

A Classification for Electronic Nose Based on Broad Learning System

Yu Wang^a, Xiaoyan Peng^b, Hao Cui^b, Pengfei Jia^{b,1}

^aCollege of Electronic and Information Engineering, Southwest University, Chongqing, China

^bCollege of Artificial Intelligence, Southwest University, Chongqing, China

Abstract. The odor of citrus juice changes during the storing process. We use an electronic nose (E-nose) to detect the volatile odors released by citrus juice and use the detect results to classify citrus juices from different storage periods. In this article, a novel classifier of E-nose, namely broad learning system (BLS) is introduced. BLS is different from traditional classifier. It has a simple network model, which can greatly reduce the training time of the model. BLS can effectively combine feature extraction and classification recognition to make the model more efficient. We apply BLS to the analysis of valencia citrus juice data. The experimental results show that BLS can effectively identify the current stage of the stored valencia citrus juice. Compared with traditional classifier such as radical basis function neural network (RBFNN) and linear discriminant analysis (LDA), the results show that BLS has better performance for the storage period classification of valencia citrus juice.

Keywords. Citrus juice, Electronic nose, Broad learning system, Classifier

1. Introduction

Due to the improvement of people's living standards, citrus juice is loved by consumers in their daily life. However, the storage of citrus juice is a serious problem. We need to detect citrus juice at all times to ensure the quality of citrus juice. Because citrus juice emits some volatile odors, we can determine the storage period of citrus juice by detecting volatile gases, and thus determine the quality of citrus juice. In this article, we use the electronic nose (E-nose) for the quality identification of juice. At present, the feasibility of the identification of juice quality by the E-nose has been verified [1, 2].

E-nose is an intelligent system composed of sensor arrays and pattern recognition algorithm. It can be used to simulate human olfactory organs to identify odors or gases [3, 4]. In the current research, E-nose has been widely used in the detection of gas or odor, and it also has certain advantages compared with traditional methods such as human olfactory analysis and precision instrument analysis. Compared with traditional methods, E-nose has the following advantages. Firstly, it can work stably for a long time. Secondly, it is less affected by external factors during the detection process. Finally, it can make accurate quantitative analysis of the target gas or odor. At present,

¹ Corresponding Author, Pengfei Jia, College of Artificial Intelligence, Southwest University, Chongqing, China; E-mail: jiapengfei200609@126.com

it has been widely used in explosive detection [5, 6], environmental monitoring [7, 8], Food storage [9] and disease diagnosis [10].

In traditional pattern recognition algorithms, feature extraction and classification recognition are divided into two independent steps. The first step of processing the raw response data of the sensor is feature extraction. Then, we use the data after feature extraction to build a classifier model. Feature extraction will influence the recognition rate of the final classification, but the target requirements of the two stages may be different, which will lead to a unsatisfactory results. It is assumed that the combination of feature extraction and classification recognition unifies the requirements of the two stages, and whether the final classification accuracy can be improved. Therefore, it is necessary to find an algorithm that integrates feature extraction and classification recognition.

Broad learning system (BLS) is a simple flat network, proposed in 2018 by C. L. Philip Chen [11, 12]. BLS uses a random mapping method to convert the original data into mapped features and enhanced features, and then connects the two horizontally to form the actual input of the model. Through random mapping, effective feature is extracted from the original data. Therefore, no other feature extraction methods are needed to process the raw data. In addition, high-dimensional (or even infinite-dimensional) data can be processed by random mapping. Finally, the simple network structure of BLS can shorten the time to train the model. Compared with traditional classifier, when BLS processes large samples, the training time required for the model is shorter. BLS perfectly combines feature extraction and classification recognition. The advantage of doing so is that it can uniformly set parameters and modify the model according to the accuracy of the final classification recognition. So far, it has been applied to time series prediction [13], drift compensation [14], image classification [15] and so on.

In this article, we apply BLS to solve the classification problem of E-nose. In the other parts of the article, we first introduce the experimental platform and sample collection in Section 2, then, overview of BLS will be introduced in Section 3, the results and discussion are presented in Section 4, finally, the conclusion of this article is in Section 5.

2. Materials and methods

2.1. Experimental setup

In this section, we introduce the self-made E-nose system. The detailed structure of the self-made electronic nose system is shown in Figure 1. The self-made E-nose system is composed of E-nose, PC, gas flow meter, a temperature-humidity controlled chamber and air pump. The E-nose has two ports, port 1 is for injecting target gas into the air chamber, and port 2 is for injecting clean air into the air chamber. When the target gas is injected into the gas chamber to make full contact, the sensors will respond and use the analog-digital converter (A/D) to convert the response signal of the sensor into a digital signal, and use a joint test action group (JTAG) to save data to PC.

