Journal of Computational Information Systems 9: 7 (2013) 2659–2666
Available at http://www.Jofcis.com

Classification of Hundred-grass-oil Samples Using E-nose ⋆

Dehan LUO [1,*], Yu SUN [1], Jiajun ZHUANG [1], H. Gholam HOSSEINI [2]

[1] *School of Information Engineering, Guangdong University of Technology, Guangzhou 510006, China*

[2] *School of Engineering,Auckland University of Technology, Private Bag 92006, Auckland 1142, New Zealand*

Abstract

One of the differences between Traditional Chinese Medicine (TCM) and its ordinary counterpart is the composition of products. The ingredients of TCM are complex mixtures of natural products or extracts with various active constituents while in ordinary medicine, all of the active ingredients are single compounds and can easily be analyzed. Therefore, the classification of TCM with complex mixtures is not well adapted with the classical discriminative techniques which apply to ordinary medicine. Quality assessment of TCM is an important factor for analyzing those complex matrices and gives comprehensive quantitative data about the global ingredients. Electronic nose (E-nose) can be used for quality assessment and classification of TCM. The aim of study is to classify four groups of hundred-grass-oil (HGO), one of the popular TCM, from different production batches using a portable E-nose (PEN3). Principal component analysis (PCA) and linear discriminant analysis (LDA) were used to investigate the classification ability of PEN3. The results of PCA analysis showed a good separation among four groups of HGO. Loadings analysis was used to optimize the number of sensors that show a higher influence on the distribution of the PCA plot. A better classification result is obtained by LDA analysis.

Keywords: Traditional Chinese Medicine (TCM); Electronic Nose (E-nose); Identification of Hundred-grass-oil; PCA; LDA

1 Introduction

Improvements in electronic nose technology over the last 20 years have enabled us to employ E-nose in automatic identification of HGO among other TCM and classifying four groups of HGO based on their production batches. A portable PEN3 E-nose comprising an array of metal oxide semiconductor (MOS) sensors with partial specificity and an appropriate pattern recognition system, making it capable of recognizing simple or complex odors [1].

⋆Project supported by the National Natural Science Foundation of China through Project (No. 60971105) and the Guangdong Natural Science Foundation through Project (No. S2011020002906).

*Corresponding author.
Email address: dehanluo@gdut.edu.cn (Dehan LUO).

April 1, 2013

Pattern recognition techniques such as principal component analysis (PCA), linear discriminant analysis (LDA), discriminant function analysis (DFA), cluster analysis (CA), and artificial neural network (ANN) have been used for data analysis in E-nose based applications [2, 3].In this study,PCA and LDA techniques have been employed and their performance in classification of the selected samples of HGO has been compared.

Based on the concept of "nasal olfaction", one of traditional empirical methods for identifying and classifying TCM [4], we proposed a simple and quick method for classifying TCM samples by an E-nose. During the past few years, E-nose technology systems have gained remarkable development. This technology ese instruments have been widely and successfully used in different fields especially in food and beverage industry, such as identification of spoiled beef [5], detection and evaluation of fish freshness [6, 7], modeling the ripening of Danish blue cheese [8], discrimination or classification of wine [2, 9, 10], monitoring the aroma of different kinds of fruit [3, 11, 12, 13]. However, very little work has been conducted so far to classify TCM with an E-nose.The aim of this study is to classify HGO from different production batches, using a pattern recognition technique, and evaluate the discrimination capability of an e-nose system.

2 Materials and Methods

2.1 Experimental samples

This study was carried out using four groups of Luofushan HGO samples with different production batches provided by Guangdong Luofushan Pharmaceutical Co. Ltd. as described in Table 1. HGO is a kind of emerald green clarified liquid made of 79 different TCM with fragrant smell.

Table 1: Information of HGO samples used in this experiment

Group label	Sample name	Production batch	Production date
p060528	HGO 1	L06E281	20060528
p070504	HGO 2	L07E041	20070504
p080118	HGO 3	L08A181	20080118
p080304	HGO 4	L08C041	20080304

2.2 Portable E-nose

Experiments were performed with a portable electronic nose (PEN3) produced by AIRSENSE Analytics GmbH in Schwerin, Germany. PEN3 is equipped with an array of 10 different MOS sensors positioned into a small chamber (V=1.8ml), a sampling apparatus and a pattern recognition software named WinMuster. Table 2 summarizes the sensitivity list of all sensors in PEN3. The response data collected by PEN3 is defined as the ratio of conductance: G/G0. G represents the resistance of each sensor in the chamber after the exposition to the headspace gas in the vial and G0 represents the resistance while the sensors expose to the zero gas filtered by active carbon.

