Chemometrics and Intelligent Laboratory Systems 139 (2014) 42–47

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

Spectroscopy-based food classification with extreme learning machine

Wenbin Zheng [a,b], Xiaping Fu [a,*], Yibin Ying [a]

[a] College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou, China
[b] College of Information Engineering, China Jiliang University, Hangzhou, China

ARTICLE INFO

Article history:
Received 4 July 2014
Received in revised form 11 September 2014
Accepted 22 September 2014
Available online 28 September 2014

Keywords:
Spectroscopy
Food
Classification
Extreme learning machine

ABSTRACT

The combination of spectroscopy and a novel classification algorithm called extreme learning machine (ELM) was developed for food classification. The performance influence of different input dimensionalities was also investigated. The classification accuracy and speed of five chemometrics techniques, including k-nearest neighbor (KNN), partial least-squares discriminant analysis (PLS-DA), back propagation artificial neural network (BP-ANN), least-squares support vector machine (LS-SVM), and ELM were compared. It was presented that the accuracy of ELM was better than its competitors in most cases. Moreover, on the classification stage, ELM performed much faster than KNN, LS-SVM, and BP-ANN, which indicated that ELM may be a promising method for real-time food classification with a comparable accuracy based on near or mid-infrared spectroscopy.

1. Introduction

Nowadays, the classification of foods is an important issue, which has many practical applications such as brand classification [1], aging time detection [2], geographic origin discrimination [3,4] and so on. Traditionally, the wet chemical methods for classification are time-consuming and expensive and they involve a considerable amount of manual work.

As an alternative method, near or mid-infrared spectroscopy has become a powerful analytical technique because it is nondestructive, simple, and fast. It requires little or no sample preparation and can be automated easily [5,6]. For the spectroscopy-based classification, chemometricians have developed many valuable algorithms, such as k-nearest neighbor (KNN) [7], partial least-squares discriminant analysis (PLS-DA) [8,9], artificial neural networks (ANN) [10,11] and support vector machine (SVM) [12–14].

KNN is a simple unsupervised classification algorithm by a majority vote of its K-nearest neighbors to assign the category of a new sample. The parameter K has a great impact on the classification performance. Especially, if $K = 1$, then the sample is simply assigned to the class of its nearest neighbor. This method can be implemented easily but has slow running speed and its classification accuracy depends closely on the dataset.

PLS-DA is a variant of partial least-squares regression (PLS) for the optimum separation of classes. It selects latent variables in the feature spaces which have a maximum covariance within the predictor variables. A number of PLS-based classification applications have been successfully employed [3,4,15,16]. Nevertheless, PLS has sometimes difficulties in yielding satisfactory performance because of non-linearity and over-fitting [17].

ANN is a non-linear and non-parametric regression technique, which has also been used to solve classification problems. A commonly used ANN is back propagation (BP) neural network, where the data are fed forward into the network without feedback, while the output error is propagated backward from the output layer to the hidden layer and finally to the input layer. Generally, the parameters of the BP neural networks are learnt via gradient descent algorithms, which are relatively slow and have many convergence issues such as stopping criteria, learning rate, learning epochs, and local minima.

SVM, as an outstanding supervised algorithm, aims to find an optimal hyperplane to correctly separate the objects of the different classes as much as possible. When the linear boundary in the input space is not enough to separate the classes, it can convert data from the low dimensional input space to a high dimensional feature space via kernel functions to handle the non-linearity. SVM could effectively avoid the over-fitting problem because it is based on the structural risk minimum mistake rather than the minimum mistake of the misclassification on training set. Therefore, it has good generalization performance and often performs well on different datasets [18,19]. A problem of SVM is that its parameters are usually difficult to be tuned. As a reformulation of SVM, the least-squares support vector machines (LS-SVM) [20] can avoid the complex calculations via solving a set of standard linear equations [21]. Moreover, numerous researches have showed that the prediction performance of LS-SVM is better than that of ANN [22,23] or PLS [24–26].

Nevertheless, since LS-SVM is originally designed for binary classification, extending it to multi-class classification needs considerable running time, which hinders it from being applied in real-time scene. Recently, a novel algorithm for single-hidden layer feed-forward neural

* Corresponding author.
E-mail address: fuxp@zju.edu.cn (X. Fu).

http://dx.doi.org/10.1016/j.chemolab.2014.09.015
0169-7439/

network called extreme learning machine (ELM) was proposed [27,28]. It was reported that the ELM could learn fast with high generalization performance and implement the multi-class classification quickly because of its network output structure [29]. As a rapidly developing technology, a large number of applications of ELM have emerged in recent years [30–32]. However, to our knowledge, ELM has not been used for the food classification based on near or mid-infrared spectroscopy.