According to paper [16], we can find that valencia citrus juice is composed of hydrocarbons, alcohols, aldehydes, ketones, esters and other chemical components. The E-nose has a sensor array composed of 15 gas sensors, which ensures that the odor components emitted by valencia citrus juice can be fully detected, and these sensors

used have the advantages of cost-effective, low energy consumption and high measurement accuracy.

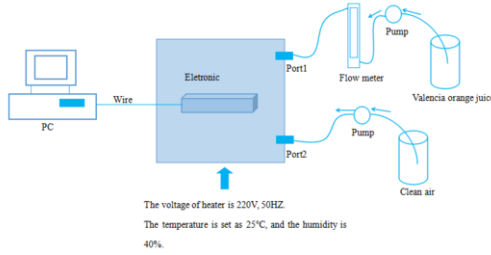


Figure 1. System diagram of self-made E-nose system

2.2. Sampling experiments

Firstly, we adjust the temperature and humidity of the air chamber, and set the temperature and humidity to 25°C and 40% respectively. Then start the sampling experiment, and one single sampling steps are as follows:

- Step 1: Restore the baseline of the sensor, inject clean air into the air chamber, and fully contact the sensor array for 5 min to obtain its baseline value.
- Step 2: Data sampling, inject the target gas into the gas chamber 7 min to obtain the response value of the sensor.
- Step 3: Cleaning phase, clean air is injected into the air chamber to fully contact the sensor array for 5 minutes to restore the sensor to the baseline value.

The response curve of the sensor array during one single sampling is shown in Figure 2.

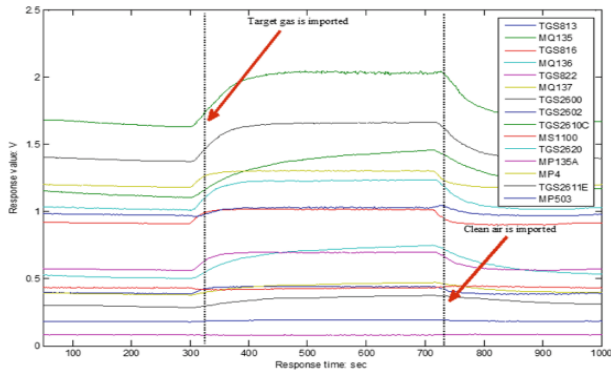


Figure 2. Response of sensor array

As shown in Figure 2, we can learn that the first 5 min is to process the sensor array to return to the baseline value, 5 min to 12 min to make the sensor array response stage, 12 min to 17 min is to inject clean air, clean the sensor array to return to the baseline value.

3. Broad learning system

BLS has a very simple network structure, which makes its training time shorter. The structure of BLS is shown in Figure 3, BLS generates mapping features from the original input data X through a random mapping method, and saves them in feature nodes Z , and generates enhancement nodes H through random mapping of feature nodes to broaden the network horizontally. Then the generated feature nodes and enhancement node are connected horizontally as the input of the model, and after input the data, get the output of the network. Finally, the ridge regression algorithm is used to calculate the connection weight between the network input and output. The detailed derivation process is as follows.

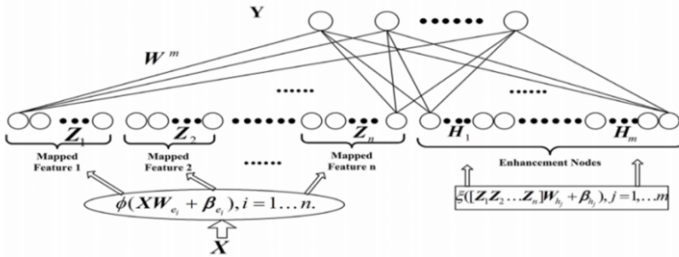


Figure 3. Structure of BLS

Assuming that the input original data is X , the i -th mapping feature Z_i is generated as follow:

$$Z_i = \phi_i(XW_{ei} + \beta_{ei}) \tag{1}$$

where W_{ei} and β_{ei} are randomly generate weights and biases respectively, ϕ_i is the linear mapping function. Mark as $Z^i = [Z_1, Z_2, \dots, Z_i]$, which means that there are i group of mapping features for horizontal connection. The way to generate the j -th group of enhancement nodes is as follows:

$$H_j = \xi_j(Z^i W_{hj} + \beta_{hj}) \tag{2}$$

where W_{hj} and β_{hj} are weights and biases respectively, ξ_j is the nonlinear mapping function. Mark as $H^j = [H_1, H_2, \dots, H_j]$, which means that there are i group of enhancement features for horizontal connection. The values of i and j are determined by the complexity of the modeling task.

