Improving the Classification Accuracy in Electronic Noses Using Multi-Dimensional Combining (MDC)

Hong Chen*, Rafik. A. Goubran*, Tofy Mussivand**

* Dept. of Systems and Computer Engineering,
Carleton University,
1125 Colonel by drive
Ottawa, ON, Canada, K1S 5B6
hchen@sce.carleton.ca
goubran@sce.carleton.ca

** University of Ottawa Heart Institute,
40 Ruskin Street
Ottawa, ON, Canada,
K1Y 4W7
TMussivand@ottawaheart.ca

Abstract

Traditional pattern recognition (PARC) methods used in Electronic noses (Enose) are either parametric (such as K-nearest neighbors (KNN), Linear Discriminant Analysis (LDA)) or non-parametric (such as artificial neural network and fuzzy logic). In this paper, Multi-Dimensional Combining (MDC) is proposed to combine the classification outputs of individual classifiers into a more robust and accurate one. Two implementations are proposed to find the individual classifiers, one is based on various feature extraction and the other is based on various dimension reduction method, with three means of combining. Six house-hold fragrances were sampled using the Cyranose 320 Enose device. The acquired data (600 measurements) was split into two sets, training and testing. Experiment were conducted at various concentrations of the sample smell, various sample numbers and various training numbers. Results show the advantage of MDC over the individual classifiers, and over the other traditional PARC methods under all conditions.

Keyword

Electronic nose, MDC, multiple classifier combining, feature extraction, dimension reduction, pattern recognition

1. INTRODUCTION

The sense of smell is a primal sense for humans as well as animals. Smell (or olfaction) allows vertebrates and other organisms to identify food, mates, predators, and provides both sensual pleasure (the odor of flowers and perfume) as well as warnings of danger (e.g., spoiled food, chemical dangers). [1] Conventional approaches of measuring smell senses involve the use of either human panels or analytical instruments (such as gas chromatographs and mass spectrometers). [2]

Due to the emerging requirement for qualitative, low-cost, real-time, and portable equipment to perform reliable, objective, and reproducible measures of volatile compounds and odors, a growing amount of research has been directed towards the development of instruments able to mimic the human olfactory apparatus. Such an instrument is called an electronic nose (Enose), an instrument used for the automated detection and classification of odors, vapors and gases. [3]

Although traditional pattern recognition algorithms perform well for simple classification tasks and under a large number of training samples, performance is degraded significantly when the available training samples are limited or when the classification task is considerably complex. Unfortunately, due to the high cost and long time duration of the sampling procedure in certain research areas (such as the sampling of the patient's breath in hospital for the purpose of medical diagnosis), only a small number of training samples are available. In addition, the smell difference between such clinical smell samples is quite subtle, therefore advanced signal processing and pattern recognition methods are required to successfully discriminate the smell samples.

2. MULTI-DIMENSION COMBINING (MDC)

Although different signal processing and pattern recognition methods have been tested on the classification problem, no single method appears superior to others in all cases. An intuitive and natural way to improve the classification accuracy is therefore to synthesize the classification output from various individual classifiers and combine this into a more robust and accurate one. A necessary and sufficient condition for a combined result to be more accurate than any of its individual members is that the individual classifiers must be both relatively accurate and independent.

There are three qualitative explanations on why multiple classifier combination in general works better than the best single classifier in the hypothesis space (say **H**). [4] The first explanation is that the limited training data may not provide sufficient information for choosing the best classifier from space **H**. Single classifiers based on these limited training samples are prone to misclassify on their decision boundary. The multiple classifier combination is unlikely to have similar decision boundaries for the different individual classifiers and therefore the performance will have less variance and be more robust. [4]

The second explanation depends on the random nature of the search in space **H**. For example, finding the best weights from all the possible candidates for a neural network consistent with a set of training examples is very difficult. Instead, neural network algorithms use random initializations and local search methods. The result is normally locally minimum and sub-optimum. Different initializations and search directions in general give rise to different suboptimal sets of weights. Multiple classifier combination could be regarded as a way of compensating for this imperfect search algorithm by trying different initialization states and search directions. [4]

The third explanation is that if the true function is outside of **H,** the multiple classifier combination may approximate it better than any individual hypothesis **h**. This is due to the fact that weighted sums of hypotheses may also lie outside of **H**, while each individual classifier is bound to the space **h**. [4]

Generally speaking, the output information that various classifiers apply can be divided into three levels: abstract level, rank level and measurement level. [5] Among the three levels, the measurement level contains the highest amount of information and the abstract level contains the lowest. As the Multilayer Perceptron (MLP) pattern recognition method provides the output at measurement level, we choose the MLP as the universal pattern recognition method for all individual classifiers.

