Smell Classification by Neural Networks

Michifumi Yoshioka[+], Toru Fujinaka[++], and Sigeru Omatu[+]
Osaka Prefecture University, Sakai, Osaka 599-8531, Japan
: omatu@cs.osakafu-u.ac.jp
[++]Hiroshima University, Higashihiroshima, 739-8524, Japan

Abastract

In this paper, a reliable electronic nose (EN) system designed from the combination of various metal oxide gas sensors (MOGS) is applied to detect the early stage of fire from various sources. The time series signals of the same source of fire in every repetition data are highly correlated and each source of fire has a unique pattern of time series data. Therefore, the error back propagation (BP) method can classify the tested smell with 99.6% of correct classification by using only a single training data from each source of fire. The results of the k-means algorithms can be achieved 98.3% of correct classification which also show the high ability of the EN to detect the early stage of fire from various sources accurately.

1. Introduction

Every year the damage from the household fire disaster brings about not only severe loss to property assets, but also physical and psychological injuries of the people. Although most of the residences have installed the fire detectors system such as smoke detectors, those devices cannot detect the early stage of fire since their warning signals are triggered by the high smoke density or the high air temperature. In this paper, the reliability of a new electronic nose (EN) system developed from various metal oxide gas sensors (MOGSs) to specify the smell from various sources of fire is presented.

Jame A. Milke [1] has proved that two kinds of MOGSs have the ability to classify several sources of fire more precisely than conventional smoke detector. However, his results can be achieved only 85% of correct classification. In this paper, a new EN that has been successfully applied to classify not only the same smell from different brands, but also the same smell at different concentration levels [2] is applied to measure smells from various sources of fire such as household burning materials, cooking smells, the leakage from the liquid petroleum gas (LPG). The time series signals of the MOGSs from the beginning to the time until the MOGSs are fully absorbed the smell from each source of fire are recorded and analyzed by the error back propagation (BP) neural networks and the k-means algorithms. The average classification rate of 99.6% can be achieved by using the BP method with only a single training data from each source of fire. The results from the k-means algorithm can be achieved 98.3% of correct classification that also confirms the reliability of this new device to be able to detect various sources of fire in the early stage much better than the results of Jame A. Mike [1].

2. Metal Oxide Gas Sensors for EN

A commercial MOGS has been developed widely for more than thirty years.

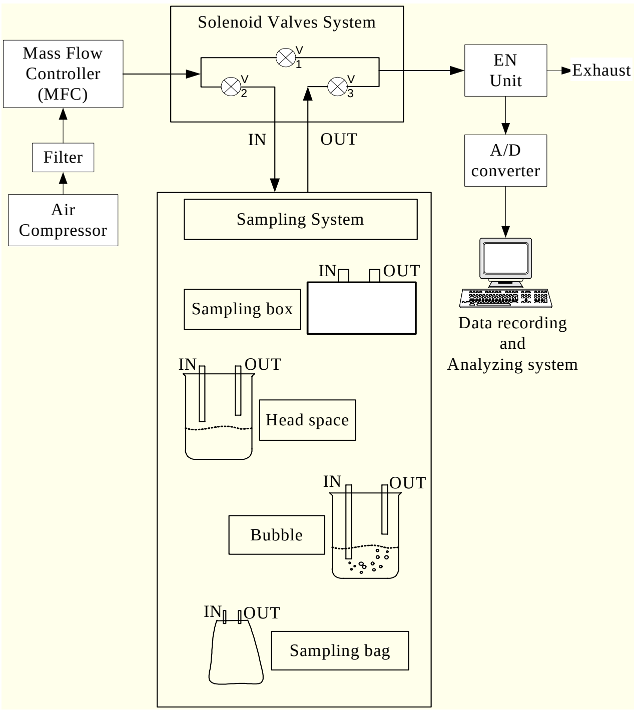

Fig. 1 Schematic diagram of the electronic nose system.

Generally, it is designed to detect some specific smell in electrical appliances such as an air purifier, a breath alcohol checker, and so on. Each type of MOGS has its own characteristics to response to different gases. When combining many MOGSs together, the ability to detect the smell is increased. An EN system shown in Fig. 1 has been developed based on the concept of human olfactory system by using the combination of MOGSs from FIS Inc. listed in Table I as the olfactory receptors in the human nose. The MOGSs unit is combined with the air flow system to flow the air and the tested smell into the MOGSs unit. The data logger converts the analog signals to digital signals and stores them in the data recording system before being analyzed by multivariate analytical methods, such as the BP method and the k-means algorithms.