The objectives of this study are (1) to develop a technique to combine the spectroscopy and ELM for food classification; (2) to investigate the performance influence when the input dimensionality is changed via the feature reduction; and (3) to compare the classification accuracy and speed of different chemometrics techniques, including KNN, PLS-DA, BP-ANN, LS-SVM, and ELM.

2. Theory and implementation

2.1. Extreme learning machine for classification

ELM is a single hidden layer feed forward network where the input weights are chosen randomly and the output weights are calculated analytically [28]. Fig. 1 is an illustration of ELM, where x is the input vector, G is an activation function, w is the weight vector connecting the hidden node and the input nodes, b is the bias corresponding to its hidden node, β is the weight vector connecting the hidden node and the output nodes, and $f(x)$ is the output vector.

For N calibration samples $(x_i, t_i) \in \mathbf{R}^d \times \mathbf{R}^m$, here t_i represents the output of $f(x)$ coded by $\{0, 1\}^m$, i.e., the category vector, then ELM with activation function $G(w, b, x)$ and L hidden nodes are mathematically modeled as:

$$\sum_{i=1}^{L} \beta_i G_i\left(w_i \cdot x_j + b_i\right) = t_j, j = 1, \ldots, N. \tag{1}$$

The N equations can be written compactly as:

$$H\boldsymbol{\beta} = T \tag{2}$$

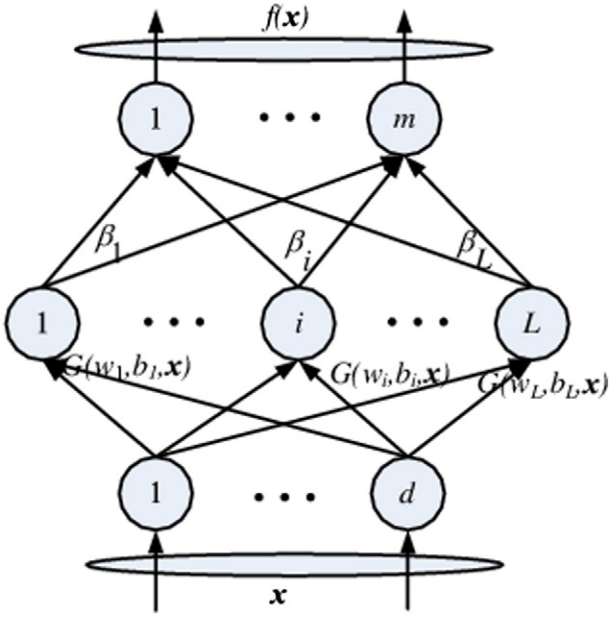

Fig. 1. An illustration of ELM.

where

$$\mathrm{H} = \begin{bmatrix} G(w_1 \cdot x_1 + b_1) & \cdots & G(w_L \cdot x_1 + b_L) \\ \vdots & \cdots & \vdots \\ G(w_1 \cdot x_N + b_1) & \cdots & G(w_L \cdot x_N + b_L) \end{bmatrix}_{N \times L} \tag{3}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_N^T \end{bmatrix}_{L \times m}, T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix}_{L \times m}. \tag{4}$$

It has been proved that the hidden node parameters can be randomly fixed and the weight matrix $\boldsymbol{\beta}$ can be estimated as: $\hat{\boldsymbol{\beta}} = H^{\dagger}T$ [29]. Furthermore, in order to obtain more generalization performance [30], $\boldsymbol{\beta}$ can be estimated as:

$$\hat{\boldsymbol{\beta}} = \left(H^T H + \lambda I\right)^{-1} H^T T \tag{5}$$

where $\lambda > 0$ is a regularized parameter. Therefore, the output weights of ELM can be obtained analytically. Theoretically, the global optimal solution can be obtained, which avoids many convergence issues as mentioned above.

For an unknown sample $\tilde{x}$, its category could be obtained by:

$$category(\tilde{x}) = \arg\max\left(\tilde{h}\hat{\boldsymbol{\beta}}\right) \tag{6}$$

where $\tilde{h} = [G(w_1 \cdot \tilde{x} + b_1) \cdots G(w_L \cdot \tilde{x} + b_L)]$.

2.2. Spectra preprocess and feature reduction

Generally, spectra data need to be preprocessed because they are often affected by many factors, such as temperature, light scattering, baseline drift, or background noise. Here, the centered normalization and first derivative were employed.

