

Chinese Spirits Identification Model Based on Mid-Infrared Spectrum

Wu Zeng¹, Zhanxiong Huo^{1,*}, Yuxuan Xie², Yingxiang Jiang¹ and Kun Hu¹

Abstract: Applying computer technology to the field of food safety, and how to identify liquor quickly and accurately, is of vital importance and has become a research focus. In this paper, sparse principal component analysis (SPCA) was applied to seek sparse factors of the mid-infrared (MIR) spectra of five famous vintage year Chinese spirits. The results showed while meeting the maximum explained variance, 23 sparse principal components (PCs) were selected as features in a support vector machine (SVM) model, which obtained a 97% classification accuracy. By comparison principal component analysis (PCA) selected 10 PCs as features but only achieved an 83% classification accuracy. Although both approaches were better than a direct SVM approach based on the classification results (64% classification accuracy), they also demonstrated the importance of extracting sparse PCs, which captured most important information. The combination of computer technology SPCA and MIR provides a new and convenient method for liquor identification in food safety.

Keywords: Mid-infrared spectra, Chinese spirits, SPCA, SVM, liquor identification.

1 Introduction

Identifying Chinese spirits is of great importance in the field of food safety. The traditional methods of identifying the authenticity of liquor are mainly chemical analysis and sensory evaluation. But the complex and varied composition of liquor directly brings a significant identification cost, which is not desirable. In recent years, because of its fast and convenient characteristics, infrared (IR) techniques have been widely applied, such as in the identification of biological safety [Rocío, Diego, Raquel et al. (2018)], detection of the state of tree growth [Wu, Xu, Long et al. (2015)], in determining food safety issues [Callao and Ruisánchez (2018)], in ensuring the accuracy of medicine [Bunaciu, Aboul-Eusin and Fleschinet (2011)] and in evaluating the deterioration degree of ancient silk fabrics [Rocío, Diego, Raquel et al. (2018)]. The IR techniques have also become an efficient way for the identification of liquor.

It is known today that many studies on applying MIR to liquor were reported. Discriminant

¹ School of Electrical and Electronic Engineering, Wuhan Polytechnic University, Wuhan, 430000, China.

² Gina Cody School of Engineering and Computer Science, Concordia University, Montreal, H3G 1M8, Canada.

* Corresponding Author: Zhanxiong Huo. Email: huosonny@gmail.com.

Received: 13 February 2020; Accepted: 03 May 2020.

analysis (DA) together with raw, 1st and 2nd derivative spectra, was developed to predict a vintage year of bottled Chinese rice wine [Yu, Ying, Sun et al. (2007)]. The calibration result for raw spectra was the best, which was up to 97.1%. And discriminant partial least-squares (DPLS) were explored to predict Australian commercial Chinese spirits of different varietal origins [Cozzolino, Smyth and Gishen (2003)]. To the validation set, the DPLS models correctly classified 100% of Riesling and up to 96% of Chardonnay wines. Then artificial neural networks (ANN) combined with partial least-squares (PLS) was employed for the discrimination of varieties of yellow wines [Liu, He and Wang (2007)]. The compressed new variables were used as the ANN input, and finally, the discrimination ratio of 100% was achieved. Similarly, Wu et al. [Wu, He and Wang (2008)] extracted 20 independent components (ICs) by independent component analysis (ICA) to employ as the input of the back-propagation (BP) neural networks for identifying the varieties of red wines. The recognition rate was up to 100%. To discriminate rice wine age, Yu et al. [Yu, Lin, Hu et al. (2008)] investigated least-squares support vector machines (LS-SVM) combined with PCA and compared with DA. Based on the calibration and validation results, LS-SVM provided more accurate prediction results.

From the above introduction of accurate analysis of the spectrum of liquor, it is better to extract features as the first step. PCA belongs to the most popular method to extract features as well as reduce dimension. As is well documented in the literature, the concept of PCA is to construct a few PCs to replace the raw spectroscopic data by capturing a maximum amount of variance, which eliminates overlap spectroscopic data without losing the main information. However, the PCs derived from PCA are still constructed by all original variables, whereas some researchers believe it may be desirable to remove unnecessary variables. Zou et al. [Zou, Hastie and Tibshirani (2006)] presented SPCA with an elastic net. It seeks a trade-off between most explaining variance and less constructing variables [Luss, D'Aspremont and Tibshirani (2010)]. Thus, the extracted information, which promotes accuracy, will be preserved.