Table 2: The sensitivity list of 10 sensors in PEN3

Number in array	Sensor name	Sensitive to	Detection Range /ppm
S1	W1C	Aromatic components	10
S2	W5S	Nitrogen oxides, very sensitive	1
S3	W3C	Ammonia and aromatic components	10
S4	W6S	Mainly hydrogen, selectively, (breath gases)	100
S5	W5C	Alkanes and aromatic components	1
S6	W1S	Propane	100
S7	W1W	Sulfur organic compounds	1
S8	W2S	Ethanol	100
S9	W2W	romatic components and organic-sulfides	1
S10	W3S	Propane(selective sometimes)	100

2.3 Experimental set up and data collection

Figure 1 shows the experimental set up for data collection which was carried out in an air-conditioned laboratory where the temperature was kept at 27±1 degrees Celsius and the humidity at 46±3%. The HGO samples with different production batches were injected into four vials (40 ml) labeled p060528, p070504, p080118 and p080304, respectively. The dose of each sample in the vial is 0.5 ml. Then four vials were hermetically capped with plastic wrap for 1 hour in order to generate a steady headspace respectively. The sampling time for each sample is 70 seconds, which is enough for each sensor to reach a stable value. The rinsing time is set as 110 seconds, during which the sensors are rinsed with charcoal filtered to force the signals of sensors to baseline. One measurement circle would last for about 6 minutes. Meanwhile, due to the high sensitivity of the second sensor W5S, the automatic dilution was set to be activated when the prime transient response (G/G0) of sensor W5S rose above 3, in order to protect the sensor array from being overloaded. Static headspace sampling (SHS) method was used because SHS is the

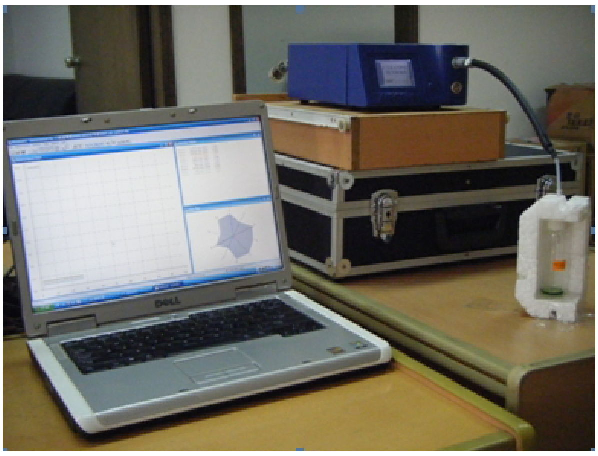

Fig. 1: Experimental Setup for headspace volatile sampling

most common technique for its accessibility [15]. During the measurement process, the headspace of each vial was pumped over the sensors in PEN3 with a constant flow speed of 400 ml/min. The collected data set would be automatically stored in a personal computer connected to PEN3 after measurement was completed. The headspace gas of each vial of HGO sample was measured 8 times continuously. Thus 32 data sets were collected for all 4 groups of HGO.

3 Results and Discussion

3.1 Sensor response

The data sets acquired were first analyzed by Matlab 7.1 to reconstruct the response curves and radial plot of the 10 sensors. Figure 2 shows the typical transient responses of the 10 sensors to the four groups. The response curves represent the ratio of the conductance of each sensor versus sampling time when the volatile gas of one sample group reached the sensor chamber. Figure 2 shows rapid change at the beginning of the sampling time particularly for sensor 2 (W5S). But the curves reach to the steady state soon due to the automatic dilution. After approximately 40 seconds all the sensors reached stable values except sensor 2, which changed evenly after 60 seconds. The contributions of the 10 sensors to the four groups of HGO are similar by comparing the response curves of the latter 30 seconds in the four graphs below. Some static features might

