Journal of Sensor Science and Technology
Vol. 21, No. 1 (2012) pp. 59-67
http://dx.doi.org/10.5369/JSST.2012.21.1.59
pISSN 1225-5475/eISSN 2093-7563

생물학적 후각 시스템을 모방한 대규모 가스 센서 어레이에서 코사인 유사도와 퍼지 클러스터링을 이용한 중복도 제거 방법

김정도[1] · 김정주[1] · 박성대[1] · 변형기[2] · K.C. Persaud[3] · 임승주[1,+]

The Redundancy Reduction Using Fuzzy C-means Clustering and Cosine Similarity on a Very Large Gas Sensor Array for Mimicking Biological Olfaction

Jeong-Do Kim[1], Jung-Ju Kim[1],Sung-Dae Park[1], Hyung-Gi Byun[2], K.C.Persaud[3], and Seung-Ju Lim[1,+]

Abstract

It was reported that the latest sensor technology allow an 65536 conductive polymer sensor array to be made with broad but overlapping selectivity to different families of chemicals emulating the characteristics found in biological olfaction. However, the supernumerary redundancy always accompanies great error and risk as well as an inordinate amount of computation time and local minima in signal processing, e.g. neural networks. In this paper, we propose a new method to reduce the number of sensor for analysis by reducing redundancy between sensors and by removing unstable sensors using the cosine similarity method and to decide on representative sensor using FCM(Fuzzy C-Means) algorithm. The representative sensors can be just used in analyzing. And, we introduce DWT(Discrete Wavelet Transform) for data compression in the time domain as preprocessing. Throughout experimental trials, we have done a comparative analysis between gas sensor data with and without reduced redundancy. The possibility and superiority of the proposed methods are confirmed through experiments.

Keywords : Redundancy, Cosine Similarity, Fuzzy Clustering, Very Large Gas Sensor Array

1. 서 론

후각이란, 냄새가 있는 화학물질의 분자에 의하여 비강 내의 후세포가 자극되어 일어나는 냄새의 감각이다. 후각은 다른 감각에 비해 많은 연구가 이루어져 있지 않으며 이론적 접근은 100년도 되지 않은 역사를 가지고 있다. 후각에 대한 관심은 신경세포가 발견되고 신경에 관한 연구가 활발해 지면서 높아졌고, 후각 신호 전달에 관한 연구는 전기 생리학 기술과 더불어 발전하였다[1-3].

1982년 Persuad가 생물학적 후각 시스템을 모사하기 위해 전도성고분자를 이용한 인공 후각 시스템을 제안한 이래로[4], 전도성고분자(conducting polymer) 센서, QMB(Quartz Microbalance) 센서 및 MOS(Metal Oxide Semiconductor) 센서 등 다양한 센서를 사용하는 전자코 시스템이 제안되고 연구된 바 있다[5-7]. 이러한 전자코 시스템의 경우 수 개 ~ 수십 개의 개별적 센서를 어레이로 구성한 후, 퍼지 알고리즘 및 신경망 등의 신호처리 기술을 이용하여 냄새를 분류한다[8, 9]. 생물학적 후각 시스템처럼 다양한 종류의 냄새를 구별하기 위해서는 냄새 분자에 따라 다른 전기적 신호를 발생시키는 많은 종류의 센서 물질들이 필요하며, 많은 센서들의 중복성이 허용되어야 한다. 이러한 생물학적 후각 시스템의 특성을 반영하여 인공 후각시스템을 설계하기 위하여 NEUROCHEM 프로젝트에서는 수십 개의 전도성 고분자를 이용하여 65,536개의 센서 어레이를 작은 크기의 칩 속에 집적하였으며, 수십 ~ 수백 개의 센서들이 중복성을 가질 수 있는 구조로 센서 어레이 칩을 성공적으로 설계한 바 있다[10]. 그러나 이러한 생물학적 후각 시스템을 모방한 대규모 가스 센서 어레이 시스템은 크게 두 가지의 문제점이 존재한다. 첫 번째, 작은 칩 속에 화학적 고분자 물질을 도핑하여 센서를 만드는 것은 불안정한 센서들을 만들어 낸다는 것이다. 생물학적 후각 시스템의 경우 냄새의 독성과 시간의 흐름에 따라 일부 후각세포들이 불안정해지거나 죽더라도 크게 영향을 받지 않지만, 인공 후각 시스템에서는 이러한 불안정한 센서들이 상당한 영향을 줄 수 있다. 두 번째, 불안정한 센서들을 제거한다 하더라도 나머지 센서들 전부를 다 사용할 수 없다는 것이다. 나머지 센서 전부를 다 사용할 경우, 감당할 수 없는 계산 시간이 필요할 뿐만 아니라, 계산 오차의 누적에 의한 오류를 발생시킬 수 있다. 실제 생물학적 후각 시스템

[1]호서대학교 전자공학과 (Department of Electronic Engineering, Hoseo University)
[2]강원대학교 정보통신공학과 (School of Electronic, Information and Communication Engineering, Kangwon National)
[3]맨체스터 대학교 (SCEAC University of Manchester, UK)
[+]Corresponding author: seungju123@naver.com
(Recived:, Nov. 3, 2011, Revised : Dec. 29, 2011, Accepted : Jan. 2, 2012)

