

유전 알고리즘과 러프 집합을 이용한 계층적 식별 규칙을 갖는 가스 식별 시스템의 설계

Design of Gas Identification System with Hierarchical Rule base using Genetic Algorithms and Rough Sets

방 영 근* · 변 형 기** · 이 철 희†

(Yonug-Keun Bang · Hyung-Gi Byun · Chul-Heui Lee)

Abstract - Recently, machine olfactory systems as an artificial substitute of the human olfactory system are being studied actively because they can scent dangerous gases and identify the type of gases in contamination areas instead of the human. In this paper, we present an effective design method for the gas identification system. Even though dimensionality reduction is the very important part, in pattern analysis, We handled effectively the dimensionality reduction by grouping the sensors of which the measured patterns are similar each other, where genetic algorithms were used for combination optimization. To identify the gas type, we constructed the hierarchical rule base with two frames by using rough set theory. The first frame is to accept measurement characteristics of each sensor and the other one is to reflect the identification patterns of each group. Thus, the proposed methods was able to accomplish effectively dimensionality reduction as well as accurate gas identification. In simulation, we demonstrated the effectiveness of the proposed methods by identifying five types of gases.

Key Words : Machine olfactory system, Dimensionality reduction, Genetic algorithms, Rough set theory

1. 서 론

최근, 위험한 현장이나 환경 하에서 인간의 인지 능력에 비해 보다 높은 항상성과 안정성, 그리고 정확성을 제공할 수 있는 인공 감각기관 개발에 관한 연구가 활발히 진행되고 있는데, 그 중에서 인간의 후각을 모사한 전자코 시스템은 오염 지역에서의 가스 식별, 음식물의 신선도 측정, 박테리아 서식의 식별 등과 같이 여러 분야에서 응용되고 있다 [1-2].

가스 식별(전자코) 시스템은 기본적으로 센서에서 검출된 화학신호를 전기신호로 변환하고, 변환된 신호들을 전처리한 후, 패턴을 분석하기 위한 기법들과 결합되어 구현된다[3]. 성능이 우수한 가스 식별 시스템을 위해서는 센서 측정 신호로부터 다양하게 나타나는 패턴들을 분석하고 가스의 종류를 식별할 수 있는 기술이 화학신호를 검출하는 센서 기술 이상으로 중요하다[4-5]. 가스 검출을 담당하는 센서부에는 가스 감지와 식별의 정확성을 높일 목적으로 여러 개의 센서를 조합한 센서 어레이를 사용하는 것이 일반적인데, 그렇게 함으로써 개별 센서의 선택성 결여와 장기간 사용 시의 데이터 유동 영향을 줄일 수 있을뿐더러, 여러 가스가 혼합되어 있을 경우 어레이를 구성하는 각 개별 센서의 응답

패턴을 종합적으로 분석하여 특정 가스의 존재와 정량을 보다 신뢰성 있게 판별할 수 있게 된다. 그런데, 센서 개수가 늘어날수록 패턴 분석을 위한 데이터의 양과 차원이 커져서 어려움을 겪게 되므로, 대부분의 기존 연구들에서는 패턴 인식에 유용한 정보들만으로 원시 데이터를 재구성하여 차원을 축소하고 효율적인 인식이 가능하도록 하는 특징 추출 과정을 거친 뒤에 군집화(clustering)나 분류(classification) 기법을 적용하여 최종적인 가스 식별을 수행하고 있다. 하지만, 일반적으로 특징 추출 과정은 데이터 구조 분석을 위한 복잡한 수학적 절차가 요구되며, 차원의 축소에는 효과를 보지만 센서의 유동 특성 등 전체적인 시스템의 특성을 제대로 반영하지 못하는 한계를 보이는 경향이 있다 [6].

최근에 센서의 유동 현상과 데이터의 고차원 문제도 보다 유연하게 대처할 수 있는 소프트 컴퓨팅 기법의 활용이 시도되고 있으며, 그 중 대표적인 예가 퍼지이론이나 신경망을 이용한 식별 시스템의 구현 기법이다. 이들은 효과적인 차원축소와 동시에 데이터의 불규칙한 패턴을 능동적으로 분석할 수 있도록 하여 센서 데이터의 유동 현상에 따른 성능 저하를 최소화 할 수 있도록 하고 있다 [7-9].

본 논문에서는 정확도와 신뢰성이 개선된 가스 식별 시스템의 설계를 위한 새로운 방법을 제시한다. 먼저, 기존 식별 시스템들에서 발생하는 차원의 저주로 인한 패턴 분석의 어려움을 피하면서도 센서 어레이에 구성되어 있는 모든 센서들의 측정 패턴을 고려할 수 있도록 유전알고리즘[10-11]을 이용한 그룹화 과정을 적용한다. 이는 다수의 센서들을 소수의 센서들로 구성된 그룹으로 나눔으로써 차원을 축소할 수 있는 구조를 가지게 되며, 또한 유전알고리즘의 최적화 특성으로 서로 유사한 측정 패턴을 가지는 센서들끼리 같은

* 정 회 원 : 강원대학교 신재생에너지 연구소

** 비 회 원 : 강원대학교 전자 정보통신공학부 교수

† 교신저자 : 강원대학교 전기전자공학과 교수

E-mail : chlee@kangwon.ac.kr

접수일자 : 2012년 3월 2일

최종완료 : 2012년 7월 24일

그룹으로 묶게 되어 그룹 내의 패턴 분석이 보다 효과적으로 이루어질 수 있게 한 것이다. 유전알고리즘의 최적화 성능을 개선하기 위해 본 논문은 통계적 최적화 기법과 유전알고리즘의 진화적 최적화 기법이 접목된 하이브리드 유전알고리즘(Hybrid Genetic Algorithm : HGA)[12-13]을 적용하여 탐색된 해의 적합성을 개선하고 알고리즘의 수렴성능을 높일 수 있도록 하였다. 그룹화의 과정이 끝나면, 각 그룹마다 가스의 종류를 식별하기 위한 규칙기반을 생성하게 되는데, 유효한 규칙의 처리 및 규칙기반의 감축에 강력한 도구가 될 수 있는 러프집합 이론[14-15]을 적용하여 보다 적은 수의 효과적인 규칙기반이 생성되도록 하였다. 그룹별로 생성된 규칙기반을 이용하여 일차적으로 가스의 종류에 대한 식별이 각각 이루어지게 되며, 이 후, 각 그룹의 식별 패턴을 분석하고 이를 토대로 최종 시스템 출력을 결정할 상위의 식별 규칙기반을 별도로 구성함으로써 센서 어레이에 포함된 모든 센서들의 측정 패턴을 반영할 수 있을 뿐만 아니라, 계층 구조형 식별 규칙을 통해 보다 정확하고 효율적인 식별이 가능할 수 있도록 하였다.