Spectra typically have hundreds or thousands of wavelength variables, and some variables may contain useless or irrelevant information for classification task like noise and background, which deteriorate the predictive ability of the model [33]. Moreover, the modeling time cost of ELM will become remarkably large if the input dimensionality is very high, since ELM usually tends to require more hidden neurons than the conventional tuning-based algorithms [34], which results in a large amount of calculation to estimate the weight matrix $\hat{\boldsymbol{\beta}}$.

Therefore, the principal component analysis (PCA) is used to extract useful information from the original data and reduce the input dimensionality. PCA can transform original variables into the new axis and calculate the principal components (PCs) as new variables. Each PC is a linear combination of the original measured variables. Usually, the first several PCs are enough to explain most of the total variances, thus these PCs can be selected as the new input variables.

2.3. Algorithm description of the proposed method

Let x_i, $i = 1,\ldots,N$ be the calibration spectral data that has been preprocessed and its features have been reduced by PCA, t_i is the category code corresponding to x_i.

From the computational point of view, the calibration procedure consists of the following steps:

(1) Generating the input weights w and bias b randomly
(2) Calculating the hidden layer output matrix **H** with all x_i using Eq. (3)
(3) Calculating the matrix T with all t_i using Eq. (4)
(4) Calculating the output weighting matrix $\hat{\boldsymbol{\beta}}$ using Eq. (5).

Thus, for an unknown sample $\tilde{x}$, its category can be obtained using Eq. (6).

3. Experiments

3.1. Spectral datasets

Four spectral datasets were chosen for this study, which are available on http://www.ifr.ac.uk/Bioinformatics/BSDataSets.html.

(1) The first dataset was the coffee samples [3] (here denoted by Coffee for convenience) obtained by Fourier transform infrared spectroscopy with diffuse reflectance sampling. It contained 56 samples (arabica and robusta species, respectively 29 and 27 of each), and each spectrum contained 286 variables in 5233–12,338 nm.
(2) The second dataset (here denoted by Meat [15]) contained 60 independent samples of fresh minced meats: chicken, pork and turkey (20 of each class), which was obtained using Fourier transform infrared spectroscopy with attenuated total reflectance sampling. Each spectrum contained 448 variables in 5353–11,123 nm.
(3) The third one was an olive oil dataset [4] (here denoted by Oil) obtained using Fourier transform infrared spectroscopy with attenuated total reflectance sampling, which contained 60 authenticated extra virgin olive oils from 4 different countries of origin: Greece, Italy, Portugal, Spain, (respectively 10, 17, 8, 25 of each), and each spectrum contained 570 variables in 5272–12,517 nm.
(3) The forth dataset (here denoted by Fruit) contains a collection of 983 Mid-infrared spectra collected from different authenticated fruit purees in one of two classes: "Strawberry" and "NON-Strawberry", respectively 351 and 632 of each class [16]. Its spectral range was truncated to 5549–11,123 nm, and each spectrum contained 235 variables.

3.2. Spectral preprocess and feature reduction

For each dataset, the spectral variable was centered (by calculating the average spectrum of the dataset and subtracted form the spectra) and then scaled by its standard deviation to remove the dependence on magnitude. The Savitzky-Golay [35] first-order derivative with a 5-point moving window (fitted by a polynomial of two degree) was also applied to eliminate the baseline drifts and enhance the small spectral difference.

For each classification model, in order to investigate the performance influence (or the stability of models) under different input dimensionalities, different numbers of input variables (i.e., the number of PCs generated by PCA) were tried. The total variance explained by these PCs varied from about 60% to 95%.

3.3. Calculation and performance evaluation

All calculations were performed in Matlab 2010a under Windows XP with 3.2 GHz CPU and 4 GB memory, and the LS-SVM algorithm was implemented with the LS-SVMlab toolbox (available on http://www.esat.kuleuven.be/sista/lssvmlab/). Here, the popular kernel, i.e., radical basis function (RBF) was chosen, and the parameters (γ, σ) of the LS-SVM was optimized by grid search with leave-one-out cross validation to avoid over fitting. For multi-class classification, the one-to-one mode was used. The Matlab statistics and neural network toolbox were employed for the implementation of PLS-DA and BP, respectively. The optimal number of latent variables in PLS-DA was searched in the interval (1 q) by cross validation, where q is equal to the minimum number between the number of samples and variables. The optimal number of hidden nodes of BP was obtained via leave-one-out cross validation in the interval from 10 to 100 (more large number of hidden nodes had no performance improvement in our experiments). For the KNN algorithm, the optimal value of K was searched in (1–30) with leave-one-out cross validation.