In order to obtain better classification results, it is necessary to have a good input feature representation. In the process of random mapping, weights and biases are randomly generated, but the randomness is unpredictable. In order to overcome the nature of randomness, by fine-tuning W_{ei} and β_{ei} through the sparse autoencoder method, more essential features can be explored. Assuming that m feature nodes and n enhancement nodes are generated, all nodes are horizontally connected to form the input of the model, namely $A = [Z^m | H^n]$. Therefore, the input Y of the BLS model can be expressed as:

$$Y = AW \quad (3)$$

where W is the connection weight of the broad network. And W can be calculated by a simple method, which can be expressed as $W = A^+Y$, where A^+ is pseudo-inverse matrix of BLS mapping matrix A . Then, the solution of the matrix pseudo-inverse problem is transformed into l_2 norm regularized BLS, which can be expressed as the following convex optimization problem:

$$\operatorname{argmin}_W \|AW - Y\|^2 + \lambda \|W\|^2 \quad (4)$$

where the second term in the above formula further restricts the weight W , λ is the regularization coefficient.

Therefore, We can use formula (5) to calculate W :

$$W = (\lambda I + AA^T)^{-1} A^T Y \quad (5)$$

Specifically

$$A^+ = \lim_{\lambda \rightarrow 0} (\lambda I + AA^T)^{-1} A^T \quad (6)$$

where I represents the identity matrix.

BLS extracts important features through random mapping. Simultaneously, the output result of the model is compared with the actual label to obtain the classification accuracy. Consequently, the BLS model perfectly combines feature extraction and classification recognition, so that parameter settings and model modifications can be unified through the obtained accuracy of classification, making the model more efficient.

4. Results and discussion

We will briefly introduce the process of sample sampling in this section. The citrus juice data set used in this article is composed of data collected in 5 storage stages of citrus juice through the sensor array in the E-nose. Each stage is sampled for 24 times, and a total of 120 sampling experiments in 5 stages, 400 points which are selected from each response curve of sensors are input into BLS, 75% of samples are used as training samples, and the rest are used as test samples. As shown in Figure 4, it is the data obtained in one sampling. The data in the boxes are randomly selected. There are a total of 400 boxes, which represent the response scatter points of 400 sensing matrices are selected.

In order to prove the availability of BLS, we compare BLS with LDA and RNFNN, and make the correct rate of the training set and the test set as indicators to evaluate its performance. Each program runs 10 times, and then takes the average as the final result. The final accuracy of classification are shown in Table 1.

From Table 1, we know that the classification accuracy of BLS training data set and test data set are higher than LDA and RBFNN, which shows that BLS has good performance in classification recognition.

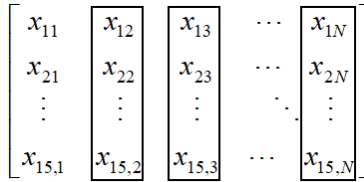


Figure 4. Schematic diagram of obtaining input BLS data in one sampling.

Table 1. Classification recognition rate of different classifiers (%)

	LDA	RBFNN	BLS
Training set	77.58	80.24	99.96
Test set	51.67	60.00	95.11

We also explored the training time of BLS and compared it with the training time of LDA and RBFNN. Each program runs 10 times, and then takes the average as the final result. The final results are shown in Table 2.

Table 2. Running time of different classifiers

Classification algorithm	Running time
LDA	0.5616
RBFNN	171.5596
BLS	8.7371

Table 2 lists the results of running time of the LDA, RBFNN and BLS. We can find that the running time of LDA is the shortest, while the running time of RBFNN is the longest. Though the running time of BLS is longer than LDA, but much shorter than RBFNN.

Combining Table 1 and Table 2, we can conclude that the BLS classifier has a greater improvement in recognition accuracy than LDA and RBFNN, and the running time of BLS is much less than RBFNN, just slightly longer than LDA with the higher recognition accuracy. Therefore, BLS has a better advantage in classification recognition.

5. Conclusion

In this paper, we propose a novel classifier based on BLS to improve the classification accuracy of the E-nose for different storage periods of citrus juice. BLS has some advantages that traditional classifier do not have. On the one hand, BLS has a simple flat network structure. The simple structure means that the model training time is short. On the other hand, BLS effectively combines feature extraction and classification recognition. The two-stage requirements can be unified, and the parameters and structure of the model can be uniformly adjusted from the final classification results, thereby further improving the classification recognition rate. Comparing BLS with LDA and RBFNN, we can find that BLS has a greater improvement in the classification accuracy of orange juice storage period, and the training time is shorter. Through the research in this article, we believe that BLS is a good classifier of E-nose.