2. 제안된 가스 식별 시스템의 구조

그림 1은 본 논문에서 제안하는 가스 식별 시스템 설계의 전체적인 가스 식별 흐름도를 보여준다.

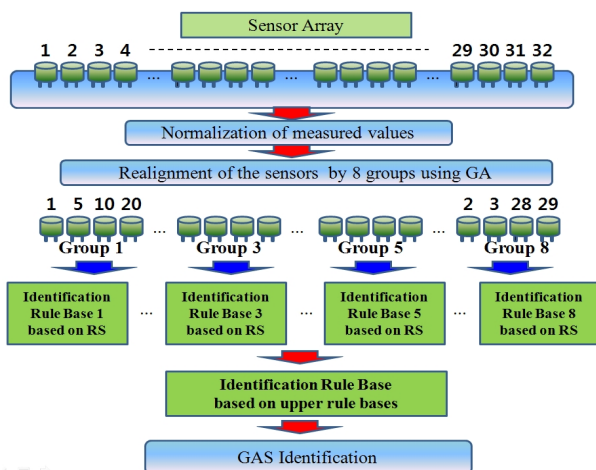


그림 1 제안된 가스 식별 시스템의 식별 흐름도
 Fig. 1 Structure of the proposed gas identification system

센서 어레이로부터 측정 변환된 데이터들은 각 센서간의 동작 편차나 동작 환경의 변화, 반복 측정에 의한 유동 등의 영향을 최소화하기 위하여 정규화 과정이 거치게 되며, 이것은 기존의 식별 시스템과 동일하다. 그러나 제안된 시스템에서는 특징 추출의 어려움과 한계점을 피하기 위하여 하이브리드 유전 알고리즘을 이용하여 센서 어레이를 구성하고 있는 센서들을 여러 개의 그룹으로 나누게 된다. 이러한 재배열 과정은 유사한 측정 패턴을 보이는 센서들 간의 조합을 통해 가스들의 종류를 판별하는 의사결정 규칙의 생성을 보다 용이하고 효과적으로 수행하기 위함이다. 그룹화를 통한 재배열이 끝나면, 각각의 그룹에 포함된 센서들의 가스 측정 데이터를 분석하여 가스의 종류를 판별하기 위한 일련

의 규칙들을 러프 집합을 이용하여 생성하게 된다. 따라서 일차적으로 가스의 종류를 판별하는 규칙기반은 각 그룹마다 독립적으로 생성되므로 그룹의 수만큼 생성된다. 이 후, 각 그룹의 가스 식별 결과를 이용하여 다시 상위의 단일한 최종 식별 규칙기반을 생성하게 되며, 이때에도 러프 집합을 이용하여 적은 수의 유효한 규칙기반이 생성될 수 있도록 한다.

이처럼 유전 알고리즘에 의한 그룹화와 러프 집합에 의한 규칙기반 생성을 통해 차원 축소를 위한 까다롭고 복잡한 특징 추출의 과정을 거치지 않으면서도 센서 어레이에 포함된 모든 센서들에서 취득한 패턴 데이터를 충분히 활용할 수 있고, 센서 동작 편차나 유동 등에 의한 데이터의 부정확성이나 불확실성에도 잘 대처할 수 있게 된다. 또한 계층 구조 식별 규칙기반으로 인해 가스 식별이 이중적으로 이루어지므로 보다 정확하고 신뢰성 있는 식별이 가능해진다.

3. 유전알고리즘을 이용한 저차원 패턴 그룹 생성

3.1 측정 데이터의 정규화

센서들로부터 얻어지는 측정 데이터들은 센서들의 민감한 특성으로 인해 같은 가스에 대해서도 측정 회수에 따라 데이터의 크기에 차이가 날 수 있으며, 측정 환경 등에도 민감하게 반응한다. 또한 동일 모델의 센서들로 교체 측정하더라도 데이터의 편차가 발생한다. 하지만, 같은 가스에 대한 각 센서별 측정 데이터의 크기는 조건에 따라 변화를 보일 수 있으나 센서 어레이를 통해 얻어지는 데이터 패턴은 거의 유사한 형태로 나타나기 때문에, 가스의 종류를 판별하는 데에는 측정값의 크기보다는 데이터 패턴이 더욱 중요한 요소로 작용한다. 따라서 패턴을 보다 명확히 분석하기 위해서는 센서 어레이로부터 측정된 데이터의 절대적인 크기를 상대적인 비교를 할 수 있도록 정규화하는 과정이 요구된다. 본 논문에서는 다음과 같은 방법으로 정규화를 수행하였다.

$$Normalized\ value = \alpha(i) \sum_{j=1}^n electric\ signal(i, j) \quad (1)$$

여기서, *Normalized value*는 정규화를 위한 목표 값, $\alpha(i)$ 는 *i*번째 반복 측정 회수에 대한 측정 신호 값들에 대한 가중치, 그리고 *electric signal*(*i, j*)는 *i*번째 측정 신호들 중 *j*번째 센서에서 추출된 전기신호를 의미한다. 따라서 이 과정을 통해 모든 센서로부터 측정된 각 가스별 신호들의 합의 크기는 모두 동일하게 되며, 따라서 동일 범위 내에서 가스들에 대한 센서들의 측정 패턴을 분석할 수 있게 된다.

3.2 센서 어레이의 그룹화

센서 어레이로부터 얻어진 측정 데이터는 센서 개수만큼의 패턴 요소들을 가지게 된다. 따라서 각 가스별 측정 패턴을 분석하기 위해서는 고차원 데이터를 분석해야만 하는 어려움이 존재하며, 어레이를 구성하는 센서의 개수가 늘어날수록 더 어려워진다.