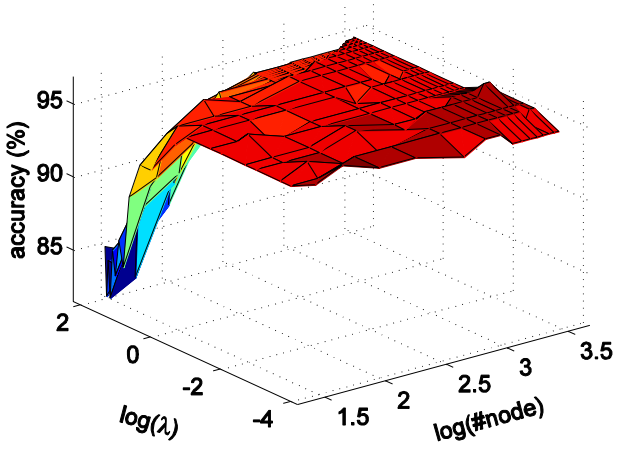

Fig. 2. A parameter search result.

The leave-one-out cross validation was used to evaluate the accuracy on the Coffee, Meat and Oil datasets. For the Fruit dataset, we used the 10-fold cross validation for the sake of computation cost. In addition, the statistical test (t-test at the 5% significance level) was also used to evaluate the accuracy comparison between the method pairs.

The averaging time cost on the classification stage was used to evaluate the running speed of the classifiers, which is defined as $t = \frac{1}{n}\sum_{i=1}^{n} t_i$, where t_i denotes the classification time cost of the i-th sample, n means the number of samples, and the unit of t was microsecond. Note that the preprocess time was not included in t_i, since all methods had same preprocess.

4. Results and discussion

4.1. Analysis of ELM parameters

The parameters of ELM include the activation function, the number of hidden nodes (denoted by #node for convenience), and the regularized parameter λ. For each input dimensionality and each selection of activation function, a grid search technique with leave-one-out cross validation was employed to obtain the optimal model parameters (#node, λ) within the region in (30–3000) and (10^{-4}–10^{2}), respectively. Fig. 2 gives a parameter search result when the input dimensionality is 3 and

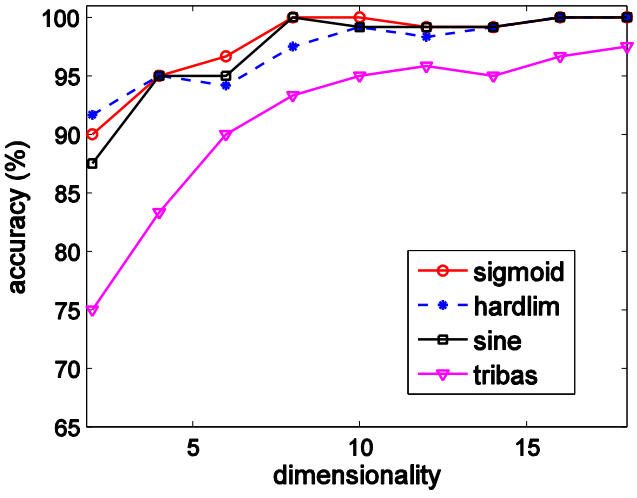

Fig. 3. Accuracy comparison of four active functions on the Meat dataset.

Table 1
Accuracy comparison of five algorithms on four datasets.

Dataset	Accuracy (%)				
	KNN	LS-SVM	PLS-DA	BP-ANN	ELM
Coffee	98.21 ± 0.00	97.50 ± 1.73	**100.00 ± 0.00**	94.82 ± 4.49	**100.00 ± 0.00**
Meat	92.31 ± 2.20	95.83 ± 2.98	97.69 ± 3.33	95.65 ± 2.76	**97.78 ± 3.41**
Oil	83.25 ± 3.23	95.08 ± 3.92	93.08 ± 9.49	90.50 ± 3.45	**97.35 ± 3.65**
Fruit	67.01 ± 1.33	**96.01 ± 1.88**	85.07 ± 9.11	95.33 ± 1.79	95.05 ± 3.15

the activation function is "sigmoid" (i.e., $f(X) = 1/(1 + \exp(-X))$) on the Meat dataset.

In Fig. 2, the accuracy is poor when #node is very small and λ is very large, while as the parameter #node increases, the accuracy increases and finally keeps relatively stable within a rather wide range, which means that the optimal parameters can be found quickly by a two-step grid search technique, i.e., crude search with a large step size and fine search with a small step size. More experiments reveal that the model is not very sensitive to λ when #node is large.

