'96 대한전기학회 하계학술대회 논문집 (1996. 7. 22~24)

퍼지 Clustering 알고리즘을 이용한 휘발성 화학물질의 분류

°변 형기* 김 갑일*
*명지대학교 전기 · 전자공학부

Classification of Volatile Chemicals using Fuzzy Clustering Algorithm

Byun, Hyung Gi and Kim, Kab Il
School of Electrical and Electronics Eng.
Myong-Ji University

Abstract *-The use of fuzzy theory in task of pattern recognition may be applicable gases and odours classification and recognition. This paper reports results obtained from fuzzy c-means algorithms to patterns generated by odour sensing system using an array of conducting polymer sensors, for volatile chemicals. For the volatile chemicals clustering problem, the three unsupervise fuzzy c-means alogrithms were applied. From among the pattern clustering methods, the FCMA2 algorithm, which updated the cluster centres more frequently, consistently outperformed. It has been confirmed as an outstanding clustering algorithm throughout experimental trials.*

1. Introduction

Odour perception within the animal kingdom is still not fully understood, and even less is known of odour recognition in the brain of the animal. It is generally accepted that a need exists for sensing systems that mimic the action of biological olfaction in applications such as quality control of foods, environmental monitoring, raw material quality, and some chemical fields. Many of these systems have been given the name "Electronic Nose" because they are capable of or intended to sense some of the gases or odours that the human nose can discriminate, and they are based on broad specificity sensors.

Over the last decade Persaud and coworkers[1] have developed organic conducting polymers as odour sensing devices, and many materials have been synthesised and characterised for odour transduction. The strategies involved are a radical departure from conventional thinking of chemical sensors since they utilise broadly selective sensors. Different polymers made from modified monomer units show broad overlapping response profiles to different volatile compounds. Sensor arrays consisting of up to twenty different conducting polymer materials are currently being commercialised. Each sensor element changes in resistance when exposed to volatile chemical species. The signals produced by an array of sensors based on conducting polymers consist of measurements of responses to odours producing different patterns that are projected into multi-dimensional space.

A variety of pattern recognition techniques may be utilised. These include neural networks that take the input patterns generated by the array of sensors, and may be trained to associate these patterns with particular classes of volatile chemicals that may be of interest to the user. Such architectures include 3-layer systems trained by back-propagation of errors and fuzzy systems.

This paper presents the use of fuzzy set theory in the task of pattern recognition generated by odour sensing systems, for volatile chemicals. Of particular interest of us is it may be possible to distinguish when there is overlap between classes. This idea is not to ask whether a volatile chemical pattern belongs to a certain class, but rather to ask what is the membership grade that a volatile chemical has relative to a class, for all possible classes. Consequently, the value of the membership grade of volatile chemical pattern is numerically equal to the value of the possibility that expresses the possibility that the volatile chemical pattern might be an instance of a certain class.

Fuzzy c-means algorithms, which cluster patterns on optimising a fuzzy objective function, introduced criteria to distinguish patterns between clusters. In the conventional clustering methods, all input patterns have the exact membership grades for the cluster centres, where the membership grades of each input pattern indicate the possibility distribution amongest the classes. For the volatile chemicals clustering problem, three different fuzzy c-means algorithms are applied, where the fuzzy c-means algorithm type 2 (FCMA2), which updated the cluster centres more frequently, consistently outperformed the two algorithms throughout experimental trails.

2. Electronic nose system based on an array of conducting polymers

One of useful set of materials that be utilised as sensors in a form of an "Electronic Nose" is that of electrically conducting organic polymers. Persaud and Pelosi[2] have investigated the sensing properties of a large number of conducting polymers. These polymers display reversible changes in electrical resistances when volatile molecules adsorb and desorb with greater sensitivity towards polar molecules. Unlike many commercially available gas sensors, rapid adsorption and desortion kinetics are observed at ambient temperatures.