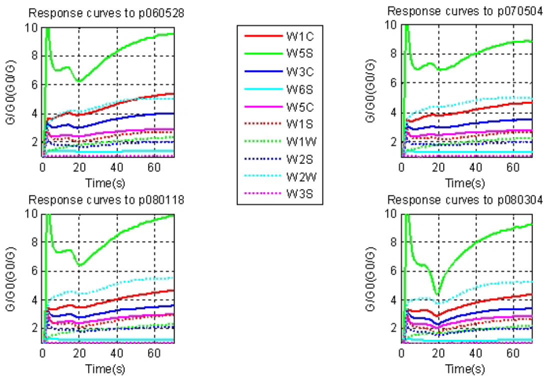

Fig. 2: Response curves of 10 sensors to the four groups against sampling time

be extracted from the data sets, such as the final conductance (G) of each sensor, conductance increment (G=G-G0), the ratio of the conductance (G/G0) of each sensor and so on.

In this study, the ratio of the conductance of each sensor between the 61st and 64th second of the sampling time was extracted as the static feature for further analysis.

3.2 Principal component analysis

Initially, the data sets were rescaled by dividing the data by the standard deviation. After analyzing by PCA [14], the dimensionality of the data sets was reduced to a lower dimension. Usually the first two principal components will carry the most information of the old variables. The corresponding plots of the first two principle components are shown in Figure 3. The principal components in the plot represent the eigenvectors of data sets processed by PCA and are presented in descending order. It is shown in Figure 4, that the first two components capture most information (88.73% totally) of the variance in the data sets. All the clusters of the groups were dispersive and stretched obviously in the direction of transverse axis except that of the group labeled p070504 which was a bit convergent. In the PCA plot the higher the interclass distance between two clusters, the higher the difference between them. The first two groups (p060528 and p070504) could be easily distinguished within the four clusters; however, the clusters of the latter two groups (p080118 and p080304) did not differ much. This small overlap joint could be due to the fact that the latter two groups are of closer production dates, making the aroma ingredient of them seem to be more similar and therefore the sensors response to them are near akin.

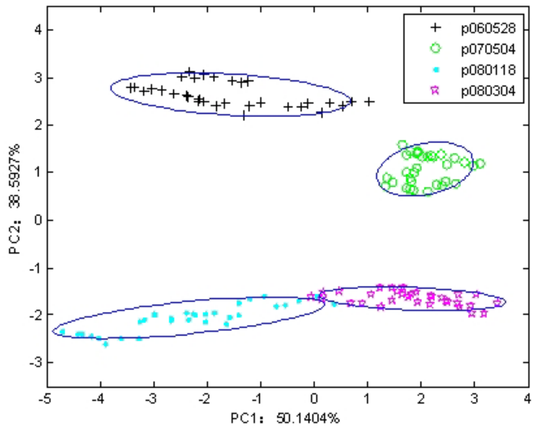

Fig. 3: The PCA score plot of four groups measured using SHS

*The figure here means that the classification rate between group p060528 and group p070504 is 98.5%.

Table 3 shows the classification rate among the four groups calculated by WinMuster. Overall, the classification rate is satisfying except that between group p080118 and group p080304 (only 72.7%).

Table 3: Information of HGO samples used in this experiment

	p060528	p070504	p080118	p080304
p060528	–	98.5%	99.4%	99.6%
p070504	98.5%	–	97.6%	99.3%
p080118	99.4%	97.6%	–	72.7%
p080304	99.6%	99.3%	72.7%	–

3.3 Loadings analysis

Single sensors may have a bad influence on the analysis result. That means, a sensor may only react on disturbing compounds, not on compounds important for the discrimination. Useful information for the selection of sensor signals to be eliminated may be obtained by the loadings analysis.