Fig. 3 shows the accuracy comparison on the Meat dataset when four different active functions were used, where the active functions include sigmoid, hard limit (abbreviated as hardlim), sine, and triangular basis function (abbreviated as tribas). For each dimensionality, the parameter pair (#node, λ) had been optimized by cross validation. It can be seen that the sigmoid function achieves the best performance among these active functions. Similar situations were obtained on the other datasets. The sigmoid function was thus used as the active function of ELM in following data analysis.

4.2. Comparison of different algorithms

Five models (i.e., KNN, LS-SVM, PLS-DA, BP-ANN and ELM) were built on each of four datasets (i.e., Coffee, Meat, Oil, and Fruit). Table 1 shows the accuracy comparison results, where the accuracy (for each model on each dataset) represents the mean and standard deviation of accuracies for all input dimensionalities (so the mean plus its standard deviation may be greater than 100%), the bold indicates the best performance on the same dataset.

On the Coffee dataset, all classification models achieve a good accuracy, which implies that the dataset can be separated easily. As Fig. 4 shows, samples of the two classes can be separated linearly, even though only two PCs are used.

However, on other two datasets (the Oil and Fruit), KNN model has a relatively large decline on accuracy. Especially, KNN performed rather poorly on the Fruit dataset, which reflected that the performance of KNN was very sensitive to different spectral data. Fig. 5 shows the score plot of PC1 and PC2 derived from the Fruit dataset. Obviously, it is more difficult to be separated with two PCs than the Coffee dataset.

Figs. 6, 7 and 8 give the detailed accuracy comparison corresponding to all input dimensionalities on the dataset Meat, Oil and Fruit respectively.

In Fig. 6, PLS-DA performs better than LS-SVM and BP-ANN (two nonlinear models), and it achieves the similar accuracy with ELM (at the 5% significance level). However, as Fig. 7 shows, the accuracy of LS-SVM is slightly superior to PLS-DA in the case of low dimensionalities, and ELM obtains the best performances in most dimensionalities (the difference between ELM and SVM is statistically significant). On the both Meat and Oil datasets, BP-ANN only performs better than KNN.

In Fig. 8, three nonlinear models (i.e., BP-ANN, LS-SVM, and ELM) obtain relatively good performances (with equal means at the 5% significance level) while the dimensionality increases. In the cases of low

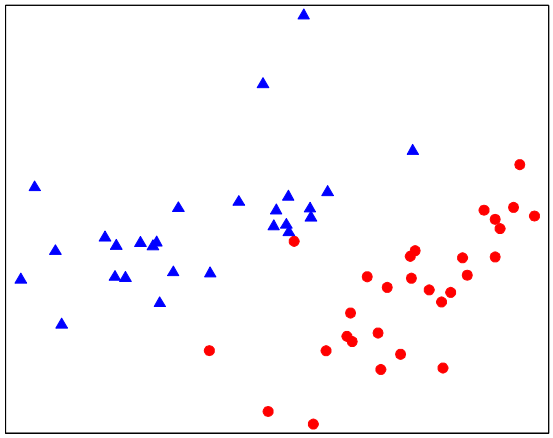

Fig. 4. 2D score plot of PC1 and PC2 derived from the Coffee dataset.

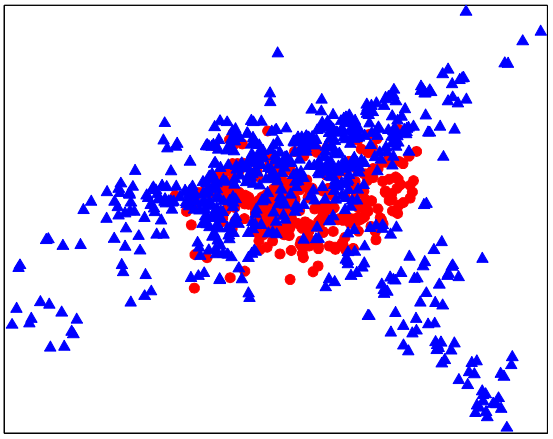

Fig. 5. 2D score plot of PC1 and PC2 derived from the Fruit dataset.

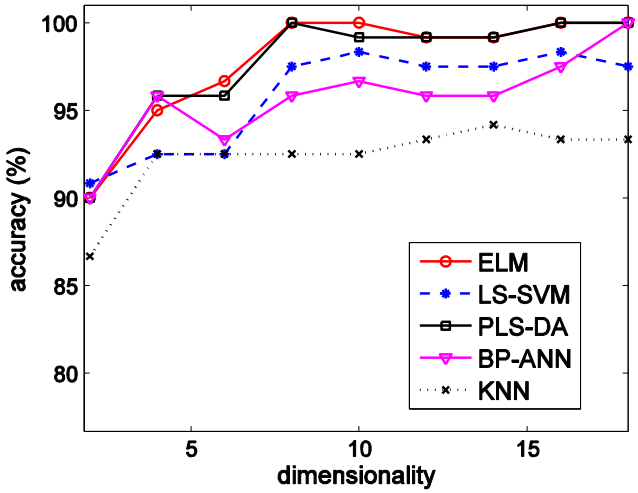

Fig. 6. Accuracy comparison of five models on the Meat dataset.