SPCA is usually used for the regression or interpretation of PCs. The approach certainly helps to address challenges in the detection liquor field, which applies the advantages of its sparse features to extract effective spectral variables and fully combine with classifiers to identify categories. Today, the common classifiers are DA, DPLS, ANN, SVM and so on. Although there is no specific classification method for liquor, and it is generally considered that SVM performs substantially superior to DPLS and ANN. Moreover, Jiang et al. [Jiang, Peng, Peng et al. (2010)] established flavor, grade, year models using SVM to detect classification performance. The results showed all of the models achieved high classification accuracy (98% to different flavor liquor, 92% to different grade liquor, and 100% to different year liquor). Therefore, MIR spectroscopy combined with SVM discrimination method was used for the rapid and simple classification of liquor in this study. Meanwhile, for classifying different tea categories, the classification accuracy of 90% was sufficient [Chen, Zhao, Fang et al. (2007)].

Thus, the advantages of this paper lie in integrating SPCA and SVM to recognize Chinese spirits. We employ the elastic net improved on least absolute shrinkage and selection operator (LASSO) to compute the sparse PCs. Then we take the sparse PCs as the input of SVM. Finally, we analyze and compare the difference between PCA and

SPCA. In essence, this study is divided into three sections: (1) introduction reference materials and methods; (2) demonstrating and analyzing the mixed method; (3) making a conclusion and prospects for the future.

2 Materials and methods

Two hundred and seventy-five bottles of Chinese spirits were derived from Baiyunbian winery of five vintage liquors in Hubei province. Tab. 1 shows the detail. They were stored in the same environment and recorded the time as the wine age. So, the flavor and alcohol of the vintages were same. Different batches from each vintage were chosen to make sure the representation of selected samples.

Table 1: The properties of selected liquor

Liquor vintage	Alcohol	Flavor	Spectrum number	
			Calibration set	Validation set
b12 (12 years)	42%	Nong jiang-flavor	30	20
b9 (9 years)	42%	Nong jiang-flavor	30	20
b6 (6 years)	42%	Nong jiang-flavor	30	20
b5 (5 years)	42%	Nong jiang-flavor	30	20
b3 (3 years)	53%	Nong jiang-flavor	30	20

Fourier transform infrared (FTIR) has high sensitivity and other characteristics by using infrared spectroscopy technology. It is widely used in various aspects of industry. Existing researches include wine product detection, honey detection, and chemical oxygen demand analysis of wastewater [Xie, Sun, Cai et al. (2019)].

Samples were scanned with a NEXUS 670 FTIR spectrophotometer (Nicollet, USA) in the reflectance mode. The spectrum used for the data analysis ranged from 4000 cm^{-1} to 6500 cm^{-1} resulting in 869 discrete variables. The liquor spectrum was obtained by subtracting the reference sample from the spectrum of an empty cell. Each liquor sample was sampled 16 times. The mean of 16 spectra samples was used in the following analyze step. Both temperature and humidity were kept steady. The 50 spectra selected randomly were divided into two parts. One part was used for training the model (calibration set), another was used to evaluate the performance of the model (validation set). And then thirty samples as a kind of calibration category were chosen from one type as will. Thus, 150 samples were used for the calibration set, and the remaining 100 samples were for the validation set. The basic theory of relevant methods is as follows.

2.1 PCA

PCA is used to extract information and solve the problem of high-dimensional data. It is one of the most widely used methods to eliminate abnormal samples, while saving calculation and storage costs. The method permits to construct new features, namely PCs, which can serve for visualization of the data.

Let X of size $n \times p$ denote the observed data matrix and standardize X without losing

generality. The PCA starts from computing the eigenvalues (λ) and eigenvectors (V) of the covariance matrix $X'X$. The corresponding loadings are the columns of V , which is according to the λ in descending order.

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p, \text{ and } V = [V_1, V_2, \dots, V_p] \quad (1)$$

Then the PCs are: $Z_{n \times k} = X_{n \times q} V_{q \times k}$

So, each PCs is the linear combination of data matrix X . The PCs can also be obtained by singular value decomposition (SVD) of X .

$$X = UDV' \quad (2)$$

where $Z = UD$ contains the PCs, and the columns of V are the loadings.

2.2 SPCA

As an improved method, SPCA focuses on the main parts of the principal components in order to make it easier to explain the principal components. The key and original intention of putting forward SPCA is to investigate the necessary of obtaining PCs with all variables. Based on PCA, Zou et al. [Zou, Hastie and Tibshirani (2006)] proposed a regression approach to achieve sparse loadings. He applied elastic net mixed l_1 and l_2 norm to constrain coefficients. That is to say while making full use of l_1 norm to take some turn to zero, it increases l_2 norm to shrinkage regression coefficients and reduces error. This combination reduces the influence of excessive compression in l_1 norm.