The loadings analysis is well correlated to the PCA; it is based on the same algorithm, but in this case, it is calculated for the sensors themselves. It is useful to check for the influence (loading) of a sensor on the distribution of data sets.The direction where the sensors in the loadings analysis are located corresponds to the direction of the PCA plot distribution. By this method we could switch off some sensors that had a minor influence on the distribution in the PCA plot [16]. Figure 4shows a loading plot of the loading factors associate to PC1 and PC2 for the samples. The points in the plot represent the sensors used in the experiment. Sensors, with loading parameters close to zero for a particular principal component, have a low contribution to the total response of the array, whereas high values indicates a discriminating sensor [12]. So we could consider switching off the sensor that has less influence on the result of PCA analysis. If a group of sensors have similar response to the samples, we could consider replacing the group with one of its member. One group of sensors including W1C, W5S, W3C and W2W evidently

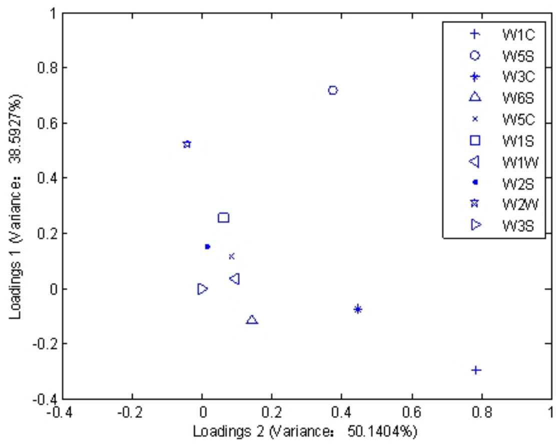

Fig. 4: Loadings analysis related to the first two principal components

have higher influence in the current pattern files than the second group including W6S, W5C, W1S, W1W, W2S, and W3S6 sensors., among which Sensor W3S has almost no influence among the second group of sensors while the other 5 sensors in this group have closer influence so that they might be represented by one of the group member. Sensor W5C has a lower detection range (1ppm) while exposure to aromatic components as shown in Table 2.

By considering the factors aforementioned, a subset of sensors including [W1C, W5S, W3C, W5C (a representative from the second group) and W2W] sensors is retained as an optimized sensor array. The corresponding PCA plot of data sets of these 5 sensors is shown in Figure 5.

3.4 Linear discriminant analysis

The response data of PEN3 for each group between the collection times (61st and 64th) were extracted and analyzed by LDA analysis. The analytic result is shown in Figure 6. As is shown in Figure 6, LDA function 1 (LD1) and function 2 (LD2) accounted for 95.017% and 3.2718% of the variance, respectively. The clusters of the data sets are obviously divided into four groups, allowing an easy discrimination of the four sample groups with different production batches. This is very satisfactory because it clearly shows a better classification by LDA analysis.

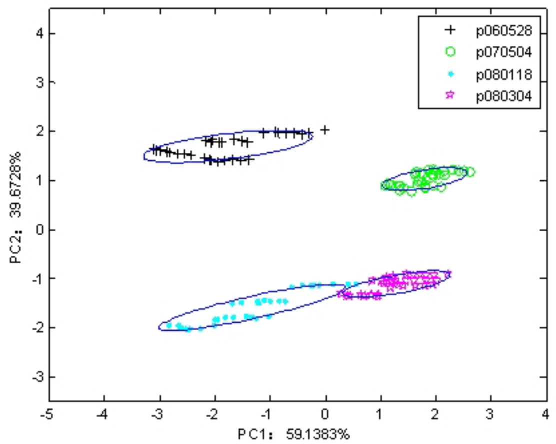

Fig. 5: The PCA score plot of the HGO groups by a subset of sensors

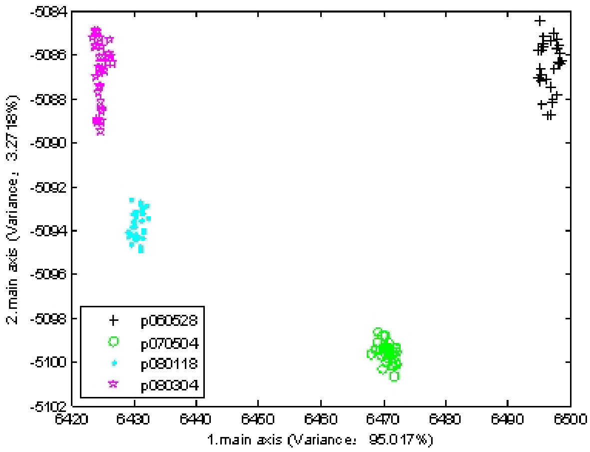

Fig. 6: Analytic result of the HGO groups by LDA

4 Conclusion

For the purpose of classification of TCMs (four groups of HGO samples with different production batches), experiments have been carried out with PEN3. The results show that it is practical and effectual to classify the congener TCMs samples by an electronic nose system and appropriate pattern recognition algorithms.