그룹화 과정은 측정 패턴이 유사한 센서들을 조합하여 여러 개의 독립된 그룹으로 분할하고, 이를 통해 고차원 센서

패턴을 저차원 패턴으로 감축하여 효과적인 패턴 분석과 식별 규칙이 생성 되도록 하는 과정이다. 하지만, 이러한 그룹화 과정에서 고려해야 될 부분은 독립된 각각의 그룹이 고유한 패턴 특성을 나타내어야 하는 것이다. 만약, 어느 한 그룹에서 우수한 측정 패턴을 나타내고 다른 그룹의 특성들이 이에 비해 의미가 없는 측정 패턴을 나타낼 경우, 이는 특징 패턴 추출과 유사한 문제가 된다. 이러한 이유는 우수한 측정 패턴 그룹을 제외하고 다른 그룹들은 활용할 수 없거나 그 영향이 미비하게 되기 때문이며, 결국 이 특징 패턴을 찾기 위한 상당한 수학적 분석을 요구하게 된다. 본 논문은 유전 알고리즘의 목적함수를 통해 이러한 패턴 특성의 편중 현상을 극복하도록 하였으며, 또한 최적화 성능을 통해 효과적으로 조합 패턴을 찾을 수 있도록 하였다. 또한, 유전 알고리즘의 수렴성과 탐색 실패 문제에 대하여 통계적 특성 기반 진화 알고리즘인 하이브리드 유전 알고리즘(HGA: Hybrid Genetic Algorithms)을 적용함으로써 보다 효과적으로 그룹화를 성취할 수 있도록 하였다. 유전 알고리즘의 진화 방향 또는 그룹들의 패턴 특성을 결정하는 목적 함수는 다음과 같이 교차 상관함수로 정의하였다.

$$\rho_{MN}^g = \frac{C_{MN}^g}{\sqrt{C_{MM}^g} \sqrt{C_{NN}^g}} \quad (2)$$

여기서, ρ_{MN}^g 는 g 번째 센서 그룹에 해당하는 목적함수의 계수, C_{MN}^g 는 g 번째 가스 센서 그룹에서 측정값들의 평균과 각각의 측정값들에 대한 교차 공분산, 그리고 C_{MM}^g 와 C_{NN}^g 는 각각 g 번째 가스 센서 그룹에서의 평균에 대한 공분산과 측정 데이터들의 공분산을 의미한다. 일반적으로 교차 상관함수의 경우 강상관성, 약상관성, 무상관성, 역상관성의 특징으로 나타나며, 따라서 그룹화를 진행하게 되면, 초기 그룹들은 강상관성의 특성을 갖는 센서들 간의 조합이 될 것이며, 중간쯤에서는 무상관적 특성을 보이는 센서들 간의 조합, 그리고 마지막으로 가면 역상관성이 강한 센서들 간의 조합으로 구성되어진다. 이러한 목적함수를 정의하는 이유는 식별 규칙기반의 생성에 있어 다양한 형태의 정보를 요구할 수 있기 때문이다. 예를 들어, 어떤 센서 그룹이 아세트론에 대하여 지속적으로 메탄으로 판별을 수행한다면 이러한 판별 패턴 또한 유용한 지식으로 활용할 수 있을 것이다. 따라서, 다양한 형태의 상관성으로 조합된 센서 그룹들의 특징 패턴들은 식별 규칙 생성을 위한 정보의 편중 현상을 극복함과 동시에 보다 정확한 식별 규칙 생성을 유도할 수 있게 된다. 그런데 식(2)의 교차상관함수로 정의된 목적함수는 음의 값도 가질 수 있으므로 다음의 식(3)과 같이 선형 스케일링하여 항상 양이 되도록 수정하였다.

$$Fit(s) = \rho_{MN}^g + C_b \quad (3)$$

식(3)은 유전알고리즘의 적합도 함수로 이 후 자손 생성에 기여하게 된다. 유전 알고리즘을 위한 초기 집단 생성은 집단의 수와 염색체의 크기가 비교적 작기 때문에 무작위 초기화를 사용하여도 별 어려움이 없다. 또한 염색체는 각 그룹에 속하는 센서의 번호들로 구성되므로 실수 코딩을 사용하며, 그 길이는 그룹 내의 센서의 개수 n_{gs} 가 된다. 자손의 생성을 위해선 기본적으로 재생산(선택), 교배, 돌연변이 연산자들이 사용되는데, 다음 세대를 선택하기 위한 재생산 연산자로는 다음과 같이 구배(gradient)와 유사한 형태의

HGA 연산자를 사용하였다.

$$\bar{s}_{ic}(G+1) = \text{round}[s_{ic}(G) + \eta_i \frac{Fit_{it}(G) - Fit_i(G)}{Fit_{best}(G)} \times [s_{it(c)}(G) - s_{ic}(G)]] \quad (4)$$

where, $(1 \leq i \leq G_{max}, 1 \leq c \leq n_s)$

여기서, $S_{ic}(G)$ 는 G 세대의 i 번째 개체의 c 번째 요소이고, $Fit_{it}(G)$ 는 G 세대의 가장 우수한 적합도이며, $Fit_i(G)$ 는 G 세대 개체들의 각각의 적합도, n_s 는 센서의 총 개수를 의미한다. 식 (4)의 재생산 연산자는 가급적 개체들의 복제는 피하면서 집단을 최적 개체 쪽으로 유도할 수 있는 구조를 가지게 된다. 그러므로 유전 알고리즘의 진화 과정에 있어 초기 초우량 개체의 출현에 따른 유전 알고리즘의 유전적 다양성의 결핍 문제를 극복하면서도 최적 개체를 탐색 할 수 있게 해준다. 다음으로 교배에는 수정 단순교배를 사용하였다 [16]. 수정 단순교배의 교배점은 $CP \in [1, n_{gs} - 1]$ 인 구간에서 무작위로 선택된다. 한 쌍의 염색체는 선택된 교배점을 기준으로 서로 교환되며 염색체의 교환 전에 먼저 염색체의 유전 정보는 다음과 같은 방법으로 전환된다.

$$\begin{aligned} \bar{v}_c^a &= \text{round}(\lambda \bar{v}_c^b + (1-\lambda) \bar{v}_c^a) \\ \bar{v}_c^b &= \text{round}(\lambda \bar{v}_c^a + (1-\lambda) \bar{v}_c^b) \quad (1 \leq cp \leq n_{gs} - 1) \end{aligned} \quad (5)$$

