

논문 2001-10-6-06

Unsupervised clustering 방법을 갖는 인공 냄새인식 시스템의 구현

최 찬석*, 김 정도*, 변 형기**

Implementation of an Artificial Odour Recognition System with Unsupervised Clustering Methods

Choi Chan-Seok*, Kim Jeong-Do*, Byun Hyung-Gi**

요 약

다양한 냄새를 인식하고 분석하기 위하여 metal oxide 형 센서어레이를 이용한 인공 냄새인식 시스템(전자 코 시스템)을 설계 제작하였다. 센서어레이로부터 측정되는 다차원 데이터를 관측자로 하여금 쉽게 구별 할 수 있도록 Euclidean distance를 기본으로 하는 unsupervised clustering 방법을 제안한다. 제안된 방법은 주성분 분석법을 Sammon의 매핑법을 시작점으로 사용한 결합방법으로 특정냄새가 속해있는 cluster들에 대한 가정이 필요하지 않으며, 주성분 분석법에서 나타나는 차원축소로 인한 오차를 최소화하고 Sammon의 매핑법 사용으로 나타나는 데이터베이스의 입력순서에 따른 cluster들의 회전현상을 제거할 수 있다. 제안된 unsupervised clustering 방법으로 구현된 인공 냄새인식 시스템은 휘발성 유기화합물과 국산양주들의 냄새 차이를 각각 평가하는데 사용되어졌고 실험을 통하여 좋은 성능을 검증하였다.

Abstract

We have been designed and constructed an artificial odour recognition system(electronic nose system) using metal oxide type sensor array for recognizing and analyzing various odours. We proposed an unsupervised clustering method based on Euclidean distances in order for human observer to examine easily multi-dimensional data, which has been measured from an array of sensors. This is a combination of Principal Components Analysis(PCA) used as a starting point for Sammon Mapping Method(SMM). No prior assumptions are made of the classes in which odour belong, and the error due to dimensional reduction at the PCA can be minimized without the disadvantages of rotation of clusters when the order of data sets in a data base was changed in the SMM. An artificial odour recognition system with the proposed unsupervised clustering method was applied to assessment of odour differences of Volatile Organic Compounds(VOCs) and Korean whiskies respectively, and demonstrated the best performances throughout the experimental trails.

1. 서 론

우리의 삶과 환경에 영향을 주는 냄새들을 해

석하고 이를 감지하기 위한 인간의 후각기관을 모방한 인공 후각인식 시스템의 구현은 매우 어려운 과제였다. 그러나, 최근 전자재료 및 컴퓨터의 비약적인 발전에 힘입어 냄새를 인간의 후각기관과 비슷하게 감지하고 분석할 수 있는 인공 냄새인식 시스템(Artificial Odour Recognition System) 개발 가능성이 현실화되고 있다^(1,2,3). 인공 냄새인식 시스템은 냄새를 인식하고 분류하기 위하여 냄새에 대하여 서로 다른 응답을 보이는 센서들을 어레이화 하고 마이크로프로세서를 기반으로 하는 전자회로를 구성하며 센서어레이

* 삼척대학교 제어계측공학과(Dept. of Control & Instrumentation Engineering, Samchok National University)