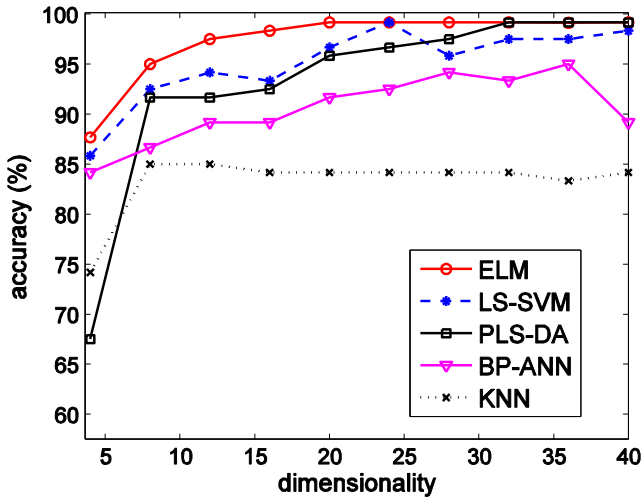

Fig. 7. Accuracy comparison of five models on the Oil dataset.

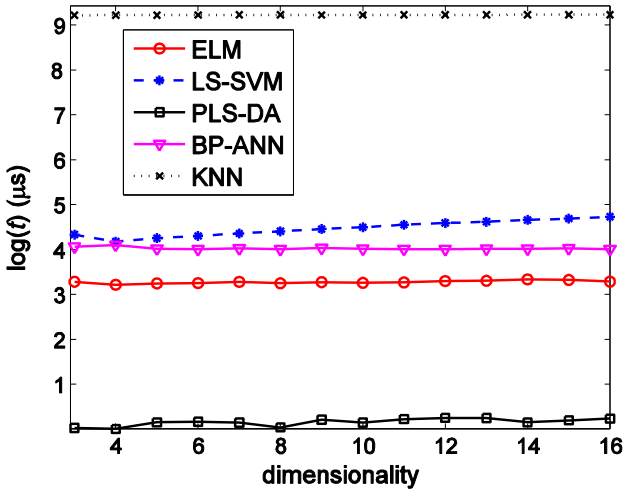

Fig. 9. Classification time comparison of five models on the Fruit dataset.

dimensionalities, LS-SVM performs better than others, and ELM achieves the same accuracy when the dimensionality is lager than 8. In contrast, the accuracy of PLS-DA and KNN is rather poor especially when the dimensionality is less than 6. The difference between ELM and PLS is also statistically significant.

In summary, ELM achieved a relatively high accuracy on all four datasets. It should be noted that its accuracy deviation of all dimensionalities is commonly large, since the accuracy of ELM is usually poor in a very low dimensionality situation, but it will reach a satisfactory performance and remain stable as the dimensionality increases (this can be observed in Figs. 6, 7 and 8). These results reveal that ELM may need more input variables to obtain more satisfactory and robust performance than its competitors.

Classification speed comparison on the Fruit dataset is also plotted in Fig. 9, in which the classification times of all models are presented. It needs to be clarified that the model parameters are optimized according to the accuracy for each dimensionality, so the parameters under different dimensionalities may be different. In other words, a model with a high dimensionality might calculate faster than the same model in a low dimensionality situation. In spite of this, the ELM models always run much faster than KNN, LS-SVM and BP-ANN but slower than PLS-DA on all dimensionalities. Similar situations also can be observed on other datasets.

Overall, considering all comparisons presented above, ELM actually achieves a good trade-off between accuracy and speed. Moreover, classification with ELM may be implemented via parallel computations because of its network structure, so it has more potential for real-time applications with a comparable accuracy.

