Fast Discrimination of Mature Vinegar Varieties with Visible_NIR Spectroscopy

Yanru Zhao, Shujuan Zhang, Huamin Zhao, Haihong Zhang, Zhipeng Liu

Fast Discrimination of Mature Vinegar Varieties with Visible_NIR Spectroscopy

Yanru Zhao, Shujuan Zhang * , Huamin Zhao ,
Haihong Zhang, Zhipeng Liu

College of Engineering , Shanxi Agricultural University，Taigu 030801, Shanxi, China
xiaoru19863804@sina.com

Abstract. In order to achieve non-destructive of mature vinegar varieties, a fast discrimination method was put forward based on Visible_near infrared reflectance (NIR) spectroscopy. The FieldSpec3 spectrometer was used for collecting 20 sample spectra data of the three kinds of mature vinegar separately. Then principal component analysis (PCA) was used to process the spectral data after pretreatment using average smoothing method and multiplicative scatter correction (MSC) method, and principal components(PCs) were selected based on accumulative reliabilities. A total of 60 mature vinegar samples were divided into calibration sets and validation sets randomly, the calibration sets had 45 samples and the validation sets had 15 samples. The stepwise discriminant analysis was trained with five PCs in calibration sets as the inputs,and mature vinegar varieties as the outputs. The stepwise discriminant analysis model was built for discrimination of mature vinegar variety ,and the model contains 15 samples in the validation sets. The result showed that a 100% recognition ration was achieved.The BP-ANN model for discrimination of mature vinegar varieties were built based on PCA and the stepwise discriminant analysis , then the model was tested with the 15 sample in the validation sets. The result showed that a 100% recognition ration was achieved with the threshold predictive error ±0.027. It based on five principal components had a higher prediction accuracy and efficiency more than the BP neural network model. It could be concluded that PCA combined with stepwise discriminant analysis and BP-ANN was an available method for varieties recognition of mature vinegar based on NIR spectroscopy.

Keywords: Vis_NIR spectroscopy, Mature vinegar variety, Principal component analysis, Stepwise discriminant analysis, BP neural network

1 Introduction

Vinegar is a kind of lives condiment which is related to people's daily life closely, it not only has an important food value, but also has the health care and the effectiveness of sterilization. In order to ensure the quality of vinegar, to protect the regular brands, and to avoid fish in shoddy, undoubtedly, it is important and essential

to develop a simple, rapid and non-destructive technology for species identification and quality of vinegar detection technology and research.

Near-infrared diffuse reflectance spectroscopy can make full use of all-band spectral data underway the qualitative and quantitative analysis. With many advantages of the spectra , it has been widely used in identification of quality of agricultural products and species [1-11]. Currently, domestic scholars have been studied on the sauce brand [1], bayberry juice varieties[2] and tea varieties[3] with near infrared spectroscopy ,but no identification was on the varieties of vinegar.

Principal Component Analysis (PCA) is a Data Mining technology in Multivariate Statistical. It is based on without losing the main information, and the more original variables were replaced by new variables. In the spectral data analysis, the difficulty that was caused by overlapping bands has been solved. It has became one of the widest applied spectrum mathematical methods.

Discriminant analysis is a widely used multivariate statistical method. It was derived according to the characteristics of things that the variables values and the class they belonged the classification of objects of unknown classification of an analysis method. Stepwise discriminant analysis with the algorithm was developed based on discriminant analysis, strong distinguish ability of the variables was pulled into discriminant, while weak capacity of it was removed. Thus the aim was achieved for vector dimension reduction and classification of unknown things.

BP is neural network based on back propagation algorithm, which has highly non-liner mapping capability . The problem about the complex non-liner pattern can be solved easily with it.

In this paper, principal component analysis, stepwise discriminant analysis and the BP neural network method were combined to establish the different species of vinegar - Near Infrared Spectroscopy models, to achieve the identification of different varieties mature vinegar.