After all, it is desired to obtain sparse loading, it only needs to restrict loadings in PCA under ensuring requirements. To address the concern, the objective function can be defined as follows at this point.

$$\hat{\beta} = \arg \min_{\beta} (\|Z_i - X\beta\|^2 + \lambda \|\beta\|^2 + \lambda_1 \|\beta\|_1) \quad (3)$$

where β is the regression coefficient. Then according to Luss et al. [Luss, D'Aspremont and Tibshirani (2010)], let $A = [\alpha_1, \alpha_2, \dots, \alpha_k]$, $B = [\beta_1, \beta_2, \dots, \beta_k]$. Then for any $\lambda > 0$, Eq. (3) can turn to Eq. (4).

$$(\hat{A}, \hat{B}) = \arg \min_{A, B} \sum_{i=1}^n \|x_i - AB'x_i\|^2 + \lambda \sum_{j=1}^k \|\beta_j\|^2 + \sum_{j=1}^k \lambda_{i,j} \|\beta_j\|_1 \quad (4)$$

$$s.t. A'A = I_{k \times k}$$

where k is the desired number of PCs.

The goal in an elastic net is to produce a sparse feature space with the l_1 penalty, but improve stability and retain correlated features with the l_2 penalty [Kelly, Degenhart, Siewiorek et al. (2012)]. This is not a computationally simple problem, but efficient methods for solving it have been developed. There is also an additional free parameter in

α , which determines the relative strength of the penalties. Previous studies have shown this technique is effective in the classification of functional magnetic resonance imaging (fMRI) data [Carroll, Cecchi, Rish et al. (2009); Ryali, Supekar, Abrams et al. (2010)].

To solve the Eq. (4), Zou et al. [Zou, Hastie and Tibshirani (2006)] adopted an alternating algorithm to obtain the minimum value. That is to say, there are two steps to estimate the sparse loading:

Step 1. Given A a fixed value and then to solve B .

Step 2. Conversely, given B , then to solve A .

The initial estimate of A can be gained directly by PCA, namely, A equals V in Eq. (2). For each j , let desired $Y_j^* = Xa_j$, then we analyze and get $\hat{B} = [\hat{\beta}_1 : \hat{\beta}_k]$, where each $\hat{\beta}_j$ is an elastic net estimate:

$$\hat{\beta}_j = \arg \min_{\beta_j} \|Y_j^* - X\beta_j\|^2 + \lambda \|\beta_j\|^2 + \lambda_{1,j} \|\beta_j\|_1 \quad (5)$$

Then the second step, we fixed B and solve A . At that time, the problem is simpler than the first one, because we can ignore the penalty part and only to solve.

$$(\hat{A}, \hat{B}) = \arg \min_{A, B} \sum_{i=1}^n \|x_i - AB'x_i\|^2 \quad (6)$$

$$s.t. A'A = I_{k \times k}$$

By computing the SVD of $(X'X)B$, we can gain the solution of \hat{A} .

$$(X'X)B = UDV' \quad (7)$$

$$\hat{A} = UV' \quad (8)$$

Thus, the detailed algorithm is as follows.

Step 1: Let A start at V .

Step 2: Solve the problem: $\beta_j = \arg \min_{\beta} (a_j - \beta)' X'X(a_j - \beta) + \lambda \|\beta\|^2 + \lambda_{1,j} \|\beta\|_1$,

where given the value for $A = [a_1 : a_k]$, $j = 1 : k$.

Step 3: For a fixed $B = [\beta_1 : \beta_k]$, compute the SVD of $X'XB = UDV'$, then update $A = UV'$.

Step 4: Repeat Steps 2 and 3, until convergence.

Step 5: Normalization: $\hat{V}_j = \frac{\beta_j}{\|\beta_j\|}$, $j = 1 : k$.

2.3 SVM

For a sample set, there exists an optimal separating surface, or hyperplane. SVM attempts to find the optimal hyperplane in high-dimensional space by maximizing the separation distance between classifications. The primary advantage of SVM over the traditional learning algorithm is that the solution of SVM is always globally optimal and avoids local minima and over-fitting in the training process. This method is also used in biomedical imaging, using SVM under a large amount of data can quickly and accurately obtain prediction results [Yuan, Yao and Tan (2018)]. For further details on SVM, we combined with a concrete liquor example to illustrate.

Assuming the training data with l number of samples is represented by $\{x_i, y_i\}$, $i = 1:l$, where $x \in R^n$ shows each sample has n -dimension and y is the class label. Under the linearly separable or approximately to training samples, we can find an optimal hyperplane, which meets the maximal distance to the closest point, to divide the data. The concrete representation is:

$$y_i [(w \times x_i) + b] \geq 1 - \xi_i, \quad i = 1:l \quad (9)$$

ξ_i is a non-negative slack variable. At this time, the optimal problem is to minimize the function of Eq. (10) under constrains Eq. (9).

$$\phi(w, \xi) = \frac{1}{2}(w \times w) + C \sum_{i=1}^l \xi_i \quad (10)$$

The positive C is a penalty factor, which controls the number of misclassified points. By introducing Lagrange multipliers, Eq. (10) is transformed to be Eq. (11).

$$\max W(a) = \sum_{i=1}^l a_i - \frac{1}{2} \sum_{i,j=1}^l y_i y_j a_i a_j (x_i \times x_j) \quad (11)$$

Under constrain $\sum_{i=1}^l a_i y_i = 0$, $0 \leq a_i \leq c$.