The main part of the MOGS is the metal oxide element on the surface of the sensor. When this element is heated at a certain high temperature, the oxygen is absorbed on the crystal surface with the negative charge. The reaction between the negative charge of the metal oxide surface and deoxidizing gas makes the resistance of the sensor vary as the partial pressure of oxygen changes [3]. Based on this characteristic, we can measure the total voltage changes during the sensors absorbing the tested odor.

Table I List of MOGSs from the FIS Inc.

Sensor Model	Main Detection Gas
SP-53	Ammonia, Ethanol
SP-MW0	Alcohol, Hydrogen
SP-32	Alcohol
SP-42A	Freon
SP-31	Hydrocarbon
SP-19	Hydrogen
SP-11	Methane, Hydrocarbon
SP-MW1	Cooking vapor

Since the MOGS is sensitive to the temperature and the humidity, the MOGSs unit is put in a small chamber that has a heating system to increase the air temperature during winter season. The heating unit can also decrease the air humidity in the chamber. The clean water is manually sprayed into the chamber when the humidity drops lower than the control level. In this experiment the temperature in the chamber is kept between 20-30°C and the humidity is kept between 30-40%RH. The tested smell is sucked to mix with the fresh air before passing to the MOGSs unit. The distance from the tested smell to the MOGSs unit is approximately 1.5 m.

3. Experimental Data Collection

The smell from twelve sources of fire listed in Table II are measured by the EN system explained in previous section. Each source of fire has been tested with forty repetition data measured in different days in order to check the repeatability response of the MOGSs to the same smell.

For each data, the voltage signal of the normal air is measured every second for one minute and its average value, $\bar{v}_{air}$, is used as an air reference point. After that, the voltage signals of the sensors when absorbing tested smell, $v_{smell,t}$, are collected every two minutes for each smell sample. Finally, the total change in signals at each period, $V_{smell,t}$, is calculated by

$$V_{smell,t} = v_{smell,t} - \bar{v}_{air}$$

where t is the time from 1 to 120s.

After testing one smell the MOGSs need to be cleaned by removing the tested smell and supplying only the fresh air until the MOGSs return to stable point before testing the new sample. This process is just like the human nose which need to breath the fresh air before able to recognize the new smell accurately. Some time series data from the experiment in Fig.2 show that all smells approach the saturation stages within the measuring periods. The signals from the same source of fire in every repetition data are similar in most data sources. The results using the BP method and the k-means algorithm to analyze the time series data from each source of fire every two seconds and the average signals during the saturation stages(time 100-120s) are discussed in Section V.

Table II List of Burning Materials in the Experiment

Sources of fire	Abbreviation
Steam from boiling water	Steam
Burning joss stick	Joss
Burning mosquito coil	Mos
Aroma oil	Aroma
Aroma candle	Candle
Flame from liquid petroleum gas(LPG)	Flame
Leakage of LPG	LPG
Steam from Japanese soup called "oden"	Oden
Boiling vegetable oil	Oil
Toasted bread	Toast
Burning paper	Paper
Burning wood	Wood

4 Correlation of the Experimental Data

Before classifying each source of data, the correlation of each data source is investigated by using the similarity index (SI) and the principal components analysis (PCA).

4.1. Similarity index

In the statistical application, the correlation value developed mainly by Karl Pearson is widely used to find the relationship between two random variables. In this paper, we call the correlation value as a similarity index (SI). The SI value varies from -1 to 1. Two random variables with a SI of either 1 or 1 are highly correlated because knowledge of one provides precise knowledge of the other. However, the SI provides information only about linear relationships between random variables.

Random variables could have a nonlinear relationship but still have a SI close to 0 [4]. Therefore, we make an assumption on this application that each data pattern has nearly linear relationship to the other data patterns. The SI value between two data is calculated by