여기서, $\lambda \in [0, 1]$ 의 값을 가지며, 식 (5)은 교배를 통해 생성되는 자손들이 항상 적합한 형태로 생성될 수 있게 변환하는 과정이다. 돌연변이 연산자에는 동적 돌연변이 연산자를 사용하였으며, 우수한 유전정보를 포함한 부모가 자손의 생성에 항상 기여할 수 있도록 엘리트 전략을 사용하였다 [23]. 본 논문에 사용된 HGA 연산자들은 유전 알고리즘의 최적해 탐색과정에서 국부 수렴으로 인한 최적해의 탐색 실패 및 수렴 시간의 지연 등을 극복하면서도 최적해를 탐색할 수 있게 한다. 유전 알고리즘을 통해 하나의 센서 그룹이 결정되면, 그 센서 그룹에 포함된 센서들을 제외한 나머지 센서들에 대해 그룹화를 반복 수행하여 원하는 개수(N)의 그룹을 형성하게 된다.

4. 러프 집합을 이용한 계층적 식별 규칙기반 생성

4.1 하위 식별 규칙의 생성

센서 데이터로부터 어떤 가스인지를 판별해내는 식별 규칙의 생성에는 러프 집합을 사용하며, 그룹화에 의해 얻어진 N 개의 센서 그룹마다 각각 독립적으로 일차적인 판별을 하는 N 개의 하위 규칙기반과 이들의 식별 패턴을 종합하여 최종적인 판별 결과를 만들어내는 1개의 상위 규칙기반으로 구성된 계층 구조로 설계한다. 계층 구조에 의한 정보 표현은 지식의 연계성이 높아 효율적으로 정보를 이용할 수 있고, 복잡성을 최소화함으로써 분석의 어려움을 완화할 수 있는 장점을 갖는다. 러프집합을 이용하여 식별 규칙을 생성하려면 우선 주어진 데이터로부터 표 1과 같은 의사결정표를 만들어야 한다. 의사결정표는 조건부와 의사결정부로 구성되어 있으며, 각 규칙에 대하여 조건부의 속성 값들과 의사결정부의 속성 값들로 정의되어진다. 표 1은 본 논문의

시물레이션 과정에서 얻어진 초기 의사 결정표의 한 예로서, 의사결정부의 속성 값들은 5 종류의 가스에 각각 부여된 표지 번호이고, 조건부의 속성 값들은 각 센서별 정규화 측정 데이터의 최소값, 최대값, 평균을 이용하여 나눈 5개 구간의 표지 번호이다. 예를 들어, 1번 규칙은 제1그룹 내의 첫 번째, 두 번째, 네 번째 센서의 조건부 속성값이 1이므로 이들 센서의 측정값이 정규화 측정 데이터의 최소값 구간에 분포하고 있음을 의미하며, 세 번째 센서의 조건부 속성값이 2이므로 측정값이 최솟값과 평균값 사이의 구간에 위치함을 의미한다. 그리고 의사결정부 속성값이 1인 것은, 센서 측정값의 분포 패턴이 조건부와 같을 경우 아세톤을 측정할 결과임을 의미한다. 이러한 방법으로 전체 데이터에 대한 속성값을 부여하면 표 1과 같은 초기 의사결정표를 생성할 수 있게 된다. 이와 같은 초기 의사결정표에는 가스 식별에 필요한 정보와 불필요한 정보가 섞여 있으므로, 규칙을 생성하기 전에 우선 불필요한 정보들을 제거하여 의사결정표를 간략화하는 것이 필요하다.

표 1에서 보면, 규칙 4와 규칙 7은 조건부 속성값들이 같음에도 불구하고 서로 다른 가스로 식별되고 있다. 규칙 12와 14도 마찬가지이다. 이러한 규칙들은 의사결정이 일관적이지 못하기 때문에 삭제한다.

표 1 그룹별 초기 의사결정표의 예

Table 1 example of initial decision table for each group

제 1 그룹					
규칙	조건부 속성(구간 값으로 정의)				의사결정부 속성 가스
	센서 1	센서 2	센서 3	센서 4	
1	1	1	2	1	1(아세톤)
2	1	1	3	2	
3	2	1	1	2	
4	2	1	3	1	2(부탄)
5	2	2	2	1	
6	2	1	2	2	
7	2	1	3	1	
8	3	2	3	2	3(메탄)
9	3	3	4	3	
10	3	3	2	3	
11	4	3	5	4	4(프로판)
12	4	4	4	3	
13	3	5	4	4	
14	4	4	4	3	
15	5	4	4	4	5(물)
16	5	3	5	4	

이런 식으로 비일관적 규칙들을 찾아서 삭제한 후에는, 식별 규칙 생성에 전혀 도움이 되지 않는 속성들을 찾아내어 삭제하게 된다. 표 1에서 '센서 1'의 속성값들을 배제하게 되면 규칙 11과 규칙 16은 같은 조건부 속성값에 대해 다른 의사결정을 내리게 되므로 속성 '센서 1'을 배제하고는 정확한 의사결정을 내릴 수 없게 된다. 다시 말해, '센서 1'은 의사결정을 위해 꼭 필요한 속성이므로 삭제될 수 없다. 반면에 표 1에서 '센서 2'의 속성값들을 배제하더라도 비일관적인 규칙이 발생되지 않으므로, 속성 '센서 2'는 의사결정에 도움이 되지 않는 불필요한 속성이 되어 제거할 수 있다.

표 2 불필요한 속성을 제거한 의사결정 표

Table 2 decision table without unnecessary attribute

제 1 그룹				
규칙	조건부 속성(구간 값으로 정의)			의사결정부 속성 가스
	센서 1	센서 3	센서 4	
1	1	2	1	1(아세톤)
2	1	3	2	
3	2	1	2	
4	2	2	1	2(부탄)
5	2	2	2	
6	3	3	2	3(메탄)
7	3	4	3	
8	3	2	3	
9	4	5	4	4(프로판)
10	3	4	4	
11	5	4	4	5(물)
12	5	5	4	

이상과 같이 비일관적인 규칙과 불필요한 속성을 제거하여 재구성하면, 표 1의 의사결정표는 표 2와 같이 간략화 된다. 규칙 생성을 위한 다음 단계는 표 2와 같이 간략화된 의사결정표에서 규칙 생성에 필수불가결한 핵심 속성(코어)과 이에 더해 운전한 규칙 생성을 가능하게 해주는 속성(리덕트)을 찾아내는 일이다. 예를 들어, 표 2에서 규칙 1과 2를 살펴보면 다른 속성값들에 상관없이 '센서 1'의 속성값이 1이면 아세톤으로 식별할 수 있다. 따라서 '센서 1'의 속성값이 규칙 1과 2의 가장 중요한 속성값이고, '센서 1'이 코어 속성이 된다. 또한, 규칙 3의 경우에는 '센서 3'의 속성값이 1이면 아세톤으로 식별 가능하므로 '센서 3'이 규칙 3의 코어 속성이 된다.