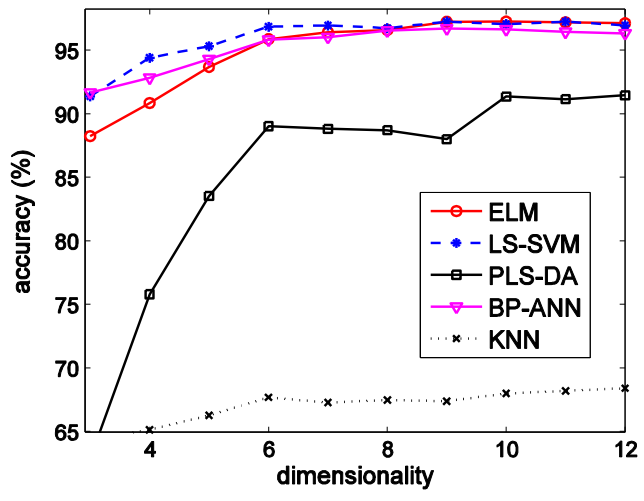

Fig. 8. Accuracy comparison of five models on the Fruit dataset.

5. Conclusions

In this work, the feasibility of the combination of spectroscopy and ELM for food classification was explored. Also the performance influence of different input dimensionalities and the selection of model parameters were investigated. The results suggested that "sigmoid" should be considered as a prior choice of the activation functions for this application. The experimental results on four datasets also revealed that the ELM model may need more input variables to obtain more satisfactory and robust performance than its competitors. Moreover, on the classification stage, the ELM model can be performed much faster than KNN, LS-SVM and BP-ANN on all dimensionalities. So it may be a promising method for spectroscopy-based real-time food classification with a comparable accuracy.

Conflict of interest

The authors declare that there are no conflicts of interest.

Acknowledgments

The authors are grateful for the financial support provided by the National High Technology Research and Development Program of China (No. 2012AA10A504), and the Natural Science Foundation of Zhejiang Province (No. LY14F020041).

References

[1] C. Tan, M. Li, X. Qin, Study of the feasibility of distinguishing cigarettes of different brands using an Adaboost algorithm and near-infrared spectroscopy, Anal. Bioanal. Chem. 389 (2007) 667–674.
[2] F. Shen, Y. Ying, B. Li, Y. Zheng, Q. Zhuge, Multivariate classification of rice wines according to ageing time and brand based on amino acid profiles, Food Chem. 129 (2011) 565–569.
[3] G. Downey, R. Briandet, R.H. Wilson, E.K. Kemsley, Near-and mid-infrared spectroscopies in food authentication: coffee varietal identification, J. Agric. Food Chem. 45 (1997) 4357–4361.
[4] H.S. Tapp, M. Defernez, E.K. Kemsley, FTIR spectroscopy and multivariate analysis can distinguish the geographic origin of extra virgin olive oils, J. Agric. Food Chem. 51 (2003) 6110–6115.
[5] J. Chen, M.A. Arnold, G.W. Small, Comparison of combination and first overtone spectral regions for near-infrared calibration models for glucose and other biomolecules in aqueous solutions, Anal. Chem. 76 (2004) 5405–5413.
[6] H. Cen, Y. He, Theory and application of near infrared reflectance spectroscopy in determination of food quality, Trends Food Sci. Technol. 18 (2007) 72–83.