의 경우도, 중복성을 갖는 수많은 수용체가 하나의 사구체에 연결되어 한 개의 신호만을 출력하기 때문에 실제 냄새의 분석을 위해서 수천 개의 사구체 신호만이 이용된다. 이러한 문제점을 해결하기 위해서 본 논문에서는 대규모 가스 센서 어레이 신호 중 불안정한 센서 신호들을 검출하고 제거하는 방법을 제안한다. 또한, 중복성을 보이는 센서들을 대표할 수 있는 대표 센서의 선정 방법을 제안함으로써 생물학적 시스템이 갖는 특성을 반영할 수 있도록 한다. 전처리 과정으로는 우선 센서 데이터에 포함되어있는 잡음을 제거하기 위해 3차 버터워스 저역통과 필터(butterworth low-pass filter)를 사용하였으며 알고리즘의 연산 시간 단축을 위해 6차 이산 웨이블렛 변환(discrete wavelet transformation : 이하 DWT)을 사용하였다. 불안정한 센서의 제거를 위해 코사인 유사도 알고리즘을 이용하여 센서간의 유사도 분석을 하였으며, 이를 통해 불안정 센서를 파악하고 제거하는 방법을 제안하였다. 또한, 안정된 센서만을 대상으로 하여 퍼지 클러스터링(fuzzy c-means : 이하 FCM) 알고리즘을 적용하여 중복성을 갖는 센서의 특성을 대표할 수 있는 대표 센서를 선택하는 방법을 제안하였다. 제안된 방법의 입증을 위해, 가스 혹은 농도가 다른 8개의 데이터군(dataset)이 사용되었으며, 각 군은 각각 4번의 실험이 이루어져 총 32개의 데이터가 제안된 방법의 효율성을 증명하기 위해 사용되었다. 주성분 분석법(principal component analysis : 이하 PCA)을 통해 제안된 방법이 유용성이 입증되었다.

2. 대규모 가스 센서 어레이

2.1 생물학적 후각 시스템

동물의 후각 시스템에서 후각 수용체 뉴런(olfactory receptor neurons)은 냄새 정보를 전기적 신호로 변환하고, 변환된 전기적 신호를 후각 신경구(olfactory bulb)로 전달한다. 대략적으로 인간의 코에는 약 1000만개의 후각 수용체가 있으며, 수용체의 구조에 따라 약 1000 종류의 수용체로 구분된다. 각각의 종류들은 2 ~ 3개 정도의 특정 냄새에 민감하게 반응한다.

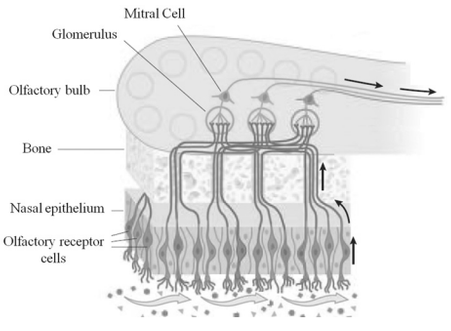


Fig. 1. The structure of biological olfactory system.

후각 수용체 뉴런들은 후각 신경구에 있는 사구체와 연결된다. 각 사구체는 거의 같은 분자에 반응하는 같은 종류의 후각 수용체 뉴런들과 연결되어 있으며, 같은 종류의 후각 수용체 뉴런은 1 ~ 2개의 사구체와 연결된다. 인간의 경우, 약 1800개 정도의 후각 신경구가 존재한다. 후각 신경구에 있는 사구체는 냄새지도(olfactory map)의 기본 단위이며, 냄새가 인가되었을 때 사구체들의 활성상태를 분석한다면 이론적으로 냄새의 정체를 해독할 수 있다. 이러한 사구체의 출력은 후각 신경구 안에 있는 회로에 의해 수정되어 뇌에 전달된다. 사구체와 연결되는 수많은 후각 수용체 뉴런들이 거의 비슷한 냄새에 반응하고, 비슷한 민감도를 가지기 때문에, 실제 생물학적 후각 시스템은 수많은 중복성(redundancy)을 가진다고 할 수 있다.

2.2 대규모 가스센서 어레이

대규모 가스 센서 어레이는 이러한 생물학적 시스템의 감각기를 모방하는 것이 목표이다. 이를 위해 유럽의 NEUROCHEM 프로젝트에서는 (=65,536)개의 전도성 고분자를 갖춘 어레이를 구축한 바 있다. 65536 개의 센싱요소는 각각 64×64개로 구성된 16개의 모듈로 나뉘어져 있다. 각각의 모듈은 센서 재료에 따라 크게 7개의 영역으로 크게 구분되어 있으며, 24 ~ 31개의 전도성 고분자 물질을 사용하여 생물학적 후각 시스템처럼 높은 중복성을 갖도록 설계하였다.

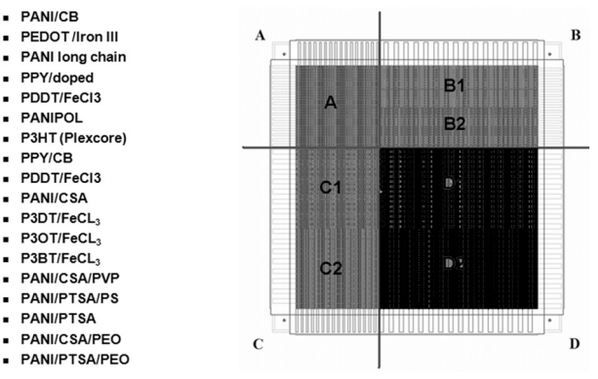


Fig. 2. Design of the whole sensor die.

전도성 고분자는 휘발성 화합물을 감지 할 수 있다. 또한 빠른 응답시간과 낮은 전력, 좋은 감도, 그리고 폭넓은 가스 인식을 보이기 때문에 전도성 고분자가 대규모 가스 센서 어레이를 위한 재료로 선정 되었다. 선택된 전도성 고분자 센서 어레이를 이용하여 NEUROCHEM에 의해 구현된 스캔회로 구조는 두 개의 아날로그 멀티플렉스, 한 개의 아날로그 피드백 버퍼로 구성되었으며, 한 번에 하나의 센서가 순차적으로 출력 노드와 접지에 연결 되어있는 이상적인 전압 분배기로 동작한다. 총 65536개의 센서 어레이는 16개

의 모듈로 구성되어 있으며, 각 모듈이 병렬로 데이터를 취득하기 때문에, 64×64개로 구성된 한 개 모듈의 센서 데이터를 취득하는 시간이 곧 전체 센서 어레이의 데이터 취득시간이 된다. 데이터 취득 시간은 약 1.24 sec이며, 한 개 센서데이터의 샘플링 속도는 3.3 kHz이다[11].