The PCA analysis had a poor performance in clustering of the groups labeled p080118 and p080304. However, a better classification result is obtained by LDA analysis. Besides, we could use loadings analysis as a means of sensors optimization.

As a result of loading analysis it was found that with a subset of sensors including 5 most sensible sensors, the clustering of the selected four groups seemed to be more convergent than similar clustering approach using all sensors.

Acknowledgement

The authors would acknowledge the Guangzhou University of Chinese Medicine for providing the Atractylodes samples.

References

[1] J. Lozano, T. Arroyo, J. P. Santos, et al. Electronic nose for wine ageing detection. Sensors and Actuators B, 133 (2008): pp. 180 – 186.

[2] Jesús Lozano, José Pedro Santos, Manuel Aleixandre, et al. Identification of typical wine aromas by means of an electronic nose. IEEE Sensors Journal, 6 (2006): pp. 173 – 178.

[3] Amalia Z. Berna, Jeroen Lammertyn, Stijn Saevels, et al. Electronic nose systems to study shelf life and cultivar effect on tomato aroma profile. Sensors and Actuators B, 97 (2004): pp. 324 – 333.

[4] Zhang Lixin, Zhao Yannan, Yang Zehong, et al. Classification of traditional Chinese medicine by nearest-neighour classifier and genetic algorithm. In Proceedings of the Fifth International Conference on Information Fusion, 2002, vol. 2, Page (s): pp. 1596 – 1601.

[5] S. Panigrahi, S. Balasubramanian, H. Gu, et al. Neural-network-integrated electronic nose system for identification of spoiled beef. LWT-Food Science and Technology, 39 (2006): pp. 135 – 145.

[6] H. GholamHosseini H. , D. Luo Dehan, H. Liu Hongxiu, et al. Intelligent processing of E-nose information for fish freshness assessment. Proceedings of the 2007 International Conference on Intelligent Sensors, Sensor Networks and Information Processing, 2007, Page (s): pp. 173 – 177.

[7] N. El Barbri, E. Llobet, N. El Bari, et al. Application of a portable electronic nose system to assess the freshness of Moroccan sardines. Materials Science and Engineering: C, 28 (2008): pp. 666 – 670.

[8] J. Trihaas, L. Vognsen, P. V. Nielsen. Electronic nose: new tool in modeling the ripening of Danish blue cheese. International Dairy Journal, 15 (2005): pp. 679 – 691.

[9] M. Pilar Martí, Olga Busto, Josep Guasch. Application of a headspace mass spectrometry system to the differentiation and classification of wines according to their origin, variety and ageing. Journal of Chromatography: A, 1057 (2004): pp. 211 – 217.

[10] J. Lozano, J. P. Santos, J. Gutiérrez, et al. Comparative study of sampling systems combined with gas sensors for wine discrimination. Sensors and Actuators B, 126 (2007): pp. 616 – 623.

[11] Jesús Brezmes, Ma. Luisa López Fructuoso, Eduard Llobet, et al. Evaluation of an electronic nose to assess fruit ripeness. IEEE Sensors Journal, 5 (2005): pp. 97 – 108.

[12] Antihus Hernández Gómez, Jun Wang, Guixian Hu, et al. Electronic nose technique potential monitoring mandarin maturity. Sensors and Actuators B, 113 (2006): pp. 347 – 353.

[13] Simona Benedetti, Susanna Buratti, Anna Spinardi, et al. Electronic nose as a non-destructive tool to characterise peach cultivars and to monitor their ripening stage during shelf-life. Postharvest Biology and Technology, 47 (2008): pp. 181 – 188.

[14] Huichun Yu, Jun Wang. Discrimination of LongJing green-tea grade by electronic nose. Sensors and Actuators B, 122 (2007): pp. 134 – 140.

[15] M. Pilar Martí, Ricard Boqué, Olga Busto, et al. Electronic noses in the quality control of alcoholic beverages. Trends in Analytical Chemistry, 24 (2005): pp. 57 – 66.

[16] Antihus Hernández Gómez, Guixian Hu, Jun Wang, et al. Evaluation of tomato maturity by electronic nose. Computers and Electronics in Agriculture, 54 (2006): pp. 44 – 52.