IMTC 2005 – Instrumentation and Measurement
Technology Conference
Ottawa, Canada, 17-19 May 2005

Multi-Dimension Combining (MDC) in abstract Level and Hierarchical MDC (HMDC) to Improve the Classification Accuracy of Enoses

Hong Chen[1], Rafik. A. Goubran[1], Tofy Mussivand[2]

[1] Department of Systems and Computer Engineering, Carleton University,
1125 Colonel by Drive, Ottawa, ON, Canada, K1S 5B6
Email: hchen@sce.carleton.ca
[2] University of Ottawa Heart Institute, 40 Ruskin Street
Ottawa, ON, Canada, K1Y 4W7

Abstract *– This paper proposes a new classification algorithm for improving the accuracy of Electronic Noses. The algorithm extends the conventional Multi-Dimension Combining (MDC) of measurement level (PARC method as Multi-layer perceptron, or MLP) into abstract level (PARC methods as K-Nearest Neighbor (KNN), Linear Discriminant Analysis (LDA) and Probabilistic Neural Network (PNN)) and Hierarchical level (HMDC, or Hierarchical Multi-Dimension Combining). The performance of the proposed algorithm is evaluated using experimental data and Enose device of Cyranose 320.*

Keyword – MDC (Multi-Dimension Combining), HMDC (Hierarchical MDC), electronic nose, pattern recognition (PARC), feature extraction, dimension reduction

I. INTRODUCTION

Electrical nose (Enose) is an instrument used for the automated detection and classification of odors, vapors and gases. [2] It has four components, a smell sample delivery system, a sensor array, signal processing techniques and pattern recognition (PARC) methods. While most PARC methods perform well for simple classification tasks and large training sets, the performance degrades significantly for more complex classification tasks or under smaller training sets. For smell classification tasks such as in clinical experiments, the inconvenience, high-cost and long time duration of the sampling process practically eliminates the possibility of obtaining large training sets, also the difference between the clinical smell samples is subtle, consequently the improvement of classification accuracy under such circumstances will have considerable practical importance in Enose application.

MDC (Multi-Dimension Combining) in measurement level (Multi-layer perceptron, or MLP) have been proved effective to improve the classification accuracy of electronic noses (Enoses). [1] In this paper, MDC is experimented on abstract level, or specifically on the three popularly employed PARC methods of K-Nearest Neighbor (KNN), Linear Discriminant Analysis (LDA) and Probabilistic Neural Network (PNN). To cope with the difficulty of abstract level, each individual classification output is converted into binary representation, and MDC method proposed in [1] are employed to determine the final decision. Further experiments are conducted on the Hierarchical MDC (HMDC) which combines the classification outputs of the three above mentioned PARC methods in abstract level (KNN, LDA, PNN), as well as the PARC method in measurement level (MLP) into a more robust and accurate classifier.

II. MDC IN ABSTRACT LEVEL AND HMDC

When MDC is applied in measurement level (such as in MLP PARC method) [1], the final decision of each individual classifier, either feature extraction (FE) or dimension reduction (DR) method, represents the confidence level of that specific method – or a numeric valued between [0,1], therefore a simple math average method (mathematic mean, geometric mean, and squared mean) could be employed to combine the individual classification result. Unfortunately, the same idea can not be easily applied to abstract level (PARC methods such as KNN, LDA and PNN), as the outcome is not a confidence level numeric but a class index integer. Therefore simple math average would not be appropriate to represent the final combined decision.

To overcome the difficulty, each outcome from the individual feature extraction or dimension reduction method is first encoded in its binary representation, and then the MDC method proposed in [1] is applied on these binary digits. The essence is to use majority vote of individual methods to maximally reduce the classification error.

Furthermore, convinced by the fact that each PARC method, whether in measurement level or in abstract level, have their own advantages and none of them could claim to be superior to others on all occasions, it is reasonable to combine the final outcomes of these individual PARC methods in both measurement level and abstract level into a more accurate and efficient one.

III. EXPERIMENT SETUP

Six house-hold fragrances (same as used in [1]) are sampled using Enose device Cyranose 320. Same signal processing techniques (including baseline manipulation, low pass filtering, feature extraction and dimension reduction) are used as [1] for performance comparison purpose. However to focus on the small training set classification problem, the dataset split is a little different from [1]. From total 100 measurements for each class, variety number of samples (from 10 to 40) are used for training purpose, another 20 samples are used for optimizing the signal processing and pattern recognition parameters, and the left 40 samples are used for final classification verification and performance comparison.

The parameters to be optimizes include parameters for Savitzky-Golay (polynomial) FIR filter, PCA (Principle component Analysis) number and ICA (Independent Component Analysis) number, as well as hidden layer number and number of neurons in each layer in Multi-Layer Perceptron (MLP) analysis. Both high and low concentration of the six fragrances were sampled and investigated for the performance evaluation.

Figure.1 and Fig.2 show block diagram of MDC and Hierarchical MDC (HMDC), respectively.