Two different methods to retrieve qualified candidate classifiers are employed. The first one creates the individual classifier from different feature extraction methods (figure 1) and the second one from different dimension reduction methods (figure 2), as illustrated below.

Feature extraction (FE) methods include: [6]

FE1: steady state, the maximum value from the total response curve,

FE2: integral of the transient during the rising time,

FE3: Windowed time slice, where the transient response is multiplied by several smooth, bell-shaped windowing functions and integrated with respect to. The idea is to capture some information about the dynamic characteristics of the response [7].

FE4: slope of the rising time, or the steady state divided by the rising time,

FE5: slope of the falling time,

FE6: integral of the falling time,

Dimension reduction (DR) methods include:

DR1: Principle Component Analysis (PCA), a multivariate statistical method expressing the response vectors in terms of linear combinations of orthogonal vectors along a new set of coordinate axes, particularly useful in representing the high dimensional data into low dimension. [3]

DR2: Independent Component Analysis (ICA), a statistical method to find a new coordinate system that makes the original signals as independent as possible, using higher order statistics from the probability densities of the data. [8]

DR3: Multiple Linear Discriminant (MLD), a statistical method trying to yield the maximum ratio of between-class scatter to within-class scatter in more than two classes. [3]

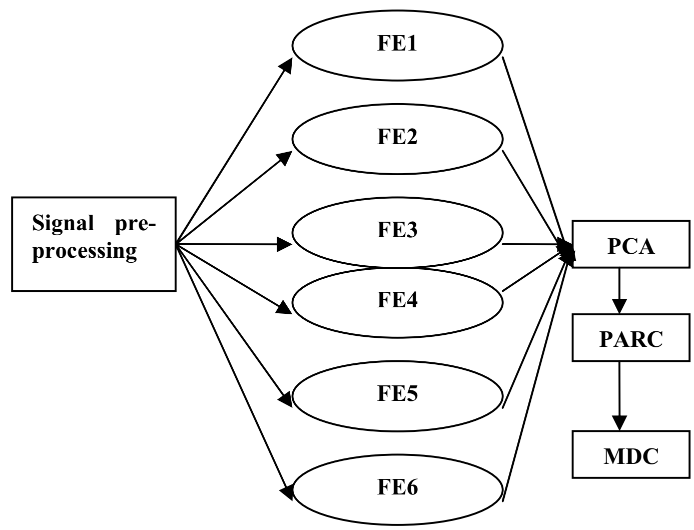

Figure 1 MDC using different feature extraction methods

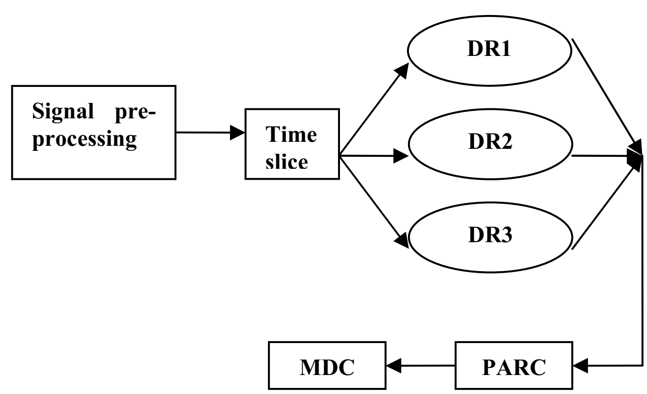

Figure 2 MDC using different dimension reduction methods

Three combination methods are proposed, namely arithmetic mean average (MDC1), geometric mean average (MDC2) and squared mean average (MDC3).

3. EXPERIMENTAL SETUP

Six common household fragrances, as shown in Table 1, were employed as the smell samples in the experiment. Cyranose 320, a handheld Enose instrument made by Cyrano Sciences Inc (USA), was used to collect the raw sensor responses. Each smell was sampled 100 times in random sequence. The raw data were transferred and stored in PC for further signal processing and pattern recognition. The entire dataset (600 sample files, each file have 32 sensor responses, and each sensor response contains 101values) were split into training and testing sets and the classification accuracy of the testing set was used as the criteria for performance evaluation.

Table 1 Fragrances samples used in the experiment

Sample #	Brand	Product
#1	Brut	Splash-on lotion
#2	Colt	Toilet spray
#3	Ice	Sport after-shave
#4	Williams	Electric shave
#5	Lady speed stick	Antiperspirant
#6	Nivea	Body moisturizer

4. EXPERIMENTAL RESULT

The first experiment is on the sensitivity improvement of the MDC method. Figures 3 and 4 show the performance evaluation under both high and low concentration of the smell samples for feature extraction and dimension reduction methods, respectively. Under all circumstances, MDC outperforms the individual classifier with an average 5%-20% increase in the classification accuracy.