Thus, the optimal hyperplane is Eq. (12).

$$f(x) = \text{sign} \left(\sum_{i=1}^l y_i a_i^* (x_i \times x) + b^* \right) \quad (12)$$

The a^* is the optimal solution of Eq. (11) and b^* is the bias.

Facing to the non-linearly problems, SVM maps input data into a high dimensional feature space through nonlinear transformation and achieves classification by constructing linear discrimination function in high dimensional space. It is convenient to compute scale product by the kernel function $K(x_i, x_j)$. Eq. (13) turns to be the optimal problem, whose corresponding optimal hyperplane is Eq. (14).

$$\max W(a) = \sum_{i=1}^l a_i - \frac{1}{2} \sum_{i,j=1}^l y_i y_j a_i a_j K(x_i, x_j) \quad (13)$$

$$f(x) = \text{sign} \left(\sum_{i=1}^l y_i a_i^* K(x_i, x_j) + b^* \right) \quad (14)$$

2.4 The whole idea

The paper desires to extract features of liquor magnetic resonance imaging (MRI) spectroscopy with SPCA, and then coupled with SVM classifier to realize the perfect classification of Chinese spirits. The main involved idea is displayed as follows.

In the process of collecting spectra, the spectral data would be affected to some degree due to various reasons, whether man-made or the spectrophotometer itself. Because it is important that the PCs accurately represent the true spectral information, abnormal sample points were removed and not used in the modeling procedure. While introducing PCA, there are instructions that it can reflect the effect of clustering. Therefore, the points deviated far in the score plot can be classified as outliers. However, in allusion to eliminate the scattering effect on the spectrum, z-Score standardization is adopted. Namely, each raw spectrum datapoint is subtracted by the sample population mean and then divided by the sample population variance to obtain a standardized value.

New data obtained after the pre-treatment can be carried to extract features. The key is to calculate the loading, and then the PCs can directly be the product between loading and new data. For calculations we have ensured that the explained variance is large enough, thus guaranteeing minimal information lose. The PCs derived from PCA are uncorrelated and their loadings are orthogonal. The total variance is calculated directly by the eigenvalues. However, in general, PCs derived from SPCA are explicitly correlated. So, calculating $\text{tr}(Z'Z)$ as the total variance in SPCA is not reasonable when Z is correlated.

Through quadratic programming (QP) decomposition to the residual after projection on Z , the square sum of obtained R is the desired explained total variance in SPCA. While meeting the explained variance, the classification accuracy is the ultimate requirement to determine the optimal number of PCs.

Subsequently the dominant PCs were selected as features for the SVM classifiers, and the optimal classification model was established after adjusting parameters. Thus, we can predict the sample by using this model. However, all of the data processing is aiming at a calibration set. How to handle any unknown sample in a validation set? Certainly, the validation set must take the same process to a calibration set. But it is worth paying more attention that it does not allow to process them simultaneously. The perfect measure to gain PCs of validation set is based on mean, variance and loadings of a calibration set. Then extracted PCs directly into the model can get the predicted value of a validation set.

3 Results and discussion

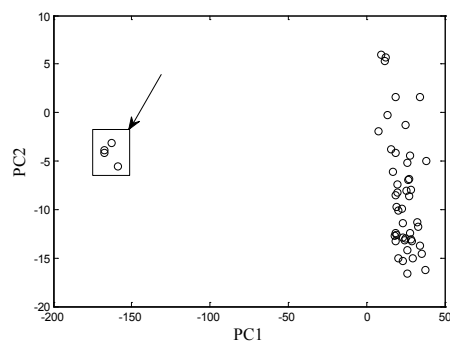
3.1 Software

All algorithms were implemented with MATLAB (R2019a), and the implementation of a

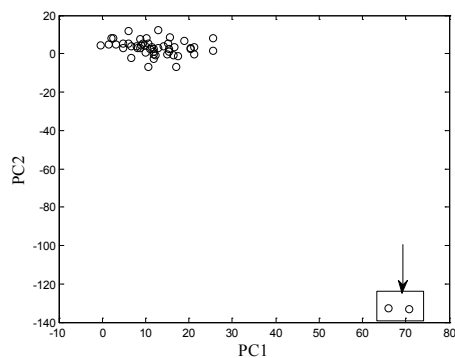
multi-class SVM algorithm, which was fitted using the functions provided in the libsvm-mat-2.89-3 toolbox, was used for the classification of liquor in all experiments. For the spectral acquisition, OMNIC 6.0a was used.