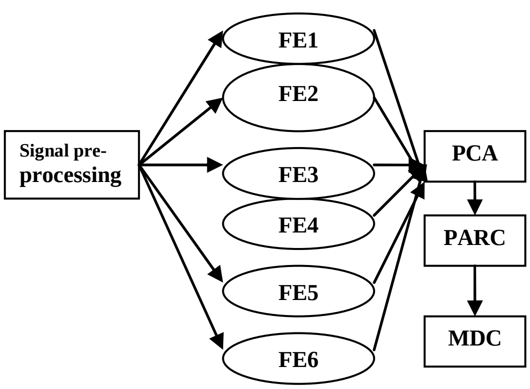

Figure.1. Block diagram of MDC in abstract level

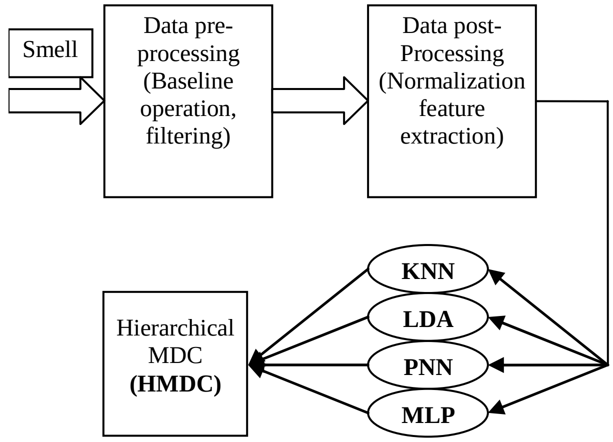

Figure.2. Block diagram of HMDC

IV. EXPERIMENT RESULTS

Figure.3 and Figure.4 are performance of MDC and HMDC compared with each individual classifier (either feature extraction or dimension reduction methods, respectively), under high concentration of the smell samples. Figure.5 is the performance of MDC and HMDC for various dimension reduction methods under low concentration of the smell samples.

Feature extraction (FE) methods include: [3]
FE1: steady state, the maximum value from the total response curve,
FE2: integral of the transient during the rising time,
FE3: integral of the falling time,
FE4: 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 [4].
FE5: slope of the rising time, or the steady state divided by the rising time,
FE6: slope of the falling time,

Dimension reduction (DR) methods include:
DR1: Principle Component Analysis (PCA), [5]
DR2: Independent Component Analysis (ICA), [6]
DR3: Multiple Linear Discriminant (MLD), [5]

Several observations could be made from these plots:
a) In general MDC in abstract level shows only little advantage over individual classifier, or not as successful as in measurement level. This is partly because we deliberately use limited training samples (10%--40% of total samples) to predict the classification accuracy of each classifier, thus decreasing the accuracy of each individual classifier. It may also due to the fact that MDC is more accurate in

measurement level than abstract level since measurement level represents the confidence level of each individual classifier, while abstract level is only a representation symbol.

b) Under higher concentration of smell samples, HMDC (line) usually outperform the individual classifiers, or at least at the same level as the best of the individual classifier, in both feature extraction and dimension reduction methods. Since the individual classifier with the best performance is not fixed in all the PARC methods, it is usually hard to determine the best classifier before hand, while HMDC in general guarantee at least the best performance of all the individual classifier, if not better.

c) The most significant improvements occur in MLP with HMDC methods. This is because Multi-Layer Perceptron achieves the least accuracy in the four PARC methods (therefore improves the most from all other higher accuracy PARC methods), it may also due to the fact that MDC is more accurate in measurement level as long as accuracy and independency of each individual classifier is provided.

d) The advantage of MDC and HMDC is more obvious in feature extraction (FE) methods than in dimension reduction (DR) methods. This might be the result of more feature extraction methods (6 in total) are available than dimension reduction methods (3 in total), it also might be the result that individual accuracy and independency in dimension reduction is not as high as that in feature extraction.

e) Under lower concentration of sample smell (Fig 5), the improvement in performance is not as obvious as in high concentration. This is not out of over expectation, as individual accuracy and independency is the requirement for the success of the MDC over individual classifier, which is obviously not obtained in the low concentration circumstances.

V. CONCLUSION AND FUTURE WORK

This paper continues the experiments in MDC in abstract level as well as the Hierarchical MDC (HMDC) from [1], to further improve the classification accuracy of the Enose system. It is observed that under any concentration of smell samples, HMDC could achieve the best classification accuracy, although under higher concentration the performance achievement is more obvious. While MDC in abstract level shows only little advantage over individual classifier, and is not as successful as that in measurement level, it is the principle component in HMDC and therefore has its own significance.

Future work to improve the performance of the Enose system could rely on two developments. One possible solution is to use adaptive filter instead of Savitzky-Golay (polynomial) FIR filter, as the variability of environment has significant effect on detection ability of the Enose. It is predicted that continuously and adaptively remove the variable noise from the measurement will increase the final classification accuracy. Another interest will be on the feature selection methods, which could reduce the high dimension of the final features to avoid the “curse of dimension”.

