

## 패턴분류 기술을 이용한 후각센서 어레이 개발

박 태 원, 이 진 호<sup>†</sup>, 조 영 충\*, 안 철\*\*

연세대학교 기계공학과, \*연세대학교 CT연구단, \*\*(주)수도프리미엄엔지니어링

### Development of Odor Sensor Array using Pattern Classification Technology

Tae-won Park, Jinho Lee<sup>†</sup>, Young-Chung Cho\*, Chul Ahn\*\*

School of Mechanical Engineering, Yonsei University, Seoul 120-749, Korea

\*\*Sudo premium Engineering Co.,Ltd, Seoul 150-093, Korea

**ABSTRACT:** There are two main streams for pattern classification technology. One is the method using PCA (Principal Component Analysis) and the other is the method using Neural network. Both of them have merits and demerits. In general, using PCA is so simple while using neural network can improve algorithm continually. Algorithm using neural network needs so many calculations rendering very slow response. In this work, an attempt is made to develop algorithms adopting both PCA and neural network merits for simpler, but faster and smarter.

**Key words:** Sensor Array(센서어레이), Sensor Fusion(센서 융합), Oder Sensor(후각센서), Patter Classification(패턴분류)

#### 1. 서 론

현대인들은 하루의 거의 모든 시간을 실내에서 보내고 있지만 현대 건축물의 실내 환경은 과거 인간들이 향유하던 환경과는 상당한 차이가 있다. 좁은 땅 위에 많은 인구가 거주하기 위해 건물은 높아지고 구조도 점점 폐쇄적으로 변해버렸다. 그 결과 실내 환경은 자연법칙이 아니라 대부분 기계 동력에 의존할 수밖에 없는 실정이다. 그러나 기계가 만들어낸 실내 환경은 열악하다. 유해한 건축자재 등을 사용하면서 실내의 오염물질이 증가하여 실내공기가 오염됐지만 환기는 불충분하다. 이런 이유 때문에 '건물병증후군', '새집

증후군'등의 신조어가 등장하기 시작하였다. 또한 땅으로부터 건물로 들어오는 기체 형태의 방사선 라돈이나 세균, 진드기와 같은 생물학적 유해 요인으로 인해 호흡기나 피부질환, 내과나 소아과 질환이 유발되고 있다. 실내 공기 오염물질의 발생원은 이 이외에도 의류, 화장품, 스프레이제품, 방충제, 세정제, 연소 기구, 흡연, 애완동물, 쓰레기, 자동차 배기가스, 연소물질까지 다양하다. 따라서 이러한 오염 물질을 정확하고 신속하게 측정하는 것은 건강한 실내 환경을 위해 매우 필수적이다.

여러 가지 오염 물질을 냄새로서 측정하는 장치인 후각 센서 어레이(odor sensor array)를 이용한 전자코 시스템(electric nose system)은 그 특성상 다양한 물질을 더 높은 신뢰성으로 측정할 수 있으나 센서의 종류와 어레이에 따라 성능이 상이하다. 하지만 기존 시중의 센서 제품들은

<sup>†</sup>Corresponding author

Tel.: 02-2123-2816; fax: 02-312-2159

E-mail address: jinholee@yonsei.ac.kr

신뢰도와 민감도, 재연성이 낮은 한계가 있다. 이를 개선하기 위해서는 새로운 센서 개발도 한 방법이 될 수 있으나 본 연구에서는 Sensor Fusion을 이용하여 신뢰도와 민감도, 재연성 및 반응속도를 높이는 것을 목표로 한다.

케미컬 가스 센서는 공기 중 특정 성분의 존재나 그 양을 검지 정량하는 장치이다. 따라서 여러 가지 가스를 동시에 검지, 정량할 수 있는 고성능 복합가스 측정 장치를 개발하기 위해서는 센서의 소형화 및 성능 향상과 더불어 최근 개발에 활기를 띄고 있는 센서 어레이에 관한 기술 개발이 필수적이다. 가스센서가 갖고 있는 기능이 어떤 특정 가스를 검지 정량하는 것이나 대부분의 경우 측정대상가스 이외의 가스가 혼합되어 있을 경우 하나의 개별소자를 이용하는 것보다 여러 개의 센서를 조합한 어레이를 쓰게 되면 각 개별소자가 반응하는 응답패턴으로부터 어떤 특정가스의 존재를 보다 확실하게 알 수가 있고 동시에 그 가스의 정량을 보다 신뢰성 있게 할 수 있다 이와 같이 센서 어레이를 사용하면 가스센서의 생명인 선택성을 크게 높일 수가 있을 뿐만 아니라 복합가스를 인식하는데 유용한 장치의 구현이 가능하다.