** 삼척대학교 정보통신공학과(Dept. of Information & Communication Engineering, Samchok National University)

<접수일자 : 2001년 3월 23일>

로부터 신호를 검출하여 특정한 냄새에 대한 고유의 특징을 갖는 패턴데이터를 출력한다. 이러한 패턴데이터들은 일반적으로 다차원의 패턴들을 갖게되고, 다차원 고유패턴들을 unsupervised clustering 하기 위한 방법으로 다차원의 패턴 데이터들을 그들이 가지고있는 고유정보 유출 없이 저차원 (주로 2차원)패턴으로 만들어 시각적으로 쉽게 clustering하는 방법을 사용하게된다. 흔히 사용되는 통계적인 패턴데이터들의 clustering 방법으로는 선형투영법 (linear projection method)과 비선형 매핑법(non-linear mapping method)으로 나뉘어질 수 있다. 주성분 분석법(Principal Component Analysis : PCA)으로 대표되는 선형투영법은 간편성(simplicity)과 일반성(generality) 때문에 다차원 데이터를 저차원 데이터로 차원축소가 필요한 데이터분석에 주로 사용되어졌다^[4]. 이에 반하여, Sammon에 의하여 제안된 Sammon Mapping Method(SMM)를 사용하는 비선형 매핑법은 비 선형적인 요소를 많이 갖는 실제 데이터 분석에 사용되었다^[5]. 선형투영법인 PCA 방법은 다차원 데이터를 저차원 데이터로 투영하는 과정에서 작은 고유치(eigenvalue)를 갖는 고유벡터(eigenvector)값을 제거하게되어 필연적으로 오차를 수반하게된다. 반면에 SMM은 다차원 데이터를 저차원데이터로 매핑 하는 방법으로 Euclidean distance를 기본으로 반복계산을 통하여 오차를 줄이는 gradient descent방법을 제안함으로써 PCA방법보다는 상대적으로 원래 데이터에 관한 고유정보 유실을 줄일 수 있다. 그러나, SMM은 랜덤한 초기값을 사용하고, 입력 데이터베이스의 배열순서가 다르면 clustering 분석을 행할 때마다 cluster들이 회전하는 현상을 보여 관측자가 cluster들을 판별하는데 어려운 점을 보인다. 이러한 문제점들을 개선하기 위하여 본 논문에서는 PCA방법으로 얻어지는 결과를 SMM의 시작점으로 사용하는 선형투영법과 비선형 매핑법의 결합방법을 제안하였다. 제안된 unsupervised clustering방법은 선형투영법에서 발생하는 오차를 최소화하고, 기존 비선형 매핑법에서 발생하는 cluster들의 회전현상을 제거하는 좋은 특성을 보여주었다. 본 연구에서는 7개의 metal oxide형 상용가스센서(Figaro Inc. Japan)들로 구성된 센서어레이와 one-chip 마이크로프로세서를 이용한 전자회로, 그리고 시스템

자체의 온도와 습도 보상회로를 갖는 실제 산업현장에서 사용될 수 있는 간편한 인공 냄새인식 시스템을 설계 제작하였으며, 제작된 시스템을 사용하여 측정된 4개의 VOC 가스들과 국산양주들의 데이터를 제안된 unsupervised clustering 방법을 이용하여 각각 정확하게 분류함으로써 unsupervised clustering 방법을 갖는 인공 냄새인식 시스템이 구현됨을 검증하였다.

II. 인공 냄새인식 시스템

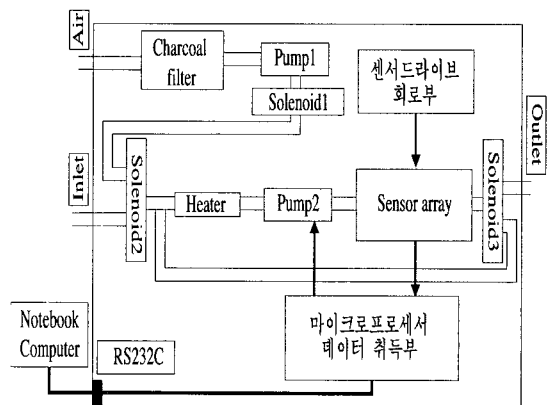


그림 1. 인공 냄새 인식시스템의 전체 시스템도.

Fig. 1. Schematic diagram of artificial odour recognition system.

그림 1은 인공 냄새인식 시스템의 하드웨어 구조를 보여준다. 7개의 일본 Figaro사의 metal oxide형 가스센서(TGS823, 831, 2610, 2620, 2181, 826, 816)들을 이용하여 센서어레이를 구성하였고, 센서들의 응답데이터를 취득하고 전체 하드웨어 시스템을 구동제어 하기 위하여 멀티플렉서(MUX)와 80C196KC one-chip 마이크로프로세서를 사용하였다. 일반적으로 화학 가스센서는 온도와 습도의 변화에 따라 센서응답이 변한다. 따라서, 시스템 내부에 일정한 온도와 습도를 유지하기 위하여 별도의 히터와 필터를 제작 설치하였다. 히터는 측정가스가 시스템 안으로 들어와 외부 온도변화에 영향을 받지 않고 일정한 온도를 유지하도록 설계되었으며, 필터는 시스템 내부에 유입되는 측정가스 및 기준 가스(reference gas)에 일정한 상대습도(relative humidity)로 유지시키는데 사용된다. 전체시스템은 솔레노이드 밸브와 펌프를 이용하여 센서부

로 가스를 주입하게 되고 이때 센서 어레이는 주입가스에 대한 반응을 센서들의 저항 변화로 나타내게 되며 이를 전압으로 환산하여 사용한다. 측정가스에 대한 응답은 기준가스(reference gas)일 때의 전압과 측정가스일 때의 전압의 차로 표시하며, 측정가스들의 분류를 위한 unsupervised clustering 방법을 위한 입력데이터로 사용된다.