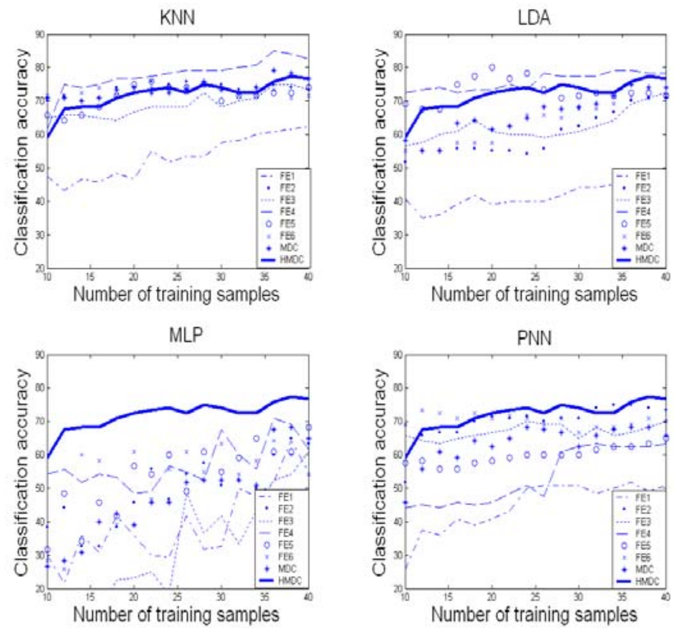

Figure.3. Performance of HMDC, MDC with various feature extraction (FE) methods in high concentration

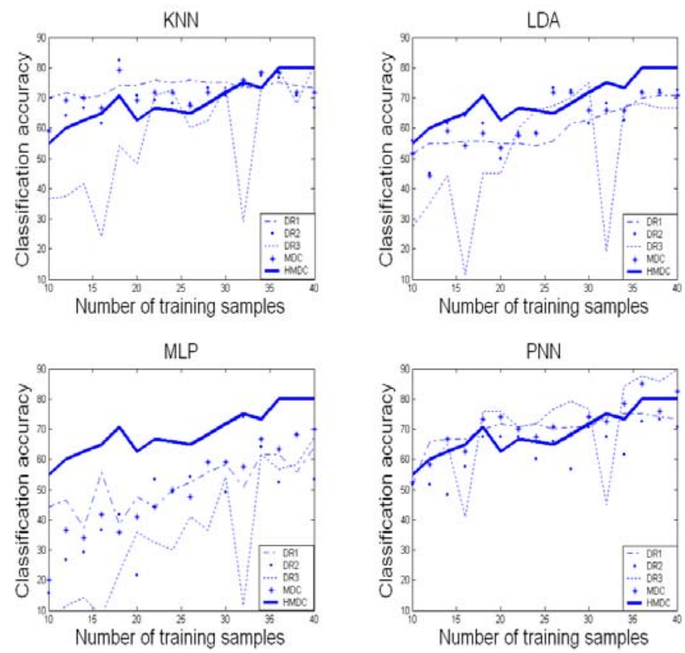

Figure.4. Performance of HMDC, MDC with various Data Reduction (DR) methods in high concentration

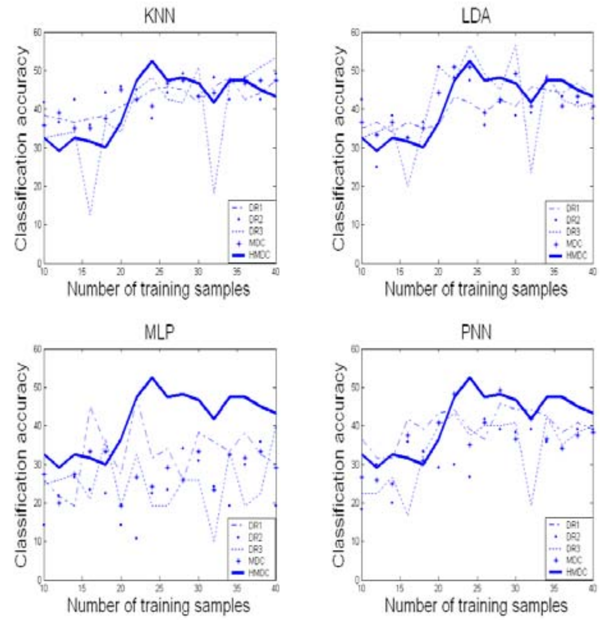

Figure.5. Performance of HMDC, MDC with various Data Reduction (DR) methods in low concentration

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] Hong Chen, Rafik. A. Goubran and Tofy Mussivand, "Improving the Classification Accuracy in Electronic Noses Using Multi-Dimensional Combining (MDC)," IEEE sensors 2004, Oct 2004

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

[3] 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.

[4] 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.

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

[6] 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.