2 Materials and methods

The equipment was composed by computer, spectrometer, halogen light, correction whiteboard etc. FieldSpec3 spectrometer which made by America ASD (Analytical Spectral Device) company was used. The interval of spectral sampling was 1nm, the range of sampling was 350 ~ 2500nm, the times of scanning were 30, resolution was 3.5nm and the probe field of view angle was 25 °, diffuse reflection was selected to carry on the sample spectrum sampling.The halogen light of 14.5V is used to match with the spectrometer. The spectrum data was exported into the form of ASCII code for processing and the analysis software was ASD View Spec Pro V5.0, Unscramble V9.7 and DPS (Data Procession System For Practical Statistics).

“DongHu Vinegar”, “DoubleToo vinegar” and “PurpleForest vinegar”were bought from market.Each species of vinegar which were produced in the same batch by the same manufacturer were collected 20 samples, and the total was 60. All samples were randomly divided into the model sets including 45 samples and the prediction sets including 15 samples.The vinegar was set into a petri dish which the attitude was 1.4cm ,the diameter was 6.5cm and the liquid height was 10mm, then spectrometer

probe was put above the vinegar surface, the distance between both was 50 mm, the angle between both was 45 °, light from the sample container containing center was 350mm and into a 45 °, each sample was scanned and averaged.

The noise was eliminated by spectral pretreatment to removed the high-frequency random noise, baseline drift etc. Move average smoothing method was applied in this paper, smooth points was 9 , then MSC (Multiplicative Scatter Correction) processing [3]was made.

3 Spectral analysis of vinegar data

3.1 Analysis of spectral map

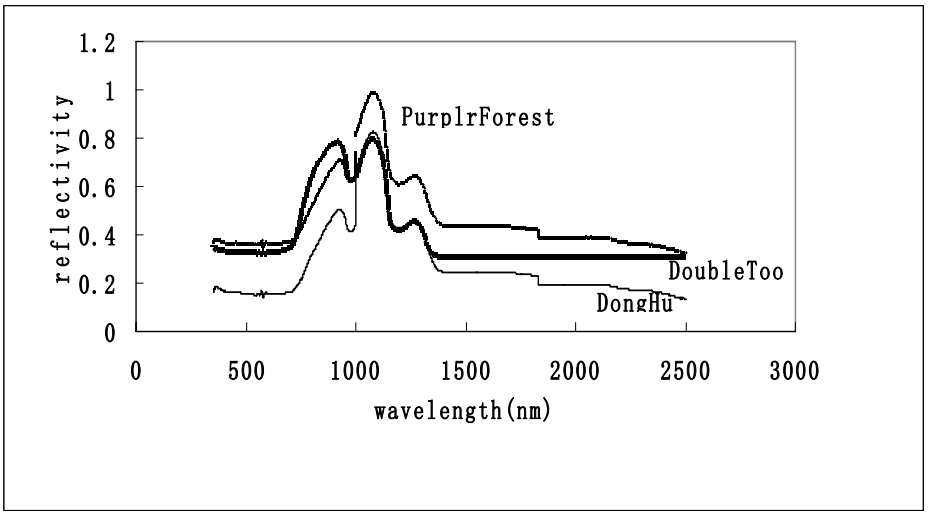

Fig.1. Visible_Near infrared reflectance spectroscopy of three different varieties of Vinegar

The typical near-infrared spectral curves of vinegar were showed in Figure 1, abscissa was the wavelength and the range of it was 350 ~ 2500nm, vertical coordinates was the spectral reflectance. It could be expressed in figure 1 that different varieties of vinegar had different spectrums; they had certain characteristics and fingerprint. Spectral reflectance data of samples were obtained, averaged and converted into ASCII code to export by ASD View Spec Pro software, then were advanced principal component analysis through Unscramble V9.7 software.