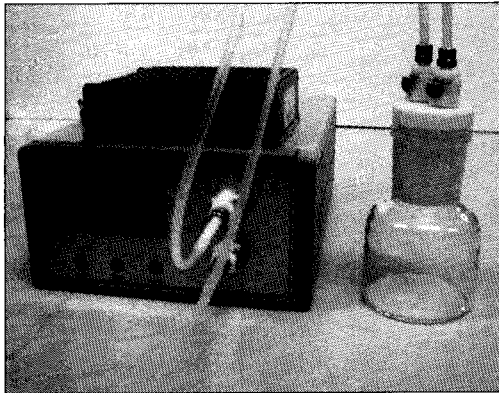


그림 2. 인공 냄새 인식 시스템의 사진.
Fig. 2. Picture of artificial odour recognition system.

그림 2는 본 연구를 통하여 설계 제작된 인공 냄새인식 시스템의 실제모습을 보여준다.

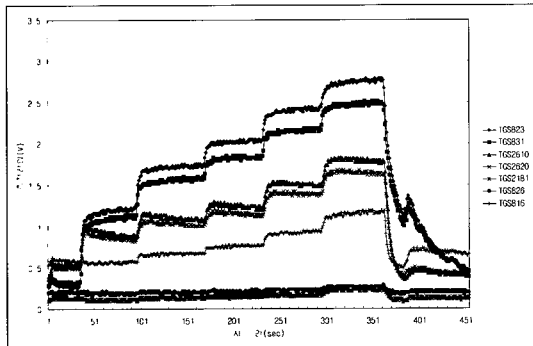


그림 3. 벤젠 가스에 대한 출력 데이터.
Fig. 3. System output data for Benzene gas.

그림 3은 설계 제작된 인공 냄새인식 시스템으로 측정된 VOC계열인 벤젠(Benzene)가스에 대한 ppm 레벨에 따른 센서어레이의 응답을 보여준다. 각각의 센서들이 ppm레벨에 따라 서로 다른 응답을 보여주는 것은 unsupervised clustering 방법을 이용하여 특징 추출 및 패턴

매칭이 가능하게되는 기초를 제공한다.

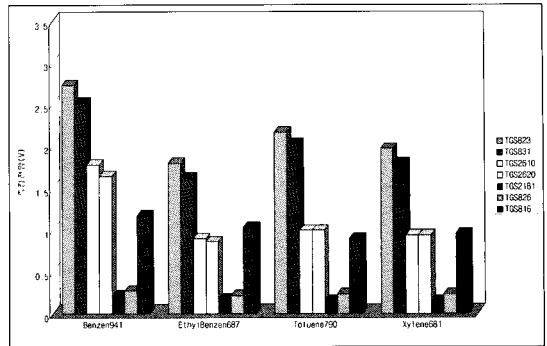


그림 4. VOC 가스에 대한 센서어레이의 출력.
Fig. 4. Sensor array output of VOC gases.

그림 4는 측정된 4개의 VOC가스 데이터들에 대한 패턴의 특징을 히스토그램으로 보여준다. 이러한 패턴 데이터들은 센서를 어레이화 함으로서 얻어질 수 있는 장점으로 제안된 unsupervised clustering 방법을 이용하여 측정된 가스들을 분류할 수 있다.

III. 제안된 Unsupervised Clustering방법

설계 제작된 인공 냄새인식 시스템으로부터 얻어진 VOC가스들과 국산양주들의 데이터 분류를 위하여 사용된 선형투영법인 PCA와 비선형 매핑법인 SMM, 및 본 논문에서 제안하는 PCA와 SMM의 결합방법인 unsupervised clustering 방법들을 소개한다.