본 연구에서는 센서 어레이에서의 간섭을 제거하기 위해 사용되었던 주성분 분석(PCA - principal component analysis)과 신경회로망(neural network)을 융합하여 신경회로망의 지능적이고 정확한 장점은 살리면서도 약점으로 지적되던 많은 계산에 의한 느린 반응 속도를 개선하고자 한다.

## 2. 실험장치 및 방법

### 2.1 실험장치

실험 장치는 다음과 같이 가스감도 시험조와 알고리즘 개발용 가스 측정기로 구성되어 있다.



Fig 1. Gas chamber



Fig 2. Gas detectors for developing algorithm

### 2.2 실험방법

실험은 KS C 6566에 의거하여 온도  $20 \pm 3^\circ\text{C}$ , 습도  $60 \pm 7\%RH$ , 센서 당 용적 1l 이상, 풍속 0.5 m/s이내의 환경에서 수행되었다. 사용 가스는 포름알데히드(HCHO), 암모니아( $\text{NH}_3$ ), 황화수소( $\text{H}_2\text{S}$ )이며 구체적인 실험 방법은 다음과 같다.

- 1) 측정 장비를 구성하여 가스감도 시험조에 넣은 후 정격 전압을 인가하고 30분 동안 소자 안정화 시간을 둔다.
- 2) 가스투입 전 청정대기 출력이 0인 것을 확인하고 시험가스를 발생시킨다.
- 3) 1시간 동안 출력을 측정, 기록하고 환기한다.

### 2.3 실험결과 처리 방법

실험 데이터는 주성분 분석을 이용한 알고리즘 I, 신경회로망을 이용한 알고리즘 II, 그리고 주성분 분석과 신경회로망을 동시에 이용한 알고리즘 III의 개발에 이용되었다.

#### 2.3.1 주성분 분석

주성분 분석은 서로 상관관계가 높은 여러 개의 변인들을 조합해서 그 변인들의 정보를 가능한 한 많이 함축하고 있는 더 작은 수의 새로운 변인들을 만들어내는 방법으로서, 이 때 새롭게 만들어진 변인을 “주성분”이라고 부른다. 주성분 분석은 원래 변인들의 “선형 조합”을 통하여 여러 개의 변인을 요약, 정리한다. 가장 대표적인 방법은 원래 변인들의 상관관계 행렬을 이용하여 고유값(eigenvalue), 고유벡터(eigenvector)를 도출하는 것이다.

주성분 분석을 사용하면, 상호 상관관계를 가지고 있는 여러 개의 변인들을 상호 독립적인 몇 개의 주성분으로 변환시키게 된다. 일반적으로 5

개 이하의 주성분을 가지고 원래 변인들의 분산의 75% 이상을 의미있게 설명할 수 있다면, 주성분 분석이 매우 성공적이었다고 할 수 있다.

### 2.3.2 신경회로망 오류역전파 학습 알고리즘

신경 회로망이란 디지털 컴퓨터가 순차적으로 업무를 처리하는 반면, 병렬로 업무를 처리하고, 어떤 일을 수행하기 위해서는 미리 학습이 필요한 인간의 뇌를 모사한 것으로 결합 내구성과 합리성, 적용성의 우수성을 지니고 있다. 구성상의 특징은 처리 요소들간의 연결 가중치(weight)를 사용, 다층망(multilayer network)을 구현한다는 것이다.

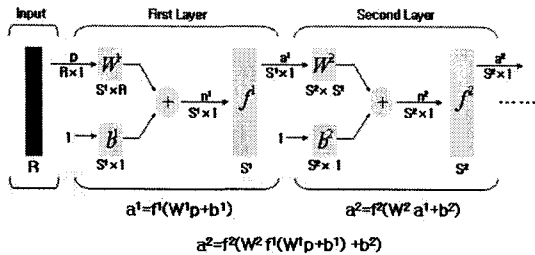


Fig 3. Neural Network

다층(multilayer), 지도학습(supervised learning)의 경우 다음과 같이 학습한다.