규칙 생성을 위한 다음 단계는 표 2와 같이 간략화된 의사결정표에서 규칙 생성에 필수불가결한 핵심 속성(코어)과 이에 더해 운전한 규칙 생성을 가능하게 해주는 속성(리덕트)을 찾아내는 일이다. 예를 들어, 표 2에서 규칙 1과 2를 살펴보면 다른 속성값들에 상관없이 '센서 1'의 속성값이 1이면 아세톤으로 식별할 수 있다. 따라서 '센서 1'의 속성값이 규칙 1과 2의 가장 중요한 속성값이고, '센서 1'이 코어 속성이 된다. 또한, 규칙 3의 경우에는 '센서 3'의 속성값이 1이면 아세톤으로 식별 가능하므로 '센서 3'이 규칙 3의 코어 속성이 된다. 코어와 리덕트는 러프 집합의 상한 근사와 하한 근사를 이용하여 쉽고 간단하게 탐색할 수 있는데, 자세한 과정을 표 2의 규칙 1을 예로 살펴보도록 하자. 규칙 1(R_1)의 조건부 속성값들은 '센서 1'(A_1)=1, '센서 3'(A_3)=2, '센서 4'(A_4)=1인데, 표 2에서 '센서 1'의 속성값이 규칙 1의 속성값과 같은 1을 갖는 규칙들의 집합, 즉 규칙 1의 '센서 1'=1에 대한 상한 근사는 다음과 같다.

$$S(R_1, A_1) = \{R_1, R_2\} \quad (6)$$

마찬가지로 규칙 1의 '센서 3'=2에 대한 상한 근사는 다음과 같고

$$S(R_1, A_3) = \{R_1, R_4, R_5, R_6\} \quad (7)$$

마지막으로 규칙 1의 '센서 4'=1에 대한 상한 근사는 다음과 같이 구해진다.

$$S(R_1, A_4) = \{R_1, R_4\} \quad (8)$$

또한 규칙 1과 동일한 의사결정부 속성(=1)을 갖는 규칙들의 집합, 즉 상한 근사는 다음과 같다.

$$D(R_1) = \{R_1, R_2, R_3\} \quad (9)$$

코어를 탐색하는 과정은 규칙 1의 조건부 속성에 대한 상한 근사들의 조합이 의사결정부 속성의 상한 근사를 만족시키는가를 확인하면 된다. 먼저 $S(R_1, A_1)$ 과 $S(R_1, A_2)$ 에 대하여 판별하면

$$S(R_1, A_1) \cap S(R_1, A_2) = \{R_1, R_2\} \cap \{R_1, R_4, R_5, R_8\} = \{R_1\} \subseteq D(R_1) \quad (10)$$

가 되고, 같은 방법으로 $S(R_1, A_1)$ 과 $S(R_1, A_3)$ 및 $S(R_1, A_2)$ 과 $S(R_1, A_3)$ 에 대하여 판별하면

$$S(R_1, A_1) \cap S(R_1, A_3) = \{R_1, R_2\} \cap \{R_1, R_4\} = \{R_1\} \subseteq D(R_1) \quad (11)$$

$$S(R_1, A_2) \cap S(R_1, A_3) = \{R_1, R_4, R_5, R_8\} \cap \{R_1, R_4\} = \{R_1, R_4\} \quad (12)$$

이 된다. 식 (10), (11), 그리고 (12)를 살펴보면, 식 (12)는 $D(R_1)$ 을 만족시키지 못함을 알 수 있다. 따라서 식(12)를 제외하고 식 (10)과 (11)이 나타내는 속성 조합 $A_1=1, A_2=2$ 와 $A_1=1, A_3=1$ 이 규칙 1의 의사결정을 수행할 수 있는 속성 조합이다. 이때 식 (10)과 (11)의 속성 조합에 모두 포함된 속성 A_1 이 규칙 1의 코어 속성이 되며, 식(10)과 (12)의 각 조합에 대해서는 A_2 와 A_3 가 각각 리덕트 속성이 된다. 같은 방법으로 모든 규칙에 대하여 코어 속성을 탐색하게 되면 표 2의 진하게 표시된 부분들이 되며, 또한 리덕트 속성도 찾을 수 있게 된다. 따라서 코어 속성을 이용하여 의사결정을 수행하고, 만약 코어 속성이 중복될 경우 리덕트 속성을 탐색함으로써 정확한 의사결정을 내릴 수 있다.

4.2 상위 식별 규칙의 생성

N 개의 그룹에 대해 러프 집합을 이용하여 각각의 식별 규칙기반 생성이 끝나면, 센서 데이터에 대해 각 그룹별로 이들 규칙기반에 의한 가스 식별 결과를 낼 수 있을 것이다. 이를 다시 데이터로 사용하여 전체 시스템을 아우르는 상위의 최종 식별 규칙 생성을 위한 초기 의사결정표를 작성하게 된다.

표 3은 이의 한 예이다. 표 3을 보면, 앞에서 설명한 방법으로 생성된 각 그룹별 하위 식별 규칙기반에 의한 가스 판별 결과 값들이 조건부 속성값으로 정의되며, 그룹별 하위 식별 규칙의 생성 과정과 동일한 방법에 의해 축소 및 간략화된 최종 식별 규칙을 생성할 수 있게 된다. 이상과 같이, 센서어레이를 묶어서 단일한 데이터원으로 취급하지 않고 그룹화를 통해 여러 개의 작은 그룹으로 나누어 각 그룹별 식별 규칙을 만든 뒤 다시 이들의 결과를 종합하여 최종적인 식별 규칙을 생성함으로써 패턴 분석을 위한 데이터의 고차원화 문제를 극복할 수 있을 뿐만 아니라 모든 센서들의 정보를 효과적으로 활용할 수 있다. 물론 식별의 이중화를 통해 식별 정확도를 향상하는 효과도 가져온다.