3. 잡음 제거 및 데이터 압축을 위한 전처리 과정

대규모 가스 센서 어레이는 많은 신호잡음이 포함될 수 있다. 이보다 더 큰 문제는 센서 데이터의 크기가 너무 크다는 것이다. 이러한 대규모 데이터를 이용하여 알고리즘에 적용하기 위해서 데이터를 변형시키지 않고 압축하는 알고리즘이 필요하다.

Fig. 3은 대규모 가스 센서 데이터의 잡음 제거 및 압축을 위한 전처리 과정의 블록 다이어그램이다.

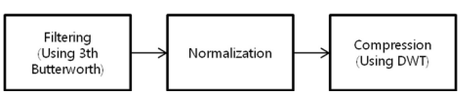
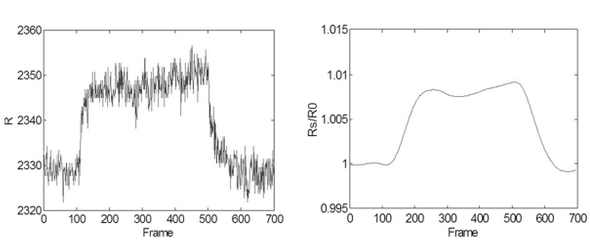


Fig. 3. The preprocessing for noise removal and compression of data on large gas sensor array.

3.1 IIR 버터워스 필터를 이용한 잡음 제거

대규모 가스 센서 어레이의 원 신호를 측정할 때, 멀티플렉서를 약 3.3 kHz로 전환해야 하기 때문에 많은 신호잡음이 포함될 수 있다. 잡음은 전원 잡음, 환경에 의한 잡음 등 많은 이유로 발생할 수 있다. 디지털 필터는 이러한 잡음을 제거하여 시스템의 성능과 신호대 잡음비를 향상 시킬 수 있다. 대표적인 디지털 필터로는 IIR 필터와 FIR필터가 있다. IIR 필터는 FIR 필터에 비해 차수가 적으며, 계산양이 적은 장점 때문에 대규모 가스센서 데이터의 필터링에 유리하므로 IIR 필터를 사용하였다. 본 논문에서는 3차 버터워스 저역통과 필터를 사용하였으며, 0.006 Hz의 차단 주파수가 사용되었다.



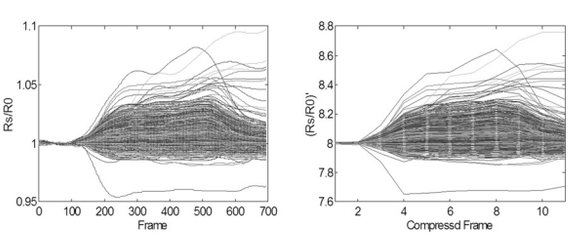
(a) (b)

Fig. 4. The result after 3rd order Butterworth filtering on large gas sensor array; (a) Data of 5th sensor, (b) Data of 5th sensor data after 3rd order Butterworth filter.

대규모 가스 센서 데이터에 존재하는 여러 가지 잡음을 3차 버터워스 저역통과 필터를 사용하여 필터링 한 것을 Fig. 4에 나타내었다. Fig. 4의 (a)는 4096개 센서 중 5번 센서 데이터의 필터 통과 전 데이터, (b)는 3차 버터워스 필터를 통과 한 5번 센서의 데이터이다. 단, (b)는 데이터의 정규화를 위해 로 표현하였다. 여기서 는 저항으로 표현된 측정 데이터이며 는 가스가 인가되기 전의 데이터의 평균값이다. 또한, Frame은 센서의 샘플 수를 의미하며, 압축 시 샘플의 수가 압축되었기 때문에 Compressed Frame 으로 표시 하였다.

3.2 이산 웨이브렛 변환을 이용한 데이터 압축

본 논문에서의 대규모 센서 어레이는 총 65536개의 센서를 사용하며 센서 데이터를 2배정밀도의 실수형으로 데이터 취득한다면 한 샘플링 당 512 k 바이트가 되고, 약 600개의 샘플(650 sec)의 데이터를 취득한다면 약 300 M 바이트가 된다. 이러한 대규모 데이터를 이용하여 신호처리 알고리즘을 적용하는 것은 쉬운 일이 아니다. 이러한 문제를 해결하기 위해 대규모 센서 어레이의 데이터를 시간축에서 압축한 후, 각종 신호처리 알고리즘을 적용하기 위해 DWT를 사용하였다. DWT는 원신호를 저주파 영역의 성분인 Approximation과 고주파 영역의 성분인 Detail로 분해할 수 있으며, 다중 해상도 분석이 가능하다. DWT는 1회 분해 때마다 2:1의 다운 샘플링을 수행한다. 본 논문에서는 6차 DWT를 사용하였으며, 모 웨이블렛으로는 Daubechies를 사용하였다.



(a) (b)

Fig. 5. The result of Time compression by DWT(4096 sensor array); (a) Uncompressed data, (b) Compressed data by DWT.

Fig. 5는 에탄올을 측정한 데이터이며, (a)는 3차 버터워스 필터를 통과한 정규화 된 가스센서 데이터, (b)는 6차 DWT에 의해 압축된 데이터의 결과를 보여주고 있다. 시간 축의 샘플수가 1/64로 압축되고, 특성을 그대로 유지하고 있음을 알 수 있다. 따라서 데이터 압축을 통해 과도한 메모리 사용과 계산 속도의 문제를 해결할 수 있다.