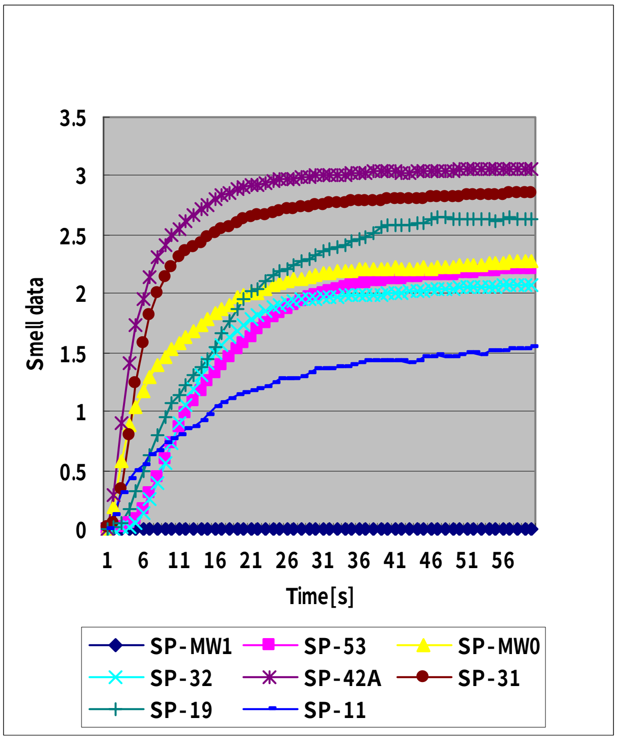

Fig.2 Time series data from some sources of fire in the experiment.

$$r_{xy} = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sqrt{(\sum_{i=1}^{n} x_i^2 - \bar{x}^2)(\sum_{i=1}^{n} y_i^2 - \bar{y}^2)}}$$

where r_{xy} is the SI value, x and y are the comparing data, $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$, and n is the size of each data which equals 480 (60 periodsx8 sensors).

By using the SI to find the relationship between the repetition data of each data source, we found that all data sources except the paper and the wood have high average SI values above 0.99. During the experiment, the paper and the wood have inconsistent burning rates, therefore the signals from the repetition data of these sources are more fluctuated than the other sources that have better consistent burning rate.

4.2 Principal components analysis

In this paper, the well known PCA is applied to analyze two cases of the experimental data. The full time series data case uses the data signals every two seconds, but the saturation stage data case uses only the average data from time 100 to 120s for analyzing.

The plots of two main components are shown in Fig.3. The distribution of the paper and the wood burning smell are more scattered than the other kinds of smells especially in the case of saturation stage data. Most of the tested data are separated into their own clusters with some overlap zones between different data source.

5. Experimental Results and Discussion

5.1 Experimental result

Two case of data are analyzed by the BP method and k-means algorithm. The full time series data (TSD) case uses the data from all MOGSs every two seconds as the input data. The saturation stage data (SSD) case uses only the average value from time 100 to120s of all sensors as the input data.

The BP structure contains three layers. The input layer of the TSD case, and the SSD case consists of four hundred eighty nodes (8 sensors x 60 periods), and eight nodes(average signal from 8 sensors), respectively. For the hidden layer, we have tried with several values and the size that gives a good accuracy and a reasonable training time for both data cases is forty nodes. The output layer contains twelve nodes, each node represents one data source. The learning rate, the momentum rate, and the minimum mean square error (MSE) during the training period are set by trial and error method to 0.1, 0.001, and 0.0003, respectively.

Based on the information during investigating the correlation of the data, most data sources are highly correlated to their repetition data with high SI values. Therefore, only one data that has the highest average SI value to the other repetition data from each sources of fire are used as the training data for the BP and the rest of the data are used as the test data. We assume that a pattern is classified correctly if (output $\leq$ 0.7 and target=1) or (output $\leq$ 0.3 and target =0). For the k-means algorithm, the training data of the BP method are used as the initialize data and then assigns the data patterns to the nearest cluster center by calculating the Euclidean distance. After that, the new cluster center is recalculated. The process continues until the position of the cluster center is not changed. The final results of this experiment are shown in Table III.

The results using the TSD from both the BP method and the k-means algorithm are sufficiently effective. The data signals from the MOGSs are affected by many factors, such as the sampling condition, the inconsistency burning rate, the fluctuation from the standard air, and so on. Therefore, the saturation stages of the data are varied by those factors. By including the signal before approaching the saturation stage, the accuracy to classify all smell is increased.

5.2 Discussion

Although the distribution of PCA shown in Fig. 3 cannot clearly separate similar smell such as the aroma oil and the aroma candle, the BP method and the k-means algorithms are able to classify them perfectly as shown in Table III. The results of TSD using the BP method have only two incorrect classified data. These two data are not misclassified as the other smells. Only the output values of their paper node are not high enough to classify them as the paper. The output values of these two data on the paper node are only 0.4951, and 0.4799, respectively and the output of the others output nodes are nearly 0. The results are much better than the results from [1] which used two kinds of MOGSs to classify several sources of fire into three fire condition levels, flaming, smoldering, and nuisance, with only 85% of correct classification. The smoke density of the tested data is not high enough to trigger the alarm of the smoke detector. In case of unusual burning smells in the residences such as the wood burning, flaming from the LPG, or the leakage LPG, it is necessary to have a proper device to detect these sources before unable to stop the fire. We can conclude that the new EN system shown in this paper is a proper device for this application.