3.2 Eliminate abnormal and spectra investigation

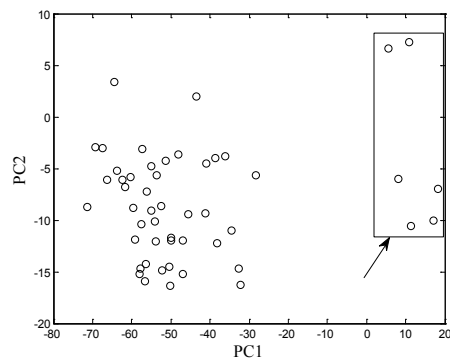
Fig. 1 demonstrates the pretreatment of spectrum data for Principal Component 1 (PC1) and Principal Component 2 (PC2). The rectangle shows the deviated further point, which defined as abnormal points. After eliminating outliers, we can draw the spectra to analyze the feature of liquor.



(a) The abnormal data of 'b12'



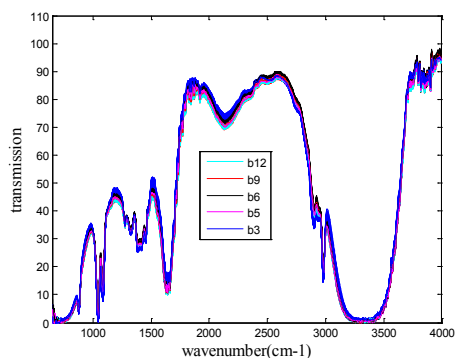
(b) The abnormal data of 'b9'



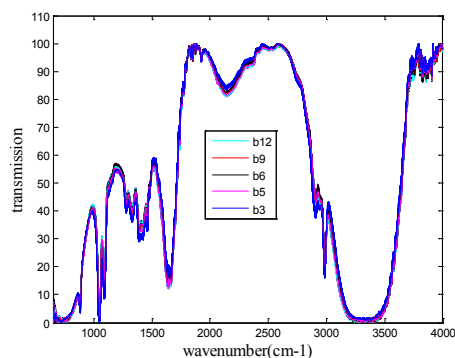
(c) The abnormal data of 'b3'

Figure 1: Finding the abnormal data of samples according to the score plot

As seen in Fig. 2(a), the spectral differences between the 'b3' with 53 degrees and other vintage liquors can be observed from 2780 cm^{-1} to 3010 cm^{-1} , which is likely related to C-H stretch overtones of ethanol. However, through the observation and analysis, we find that the liquor spectrum is similar due to the various material, uncertainty component concentration and inconspicuous difference, especially the large amount of water in liquor causing peaks in saturation state. Thus, it is difficult to recognize and distinguish directly. So, the chemometrics method is needed to help establish the model to achieve the purpose of the classification and identification of different vintage liquor. The baseline was automatically adjusted by the software OMNIC. Fig. 2(b) illustrates the corrected spectrum.



(a) The transmission spectra



(b) The corrected transmission spectral

Figure 2: The spectral of each class (color figure available online)

3.3 PCA and SPCA

High-dimensional standardize data can be visualized through PCA or SPCA. The score plot using the first two PCs can indicate the data trend in two-dimension space. Both Fig. 3 depict the score plot of PCA and SPCA, which reflect cluster relation to some degree. Comparing them, we can discover that there are divided into two categories. But this is not a desired result, and the advantages of SPCA and PCA are not well demonstrated. So,

separating the four vintage liquors with the same ethanol concentration requires SVM.