4. 중복도 제거 알고리즘

대규모 가스 센서 어레이에는 불안정하게 동작하는 센서가 존재하며, 비슷한 동작 특성을 보이는 센서들의 중복성을 볼 수 있다. 본 장에서는 코사인 유사도를 이용한 불안정 센서의 선별 방법과 센서들의 중복성을 제거하기 위해 FCMA를 이용한 대표 센서 선정 방법을 제시한다.

4.1 코사인 유사도를 이용한 불안정 센서 제거

고분자 화학물질이 작은 칩 속에 도핑되어야 하는 65536개의 대규모 가스 센서어레이 시스템의 경우, 실질적으로 모든 센서가 안정한 동작을 하는 것은 거의 불가능하다. 실제 생물학적 후각 시스템의 경우에서도, 전체 후각 수용체들이 정상 동작을 하는 것이 아니고, 화학적 냄새 및 바이러스 등에 의해 이상동작을 하는 경우가 있다. 물론 충분한 중복을 허용하기 때문에 이상동작을 하는 세포 신호를 적절하게 필터링 할 수 있다. 그러나 아직 생물학적 후각 시스템을 완벽하게 모사하지 못하는 인공 후각 시스템의 경우에서, 불안정 센서는 시스템의 특성을 혼란시키는 치명적 요인이 될 수 있다.

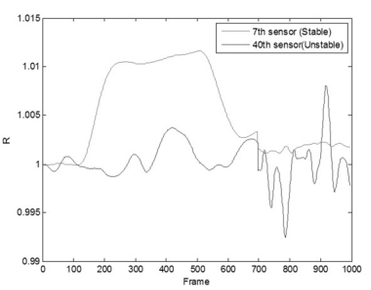


Fig. 6. Examples of unstable sensor in large-scale gas sensor array.

Fig. 6는 대규모 가스 센서 어레이에 존재하는 불안정 센서의 예를 보여주고 있다. 약 100 프레임(100 sec) 부근에서 가스를 인가하고 500프레임 부근에서 클리닝을 시작하였을 때, 7번째 센서의 경우 상당히 안정한 동작을 보이고 있으나 40번째 센서의 경우 불안정한 동작을 보임을 알 수 있다. 4096개로 구성된 16개의 센서 어레이 중에서 4096개 하나의 모듈 내에 존재하는 불안정 센서들을 파악하기 위해 Fig. 5의 (b)와 같이 DWT를 취한 데이터 중, 1번부터 7번 프레임까지의 데이터를 이용하였다. 그러나 7차원의 데이터는 분석이 어려운 고차원 데이터이기 때문에, PCA를 이용하여 2차원의 데이터로 변환을 하였다. PCA 방법은 확률적인 접근 방식을 이용한 것으로 주로 다루기 힘든 고차원의 신호를 낮은 차원으로 줄여 다루기 쉽게 해주는 통계적 방법을 말한다.

Fig. 7은 PCA를 사용하여 7차원의 특징을 2차원으로 차원 축소한 것이다. Fig. 7에서 보면 중첩성을 보이지 않고 별도로 떨어져 있는 특성을 보이는 센서들의 경우, 대부분 불안정한 센서들이다. 또한, 센서마다 분포가 다르지만 어느 일정 구간에서 비슷한 특성을 보이는 센서들이 밀집되어 있는 것을 볼 수 있다.

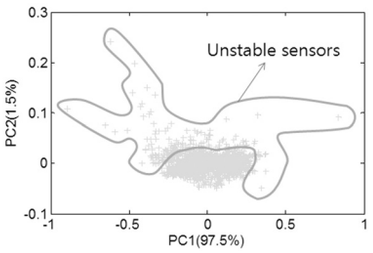


Fig. 7. Unstable sensors of 4096 sensors which depicted on 2 dimension by PCA.

또한, 센서마다 분포가 다르지만 어느 일정 구간에서 비슷한 특성을 보이는 센서들이 밀집되어 있는 것을 볼 수 있다. 불안정한 센서를 제거하기 위해서는 밀집되지 않은 다른 특성을 보이는 센서들을 제거하면 된다. 따라서 하나의 센서를 선택하여 그 센서와 유사한 센서가 얼마나 많이 존재하는지를 파악하였다. 즉, 각 센서들 사이에 유사도를 구하기 위해 코사인 유사도 알고리즘을 사용하였다. 코사인 유사도 알고리즘은 각 벡터 간의 거리를 계산하여 그 결과를 유사도로 표현하는 기법으로 다양한 분야에 적용되는 알고리즘이다. 코사인 유사도는 두 개의 데이터가 일치할 수 록 '1'에 가까워지며 데이터가 불일치할 수 록 '0'에 가까워진다. 센서간의 유사도를 구하기 위해 코사인 유사도 알고리즘을 이용하여 4096개의 센서를 일 대 일로 각각 비교한다. 만약, 유사도 값이 문턱치 보다 작으면 두 데이터를 중복성이 있는 센서로 간주한다. 본 논문에서는 센서의 중복성을 보기 위한 문턱치를 0.999995로 놓고 실험을 진행하였다. 이렇게 비슷한 반응을 보이는 센서의 수가 실험을 통해 정해 놓은 기준 중복 횟수(standard redundant number : 이하 SRN)보다 많으면 안정 센서, 적으면 불안정 센서로 간주한다. Fig. 8은 코사인 유사도 알고리즘을 이용하여 불안정 가스 센서를 제거하는 순서도이다.

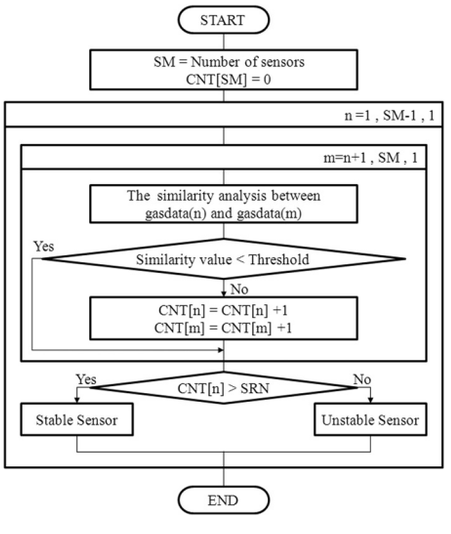


Fig. 8. The similarity analysis method using cosine similarity for the removal of unstable sensors.