6. Conclusions

We have proposed a new EN system designed from various kinds of MOGSs. The EN has the ability to identify various sources of fire in the early stage with more than 99% of accuracy by using only a single training data in the BP case. The results from the k-means algorithm are also able to predict the sources of fire with more than 98% of accuracy. It can be concluded that the EN is suitable for detecting the early stage of fire.

References

[1] Jame A.Mike, "Applicaiton of Neural Networks for Discriminating Fire Detectors", International Conference on Automatic Fire Detection, AUBE'95, 10th, Duisburg, Germany, pp.213-222, April 4-6, 1995.

[2] Bancha Charumporn, Michifumi Yoshioka, Toru Fujinaka, and SigeruOmatu, "An Electronic Nose System Using Back Propagation Neural Networks with a Centroid Training Data Set", proceeding of the Eight International Symposium on Artificial Life and Robotics, Japan, pp.605-608, January 24-26, 2003

[3] General Information for TGS sensors, Figaro Engineering, available at www.figarosensor.com/products/general.pdf. M. Tanaka, "Modeling of Mixtures of Principal Component Analysis Model with Genetic Algorithm," Proc. 31st Int'l Symposium on Stochastic Systems Theory and Its Applications, pp. 157-162, Yokohama, Japan, 1999.

[4] William L. Carlson and Betty Thorne, Applied Statistical Methods, Prentice Hall International, 1997

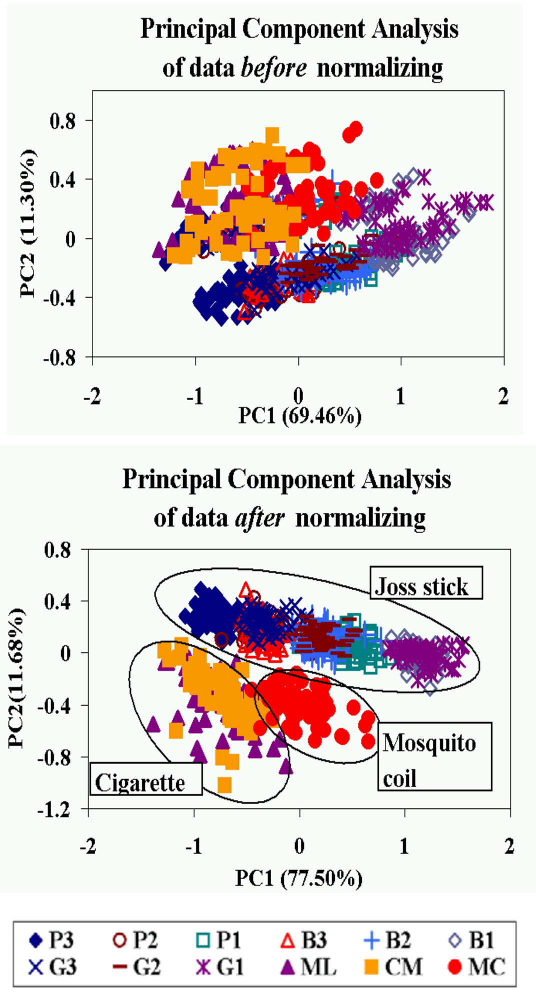

Fig.3. Two main components of the experimental data using the PCA.

Table III Experimental Results

Sources	TSD				SSD			
	BP		k-means					
	True	%	True	%	True	%	True	%
Steam	39/39	100	40/40	100	38/39	97	39/40	98
Joss	39/39	100	40/40	100	39/39	100	40/40	100
Mos	39/39	100	40/40	100	39/39	100	40/40	100
Aroma	39/39	100	40/40	100	39/39	100	40/40	100
Candle	39/39	100	40/40	100	39/39	100	40/40	100
Flame	39/39	100	40/40	100	39/39	100	40/40	100
LPG	39/39	100	40/40	100	39/39	100	40/40	100
Oden	39/39	100	40/40	100	39/39	100	40/40	100
Oil	39/39	100	40/40	100	38/39	97	37/40	93
Toast	39/39	100	40/40	100	38/39	97	40/40	100
Paper	38/39	95	35/40	88	31/39	80	28/40	70
Wood	39/39	100	37/40	93	32/39	82	28/40	70
Average		99		98		96		94