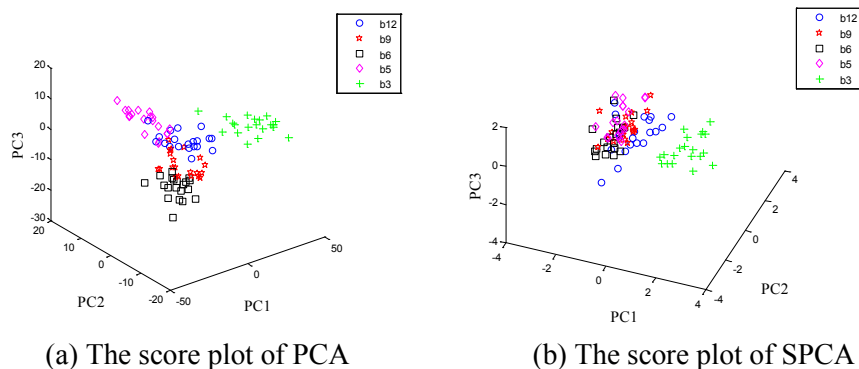


Figure 3: The score plots of PCA and SPCA

3.4 SVM

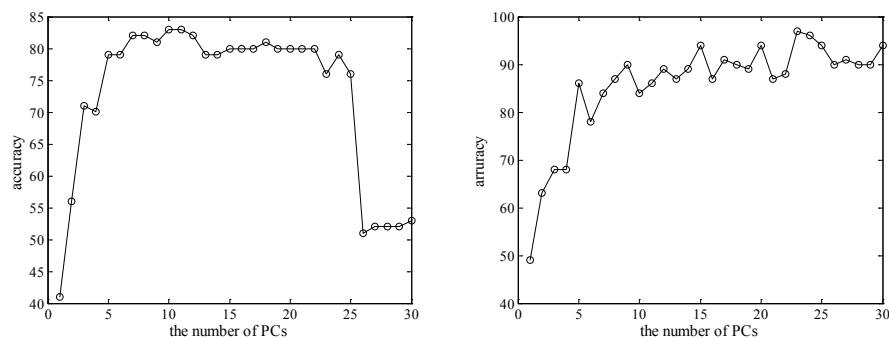
To obtain acceptable performance, two important parameters in SVM have to be chosen carefully. The regularization parameter c , which determines the trade-off between minimizing the training error and minimizing model complexity, and the parameter g of the kernel function that implicitly defines the non-linear mapping from input space to some high-dimensional feature space.

As long as some necessary conditions are met (Mercer conditions) [Jiang, Peng, Peng et al. (2010)], any one of many available kernel functions can be used, such as linear, Gaussian radial basis function (RBF), polynomial, sigmoid, etc. Although a new algorithm for online incremental decrement learning based on SVM [Chen, Xiong, Xu et al. (2019)] and an improved method for Support Vector Domain Description (SVDD) [El Boujnouni and Jedra (2018)] have been proposed, RBF is considered to be suitable for implementing nonlinear classification tasks, and we focus on RBF in this paper. As explained by Han et al. [Han, Cao and Han (2012)], the parameters of RBF kernel function have important effect for SVM classification performance, it is very necessary to select appropriate parameters.

As we know, too much or too little PCs will have a bad effect. Aiming to choose the optimal number of PCs, we must rely on variance contribution on request. The result tells us the top thirty latent variables contain more than 96% of the total variance. Then under the case, the ultimate accuracy becomes the most key factor. Fig. 4 demonstrates the link to the number of PCs and classification accuracy.

From Fig. 4(a), when only choosing the top 10 PCs derived from PCA are used, the final accuracy is the highest, up to 83%. Simultaneously, these 10 PCs contain 85.56% of the total variance, which fully meets the requirement. Similarly, from Fig. 4(b), the top 23 PCs can achieve up to 97% classification accuracy. Although the total percentage of explained variance is only 69%, the graphically displayed information of the data is practically the same as that obtained from the first 23 original PCs, which reconstruct together 90.02% of the total variance. Thus, the top 10 and 23 PCs are extracted respectively by PCA and SPCA as the

input of SVM classifiers. Since the correct selection of both parameters is critical to SVM predictive performance, we calculate c and g using a way of ‘heuristic search’, where the parameter combination with the best k-fold cross-validation accuracy is selected.



(a) Classification accuracy of PCA

(b) Classification accuracy of SPCA

Figure 4: The link of the number of PCs and classification accuracy

Figs. 5-7 show the results of the calibration and validation set classification for three models, including the SVM with raw, PCA and SPCA. The prediction results of the above three models are revealed in Tab. 2. For the calibration set, the classification accuracy is nearly 100%, but it is different from the validation set. Obviously, for the validation set, the SPCA+SVM model correctly classified 97%, which is higher than the 83% of PCA+SVM model and 64% for SVM only. In terms of numerical comparison, it is clear that the PCs extracted by SPCA is effective, and provides a new way for liquor identification.