The changes in resistance of each sensor element following odour stimulation were recorded through a microcontroller based circuit and transmitted through a serial interface to an external personal computer used for data display and information processing. The array response is normalised to represent relative changes in resistance and thus concentration-independent patterns can be produced. These patterns show the relative response of individual sensor elements to specific gases and odours, and are used as inputs to discriminate between gases and odours using pattern recognition technique based on fuzzy system.

3. Fuzzy cluster analysis

A pattern is "between" clusters in the sense that it could have been classified as belonging to one cluster almost as well as to another. In this case, a fuzzy clustering method can give an increasingly accurate representation of the pattern. The most popular fuzzy clustering algorithm is based on the work of Bezdek[3] and Zimmermann[4], while two modified algorithms for fuzzy clustering, proposed by Kamel and Selin[5] show improved speed and convergence characteristics.

The fuzzy clustering problem can be started as the partitioning of a finite set of patterns

$$X_i = \{x_1, x_2, \ldots, x_n\} \subset R^s, \quad 1 \le i \le n \qquad \text{(Eq.1)}$$

into fuzzy clusters, where R^s is the real space and n is the number of input patterns.

Each partition of the set of patterns X_i into fuzzy cluster centres

$$Z_j = \{z_1, z_2, \ldots, z_c\} \subset R^s, \quad 1 \le j \le c \qquad \text{(Eq.2)}$$

where c is number of cluster centre patterns, can fully be described by W_{ij} the membership grade of pattern i with cluster j subject to the following constraints:

$$(1)\ W_{ij} \in [0, 1]; \qquad 1 \le i \le n, \qquad 1 \le j \le c$$

$$(2)\ \sum_{j=1}^{c} W_{ij} = 1; \qquad 1 \le i \le n$$

(Eq.3)

$$(3)\ 0 < \sum_{i=1}^{n} W_{ij} < n \qquad 1 \le j \le c$$

From the above constranits, patterns can belong to several clusters and to different membership grades. Conditions (2) and (3) in Eq.3 just require that the total membership grade of a pattern is normalised to 1 and that it cannot belong to more clusters than exist.

With constraints of membership grade for each pattern, the fuzzy clustering problem attempts to cluster patterns by searching for local minima of the following objective function:

$$J_m(W,Z)=\sum_{i=1}^{n}\sum_{j=1}^{c} W_{ij}^{m} d_{ij}^{2} \qquad \text{(Eq.4)}$$

where $d_{ij}=\|X_i-Z_j\|$, the Euclidean distance between the input patterns and cluster centres. It measures the similarity (or dissimilarity) of each pattern. Other distance measurements can be used, however the Euclidean distance is used more frequently. Also, $m \in [1, \infty)$ is called the membership weighting exponent. It is usually chosen as $m>1$ for the fuzzy clustering problem, and more details of m will be discussed at a later time. Bezdek[3] proposed the conventional fuzzy c-means algorithm(FCMA1) to solve the Eq.4 as follows:

$$W_{ij}=1\Big/\sum_{l=1}^{c}\left[\left(d_{ij}/d_{il}\right)^{2/(m-1)}\right] \qquad \text{(Eq.5)}$$

$$\text{for } d_{il}>0, \forall i, j$$

$$\text{if } d_{il}=0 \text{ then } W_{ij}=1 \text{ and } W_{ij}=0 \text{ for } j\neq l$$

and

$$Z_j=\sum_{i=1}^{n} W_{ij}^{m} X_i \Big/ \sum_{i=1}^{n} W_{ij}^{m} \qquad \forall j \qquad \text{(Eq.6)}$$

The classical(crisp) problem is involved at $m = 1$, whereas if $m > 1$, the clusters become fuzzy. More fuzziness can be obtained by increasing the value of m. No theoretically justified rule for choosing m exists. Usually m=2 is chosen for most applications.

Unfortunately, the above algorithm takes a long convergence time to fine the best centres. To solve this problem, Kamel and Selim[5] proposed two new fuzzy c-means algorithms for giving fast convergence time. The proposed algorithms update either membership grades or cluster centres more frequently.