1. PCA(주성분분석)

PCA 방법은 다차원 데이터를 관측자가 가지적으로 분류를 확인할 수 있는 2차원 데이터로 나타낸 데이터 패턴들을 선형적으로 투영하는 알고리즘이다.^[2]

$$X = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1L} \\ X_{21} & X_{22} & \dots & X_{2L} \\ \vdots & \vdots & \dots & \vdots \\ X_{M1} & X_{M2} & \dots & X_{ML} \end{bmatrix} \quad (3-1)$$

여기서 X는 N×L의 입력 데이터 패턴으로 N은 패턴 수, L은 센서어레이의 센서수로 표시된다.

(3-1)식으로부터 차원 축소되어 나타나는 2-차원투영 데이터 패턴은 다음과 같이 표현된다.

$$Y = \begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \\ \vdots & \vdots \\ Y_M & Y_{N2} \end{bmatrix} \quad (3-2)$$

여기서 Y는 2차원으로 선형 투영된 데이터 패턴으로 N은 패턴 수를 표시한다.

센서들에 대한 고유치(eigenvalue) λ_k 와 고유벡터(eigenvector) μ_k 는 다음과 같은 식을 풀어 구할 수 있다.

$$C \mu_k = \lambda_k \mu_k \quad (3-3)$$

여기서 공분산 행렬 C의 고유치는 입력 패턴들의 변위(variance)와 평균(means)의 함수들이다. 2차원으로 데이터들을 표시하기 위하여 계산된 고유치 가운데 가장 큰 고유치 λ_1 과

λ_2 를 선택하고 이와 연관된 고유벡터 μ_1 과 μ_2 를 주성분으로 하여 다차원 입력 데이터 패턴들을 2차원으로 선형적으로 투영할 수 있다. 앞에서 언급되어진 바와 같이 PCA방법은 가장 보편적으로 사용되는 다차원 데이터 패턴들을 저차원 데이터 패턴으로 투영시키는 방법이다. 하지만, 다차원 데이터를 저차원으로 투영하는 과정에서 원하는 차원까지의 고유벡터만을 사용하기 때문에 때때로 다차원 데이터들의 고유정보가 손실되어질 수 있다.

2. SMM(Sammon Mapping Method)

Sammon mapping 방법은 Euclidean distance 등과 같은 거리 측정방법을 기반으로 다차원 데이터 패턴들을 원하는 저차원 데이터 패턴들로 차원을 축소하는 Sammon^[5]에 의하여 제안된 비선형 매핑 알고리즘이다.

SMM은 기준이 되는 매핑 데이터로부터 입력 데이터 사이의 거리를 오차(Error)로 규정하고 이를 감소시키기 위해 gradient descent 방법을 사용하여 최적의 매핑 데이터를 찾아가는 방법을 사용하였다. SMM의 매핑 오차 $E(m)$ 은 모든 입력 패턴사이의 상호거리에 대한 입력 패턴과 매핑 데이터의 상호거리에 대한 비로 나타낸다.

$$E(m) = \frac{1}{c} \sum_{i,j}^N \frac{[d_{ij}^*(m) - d_{ij}(m)]^2}{d_{ij}^*(m)} \quad (3-4)$$

$$c = \sum_{i,j}^N [d_{ij}^*(m)]^2 \quad (3-5)$$

여기서 $d_{ij}^*(m)$ 은 m번 반복한 후의 입력 패턴 간의 상호 거리를 나타내고, $d_{ij}(m)$ 는 매핑 데이터간의 상호거리를 나타낸다.

SMM은 매핑과정에서 일어나는 오차를 반복계산으로 최소화 할 수 있어 PCA방법 보다 상대적으로 원래 입력 패턴데이터에 관한 고유정보손실을 줄일 수 있다. 그러나, 초기의 매핑 데이터를 임의로 설정하고, 입력 데이터베이스의 배열 순서가 다르면 clustering 분석을 행할 때마다 cluster들이 회전하는 현상을 보여 가시적으로 패턴들을 분석하고자 하는 unsupervised clustering 방법에서 관측자가 cluster들에 대한 기준을 세우기가 어렵다.

3. PCA방법과 SMM을 결합

앞서 언급된 PCA방법과 SMM의 결점들을 보완하고 보다 효과적인 패턴 clustering 방법으로 본 논문에서는 PCA방법으로 얻어진 결과를 SMM의 시작점으로 사용하는 선형투영법과 비선형 매핑법의 결합방법을 제안한다. SMM에서 매핑데이터의 초기치를 임의로 설정하는 대신 선형투영법인 PCA방법으로 구한 주성분 요소를 초기치로 설정하여 매핑 데이터의 초기치가 임의로 설정되지 않으므로 매핑 후 축의 회전 현상이 제거되고, 또한 SMM의 반복계산에 의한 오차감소 알고리즘을 사용함으로써 PCA방법에서 나타나는 오차를 최소한으로 만들어 앞서 언급한 PCA방법이나 SMM을 단독으로 unsupervised clustering방법으로 사용 할 때보다 차원축소에 뛰어난 성능을 가능하게 하였다.