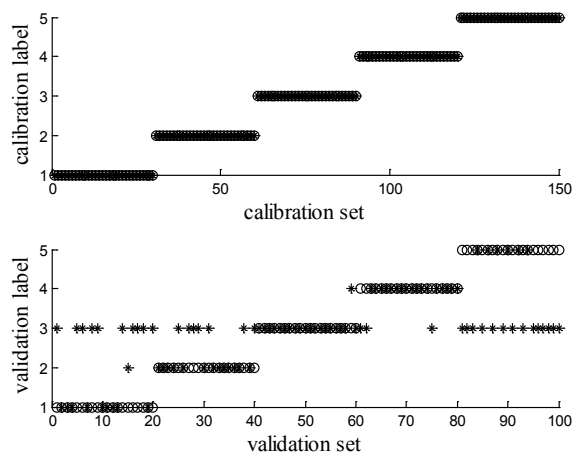


Figure 5: The result of direct SVM

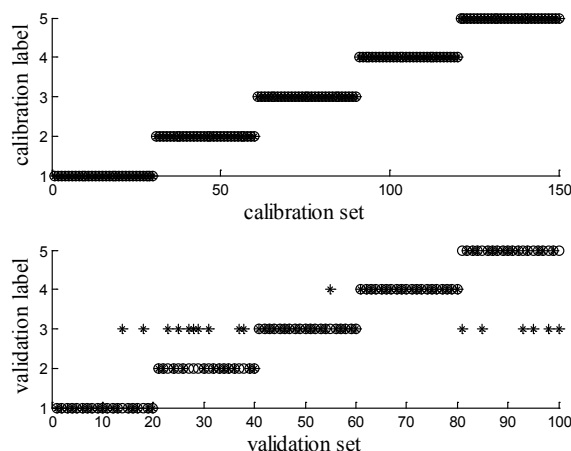


Figure 6: The result of PCA+SVM

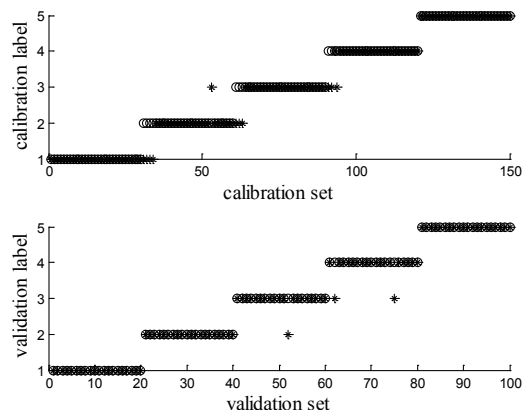


Figure 7: The result of SPCA+SVM

(The circles represent the actual label, and stars are the predicted label.)

Now we integrate the properties of liquor (see Tab. 1) to illustrate in detail. As we know, the alcohol of 'b3' is different from others. SPCA can easily extract features to identify different alcohol. However, we observe that all misclassified samples in the validation set are 5 years and 6 years aged samples. Although the four vintage liquors still mix in the score plots, those can be effectively classified using SVM. We investigated and concluded that the large amount of water in liquor may be the reason for the peak saturation state.

From the accurate numeral of classification, the best accuracy is only 97%. However, many methods referred in Section 1 can acquire 100% identification precision. Considering the potential reason, we may not have chosen the appropriate penalty coefficient of SPCA. It follows that SVM integrated SPCA is not effective enough. Moreover, the hybrid method is related to the collected data, which derived from

different collecting time.

Table 2: The results for the calibration and validation set using hybrid SVM model

Data set	Method	Corresponding label	Misclassified samples		
			SVM	PCA+SVM	SPCA+SVM
b1	Calibration set	1	0	0	0
	Validation set		11	2	0
b9	Calibration set	2	0	0	5
	Validation set		7	8	0
b6	Calibration set	3	0	0	3
	Validation set		1	1	1
b5	Calibration set	4	0	0	3
	Validation set		3	0	2
b3	Calibration set	5	0	0	0
	Validation set		14	6	0
Validation accuracy			64%	83%	97%

4 Conclusion

This paper describes a new method to identify quality of Chinese spirits with infrared spectroscopy. By focusing on the redundancy of the liquor spectrum, we extracted the most refined PCs as the typical representative of spectral information. Then a SVM classifier is integrated to build a classification model to improve the prediction accuracy. Since SPCA is an improvement of PCA, PCs derived from SPCA is sparse by comparison with PCA. Therefore, it can eliminate redundancy and increase classification accuracy effectively. By analyzing the numerical results, we conclude that the infrared spectroscopy based on SVM and SPCA is effective at identifying the liquor category rapidly and accurately without destroying liquor samples.

Although the new binding mode of SPCA+SVM cannot implement the correct result entirely, the SPCA regression approach still shows the superiority and prospects in feature extraction. However, the value of penalty coefficients relies on experience, which is not based on current data. Thus, our next plan is to achieve an adaptive parameter selection and strive to reach 100% of classification accuracy.

Acknowledgement: The authors gratefully acknowledge the laboratory of electronic and information engineering, Huazhong University of Science and Technology (HUST), as well as Fourier transform infrared spectrometer provided by analysis and testing center in Wuhan Polytechnic University, Hubei Province, China. The authors thank Dr. Rong Wang, Dr. Dan and Prof. Jie Li for their helpful advice.

Funding Statement: This work was financially supported by the National Nature Science Foundation of China (Grant Number: 61962010).