In the first proposed algorithm that we adapted and called FCMA2, the cluster centres are recomputed or updated as soon as the membership grades of each patterns are computed. Therefore, by the time the membership grades of all patterns are computed the cluster centres would have been updated at the same time as the number of all the patterns. For this new algorithm, the following definition is needed[5]:

Let $W_r = \{w_{r1}, w_{r2}, \ldots, w_{rc}\}$ be the r-th row of W. W_r gives the membership grades of pattern r. Define problem $P(W_r, Z)$ as follows:

$$\min J(W_r, Z)=\sum_{j=1}^{c} W_{rj}^{m} d_{rj}^{2}$$

$$\text{subject to } \sum_{j=1}^{c} W_{rj}=1,\ W_{rj}\geq 0,\ 1\leq j\leq c \qquad \text{(Eq.7)}$$

The definition means that given cluster centres Z finds the membership grades of pattern r. It can be solved by

$$W_{rj}=1\Big/\sum_{l=1}^{c}\left[\left(d_{rj}/d_{rl}\right)^{2/(m-1)}\right] \qquad \text{(Eq.8)}$$

$$\text{for } d_{rl}>0,\ 1\leq l\leq c,\ 1\leq j\leq c$$

$$\text{if } d_{rl}=0 \text{ then } W_{rl}=1 \text{ and } W_{rj}=0 \text{ for } j\neq l$$

Following each membership grades update, the cluster centres are more frequently updated.

By contrast, in the second proposed algorithm called FCMA3, the membership grades are updated or recomputed as soon as a cluster centre is computed and the problem $P(W, Z_r)$ is defined as follows:

$$\min J(W, Z_r)=\sum_{i=1}^{c} W_{ir}^{m} d_{ir}^{2} \qquad \text{(Eq. 9)}$$

and the solution is given by

$$Z_r=\left(\sum_{i=1}^{n} W_{ir}^{m} X_i\right)\Big/\sum_{i=1}^{n} W_{ir}^{m} \qquad \text{(Eq. 10)}$$

It can be said that the membership grades are more frequently updated followed by cluster centres updating in FCMA3.

These three different fuzzy clustering algorithms, FCMA1, FCMA2, and FCMA3, are used as methods for chemical volatiles classification.

4. Experimental results

When we look at the relative responses of individual sensors to each other in an array, it is observed that unique patterns are generated for each chemical species (Fig.1).

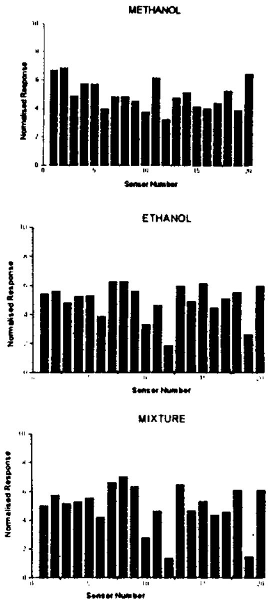

Figure 1. Patterns generated from volatile chemicals

The fuzzy c-means algorithms are applied to voaltile chemicals clustering problems to compare the capability of the algorithms. The data sets used for testing of the fuzzy c-means algorithms were the five volatile chemical patterns: methanol, ethanol, n-butyl acetate, butanone, and a mixture of ethanol and methanol. Each volatile chemicals was tested ten times to provide sample patterns for testing. All fifty sample patterns, which are the output from the twenty sensors, were used within the fuzzy c-means algorithms as input patterns.

In volatile chemicals classification, the cluster centres are initialised by small random values, and the membership grades of patterns are initialised by applied chemical volatile class number with fuzzy constraints for FCMA2. In addition, the behaviour of the stopping criteria could be determined in the fuzzy clustering problem, as the fuzzy clustering is solved by iteration methods. The criteria 0.0001 used for stopping iteration in each algorithm, and this criteria can be determined through the computational experiments.