IV. 실험 결과

PCA방법과 SMM을 결합한 제안 unsupervised clustering 방법의 평가는 본 연구를 통하여 설계 제작된 인공 냄새인식 시스템으로부터 얻어진 4종류의 VOC 가스들과 시중에 유통되는 국산양

주들의 데이터들을 이용하여 실험하였다. 실험에 사용된 VOC 가스들의 종류와 농도는 표 4.1과 같다.

표 1. 실험에 사용된 VOC 가스들의 종류와 농도
Table 1. VOC gases and its concentration levels or experimental trail

가스의 종류	농도(ppm)
Benzene	588(BZ1), 941(BZ2)
Ethyl Benzene	429(EB1), 687(EB2)
Toluene	493(TL1), 790(TL2)
Xylene	426(XL1), 681(XL2)

* 예: 588(BZ1) : 588 ppm, (벤젠1의 그래프 표기)

다른 실험은 같은 양의 국산양주들을 서로 비교해 보았다. 실험에 사용된 국산양주는 칼튼힐(알코올 농도: 40%, 표기: KH)과 캡틴큐(알코올 농도: 35%, 표기: CQ)이다.

먼저, 표1.에 나타난 VOC 가스들에 대한 인공 냄새인식 시스템을 통한 실험결과를 다음의 그림들을 통하여 보여준다. 그림 5는 VOC 가스들을 PCA방법으로 clustering한 결과이다.

그림 5에 나타난 화살표는 같은 종류의 VOC 가스들이 농도에 따라 다른 cluster를 형성함을 보여준다. 관측자가 눈으로 쉽게 구별할 수 있는 cluster들을 형성함을 볼 수 있으나, Ethyl Benzene과 Xylene등은 농도에 따른 구별이 쉽지 않음을 보여준다. 1차 주성분(PC1 = λ_1)은 0.524이고 2차 주성분(PC2 = λ_2)는 0.0015임으로, PCA방법의 신뢰도를 정의하는 식(4-1)을 통하여 이 실험의 신뢰도를 구하면 %K = 98.533으로 본 실험을 통하여 얻어진 clustering의 결과는 1.467%의 오차를 갖고있음을 알 수 있다.

$$\%K = \frac{(\lambda_1 + \lambda_2)}{\sum_{m=1}^M \lambda_m} \times 100 \quad (4-1)$$

여기서 M은 고유치의 총 수이다.

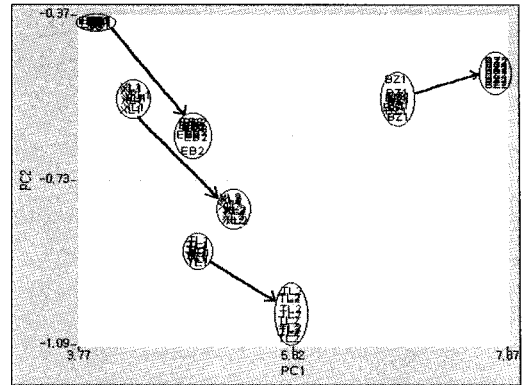


그림 5. PCA방법을 이용한 VOC 가스들의 clustering 분석결과.

Fig. 5. Clustering analysis result for VOC gases using PCA method.

그림 6에 나타난 결과는 임의의 초기치를 매핑 데이터로 사용한 SMM으로 구현된 VOC 가스들의 clustering 분석 결과이다. PCA방법보다는 더욱 잘 clustering 된 데이터들을 볼 수 있고, 매핑 오차는 0.000098로 원래 입력 패턴 데이터들의 정보가 거의 오차 없이 2차원으로 매핑되어졌음을 알 수 있다. 그러나, PCA보다는 많이 다른 축의 회전이 발생되었음을 볼 수 있다. 만일, 입력 데이터베이스가 달라지면 완전히 다른 축의 회전이 임의의 초기치를 적용함으로써 발생할 수 있는 문제가 여전히 존재한다.