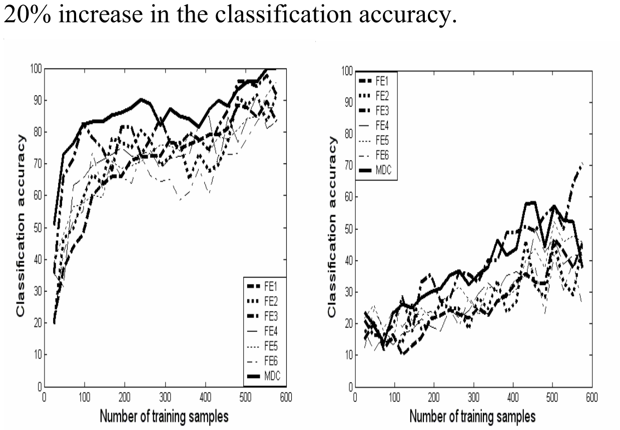

Figure 3 MDC compared with various feature extraction methods, left --high concentration, right -- low concentration

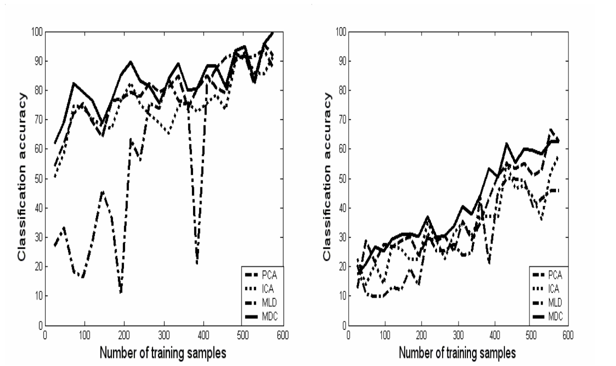

Figure 4 MDC compared with various dimension reduction methods, left – high concentration, right – low concentration

Another experiment was conducted in the selectivity improvement of the MDC method. Figures 5 and 6 are the performance evaluation under various numbers of smell samples for feature extraction and dimension reduction methods, respectively. The classification accuracy increases as the number of smell samples decreases. Still, MDC outperforms each of the individual classifier.

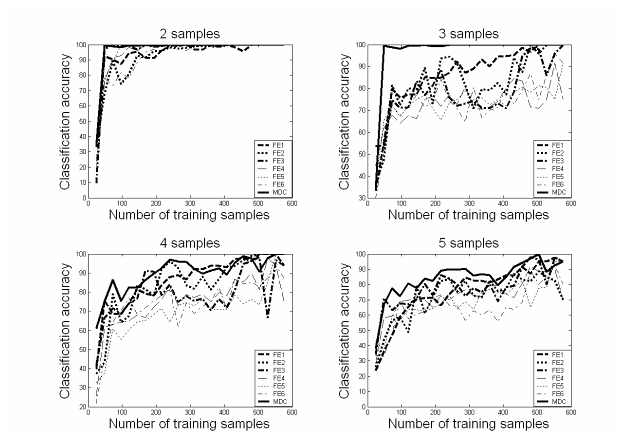

Figure 5 Performance of selectivity (feature extraction methods)

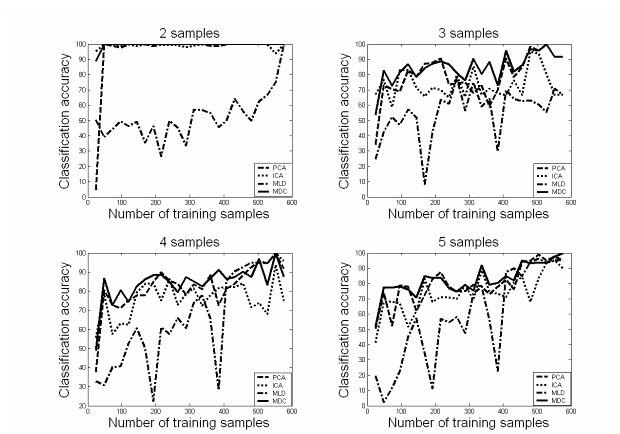

Figure 6 Performance of selectivity (dimension reduction methods)

Further experiments compare the performance of MDC with three popularly used pattern recognition (PARC) methods, namely KNN (K Nearest Neighbors), LDA (Linear Discriminant Analysis) and PNN (Probabilistic Neural Network). Figure 7 illustrates that the performance of MDC is superior to PNN and LDA and slightly improved over KNN. It is concluded that MDC techniques in MLP pattern recognition method can increase the overall classification accuracy, which could not be achieved by using any single individual classifier.