3.2 The main component analysis of vinegar

The purposes of PCA were data reduction and eliminate the section of each overlapping that co-existed in lots of information. A large number of original

variables spectrum were converted, small number of new combinations were composed as linear combinations of original variables, structural characteristics of the original data variables could be characterized by new variables maximum, as little as possible losing the information [1,2]. The 60 samples of three kinds of vinegar were analyzed by principal component analysis through Unscramble V9.7 software, see Figure 2.

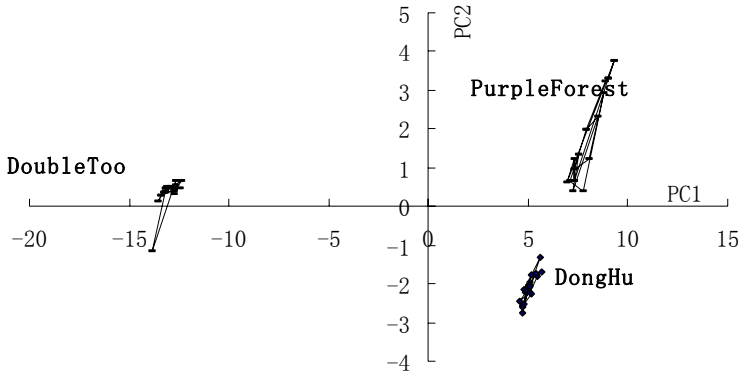

Fig.2. PCA scores plots (PC1× PC2) for Sea buckthorn juice sample across the entire spectral region

Principal components 1, 2 score image of 60 model samples was showed in Figure 2, where the horizontal axis represents each sample of the first principal component scores, the vertical axis expresses each sample value of the second principal component scores. “DongHu Vinegar”,”DoubleToo Vinegar” and ”PurpleForest vinegar” were divided into three categories, it indicated the main components 1, 2 had better clustering effect. Figure 2 showed the degree of polymerization of 20 samples of Donghu vinegar was best, distributed in the fourth quadrants and aggregate near the X axis, the degree of polymerization of 20 samples of Double Too vinegar was better, distributed in the second quadrants and aggregate near the X axis. Degree of polymerization of PurpleForest vinegar was better too, parted in the first quadrant.These images were distinguished easily. Analysis showed that the former two principal components of the 3 vinegar had better clustering effect.

3.3 The main component analysis of vinegar

3.3.1 Important components extraction based on Principal Component Analysis

There are 2150 points in sample spectral bands which from 350 ~ 2500nm, when the whole spectrum amount was calculated, the calculation was large, spectral information of samples was very weak in some regions, lacked the relationship between the composition of the samples and the nature of samples. So through PCA analysis, sensitive species of the new variable vinegar were selected as the input of stepwise discriminant analysis of model identification.

The total credibility of 7 principal component which was obtained by principal component analysis could be seen in figure 1.As the first 5 principal components of the total confidence level is 99.99%,the total credibility had little change when the principal components were added, the five principal components was chosen to explain the original wavelength variable, the original more than 2000 wavelength variables was compressed as 5 new variables which orthogonal to each other, each other independently of each other, 99.99% of the represent original variables that were included in the information.

Table 1. Accumulative reliabilities of the first 5 Principal components

Principal components	Total confidence /%
PC01	96.363
PC01	96.363
PC01	96.363
PC01	96.363
PC01	96.363

3.3.2 Establishment and Analysis of vinegar species identification model based on stepwise discriminant

The scores of 5 principal components were taken as the input of stepwise discriminant analysis; this analysis was completed by the DPS software.

The 60 samples of three bands of vinegar were divided into 45 training samples and 15 confirmation samples, DongHu vinegar, DoubleToo vinegar and the PurpleForest vinegar were represented by the digit 1, 2, 3 separately, score data of 5 principal components of 45 samples were dealt with based on the Baye criterion discriminant analysis. Finally discriminant function was established, the brand distinction models as follows：

$$Y_1(x) = -252.982 + 73.9001x_1 - 63.96x_2 - 23{,}8146x_4 - 205.208x_5$$ DongHu vinegar

$$Y_2(x) = -970.014 - 146.805x_1 + 115.2x_2 + 26.998x_4 + 332.0506x_5$$ DubleToo vinegar

$$Y_3(x) = -240.946 + 72.3922x_1 - 51.18x_2 - 3.3682x_4 - 125.207x_5$$ PurpleForest vinegar

Outcomes could be seen from the above equations, the third principal component x3 was worst to identificate the vinegar species, it was removed in the process of develop the discriminant equation and reached the purpose of dimensionality reduction .

