unit (CPU) with 8 GB random access memory (RAM), with 64-bit Windows 7 Professional operating system). All methods are implemented in the language MATLAB, 64-bit version 2010b. Figure 3 shows the computing times of the methods for the two data sets. It can be clearly seen that spare representation classification is much more efficient than the other two methods.
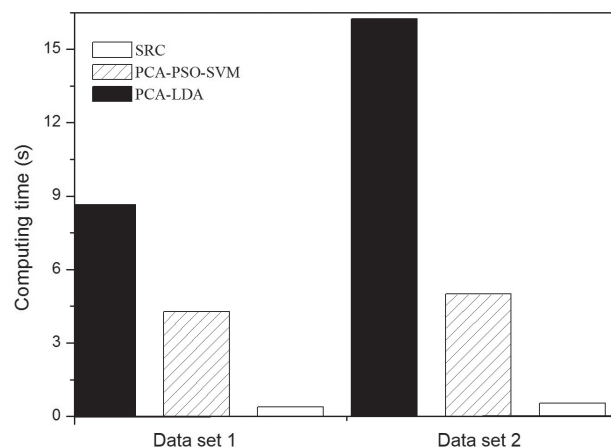
Figure 3. Computing times of the two data sets.

Therefore, it can be concluded that spare representation classification works better than PCA-LDA and PCA-PSO-SVM algorithms. Comparing with the two algorithms, the spare representation classification algorithm is robust to noise and has the higher classification accuracy and less computation time. As the result, it could be an effective method for discriminating different brands of cigarettes.

Conclusions

In this study, an effective spare representation classification method is proposed to classify high dimensional spectroscopy data. Comparing with the traditional algorithms, the method does not need any principal component analysis to reduce the dimensionality of the data and has the higher classification accuracy and less computation time. The results suggest that NIR spectroscopy technology together with sparse representation classification algorithm could be an alternative to traditional methods for discriminating different brands of cigarettes.

Acknowledgments

References

1. Hollas, J. M.; *Modern Spectroscopy*; Wiley: Hoboken, 2003.

2. Workman, J.; Springsteen, A.; *Applied Spectroscopy: A Compact Reference for Practitioners*; Academic Press: New York, 1998.

3. Osborne, B.; Fearn, T.; *Near Infrared Spectroscopy in Food Analysis*; Wiley: New York, 1986.

4. Duarte, F. J.; *Tunable Laser Applications*, 2nd ed.; CRC Press: New York, 2008.

5. Nicolai, B. M.; Beullens, K.; Bobelyn, E. J.; *Postharvest Biol. Technol.* **2007**, *46*, 99.

6. Breitkreitz, M. C.; Raimundo Jr., I. M.; Rohwedder, J. J. R.; Pasquini, C.; Dantas Filho, H. A.; José, G. E.; Araújo, M. C. U.; *Analyst* **2003**, *128*, 1204.

7. Dechter, R.; Kleinrock, L.; *Broadcast Communications and Distributed Algorithms*; IEEE Computer Society: Washington, 1986.

8. Aizenberg, I. N.; Aizenberg, N. N.; Krivosheev, G. A.; *Lect. Notes Comput. Sci.* **1999**, *1715*, 21.

9. Schmidhuber, J.; Gagliolo, M.; Wierstra, D.; Gomez, F.; *Technical Report No. IDSIA-19-05* **2005**, available at https://www.researchgate.net/publication/228831252_Recurrent_Support_Vector_Machines, accessed in February 2018.

10. Hinton, G. E.; Osindero, S.; Teh, Y. W.; *Neural Comput.* **2006**, *18*, 1527.

11. Bengio, Y.; *Neural Networks* **2012**, *1*, 437.

12. Hinton, G. E.; *Trends Cognit. Sci.* **2007**, *11*, 428.

13. Bengio, Y.; Courville, A.; Vincent, P.; *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 1798.

14. Schmidhuber, J.; *Neural Networks* **2015**, *61*, 85.

15. Bengio, Y.; LeCun, Y.; Hinton, G.; *Nature* **2015**, *521*, 436.

16. Schmidhuber, J.; *Scholarpedia* **2015**, *10*, 32832

17. Olshausen, B. A.; *Nature* **1996**, *381*, 607.

18. Cireşan, D.; Meier, U.; Masci, J.; Schmidhuber, J.; *Neural Networks* **2012**, *32*, 333.

19. Bengio, Y.; *Found. Trends Mach. Learn.* **2009**, *2*, 127.

20. Ivakhnenko, A.; *IEEE Trans. Syst. Man, Cybern.* **1997**, *4*, 364.

21. Li, Y.; Ngom, A.; *BMC Syst. Biol.* **2013**, *4*, 1.

22. Li, Y.; Ngom, A.; *Neurocomputing* **2013**, *118*, 41.

23. Rodgman, A.; Perfetti, T. A.; *The Chemical Components of Tobacco and Tobacco Smoke*, 2nd ed.; CRC Press: New York, 2013.

24. Dogantekin, E.; Dogantekin, A.; Avci, D.; Avci, L.; *Digit. Signal Process.* **2010**, *20*, 1248.

25. Vapnik, V.; *The Nature of Statistical Learning Theory*; Springer-Verlag: New York, 1995.

26. Vapnik, V.; *Statistical Learning Theory*; Wiley: New York, 1998.

27. He, Q.; Yan, J.; Shen, Y.; Bi, Y.; Ye, G.; Tian, F.; *Intell. Autom. Soft Comput.* **2012**, *18*, 967.