표 3 최종 식별을 위한 초기 의사결정표의 예

Table 3 example of initial decision table for final identification

규칙	조건부 속성(그룹별 식별 값)				의사결정부 속성 가스	
	그룹 1	그룹 2	...	그룹 N		
1	아세톤	아세톤	...	아세톤	1(아세톤)	
2	아세톤	아세톤		부탄		
3	부탄	아세톤		아세톤		
4	아세톤	부탄		아세톤		
5	아세톤	부탄		부탄	2(부탄)	
6	부탄	부탄		부탄		
9	메탄	메탄		메탄	3(메탄)	
10	메탄	부탄		메탄		
⋮	⋮	⋮		⋮	⋮	⋮

5. 시뮬레이션 및 검토

제안된 가스 식별 시스템의 성능을 검증하기 위하여 아세톤, 부탄, 메탄, 프로판, 그리고 물의 5종류의 가스를 사용하였으며, 데이터는 32개의 센서로 구성된 센서어레이로부터 각 가스별로 35회씩 측정하여 취득하였다[9-10]. 따라서 수집된 총 데이터의 크기는 32×175의 크기가 된다. 이 중에서 식별 시스템의 설계를 위한 훈련 데이터로 각 가스별로 30회씩 측정된 값들을 이용하였으며, 나머지 데이터들은 시스템의 성능 검증을 위한 검증 데이터로 활용하였다. 다음의 표 4는 제안된 식별 시스템 설계를 위한 학습 과정에서 그룹화에 사용된 하이브리드 유전알고리즘에 적용된 파라미터 값들로서, 센서어레이를 각각 4개의 센서로 구성되는 8개의 그룹으로 묶게 된다.

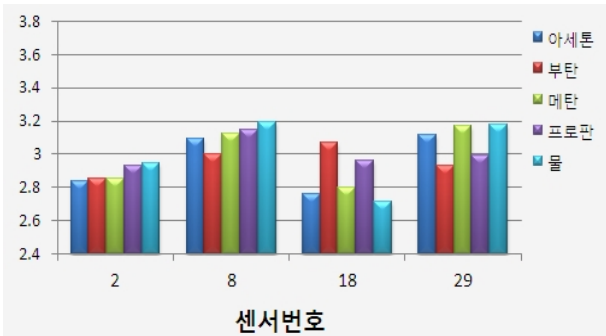
표 4 HGA에 사용된 파라미터

Table 4 parameters of HGA

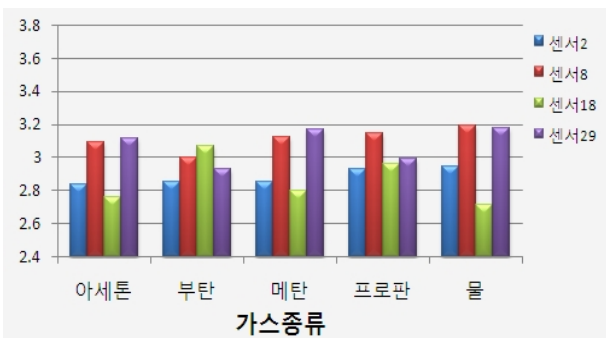
maxgen	최대 반복횟수	35
popsize	집단크기(N)	30
lchrom	염색체의 길이(l)	4
etha	재생산 연산자 변수(η)	1.7
pcross	교배확률(p_c)	0.9
pmutat	돌연변이확률(p_m)	0.1

그림 2는 8개 센서 그룹 중의 한 그룹에 대해 각 센서별 가스 측정 패턴과 각 가스별 센서 동작 패턴, 그리고 그룹화를 위한 하이브리드 유전알고리즘의 최적화 결과를 나타낸 것이다. 모든 센서 그룹에 대해 그림 2와 같이 가스 측정 데이터의 패턴을 분석해보면 각 그룹마다 다른 형태의 패턴을 나타냄을 볼 수 있는데, 이는 각 그룹별로 생성되는 식별 규칙들이 서로 다른 형태의 가스 패턴을 반영할 수 있음을 의미하며, 따라서 제안된 식별 시스템이 다양하고 비일관적인 센서들의 측정 패턴을 충분히 수용할 수 있는 구조를 이루고 있음을 뒷받침해준다.

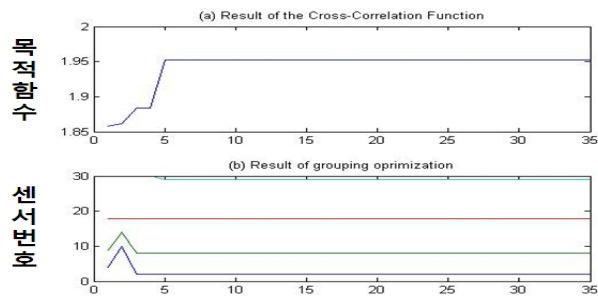
또한 그림 2(b)와 같은 가스별 센서 동작 패턴을 살펴보면 그룹 내의 센서들이 각 가스들에 대하여 서로 유사한 형태의 측정 패턴을 보임을 알 수 있는데, 결과적으로 그룹을 구성하는 센서들의 부조화에 따른 비일관적 식별 규칙 생성을 최대한 억제하는 효과를 나타낼 수 있다.