This article proposes a better classification method to deepen the theory of broad learning system and expand its application range. For our citrus data, it can better classify the different storage periods of citrus, and promote the practical application of E-nose in the field of citrus. However, the weight matrix in the BLS model is generated

randomly and has a strong uncertainty. Therefore, we can optimize the weight matrix assignment method to further improve the accuracy in future work.

Acknowledgments

This work is supported by National Natural Science Foundation of China (61906160, 61804127), Fundamental Science and Advanced Technology Research Foundation of Chongqing (Grant No. cstc2018jcyjA0867, Grant No. cstc2018jcyjAX0261), Fundamental Science on Nuclear Wastes and Environmental safety Laboratory (Grant No. 19kfhk03), Open Research Fund Program of Data Recovery Key Laboratory of Sichuan Province (Grant No. DRN19015), and the Fundamental Research Funds for the Central Universities (XDJK2020C040).

References

- [1] Cao H, Jia P, Xu D, Jiang Y, Qiao S. Feature extraction of citrus juice during storage for electronic nose based on cellular neural network. *IEEE Sensors Journal*. 2020 Apr;20(7):3803-3812.
- [2] Jiang X, Jia P, Qiao S, Duan S. Odor change of citrus juice during storage based on electronic nose technology. *International conference on neural information processing*. 2017 Nov 14-18; Guangzhou, China; p. 317-326.
- [3] Pearce T C. Computational parallels between the biological olfactory pathway and its analogue 'the electronic nose': Part II. Sensor-based machine olfaction. *BioSystems*, 1997 Jan;41(2):69-90.
- [4] Bahraminejad B, Basri S, Isa M M, Hambali Z. Application of a sensor array based on capillary-attached conductive gas sensors for odor identification. *Measurement Science and Technology*. 2010 Aug;21(8):085204(17pp).
- [5] Sukhvir K, Arvind K, Kaur RJ, Priya A, Harminder S. SnO₂-glycine functionalized carbon nanotubes based electronic nose for detection of explosive materials. *Sensor Letters*, 2016 Jul;14(7):733-739.
- [6] Brudzewski K, Osowski S, Pawlowski W. Metal oxide sensor arrays for detection of explosive at sub-parts-per million concentration levels by the differential electronic nose. *Sens. Actuators B Chem*. 2012 Jan;161(1):528-533.
- [7] Galang GM, Zarra T, Naddeo V, Belgiojorno V, Florencio JB. Artificial neural network in the measurement of environmental odours by E-nose. *Chemical engineering transactions*. 2018 Jan;68:247-252.
- [8] Pan L, Yang SX. An electronic nose network system for online monitoring of livestock farm odors. *IEEE-ASME Transactions on Mechatronics*, 2009 Jun;14(3):371-376.
- [9] Pang L, Wang J, Lu X, Yu H. Discrimination of storage age for wheat by E-nose. *Transactions of the ASABE*. 2008 Sep/Oct;51(5):1707-1712.
- [10] Saidi T, Zaim O, Moufid M, Bari NE, Ionescu R, Bouchikhi B. Exhaled breath analysis using electronic nose and gas chromatography-mass spectrometry for non-invasive diagnosis of chronic kidney disease, diabetes mellitus and healthy subjects. *Sens. Actuators B chem*. 2018 Mar;257:178-188.
- [11] Chen CL, Liu Z, Feng S. Universal approximation capability of broad learning system and its structural variations. *IEEE Transactions on Neural Networks*, 2019 Apr;30(4):1191-1204.
- [12] Chen CL, Liu Z. Broad Learning System: An effective and efficient incremental learning system without the need for deep architecture. *IEEE Transactions on Neural Networks*. 2018 Jan;29(1):10-24.
- [13] Xu M, Han M, Chen CL, Qiu T. Recurrent broad learning systems for time series prediction. *IEEE Transaction on Cybernetics*, 2020 Feb;50(4):1405-1417.
- [14] Liu B, Zeng X, Tian F, Zhang S, Zhao L. Domain transfer broad learning system for long-term drift compensation in electronic nose systems. *IEEE Access*. 2019 Sep;7:143947-143959.
- [15] Fan J, Wang X, Wang X, Zhao J, Liu X. Incremental wishart broad learning system for fast polarar image classification. *IEEE Geoscience and Remote Sensing Letters*. 2019 Dec;16(12):1854-1858.
- [16] Moshonas MG, Shaw PE. Quantitative determination of 46 volatile constituents in fresh unpasteurized orange juice using dynamic headspace gas chromatography. *J. Agric. Food Chem*. 1994 Jul;42(7):1525-1528.