The clustering result to volatile chemicals using FCMA1 is shown in Table 1. The fifty samples are almost well clustered. However, butanone patterns are divided. Also, an interesting result from Table 1 involves the mixture of chemicals where all the sample patterns of the mixture chemicals are clustered as the 3rd class, the same as ethanol.

Class	1	2	3	4	5
Methanol					10
Ethanol			10		
n-butyl acetate		10			
butanone	6			4	
mixture			10		

Table 1. Classification result of fifty volatile chemical patterns into five classes using FCMA1

Table 2 presents the result of the application of FCMA2 to volatile chemical clustering. The FCMA2, which updates cluster centres frequently, shows an impressive performance for volatile chemicals clustering. All fifty volatile chemical patterns were perfectly clustered, even all the mixture chemical patterns are clustered as the 5th class which is far from the ethanol or methanol classes.

Class	1	2	3	4	5
Methanol	10				
Ethanol		10			
n-butyl acetate			10		
butanone				10	
mixture					10

Table 2. Classification result of fifty volatile chemical patterns into five classes using FCMA2

The third applied algorithm for volatile chemicals clustering problem was the FCMA3. The result is shown in Table3. The classification result of the FCMA3, which updates membership grades frequently, is shown to be nearly the same quality as the FCMA2 result for volatile chemical sample patterns. All the sample patterns were exactly clustered into five different classes.

Class	1	2	3	4	5
Methanol			10		
Ethanol	10				
n-butyl acetate				10	
butanone		10			
mixture					10

Table 3. Classification result of fifty volatile chemical patterns into five classes using FCMA3

The other important factor of fuzzy c-means algorithms is convergence behaviour. With the volatile chemicals clustering problem, fast convergence will allow to find an optimum solution for classification of volatile chemicals among the three different fuzzy c-means algorithms. Figure 2 shows the convergence behaviour of the three fuzzy c-means algorithms for fifty volatile chemical sample patterns. It shows that the FCMA2 and FCMA3 performed better than the FCMA1. The convergence behaviour of the FCMA2 gave extremely fast result for the volatile chemicals clustering problem.

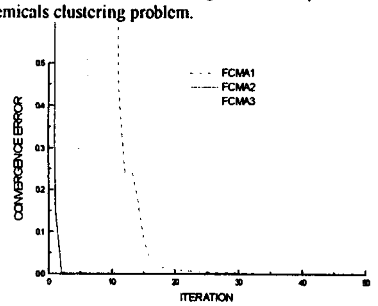

Figure 2. The convergence behaviour of the fuzzy c-means algorithms

5. Conclusions

The three fuzzy c-means algorithms were applied to the five volatile chemicals differentiation problem, and the comparison results between the algorithms showed that the proposed algorithms FCMA2 and FCMA3 provided faster convergence times than FCMA1. the use of frequent updating of membership grades and cluster centres showed its effectiveness in accelerating the solution procedure. By testing the convergence performance of these algorithms, the results indicated that the FCMA2 algorithm consistently outperformed the other fuzzy c-means algorithms. Consequently, the short iteration procedure provides a good saving method of calculation times which is important when the algorithms are required to run on large data sets.

The fuzzy c-means algorithms were investigated to find the optimum algorithm for volatile chemicals classification. Their classification capability for the five volatile chemicals was confirmed in the experimental work.

References

1. Persaud and Travers, "Multielement arrays for sensing volatile chemicals", Intelligent Instruments and Computers, Elsevier Press, pp.147-154, July/August, 1991

2. Persaud and Pelosi, "Sensor arrays using conducting polymers for an artificial nose", NATO ASI Series E: Applied Science, Vol. 212, Springer-Verlag, pp. 237-256, 1992

3.Bezdek, "Pattern recognition with fuzzy objective function algorithms", Plenum Press, pp. 65-80, 1981

4. Zimmermann, "Fuzzy set theory and its applications", Kluwer Academic Publishers, pp. 217-236, 1991

5. Kamel and Selim, "New algorithms for solving the fuzzy clustering problem", Pattern Recognition, pp. 421-428, 1994.