본 논문에서는 SRN의 값에 변화를 주면서 적절한 값을 정하는 실험을 진행 하였으며, SRN에 의해 불안정 센서가 제거된 가스 데이터를 PCA로 표현하여 Fig. 9에서 나타내었다. Fig. 9에서 나타낸 것처럼 각 SRN의 값에 따라서 안정 센서로 선택된 센서들의 범위가 다르다. 안정 센서로 선택된 센서 중에서 대표 센서를 정해야 하기 때문에 SRN을 부적절하게 결정한다면, 불안정 센서가 대표 센서에 포함되어지기 때문에 신중한 선택을 해야 한다.

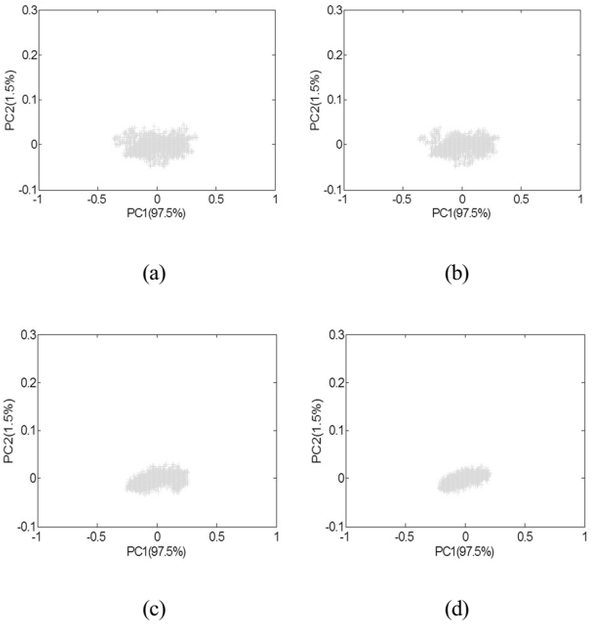


Fig. 9. Removal of unstable sensors according to standard redundant number(SRN); (a) SRN = 5, (b) SRN = 10, (c) SRN = 30, (d) SRN = 100,

4.1 코사인 유사도를 이용한 불안정 센서 제거

생물학적 후각 시스템의 경우, 비슷한 특성을 갖는 수많은 수용체들이 한 개의 사구체에 연결되어 한 개의 신호만을 출력한다. 즉, 한 개의 사구체와 연결되는 수용체들의 신호는 중복성을 가진다는 것을 의미한다. 생물학적 후각 시스템에서는 많은 중복성을 허용하고, 중복성을 갖는 수용체가 존재한다 하더라도 사구체에 의해 단 한 개만의 신호가 출력된다. 생물학적 후각 시스템을 모사하고 있는 대규모 가스 센서 어레이 시스템의 경우도, 수 십 개의 고분자 물질을 이용하여 비슷한 중복성을 갖도록 설계하였다. Fig. 10의 (a)는 에탄올이 주입되었을 때 108번 센서와 110번의 센서가 거의 비슷한 반응을 보이는 것을 나타낸다. 실제, 본 대규모 센서 어레이 시스템에서 중복도를 갖는 센서들은 수십 개 이상인 경우가 많다. Fig. 10의 (b)는 부탄올이 인가되었을 때 20번 센서와 비슷한 반응을 보이는 센서의 특성만을 나타내고 있는데, 약 96개의 센서가 비슷한 반응을 보이고 있다. 이러한 센서들은 서로 중복성을 갖는다고 할 수 있다.

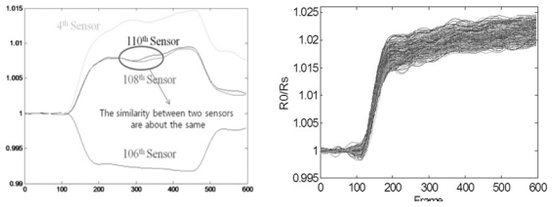


Fig. 10. Example of redundancy; (a) The example of redundancy at response from ethanol, (b) Sensor response that are similar to the response of 20th sensor from butanone.

대규모 센서 에레이 시스템에서도, 중복성을 갖는 많은 센서들이 존재하지만, 실제로 하나의 대표 센서가 출력되어야 한다. 하나의 대표 센서의 신호만을 출력하기 위해서는 특성이 비슷한 센서 중 센서를 대표할 수 있는 대표 센서 하나를 선정하고, 나머지 중복성을 갖는 센서는 신호처리 과정에서 제거되어야 한다. 코사인 유사도 알고리즘을 사용하여 불안정 센서를 제거하고 안정 센서를 선택하였지만, 안정된 센서에서도 많은 중복성이 있으며, 이 중복성은 하나의 출력으로 나타낼 필요가 있다. 따라서 비슷한 특성을 가진 센서들의 클러스터(cluster)중 가장 대표가 되는 센서만을 선택하는 방법을 사용하여 중복성을 제거하였다. 클러스터를 분할하는 클러스터링 알고리즘은 여러 가지가 존재하지만, 본 논문에서는 Fuzzy c-means algiorithm(FCMA)을 사용하였다. FCMA는 하나의 클러스터에 속해져 있는 각 데이터의 소속정도(membership grade)에 의해서 클러스터에 대한 데이터의 소속정도를 일일이 열거한 데이터 분류 알고리즘이다.