The four main components of 45 known samples were set into the discriminant function and re-classified by posterior probability, the discrimination results in table 2.Results could be seen from the table, and the discriminant function which was established on three brand vinegar discriminant accuracy was 100%, which indicated the identification model has a high credibility.

Table 2. Resubstisution rate with stepwise discriminabt analysis

Typle of vinegar	Results of returns sentences			Total	Accuracy of returns sentences /%
	DongHu	DoubleToo	PurpleForest		
DongHu	15	0	0	15	100
DoubleToo	0	15	0	15	100
PurpleForest	0	0	15	15	100

To further verify the reliability of the model establish, 15 unknown samples were replaced into the discriminant function to predict the validation. 4 principal components of 15 samples were set into the discriminant function, and the correct rate was 100%, it showed that the identification was reliable. The results were in Table 3.

Table3. Predicting rate with stepwise discriminabt analysis

Typle of vinegar	Results of returns sentences			Total	Accuracy of returns sentences /%
	DongHu	DoubleToo	PurpleForest		
DongHu	5	0	0	5	100
DoubleToo	0	5	0	5	100
PurpleForest	0	0	5	5	100

3.3.3 Construction and analysis of vinegar species identification model based on principal component analysis, stepwise discriminant analysis and BP neural network

Dentification model of vinegar was built with 4 of 5 components that were extracted by PCA and seemed as input variables of BP neural network, species value as the output, and then by adjusting the number of nodes in the hidden layer to optimize network, structure is 4(input nodes):7 (hidden nodes): 1 (output nodes).Fitting residual of 45 samples was 8.19 × 10-5, predicted results of 15 unknown samples were showed in Table 4. The threshold is set to ± 0.027 in the case, identification correct rate of the unknown sample reached 100%.

For comparison the number choice effect of principal components to BP neural network forecast, distinction model was established with 5 main components were taken as the input variables of BP neural network and the variety value as the output. Network was optimized through adjustment the nodal point numbers of concealment level, structure contioned 5 (input nodes): 9 (hidden layer): 1 (output nodes). The fitting residual of 45 modeling samples was 8.83 ×10-5, in the situation of threshold value hypothesis for ±0.03, the recognition accuracy of unknown samples achieved 100%.

The above BP neural network analysis was completed by the DPS software. The network settings training iteration number of times was 1000 times, the goal error hypothesis was 0.0001. 45 samples were selected as the modeling sets, 15 samples as

forecast sets. The 15 unknown samples were carried on to predict, the forecast results could be seen in table 4. Results indicated that insensitive variable were existed in the principal components of vinegar,this variable was removed with the stepwise discriminant analysis [11],BP neural network identification model was re-established based it and to improve the prediction efficiency.

Table 4 Prediction results for unknown samples by BP model

Sample number	Real value	Predicted value		Deviation	
		4 principal components	5 principal component	4 principal components	5 principal component
1	1	1.024	1.0165	0.0240	0.0165
2	1	1.0177	1.0139	0.0177	0.0139
3	1	1.0118	1.0085	0.0118	0.0085
4	1	1.0097	1.0070	0.0097	0.0070
5	1	1.0109	1.0078	0.0109	0.0078
6	2	1.9946	1.9885	-0.0054	-0.0115
7	2	1.9980	1.9914	-0.0020	-0.0086
8	2	1.9998	1.9943	-0.0002	-0.0057
9	2	1.9926	1.9861	-0.0074	-0.0139
10	2	1.9885	1.9810	-0.0115	-0.0190
11	3	2.9971	2.9949	-0.0029	-0.0051
12	3	2.9819	2.9787	-0.0181	-0.0213
13	3	2.9736	2.9701	-0.0264	-0.0299
14	3	2.9994	2.9989	-0.0006	-0.0011
15	3	2.9993	2.9987	-0.0007	-0.0013