[7] C. Tan, X. Qin, M. Li, Comparison of chemometric methods for brand classification of cigarettes by near-infrared spectroscopy, Vib. Spectrosc. 51 (2009) 276–282.
[8] M. Bevilacqua, R. Bucci, A.D. Magri, A.L. Magri, F. Marini, Tracing the origin of extra virgin olive oils by infrared spectroscopy and chemometrics: a case study, Anal. Chim. Acta. 717 (2012) 39–51.
[9] R. Vitale, M. Bevilacqua, R. Bucci, A.D. Magri, A.L. Magri, F. Marini, A rapid and non-invasive method for authenticating the by NIR spectroscopy and chemometrics, Chemometr. Intell. Lab. Syst. 121 (2013) 90–99.
[10] M. Gestal, M. Gómez-Carracedo, J. Andrade, J. Dorado, E. Fernández, D. Prada, A. Pazos, Classification of apple beverages using artificial neural networks with previous variable selection, Anal. Chim. Acta. 524 (2004) 225–234.
[11] A. Palacios-Morillo, A. Alcazar, F. De Pablos, J.M. Jurado, Differentiation of tea varieties using UV–Vis spectra and pattern recognition techniques, Spectrochim. Acta A 103 (2013) 79–83.
[12] J. Zhao, Q. Chen, X. Huang, C. Fang, Qualitative identification of tea categories by near infrared spectroscopy and support vector machine, J. Pharmaceut. Biomed. 41 (2006) 1198–1204.
[13] F.J. Acevedo, J. Jiménez, S. Maldonado, E. Domínguez, A. Narváez, Classification of wines produced in specific regions by UV–visible spectroscopy combined with support vector machines, J. Agric. Food Chem. 55 (2007) 6842–6849.
[14] L. Xie, Y. Ying, T. Ying, Classification of tomatoes with different genotypes by visible and short-wave near-infrared spectroscopy with least-squares support vector machines and other chemometrics, J. Food Eng. 94 (2009) 34–39.
[15] O. Al-Jowder, E. Kemsley, R. Wilson, Mid-infrared spectroscopy and authenticity problems in selected meats: a feasibility study, Food Chem. 59 (1997) 195–201.
[16] J.K. Holland, E.K. Kemsley, R.H. Wilson, Use of Fourier transform infrared spectroscopy and partial least squares regression for the detection of adulteration of strawberry purees, J. Sci. Food Agric. 76 (1998) 263–269.
[17] H.-Y. Zou, H.-L. Wu, H.-Y. Fu, L.-J. Tang, L. Xu, J.-F. Nie, R.-Q. Yu, Variable-weighted least-squares support vector machine for multivariate spectral analysis, Talanta 80 (2010) 1698–1701.
[18] V. Vapnik, The Nature of Statistical Learning Theory, Springer, 2000.
[19] N. Pochet, F. De Smet, J.A. Suykens, B.L. De Moor, Systematic benchmarking of microarray data classification: assessing the role of non-linearity and dimensionality reduction, Bioinformatics 20 (2004) 3185–3195.
[20] J.A. Suykens, J. Vandewalle, Least squares support vector machine classifiers, Neural. Process. Lett. 9 (1999) 293–300.
[21] U. Thissen, B. Üstün, W.J. Melssen, L.M. Buydens, Multivariate calibration with least-squares support vector machines, Anal. Chem. 76 (2004) 3099–3105.
[22] I.V. Kovalenko, G.R. Rippke, C.R. Hurburgh, Determination of amino acid composition of soybeans (Glycine max) by near-infrared spectroscopy, J. Agric. Food Chem. 54 (2006) 3485–3491.
[23] W.D. Ni, L. Norgaard, M. Morup, Non-linear calibration models for near infrared spectroscopy, Anal. Chim. Acta. 813 (2014) 1–14.
[24] F. Chauchard, R. Cogdill, S. Roussel, J. Roger, V. Bellon-Maurel, Application of LS-SVM to non-linear phenomena in NIR spectroscopy: development of a robust and portable sensor for acidity prediction in grapes, Chemometr. Intell. Lab. Syst. 71 (2004) 141–150.
[25] A. Borin, M.F. Ferrão, C. Mello, D.A. Maretto, R.J. Poppi, Least-squares support vector machines and near infrared spectroscopy for quantification of common adulterants in powdered milk, Anal. Chim. Acta. 579 (2006) 25–32.
[26] H. Yu, H. Lin, H. Xu, Y. Ying, B. Li, X. Pan, Prediction of enological parameters and discrimination of rice wine age using least-squares support vector machines and near infrared spectroscopy, J. Agric. Food Chem. 56 (2008) 307–313.
[27] G.-B. Huang, Q.-Y. Zhu, C.-K. Siew, Extreme learning machine: a new learning scheme of feedforward neural networks, IEEE International Joint Conference on, Neural Networks2004. 985–990.
[28] G.-B. Huang, D.H. Wang, Y. Lan, Extreme learning machines: a survey, Intl. J. Mach. Learn. Cybern. 2 (2011) 107–122.
[29] G.-B. Huang, Q.-Y. Zhu, C.-K. Siew, Extreme learning machine: theory and applications, Neurocomputing 70 (2006) 489–501.
[30] W. Zheng, Y. Qian, H. Lu, Text categorization based on regularization extreme learning machine, Neural Comput. Appl. 22 (2013) 447–456.
[31] R. Moreno, F. Corona, A. Lendasse, M. Grana, L.S. Galvao, Extreme learning machines for soybean classification in remote sensing hyperspectral images, Neurocomputing 128 (2014) 207–216.
[32] J.T. Peng, L.Q. Li, Y.Y. Tang, Combination of activation functions in extreme learning machines for multivariate calibration, Chemometr. Intell. Lab. Syst. 120 (2013) 53–58.
[33] Q.-J. Han, H.-L. Wu, C.-B. Cai, L. Xu, R.-Q. Yu, An ensemble of Monte Carlo uninformative variable elimination for wavelength selection, Anal. Chim. Acta. 612 (2008) 121–125.
[34] Q.-Y. Zhu, A.K. Qin, P.N. Suganthan, G.-B. Huang, Evolutionary extreme learning machine, Pattern Recogn. 38 (2005) 1759–1763.
[35] P.A. Gorry, General least-squares smoothing and differentiation by the convolution (Savitzky–Golay) method, Anal. Chem. 62 (1990) 570–573.