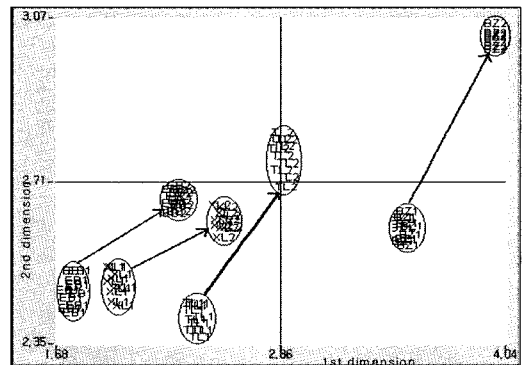


그림 6. 임의의 초기치를 갖는 SMM을 이용한 VOC 가스들의 clustering 분석결과. (반복횟수 = 61, 매핑 오차 = 0.000098).

Fig. 6. Clustering analysis result for VOC gases using SMM having arbitral initial values. (Iteration = 61, Mapping Error = 0.000098).

그림 7은 PCA방법과 SMM을 결합한 제안된 Unsupervised Clustering 방법으로 만들어진 VOC 가스들의 clustering 분석 결과이다. 그림 6에서 보는 것과 같이 SMM이 단독으로 사용되었을 경우 반복횟수를 60회까지 실행하여 오차를 줄였지만, 제안된 unsupervised clustering 방법은 20회 정도의 반복횟수로도 SMM보다 더욱 작은 매핑 오차를 얻을 수 있었다. 또한 명확한 각 가스들의 cluster들과 농도에 따른 VOC 가스들의 cluster 변화를 확실하게 볼 수 있어 VOC 가스들의 종류 및 농도에 따른 clustering을 가능하게 하였다. 덧붙여, 그림 5의 PCA방법이나 그림 6의 임의의 초기치를 갖는 SMM 분석방법들에서 관측자로 하여금 혼돈을 줄 수 있는 Ethyl Benzene, Toluene, 그리고 Xylene 등의 가시적인 분석이 더욱 확실하여졌음을 볼 수 있고, PCA의 결과를 SMM의 초기치로 사용함으로써 PCA방법과 같이 축의 회전 제거되어 관측자들이 혼돈 없이 cluster들을 확인 할 수 있었다. 이상의 실험결과들을 통하여 제안된 unsupervised clustering 방법의 좋은 성능을 검증하였다.

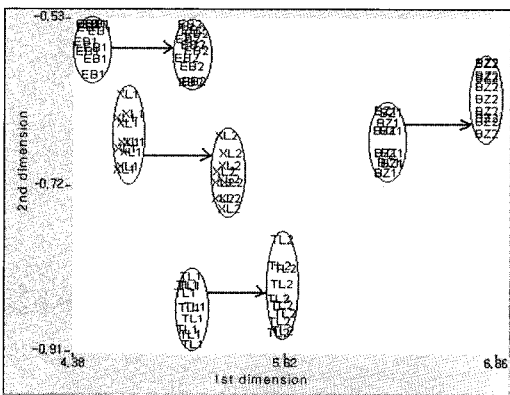


그림 7. 제안된 unsupervised clustering 방법을 이용한 VOC 가스들의 clustering 분석결과. (반복 횟수 : 22, 매핑 오차 = 0.000093).
 Fig. 7. Clustering Analysis Result for VOC gases using the proposed Unsupervised Clustering Method. (Iteration = 22, Mapping Error = 0.000093).

그림 8은 국산양주들의 분류를 제안된 unsupervised clustering 방법으로 수행한 결과를 보여준다. 그림에서 나타난 것과 같이 두 개의

실험대상 국산양주들은 거의 같은 수준의 알코올을 함유한 술이지만 각각의 제조과정에서 쓰인 향료 및 미세한 제조방법의 차이에 따른 냄새의 차이를 전문가들로 행하여지는 관능검사방법과 유사한 결과를 일반 관측자들이 확실하게 볼 수 있도록 나타낼 수 있다.