(a) 센서별 가스 측정 패턴



(b) 가스별 센서 동작 패턴



진화 세대

c) HGA의 목적 함수 변화 및 센서 조합 결과

그림 2 그룹1의 센서 조합 결과 및 특성 패턴

Fig. 2 sensor combination result and feature pattern in group 1

표 5 그룹 1의 가스 식별 규칙

Table 5 gas identification rules of group 1

규칙	지표	속성			규칙	지표	속성		
		센서2	센서8	센서29			식별가스	센서2	센서8
1			3	3	9	2			메탄
2	2	1			10	4		2	프로판
3	3		2		11	4		3	
4	1		1		12	4		4	
5	2	2			13	5		5	물
6	2		4		14	5		4	
7	3		4		15	4		5	
8	2		5						

그림 (c)는 HGA의 수렴 속도가 상당히 빠름을 보여 준다. 표 5는 그림 2의 그룹1에 대해 생성된 식별 규칙이다. 여기서, 진하게 표시한 값들이 각 규칙의 코어 값을 의미하고 나머지는 리덕트 값을 의미한다. 표 5를 살펴보면, 그룹을 구성하는 2번, 8번, 18번, 29번 센서 중 18번 센서가 불필요한 속성으로 삭제되었음을 알 수 있다. 또한 규칙기반 설계에 5 종류의 가스에 대하여 각각 30회씩 측정된 데이터를 이용하였으므로 초기 규칙의 수는 150개이지만, 러프집합을 이용한 규칙기반 생성과정을 통해 매우 적은 수인 15개의 규칙으로 간략화 되었음을 알 수 있다.

표 6은 그룹화 과정에서 마지막으로 분류된 그룹 8의 식별 규칙으로 다른 그룹들보다 더욱 간략화된 형태를 보여주고 있다.

표 5, 6을 포함한 8개 그룹들의 식별 규칙을 비교해보면, 아세톤-부탄-메탄-프로판-물 순으로 식별 규칙을 배열할 때 규칙기반의 조건부 속성값들이 그룹 1에서는 대부분 오름차순에 가깝지만 그룹 8로 갈수록 내림차순에 가까워짐을 볼 수 있는데, 이는 센서어레이의 그룹화가 초기 강상관성으로부터 역상관성 방향으로 진행되었던 것을 반영한 결과이다. 각 그룹별 식별 규칙기반이 상당히 상이한 것은 제안된 식별 시스템이 다양한 형태로 주어지는 센서 데이터들의 패턴을 충분히 반영할 수 있음을 말해주는 것이다.

그림 3은 학습 데이터에 대한 8개 그룹의 가스 식별을 보여주며, 그룹 3, 5, 6, 8에서 평균적으로 좋은 식별 성능을 내고 있으며, 나머지 그룹들은 90% 이하의 식별률을 보이고 있다. 이는 32개의 센서들을 4개씩 그룹화 하는 과정에서 발생한 차원 축소로 데이터의 부정확성이나 불확실성을 보완할 수 있을 정도의 충분한 정보가 확보되지 못한 탓으로 볼 수 있는데, 그림에서 볼 수 있듯이 식별률이 낮은 그룹에서도 모든 가스에 대한 식별이 부정확한 것이 아니라 특정 가스에 대한 식별 오류가 주된 원인이라는 사실이 이러한 분석을 뒷받침한다.

그런데, 특정 그룹에서 낮은 식별 결과를 나타내는 가스의 종류에 대하여 다른 그룹에서는 좋은 식별률을 보이고 있기 때문에 제안된 시스템에서처럼 각 그룹별 식별 결과를 종합적으로 묶는 상위의 식별 규칙이 있다면 각 그룹들이 상호 보완적 역할을 수행하여 전반적으로 우수한 식별 성능을 발휘할 수 있을 것이다. 제안된 식별 시스템은 학습 데이터에 대해서는 당연히 100%의 식별 성능을 나타내고 있

표 6 그룹 8의 가스 식별 규칙

Table 6 gas identification rules of group 8

지표 규칙	속성				지표 규칙	속성			
	센서16	센서25	센서28	식별가스		센서16	센서25	센서28	식별가스
1	4	4		아세톤	8	3	2	메탄	
2		4	3		9		2		1
3	5		3		10	1			3
4	4	3			11	1	2		
5	5	5		부탄	12	2	1	물	
6	5		4		13		1		1
7	4	5							

표 7 상위 최종 식별 규칙

Table 7 final identification rules of the system

지표 규칙	속성			지표 규칙	속성		
	그룹7	그룹8	식별가스		그룹7	그룹8	식별가스
1	아세톤	아세톤	아세톤	5	메탄	메탄	메탄
2	부탄	부탄	부탄	6	프로판	프로판	프로판
3	부탄	아세톤		7	물	물	물
4	아세톤	메탄	메탄	8	아세톤	물	

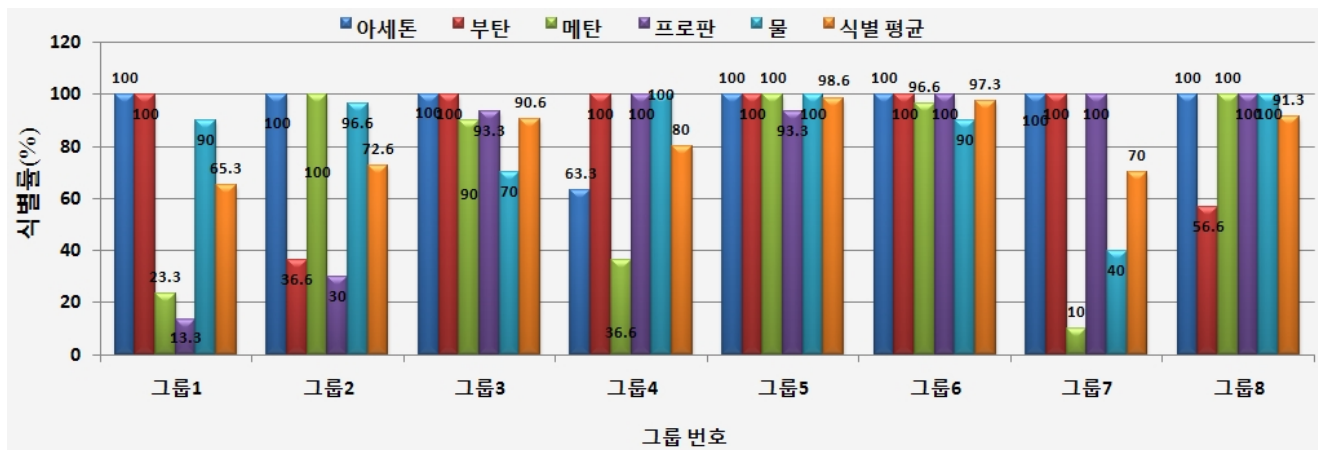


그림 3 훈련과정의 8개 그룹의 식별 패턴

Fig. 3 identification patterns of 8 groups in training mode

으며, 검증 데이터에 대해서는 프로판의 경우에 5개의 샘플에 대하여 1번의 식별 오류를 일으킨 것을 제외하고는 모두 정확히 식별하였는바, 성능이 매우 우수함을 입증하고 있다.