1. 데이터를 입력하고, 그에 따른 출력을 계산
2. 입력에 따른 출력과 원하는 출력간의 오차를 계산
3. 오차를 줄이기 위해 가중치의 증감 여부를 결정
4. 각각의 가중치를 얼마나 변화시킬 것인가를 결정
5. 4 단계에서 결정된 값으로 가중치를 갱신
6. 모든 학습 데이터에 대해 오차가 적정 수준으로 감소하기까지 1 ~ 5단계를 반복

역전파(back-propagation)라고 하는 이유는 가중치를 갱신하는 방향이 신경망의 처리 방향과 반대이기 때문이다.

## 3. 결과 및 고찰

### 3.1 센서 간섭

포름알데히드(HCHO) 가스와 암모니아(NH<sub>3</sub>) 가스의 경우 각각 포름알데히드(HCHO) 측정기, 암모니아(NH<sub>3</sub>) 측정기에만 신호가 잡혀 아무런 간섭이 없었다. 하지만 황화수소(H<sub>2</sub>S) 가스의 경우 황화수소(H<sub>2</sub>S) 측정기 뿐만이 아니라 포름알

데히드(HCHO) 측정기와 암모니아(NH<sub>3</sub>) 측정기 모두에 신호로 잡혔다.

회수	HCHO	H <sub>2</sub> S	NH <sub>3</sub>
1	0.05	0	0
2	0.1	0	0
3	0.19	0	0
4	0.51	0	0
5	0.97	0	0

Table 1. Measurement of HCHO gas (ppm)

회수	HCHO	H <sub>2</sub> S	NH <sub>3</sub>
1	1.16	1.8	0.2
2	3.62	3.6	4.8
3	6.09	7	7.6
4	7.1	8.1	9.3
5	29.12	26.4	35.8

Table 2. Measurement of H<sub>2</sub>S gas (ppm)

회수	HCHO	H <sub>2</sub> S	NH <sub>3</sub>
1	0	0	1.1
2	0	0	2.1
3	0	0	4.8
4	0	0	12.3
5	0	0	34.3

Table 3. Measurement of NH<sub>3</sub> gas (ppm)

	HCHO	H <sub>2</sub> S	NH <sub>3</sub>
실제	0.64	5.2	14.7
측정	5.31	5.2	25.6

Table 4. Measurement of mixed gas (ppm)

### 3.2 알고리즘 적용 결과

위 실험 결과에서 볼 수 있듯이 세 가스가 섞여있는 경우 측정기 3개에서 각각 측정된 값은 정확한 값이 아니라 다른 가스의 간섭에 의한 잘못된 값이 나온다. 이것을 알고리즘으로 개선한 결과는 아래와 같다.

### 3.2.1 주성분 분석 알고리즘

주성분 분석을 이용한 알고리즘은 공분산 행렬 (covariance matrix)에서 고유값과 고유벡터를 구하여 세 가지 가스와 계측기간의 관계를 도출하여 작성하였다. 공분산 행렬과 고유값, 고유벡터, 상호 관계식은 다음과 같다.

$$\begin{aligned} \text{공분산 행렬} &: \begin{bmatrix} 7.073 & 7.721 & 10.512 \\ 7.721 & 8.583 & 11.311 \\ 10.512 & 11.311 & 15.809 \end{bmatrix} \\ \text{고유값} &: [0.0007 \quad 0.3249 \quad 31.1387] \\ \text{고유벡터} &: \begin{bmatrix} 0.8751 & -0.0848 & 0.4765 \\ -0.3578 & -0.7763 & 0.519 \\ -0.3259 & 0.6246 & 0.7097 \end{bmatrix} \end{aligned}$$

$$\begin{aligned} \text{HCHO(gas)} &= \text{HCHO(meter)} - 0.9181\text{H}_2\text{S(meter)} \\ \text{NH}_3(\text{gas}) &= \text{NH}_3(\text{meter}) - 1.3665\text{H}_2\text{S(meter)} \end{aligned}$$

		HCHO	H <sub>2</sub> S	NH <sub>3</sub>
실제		0.64	5.2	14.7
측정	값	5.31	5.2	25.8
	오차	814%	0%	75.5%
알고리즘	값	0.54	5.2	18.5
	오차	15.6%	0%	25.8%

Table 5. Result of PCA algorithm (ppm)