Note:Variety value

1-DongHu vinegar；

2-DoubleToo vinegar；

3-PurpleForest vinegar

4 Concluding Remarks

(1)The method which was used visible near-infrared spectroscopy of Identify vinegar varieties is rapid, non-destructive and low cost.

(2) The principal components which were obtained from the PCA analysis was took the input of analysis. The computation was reduced greatly, in addition the useless variable was rejected in the stepwise discriminant analysis process and raised the analysis accuracy and the stability.

(3) Vinegar variety distinction model was established with this method. From the forecast effect, the unknown sample recognition rate had achieved 100%, indicated constructed of vinegar variety distinction model was reliable, stable. Therefore, the

variety distinction model of PCA analysis union stepwise discriminant analysis based on the near-infrared spectrum is feasible.

(4) It could be seen of this paper that insensitive variables were existed in principal components,it could be removed by stepwise discriminant analysis, the forecast efficiency could be raised with BP neural network distinction model based the method.

References

1. Xiaoling Tong. Yidan Bao, Yong He. Study on fast discrimination of soy sauce using near infrared spectra[J].Spectroscopy and Spectral 2008,28(3):597—601. (in Chinese with English abstract)
2. Haiyan Cen, Yidan Bao, Yong He. Fast discrimination of varieties of bayberry juice based on spectroscopy technology[J].Spectroscopy and Spectral,2007,27(3):503—506. (in Chinese with English abstract)
3. Xaoli Li, Yong,He .Zangjun Qiu. A new method to fast discrimination of tea varieties using visible/ near infrared spectroscopy[J].Spectroscopy and Spectral, 2007,27(2):279—282. (in Chinese with English abstract)
4. Xiaoli Li, Xiangyue Hu, Yong He. New approach of discrimination of varieties of juicy peach by near infrared spectra based on PCA and MDA model[J].J.Infrared Millim.Waves, 2006,25(6):417—420. (in Chinese with English abstract)
5. Yande Liu, Ji Luo, Xingmiao Chen. Analysis of soluble solid content in nanfeng tangerine with visible near infrared spectroscopy[J]. J.Infrared Millim.Waves, 2008,27(2):119—122. (in Chinese with English abstract)
6. Jiewen Zhao, Haidong Zhang. Muhua Liu. Non-destructive determination of sugar contents of app les using near infrared diffuse reflectance [J]. Transactions of the CSAE, 2005, 21 (3) : 162—165. (in Chinese with English abstract)
7. Yande Liu, Yibin Ying. Sugar content prediction of apples with near infrared diffuse reflectance technique[J] Transactions of the CSAE,2008,20(1): 189—192. (in Chinese with English abstract)
8. Guifeng Li, Guojian Zhao, Xiangdong Wang,etc. Nondestructive measurement and fingerprint analysisof apple texture quality based on NIR Spectra[J]. Transactions of the CSAE,2008,24(6):169—173. (in Chinese with English abstract)
9. Gang Wang, Shiping Zhu, Jian-Quan Kan, etc. Nondestructive detection of volatile oil content in zanthoxylum bungeagum maxim by near infrared spectroscopy[J].Transactions of the Chinese Society for Agricultural Machinery,2008,39(3):79—85. (in Chinese with English abstract)
10. Xiaobo Zou , Jiewen Zhao, Rong Xia. Near infrared determination of sugar content in apples based on multiresolution decomposition and interval Partial Least Square (iPLS) method [J]. Transactions of the Chinese Society for Agricultural Machinery,2006,37(6):79—82. (in Chinese with English abstrac