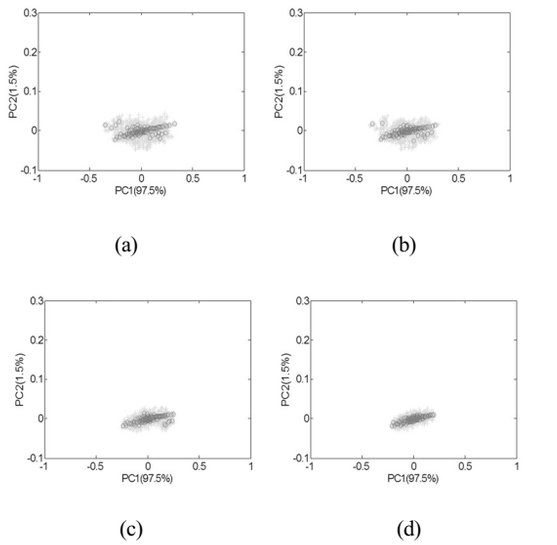


Fig. 11. Extraction of representative sensors using FCM Algorithm(The number of representative sensors = 100); (a) SRN = 5, (b) SRN = 10, (c) SRN = 30, (d) SRN = 100.

Fig. 11는 FCMA를 사용하여 불안정 센서가 제거 된 가스 센서의 데이터에서 클러스터의 중심이 되는 센서를 선택하여 나타내었다. 선택된 대표 센서는 비슷한 특성을 보이는 센서 데이터의 중복성을 하나의 대표 특성으로 나타낼 수 있다. FCMA를 적용하기 전 클러스터 중심의 수를 결정해야 한다. 클러스터 센터를 너무 많이 설정하면 각 클러스터 사이에 중복성이 존재하게 되며, 클러스터의 수를 너무 적게 설정하면, 하나의 클러스터가 포함하는 센서의 분포가 넓어져 다른 특성의 센서를 하나의 클러스터로 포함하기 때문에 결과에 좋지 않은 영향을 준다. 각 클러스터의 중심은 클러스터에 포함된 센서의 특징을 나타내는 지점이다. 따라서 클러스터 중심에서 소속도가 가장 높은 센서를 각 클러스터의 대표 센서로 선정하였다. FCM 알고리즘을 이용하여 선택한 대표 센서는 비슷한 특성을 보이는 센서 데이터의 중복성을 하나의 대표 특성을 통해서 나타낼 수 있다.

5. 유효성 입증

생물학적 후각을 모방한 대규모 가스 센서 어레이 시스템에서 데이터 압축과 중복 데이터 제거 방법의 실험에 대한 유효성을 입증하기 위해서 두 종류의 가스와 4가지 농도 그리고 각 농도에 대한 4번의 실험 데이터로 실험을 하였다. 전체 65536개의 센서 데이터는 4096센서 모듈 16개 데이터로 각각 나누어져있다. 본 논문의 실험에서는 16개의 데이터(L1, L2, L3, ... , L16)군 중 L2의 4096개의 센서 값만을 사용하여 실험을 하였다. 가스는 에탄올과 부탄올을 사용하였으며, 네 가지 농도 5 sccm, 10 sccm, 20 sccm, 40 scm를 사용하였다. 두 종류의 가스와 총 4가지 농도에서 얻은 실험 데이터 중 에탄올과 부탄올 각각 4차례의 실험의 결과를 이용하여 실험을 하였다. 그러나 유효성 입증 실험에서는 4096개의 센서 데이터를 사용하지 않고 한 농도에서 선택된 대표센서들만을 적용한다.

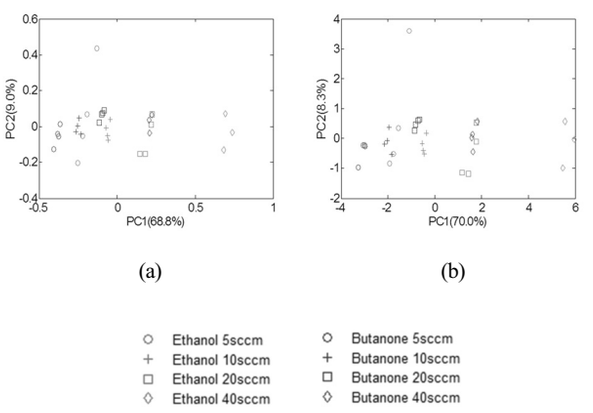


Fig. 12. The result of PCA using 4096 sensors; (a) Before the DWT, (b) After the DWT.

예를 들어, 에탄올 40 sccm에서 4차 실험에서 대표 센서 추출을 했다면 여기서 선택된 대표 센서들의 번호를 모든 데이터 군에 적용을 하는 것이다. 그러나 대표 센서를 선택할 때에는 낮은 농도 보다는 높은 농도에서 하는 것이 더 효과적이다.

낮은 농도에서는 가스별 특성이 잘 나타나지 않기 때문에 비교적 높은 농도의 대표 센서들을 선택해서 사용하는 것이 더 좋은 결과를 얻을 수 있다. 본 논문의 실험에서는 에탄올 20 sccm의 4차 실험에서 선택된 대표 센서들을 전체 데이터 군에 적용하였다. 본 논문에서 제안한 방법을 적용하지 않고 전처리 과정에서 3차 버터워스 필터와 DWT가 된 상태에서 4096 센서 데이터를 PCA를 사용하여 Fig. 12에 나타내었다.

본 논문에서 제안한 방법을 적용하지 않고 전처리 과정에서 3차 버터워스 필터와 DWT가 된 상태에서 4096 센서 데이터를 PCA를 사용하여 Fig. 12에 나타내었다. Fig. 12의 (a)는 DWT로 압축하지 않은 데이터로써 3차 버터워스 필터만을 적용한 대규모 가스 센서 어레이이다. (b)는 3차 버터워스 필터를 적용 후 DWT로 압축한 가스 센서 어레이이다. Fig. 12의 (a), (b)는 대표 센서를 선정하지 않고 4096개의 센서를 32 개의 실험으로 나타내었다. Fig. 12를 보면 부탄올과 에탄올이 구분되지 않고 중첩되어 있으며, 농도별로 군집을 형성하지 않았다. 즉, 부탄올과 에탄올의 특성을 구분해내지 못하였으며, 같은 농도의 실험에서도 비슷한 특성을 보이지 않음을 알 수 있다.