### 3.2.2 신경회로망 오류역전과 학습 알고리즘

신경회로망 오류역전과 학습 알고리즘은 첫 번째 층(first layer)에서의 함수는 log-sigmoid 함수  $\left(\frac{1}{1+e^{-n}}\right)$ 를, 두 번째 층(second layer)에서의 함수는 linear 함수(n)를 이용하였다. 65,534회에서 10<sup>-5</sup>의 오차 허용 범위 이내에 들어서 그 값을 사용하였다. 그때의 가중치의 값과 알고리즘 적용 결과는 다음과 같다.

$$\begin{aligned} W_1 &= \begin{bmatrix} 4.492 & -3.324 & -0.031 \\ -2.410 & -3.798 & -4.186 \\ 0.082 & -3.356 & 2.099 \end{bmatrix}, b_1 = \begin{bmatrix} 0.139 \\ 0.123 \\ 1.162 \end{bmatrix} \\ W_2 &= \begin{bmatrix} 1.592 & 0.596 & -3.518 \\ -15.234 & 2.696 & -0.820 \\ -0.513 & -2.042 & 3.576 \end{bmatrix}, b_2 = \begin{bmatrix} 2.711 \\ 8.534 \\ 0.284 \end{bmatrix} \end{aligned}$$

		HCHO	H <sub>2</sub> S	NH <sub>3</sub>
실제		0.64	5.2	14.7
측정	값	5.31	5.2	25.8
	오차	814%	0%	75.5%
알고리즘	값	0.284	2.71	8.53
	오차	55.6%	47.9%	42.0%

Table 6. Result of Neural Network(BP) algorithm (ppm)

### 3.2.3 주성분 분석과 신경회로망 오류역전과 학습 알고리즘의 혼합

먼저 주성분 분석을 이용하여 2개의 주성분으로 축약하였다. 공분산 행렬과 고유값, 고유벡터는 다음과 같다.

$$\begin{aligned} \text{공분산 행렬} &: \begin{bmatrix} 7.987 & 9.347 & 5.714 \\ 9.347 & 11.089 & 7.604 \\ 5.714 & 7.604 & 20.020 \end{bmatrix} \\ \text{고유값} &: [0.0429 \quad 10.0234 \quad 29.0302] \\ \text{고유벡터} &: \begin{bmatrix} 0.7503 & -0.4986 & 0.4341 \\ -0.6601 & -0.5286 & 0.5337 \\ 0.0366 & 0.687 & 0.7257 \end{bmatrix} \end{aligned}$$

따라서 다음과 같은 두 개의 주성분으로 축약할 수 있다.

$$\begin{aligned} \text{주성분1} &: -0.4986 \times \text{HCHO} - 0.5286 \times \text{H}_2\text{S} + 0.687 \times \text{NH}_3 \\ \text{주성분2} &: 0.4341 \times \text{HCHO} + 0.5337 \times \text{H}_2\text{S} + 0.7257 \times \text{NH}_3 \end{aligned}$$

이것을 3.2.2에서 언급한 신경회로망 오류역전과 학습 알고리즘과 같이 첫 번째 층(first layer)에서의 함수는 log-sigmoid 함수  $\left(\frac{1}{1+e^{-n}}\right)$ 를, 두 번째 층(second layer)에서의 함수는 linear(n) 함수를 이용하여 계산한 결과 4238회에서 10<sup>-5</sup> 수준의 오차 허용 범위 내에 들었으며 가중치의 값은 다음과 같다.

$$\begin{aligned} W_1 &= \begin{bmatrix} -6.392 & -6.551 \\ 29.013 & 55.559 \end{bmatrix}, b_1 = \begin{bmatrix} 3.926 \\ 2.226 \end{bmatrix} \\ W_2 &= \begin{bmatrix} -7.089 & 4.997 \\ -6.976 & 5.852 \end{bmatrix}, b_2 = \begin{bmatrix} 0.256 \\ 2.647 \end{bmatrix} \end{aligned}$$

위 식을 이용하여 계산한 후 다시 변환한 결과는 다음과 같다.

		HCHO	H <sub>2</sub> S	NH <sub>3</sub>
실제		0.64	5.2	14.7
측정	값	5.31	5.2	25.8
	오차	814%	0%	75.5%
알고리즘	값	0.2	3.5	9.5
	오차	68.8%	32.7%	35.4%

Table 7. Result of PCA and Neural Network(BP) algorithm (ppm)

#### 4. 결론

알고리즘 결과를 종합하면 다음과 같다.