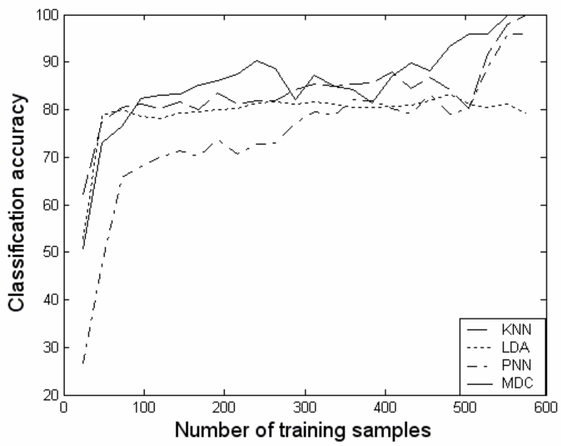

Figure 7 Performance evaluations of various pattern recognition methods

5. CONCLUSIONS AND FUTURE WORK

This paper investigated the ability of the Multi-Dimension Combining (MDC) method to improve the classification accuracy of the Enose system, and compared the performance of MDC with the performance of each individual classifier—either individual feature extraction or individual dimension reduction method. The performance of the MDC with other pattern recognition methods such as KNN, LDA and PNN was also compared. MDC improves the classification accuracy in both sensitivity and selectivity experiments, and also outperforms three normally employed pattern recognition methods.

Future work to improve the performance of the Enose system could rely on two developments. First is the optimization and improvement of the existing signal processing and pattern recognition methods. Fuzzy logic, genetic algorithm and adaptive noise canceling are among the several candidates for further research. The other alternative is to improve the proposed MDC methods. One suggested improvement is the use of nonlinear MDC methods rather than the linear ones proposed here. Another possible way is to combine the information collected from the "electronic tongue" and "electronic nose" using the proposed MDC methods. "Electronic tongue" is equipped for liquid sample discrimination based on pulse voltammetry. [9] Mimicking the human's capability of synthesizing the signal from both nose and tongue, it is expected that the fusion of "Enose" and "Etongue" using various MDC methods will achieve better performance.

Acknowledgement

This work was performed at the department of Systems and Computer Engineering (SCE), Carleton University, in cooperation with the University of Ottawa Heart Institute. Financial support from Ontario Research and Development Challenge Fund and NSERC are gratefully acknowledged.

REFERENCES

[1] T.C.Pearce, S.S.Schifffman, H.T.Nagle, J.W.Garden, *Handbook of machine olfaction—Electronic Nose Technology,* WILEY-VCH, 2003.

[2] J.Brezmes, N.Canyellas, E.Llobet, X.Vilanova, X.Correig, "Application of artificial neural networks to the design and implementation of electronic olfactory systems," NEUREL 2000. Proceedings of the 5th Seminar on, pp. 75 – 80, September 2000.

[3] Richard O. Duda, Peter E. Hart, David G. Stork, *Pattern classification*, Wiley Interscience, 2001.

[4] Matteo Pardo and Giorgio sberveglieri, "Learning from data: A tutorial with emphasis on modern pattern recognition methods," IEEE Sensors Journal, vol. 2, no. 3, pp. 203-217, June 2000.

[5] Lei Xu, Adam Krzyzak, Ching Y. Suen, "Methods of combining multiple classifiers and their applications to handwriting recognition," IEEE transaction on systems, man and cybernetics, vol.22, no.3, pp. 418-435, May/June 1992.

[6] Graham E. Searle, Julian W. Gardner; "System identification of electronic nose data from Cyanobacteria experiment," IEEE Sensors Journal, vol. 2, no. 3, pp. 219-229, June 2002.

[7] Ricardo Gutirrez-Osuna and H. Troy Nagle, "A method for evaluating data-preprocessing techniques for odor classification with an array of gas sensors," IEEE transactions on systems, man and cybernetics-Part B: Cybernetics, vol.29, no.5, pp. 626-632, October 1999.

[8] Martin Kermit and Oliver Tomic, "Independent component analysis applied on gas sensor array measurement data," IEEE Sensor Journal, vol. 3, issue 2, pp. 218-228, April 2003.

[9] Peter Wide, Frederik Winquist, Pontus Bergsten, Emil M Petriu, "The human-based multi-sensor fusion method for artificial nose and tongue sensor data," IEEE Instrument and Measurement Technology Conference, St. Paul, Minnesota, USA, May 1998.