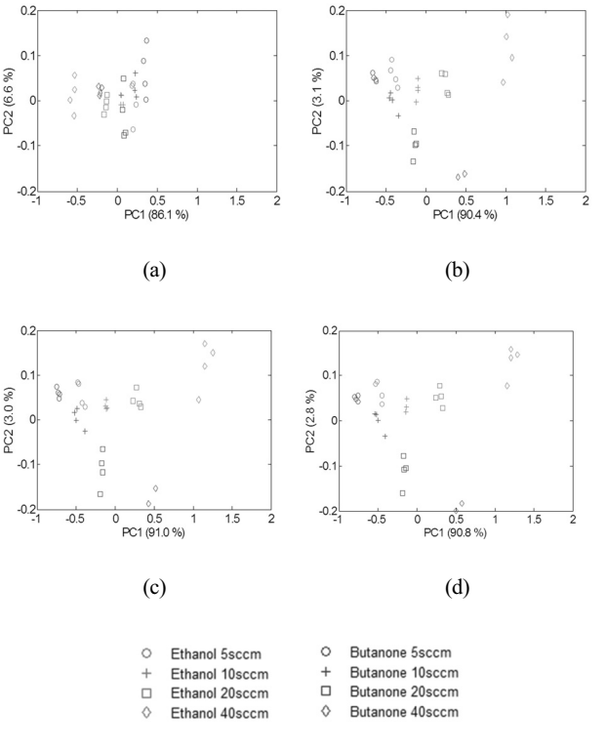


Fig. 13. The result of PCA when SRN is 50; (a) NRS = 50, (b) NRS = 100, (c) NRS = 200, (d) NRS = 300.

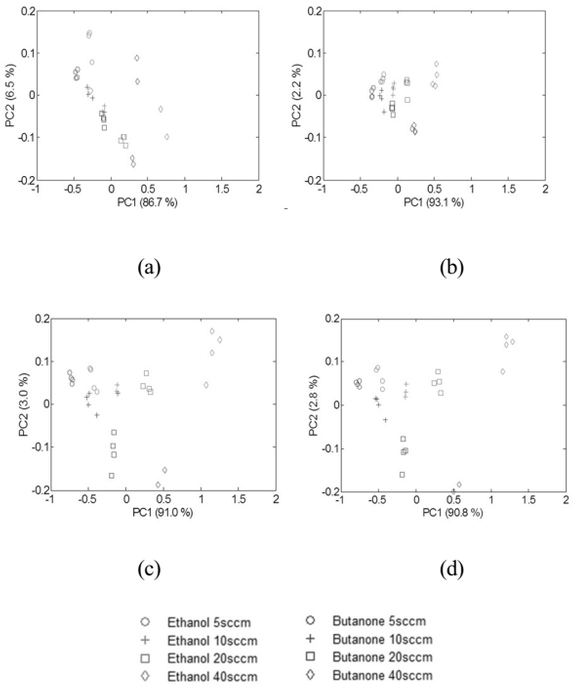


Fig. 14. The result of PCA when SRN is 100; (a) NRS = 50, (b) NRS = 100, (c) NRS = 200, (d) NRS = 300.

Fig. 13, 14에 해당하는 실험은 본 논문에서 제안하는 방법을 적용할 때, 중복도와 대표센서 수가 결과에 미치는 영향을 나타내었다. 실험을 한 결과 SRN에 변화를 주어도 대표 센서의 수가 같을 때 결과에 큰 차이가 없다. 그러나 NRS의 값에 따라 결과에 큰 영향을 주었다. 각 Fig. 13, 14의 (a), (b), (c), (d)를 보면 대표 센서가 100 이상 선택되었을 때 에탄올과 부탄올간 분류가 명확하게 되는 것을 볼 수 있으며, 각 가스별 농도끼리 군집을 이루는 것을 확인할 수 있다. 그리고 농도가 짙어 질수록 각 가스별로 뚜렷하게 나누어지는 것을 볼 수 있다. 이러한 경향성은 다른 농도가 적용되었을 때에도 그 경향을 예측이 가능하다.

본 논문에서 제안한 방법은 대규모 가스 센서 어레이 시스템에 적용하여 데이터의 크기를 줄여 연산의 속도를 높였다. 또한, 불안정 센서를 제거하여 센서의 특성을 명확히 나타내었으며, 대표 센서의 선택으로 불필요한 중복성을 제거 하였다. 그리고 대규모 가스 센서 어레이 시스템에서 가스별 특성과 농도별 특성을 표현하는데 매우 우수한 성능을 보였다.