		HCHO	H <sub>2</sub> S	NH <sub>3</sub>
실제		0.64	5.2	14.7
측정	값	5.31	5.2	25.8
	오차	814%	0%	75.5%
알고리즘 I	값	0.54	5.2	18.5
	오차	15.6%	0%	25.8%
알고리즘 II	값	0.284	2.71	8.53
	오차	55.6%	47.9%	42.0%
알고리즘 III	값	0.2	3.5	9.5
	오차	68.8%	32.7%	35.4%

Table 8. Result of Various algorithm (ppm)

H<sub>2</sub>S 측정기의 경우 다른 가스와 간섭이 생기지 않았으므로 신경회로망을 이용한 알고리즘 II 나 주성분 분석과 신경회로망을 이용한 알고리즘 III에서의 오차는 고려의 대상이 되지 않는다. HCHO 가스 측정의 경우 H<sub>2</sub>S 가스의 농도에 따라 다르지만 위의 경우 800% 이상의 오차가 나던 것을 세 가지 알고리즘 모두 50% 내외의 오차로 줄일 수 있었고 NH<sub>3</sub>의 경우 75% 이상의 오차를 25~42% 정도로 줄일 수 있었다. 다만 어느 알고리즘이 가장 우수한지에 대한 것은 이번 경우만으로 판단하기는 이르다. 이 경우에는 주성분 분석을 이용한 알고리즘 I의 경우가 가장 오차를 많이 줄이는 것으로 나왔으나 실제 센서 어레이 제작 시 적용을 포함한 다양한 경우에 항상 그러할지는 미지수이다. 신경회로망을 이용한 알고리즘 II는 비록 오차를 알고리즘 I만큼 줄이지는 못하였으나 신경회로망 특유의 학습성으로 설치 장소에 알맞은 가중치를 계속 보정할 수 있다는 장점이 있다. 다만 무수한 반복에 따른 계산 부하와 그로인해 발생하는 오랜 계산 시간의 문제가 단점으로 지적되고 있다. 따라서 이것을

보안하기 위하여 주성분 분석으로 데이터를 축약한 후 신경회로망으로 값을 찾아가는 알고리즘 III를 고안하였다. 이번 실험의 데이터를 처리하는데 99.94%의 데이터를 보존하면서 축약할 수 있어서 Table 8.에서 볼 수 있듯이 알고리즘 II와 크게 차이 없는 오차를 지니고 있다. 따라서 이 알고리즘은 여러 장점을 가지고 있음에도 그 계산 부하와 느린 속도 때문에 널리 적용되지 못했던 신경회로망 알고리즘의 개선 방법으로 활용되어 센서 어레이에의 신경회로망 적용을 더욱 가속화 시킬 수 있으리라 생각된다.

#### 참고문헌

1. Lee, D. S., Huh, J. S., and Lee, D. D., 2001, "Recognition of Inflammable Gases Using Sensor Array and Principal Component Analysis", Journal of the Korean Sensors Society, Vol. 10, No. 2, pp. 108 -116
2. Do, Y. T., 1995, "A Technique to Reduce Measurement Uncertainty Using Multisensors", Journal of the Korean Sensors Society, Vol. 4, No. 4, pp. 47-54
3. Kim, J. D., Byun, H. G., and Ham, Y. K., 2004, "Design of a Portable Electronic Nose System using PDA", Journal of the Korean Sensors Society, Vol. 13, No. 6, pp. 454-461
4. Byun, H. G., Lee, J. S., and Kim, J. D., 2004, "Implementation of a Portable Electronic Nose System for Field Screening", Journal of the Korean Sensors Society, Vol. 13, No. 1, pp. 41-46
5. Kim, B. H., Lee, J. H., and Kim, T. J., 2005, "Measurement system of multi-components in indoor air quality", Journal of the Korean Sensors Society, Vol. 14, No. 2, pp. 125-130
6. Alex M. Martinez, Member, IEEE, and Avinash C. Kak, 2001, "PCA versus LDA", IEEE Transactions on pattern analysis and machine intelligence, Vol. 23, No. 2, pp. 228-233
7. Tarik K. Alkasab, Joel White and John S. Kauer, 2002, "A computational system for

simulating and analyzing arrays of biological and artificial chemical sensors", *Chem. Senses* Vol. 27, pp. 261-275

8. Tim C. Pearce and Manuel A. Sanchez-Montanes, 2002, "Chemical sensors array optimization: Geometric and information theoretic approaches", *Handbook of Machine Olfaction*, Wiley-VCH.