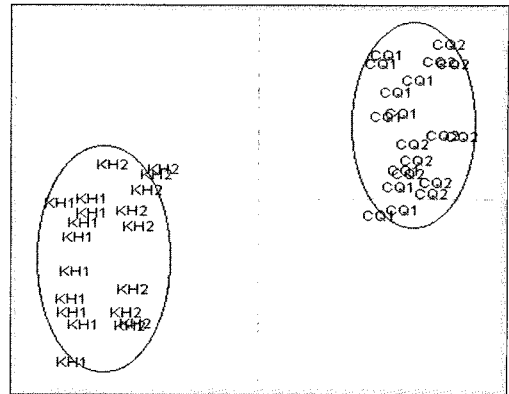


그림 8. 제안된 Unsupervised Clustering 방법을 이용한 국산양주들의 clustering 분석결과.
 Fig. 8. Clustering analysis result for Korea whiskeys using the proposed unsupervised clustering method.

V. 결론

본 연구를 통하여 7개의 metal oxide형 가스 센서를 갖은 센서어레이로 구성된 간편한 인공 냄새인식 시스템을 설계 제작하였다. 구현된 시스템을 사용하여 측정된 데이터들은 선형투영법인 PCA방법과 비선형 매핑법인 SMM을 결합한 제안된 unsupervised clustering 방법으로 cluster들을 분석하였다. 제안된 unsupervised clustering 방법이 PCA방법에서 나타나는 투영오차를 최소화 할 수 있고, SMM에서 발생하는 입력데이터베이스에 따른 축의 회전현상을 PCA방법의 결과를 SMM의 시작점으로 사용하여 제거할 수 있으며, SMM에 비하여 빠른 수렴 속도를 갖는다는 것을 실험을 통하여 검증 할 수 있었다. 본 논문을 통하여 제안된 초기단계의 인공 냄새인식 시스템이 더욱 많은 연구를 통하여 개발 발전되어진다면 환경 방지 대책에 있어 사용되어질 계측기를 비롯한 많은 분야에 응용 될 수 있을 것이라 보여진다.

감사의 글

본 연구는 한국과학재단 목적기초연구(과제번호 : 2000-1-30200-017-3)지원으로 수행되었음.

참 고 문 헌

[1] K. C. Persuad & P. vers, "Multielement arrays for sensing volatile chemicals", Intelligent Instruments and Computers, Vol. 147, 1991.

[2] H-G Byun, K. C. Persurd, S. M. Khaffaf, P. J. Hobbs, T. H. Misselbrook, "Application of unsupervised clustering methods to the assessment of malodour in agriculture using an array of conducting polymer odour sensors", Computers & Electronics in agriculture, Vol. 17, pp.233-247, 1997.

[3] D-H Lee, J. S. Payne, H-G Byun, K. C. Persuad, "Application of radial basis function neural networks to odour sensing using a board specificity array of conducting polymers", In: Lecture Notes in Computer Science, Vol. 1112, pp.299-304, 1996.

[4] J. W. Gardner, "Detection of vapour and odours from a multisensor array using pattern recognition Part I: Principal component and cluster analysis", Sensors & Actuators B, 4, pp.109-115, 1991

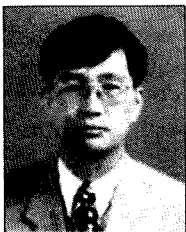
[5] J. W. Sammon Jr., "A nonlinear mapping for data structure analysis", IEEE Trans. on Computers, Vol. C-18, No. 5, pp.401-409, May, 1969.

著 者 紹 介



최 찬 석

1998년 2월 삼척대학교(공학사 - 제어계측공학)
 2001년 2월 삼척대학교(공학석사 - 제어계측공학)
 주관심분야: 센서 응용 및 신호 처리, 인공 지능 제어



김 정 도

1962년 9월 26일생
 1987년 성균관대학교 전자공학과(공학사)
 1990년 성균관대학교 대학원 전자공학과 (공학석사)
 1994년 성균관대학교 대학원 전

자공학과 (공학박사)
 1995년 - 현재 삼척대학교 제어계측공학과 부교수
 주관심분야: 센서시스템, 마이크로프로세서응용, 지능형 로봇 시스템



변 형 기

1961년 8월 1일생
 1984년 명지대학교 전기공학과(공학사)
 1990년 영국 UMIST(Univ. of Manchester Institute of Science & Technology)

(공학석사)
 1995년 영국 UMIST(Univ. of Manchester Institute of Science & Technology) Ph.D.
 1996년 - 현재 삼척대학교 정보통신공학과 조교수
 주관심분야: 센서시스템, 패턴인식