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- Bunaciu, A. A.; Aboul-Euein, H. Y.; Fleschin, S.** (2011): Recent applications of Fourier transform infrared spectrophotometry in herbal medicine analysis. *Applied Spectroscopy Reviews*, vol. 46, no. 4, pp. 251-260.
- Callao, M. P.; Ruisánchez, I.** (2018): An overview of multivariate qualitative methods for food fraud detection. *Food Control*, vol. 86, pp. 283-293.
- Carroll, M. K.; Cecchi, G. A.; Rish, I.; Garg, R.; Raob, A. R.** (2009): Prediction and interpretation of distributed neural activity with sparse models. *NeuroImage*, vol. 44, no. 1, pp. 112-122.
- Chen, Q. S.; Zhao, J. W.; Fang, C. H.; Wang, D. M.** (2007): Feasibility study on identification of green, black and oolong teas using near-infrared reflectance spectroscopy based on support vector machine (SVM). *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 66, no. 3, pp. 568-574.
- Chen, Y. T.; Xiong, J.; Xu, W. H.; Zuo, J. W.** (2019): A novel online incremental and decremental learning algorithm based on variable support vector machine. *Cluster Computing*, vol. 22, no. 3, pp. 7435-7445.
- Cozzolino, D.; Smyth, H. E.; Gishen, M.** (2003): Feasibility study on the use of visible and near-infrared spectroscopy together with chemometrics to discriminate between commercial Chinese spirits of different varietal origins. *Journal of Agricultural and Food Chemistry*, vol. 51, no. 26, pp. 7703-7708.
- El Boujnouni, M.; Jedra, M.** (2018): Improving support vector domain description by maximizing the distance between negative examples and the minimal sphere center's. *Computer Systems Science and Engineering*, vol. 33, no. 6, pp. 409-420.
- Han, S. J.; Cao, Q. B.; Han, M.** (2012): Parameter selection in SVM with RBF kernel function. *World Automation Congress*, pp. 1-4.
- Jiang, A.; Peng, J. T.; Peng, S. L.; Wei, J. P.; Li, C. W.** (2010): Research on SVM-based liquor infrared spectrum analysis method. *Computers and Applied Chemistry*, vol. 27, no. 2, pp. 233-236.
- Kelly, J. W.; Degenhart, A. D.; Siewiorek, D. P.; Smailagic, A.; Wang, W.** (2012): Sparse linear regression with elastic net regularization for brain-computer interfaces. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 4275-4278.
- Liu, F.; He, Y.; Wang, L.** (2007): Discrimination of varieties of yellow wines by using Vis/NIR spectroscopy and PLS-BP model. *IEEE International Conference on Control and Automation*, pp. 1492-1495.
- Luss, R.; D'Aspremont, A.; Tibshirani, R.** (2010): Clustering and feature selection using sparse principal component analysis. *Optimization and Engineering*, vol. 11, no. 1, pp. 145-157.
- Rocío, R. R.; Diego, L. G.; Raquel, M. C.; José, M. A.** (2018): NIR spectroscopy and

chemometrics for the typification of Spanish wine vinegars with a protected designation of origin. *Food Control*, vol. 89, pp. 108-116.

Ryali, S.; Supekar, K.; Abrams, D. A.; Menon, V. (2010): Sparse logistic regression for whole-brain classification of fMRI data. *NeuroImage*, vol. 55, no. 2, pp. 752-764.

Wu, G.; He, Y.; Wang, Y. (2008): Discrimination of varieties of red wines based on independent component analysis and BP neural network. *Congress on Image and Signal Processing*, pp. 272-276.

Wu, Z. Z.; Xu, E. B.; Long, J.; Zhang, Y. J.; Wang, F. et al. (2015): Monitoring of fermentation process parameters of Chinese rice wine using attenuated total reflectance mid-infrared spectroscopy. *Food Control*, vol. 50, pp. 405-412.

Xie, J.; Sun, D.; Cai, J.; Cai, F. (2019): Waveband selection with equivalent prediction performance for FTIR/ATR spectroscopic analysis of COD in sugar refinery waste water. *Computers, Materials & Continua*, vol. 59, no. 2, pp. 687-695.

Yu, H. Y.; Ying, B.; Sun, T.; Niu, X. Y.; Pan, X. X. (2007): Vintage year determination of bottled Chinese rice wine by VIS-NIR spectroscopy. *Journal of Food Science*, vol. 72, no. 3, pp. 125-129.

Yu, H.; Lin, H.; Xu, H.; Ying, Y. B.; Li, B. B. et al. (2008): Prediction of enological parameters and discrimination of rice wine age using least-squares support vector machines and near infrared spectroscopy. *Journal of Agricultural and Food Chemistry*, vol. 56, no. 2, pp. 307-313.

Yuan, L. F.; Yao, E. L.; Tan, G. M. (2018): Automated and precise event detection method for big data in biomedical imaging with support vector machine. *Computer Systems Science and Engineering*, vol. 33, no. 2, pp. 105-114.

Zou, H.; Hastie, T.; Tibshirani, R. (2006): Sparse principal component analysis. *Journal of Computational and Graphical Statistics*, vol. 15, no. 2, pp. 265-286.