6. 결론

생물학적 후각 시스템을 모방한 대규모 가스센서 어레이 시스템은 인간의 후각 시스템과는 다르게 하루 동안에도 지치지 않고 수천 개의 냄새를 구분할 수 있다는 장점이 있다. 그러나 대규모 가스 센서 어레이에서 생기는 불안정한 센서를 포함한다는 점과 수 만개 센서에서 얻은 데이터가 너무 크고 감당할 수 없는 계산 시간이 필요하다는 단점이 존재한다. 본 논문에서는 생물학적 후각 시스템을 모방한 가스 센서 어레이의 단점을 해결하기 위해서 첫 번째, 대규모 가스 센서 어레이 신호 중 불안정한 센서 신호들을 검출하여 제거하는 방법과 두 번째, 센서간의 중복성을 제거하기 위해서 대표 센서를 선정하는 방법을 제안하였다. 불안정한 센서를 제거하기 위해서 중복도라는 기준을 적용하였다. 이 중복도는 코사인 유사도를 사용하여 특성이 비슷한 센서의 개수이다. 중복도가 높다는 것은 주입된 가스에 같은 특성을 보이는 센서들이 많이 존재한다는 것이기 때문에 안정된 센서를 찾는 기준으로 사용하였다. 그리고 중복도에 의해서 제거된 불안정한 센서들을 PCA를 통해서 확인하였다. 또한, 불필요한 중복을 줄이기 위해 FCMA를 사용하여 대표 센서를 선정하였다. 대표 센서는 FCMA가 클러스터를 생성하고 클러스터 중심을 얻는 과정과 같다. 여기서 찾은 클러스터는 매우 유사한 특성을 보이는 센서의 군집이기 때문에 중복성을 보인다고 판단하였다. 따라서 이 클러스터의 중심에서 가장 소속도가 높은 센서를 찾아 대표 센서로 선정하였다. 이러한 대표 센서의 선정은 센서의 수를 수천 개에서 수백 개로 줄여 연산시간을 단축시켰다. 그러나 대표 센서의 수에 따라서 결과에 매우 큰 영향을 끼치는 것을 확인하였다. 따라서 대표 센서의 수를 보다 신중하게 선택되어야 한다.

본 논문에서 제시한 방법은 불필요한 중복성을 제거하고 대표 센서만을 이용하여 적은 데이터 양으로 데이터의 분류 속도를 향상 시켰다. 또한 각 가스와 농도별로 군집을 이루어 분류함으로써 제안된 방법을 적용하기 전과 비교하여 우수한 분류 및 군집 성능을 보이는 결과를 입증하였다.

REFERENCES

[1] M. H. Song, H. D. Park, K. J. Lee, and K. L. Park, “A Study on the detection of the ventricular fibrillation based on wavelet transform and artificial neural network”, *KIEE*, vol. 53, no. 11, 2004.

[2] S. W. Shin, K. S. Kim, J. W. Lee K. H. Lee, and D. J. Kim, “Noise reduction and estimating the similarity of ambulatory ECG signals”, *KIEE*. vol. 57, no. 3, 2008.

[3] M. H. Song, J. Lee, S. P. Cho, and K. J. Lee, “SVM classifier for the detection of ventricular fibrillation”, *IEEK*, vol. 42, no. 5, 2005.

[4] A. C. Douglas and M. S. Richard, “An approach to cardiac arrhythmia analysis using hidden Markov models”, *IEEE Transactions on Biomedical Engineering*, vol. 37, no. 9, pp. 826-836, 1990.

[5] A. Pachauri and M. Bhuyan, “Wavelet and energy

based approach for PVC detection", *International Conference on Emerging Trends in Electronic and Photonic Devices and Systems*, pp. 258-261, 2009.

[6] S. R. R. Alfredo, E. C. Paraiso, and C. A. A. Kaestner, "Automatic detection of arrhythmias using wavelets and self-organized artificial", *International Conference on Intelligent Systems Design and Applications*, pp. 648-653, 2009.

[7] J. A. Nasiri, M. Sabzekar, H. S. Yazdi, M. Naghibzadeh, and B. Naghibzadeh, "Intelligent arrhythmia detection using genetic algorithm and emphatic SVM(ESVM)", *UK Sim European Symposium on Computer Modeling and Simulation*, pp. 112-117, 2009.

[8] N. Srinivasan, M. T. Wong, and S. M. Krishnan, "A new phase space analysis algorithm for cardiac arrhythmia detection", *International Conference of the IEEE EMBS*, pp. 17-21, 2003.

[9] J. S. Bae and T. L. Song "Image tracking algorithm using template matching and PSNF-m", *International Journal of Control Automation and Systems*, vol. 6, no. 3, pp. 413-423, 2008.

[10] H. Sakoe and S. Chiba, "Dynamic programming alogrithm optimization for spoken word recognition", *IEEE Trans. on ASSP*, vol. 26, no. 1, pp. 43-49, 1978.

[11] R. Beccherelli., E.Zampetti, S.Pantalei, M. Bernabei, and K.C.Persaud, "Design of a very large chemical sensor system for mimicking biological olfaction", *Sensors and Actuators, B*, vol. 146, pp. 446-452, 2010.

김 정 도 (Jeong-Do Kim)

- 1987년 성균관대학교 전자공학과(공학사)
- 1990년 성균관대학교 전자공학과(공학석사)
- 1994년 성균관대학교 전자공학과(공학박사)
- 2004년 삼척대학교 제어계측공학과 교수
- 현재 호서대학교 전자공학과 교수
- 관심분야 : 오감정보처리, 센서신호처리

김 정 주 (Jung-Ju Kim)

- 2008년 호서대학교 전자공학과 (공학석사)
- 2009년~현재 호서대학교 전자공학과 박사과정
- 관심분야 : 유비쿼터스, 임베디드시스템, 센서네트워킹, HCI

박 성 대 (Sung-Dae Park)

- 2007년 호서대학교 전자공학과(공학사)
- 2009년 호서대학교 전자공학과(공학석사)
- 2009년~현재 호서대학교 전자공학과 박사과정
- 관심분야 : 스마트센서, 임베디드시스템

변 형 기 (Hyung-Gi Byun)

- 1995년 영국 UMIST, Ph.D
- 1996년~현재 강원대학교(삼척캠퍼스)정보통신공학과 교수
- 주관심분야 : 센서 시스템 신호처리 및 패턴 인식

Krishna Chandra Persaud

- 1977년 University of Warwick, UK, Ph.D
- 1988년~현재 Professor of Chemoreception in the School of Chemical Engineering and Analytical Science, the University of Manchester.
- 주관심분야 : Electronic nose, conducting polymer sensor array

임 승 주 (Seung-Ju Lim)

- 2010년 호서대학교 전자공학과(공학사)
- 2012년 호서대학교 전자공학과(공학석사)
- 관심분야 : 스마트센서, 임베디드시스템, tm마트 인터페이스