표 8 제안된 시스템의 식별 성능

Table 8 identification performances of the proposed system

학습 데이터에 대한 시스템의 성능(%)						
지표	가스	아세톤	부탄	메탄	프로판	물
가스별 식별률		100	100	100	100	100
평균 식별률		100				
검증 데이터에 대한 시스템의 성능(%)						
지표	가스	아세톤	부탄	메탄	프로판	물
가스별 식별률		100	100	100	80	100
식별회수/샘플수		5/5	5/5	5/5	4/5	5/5
평균 식별률		96(오차 : 4%)				

6. 결 론

본 논문은 유전알고리즘과 러프집합을 이용한 계층적 식별규칙을 갖는 가스 식별 시스템의 설계법을 다루었다. 제안된 방법은 다수의 가스 센서들로 구성된 센서 어레이로부터 얻어진 신호들에 따른 패턴 분석의 어려움을 극복하기 위해 유전알고리즘을 이용한 그룹화 과정을 통해 차원을 축소하였으며, 각각의 그룹에 상응하는 식별 규칙의 식별 패턴을 이용한 전체 식별 규칙을 계층적으로 구현함으로써 많은 수의 센서들이 가지는 측정 패턴을 모두 반영할 수 있도록 하여 가스 식별의 정확성을 높일 수 있도록 하였다. 센서의 그룹화에 이용된 유전알고리즘을 최적화의 정확도와 수렴 속도 개선을 위해 하이브리드 유전 알고리즘을 채택하였으며, 식별 규칙의 생성에는 러프 집합을 이용하여 비일관적 데이터의 효율적 처리와 최소의 규칙으로 식별이 가능하도록 하였다. 제안된 시스템의 효용성을 검증하기 위해 5종류

의 가스에 대한 식별 결과를 분석하였으며, 그 결과 매우 높은 식별률을 나타내 유용함을 알 수 있었다. 따라서 제안된 가스 식별 시스템의 설계 기법은 가스 식별뿐만 아니라 다양한 분야의 센서를 이용한 패턴 인식 시스템에 효과적으로 적용될 수 있을 것으로 기대된다. 향후 과제로는 제안된 식별 기법의 성능을 최대화할 수 있는 센서 배열의 구조, 그룹화의 최적화, 상위 식별 규칙 생성의 개선 등을 꼽을 수 있다.

참 고 문 헌

[1] T. D. Ginson, O. Prosser, J. N. Hulbert, R. W. Marshall, P. Corcoran, P. Lowery, E. A. Ruck-Keene, S. Heron, "Detection and simultaneous identification of microorganisms from headspace samples using an electronic nose", *Sensors and Actuators B*, Vol.44, pp.413-422, 1997.

[2] J. W. Gardner, M. Craven, C. Dow, E. L. Hines, "The prediction of bacteria type and culture growth phase by an electronic nose with a multi-layer perceptron network", *Meas. Sci. Technol*, Vol.9, pp.120-127, 1998.

[3] R. Gutierrez-Osuna, "Pattern Analysis for Machine Olfaction: A Review", *IEEE Sensors Journal*, Vol.2, No.3, pp.189-202, 2002.

[4] F. Marcelloni, "Recognition of olfactory signals based on supervised fuzzy C-means and k-NN algorithms", *Pattern Recognition Letters*, Vol.22, pp.1007-1019, 2001.

[5] E. L. Hines, E. Llobet, J. W. Gardner, "Electronic noses: a review of signal processing techniques", *Meas. Sci. Technol*, Vol. 9, pp. 120-127, 1998.

[6] A. Hierlemann, R. Gutierrez-Osuna, "Higher-Order Chemical Sensing", *Chem. Rev*, Vol.108, pp.563-613, 2008.

[7] D. Vlachos, J. Avaritsiotis, "Fuzzy neural networks for gas sensing", *Sensors and Actuators B*, Vol.3, pp.77-82, 1996 *Meas. Sci. Technol*, Vol. 9, pp. 120-127, 1998.

[8] N. Y. Kim, H. G. Byun, K. C. Persaud, "Normalization approach to the stochastic gradient radial basis function network algorithm for odor sensing systems", *Sensors and Actuators B*, Vol. 124, pp. 407-412, 2007.

[9] E. Llobet, E. L. Hines, J. W. Gardner, P. N. Bartlett, T. T. Mottram, "Fuzzy ARTMAP based electronic nose data analysis", *Sensors and Actuators B*, Vol. 61, pp. 183-190, 1999.

[10] D. E. Goldberg, "Genetic Algorithms in Search, Optimization and Machine Learning", Addison-Wesley Publishing Co. Inc., N. Y., 1989.

[11] K. F. Man, "Genetic Algorithms: Concepts and Applications", *IEEE Trans. on Industrial Electronics*,

vol. 43, no. 5, pp. 519-534, 1996.

[12] D. T. Pham, G. Jin, "Genetic Algorithm using Gradient-Reproduction Operator", *Electronics Letters*, vol. 31, no. 18, pp. 1558-1559, 1995.

[13] D. T. Pham, G. Jin, "A Hybrid Genetic Algorithm", *Proc. 3rd World Congress on Expert Systems*, Seoul, Korea, vol. 2, pp. 748-757, 1996.

[14] Z. Pawlak, "Rough set theory and its applications", *J. Telecommun. Inform. Technol.*, vol. 3, pp. 7-10, 2002

[15] Y. K. Bang, C. H. Lee, "Multiple Model Fuzzy Prediction Systems with Adaptive Model Selection Based on Rough Sets and its Application to Time Series Forecasting", *Journal of Korean Institute of Intelligent Systems*, vol. 19, pp. 25-33, 2009.

[16] Y. K. Bang, C. H. Lee, "Design of Fuzzy Prediction System based on Dual Tuning using Enhanced Genetic Algorithms", *Trans. KIEE.*, vol. 59, no. 1, pp. 184-191, 2010.

저 자 소 개



방 영 근 (方 寧 根)

1974년 4월 7일
2003년 강원대(삼척) 전기공학과(석사)
2010년 강원대학교 전기전자공학과(박사)
E-mail : b2y2c1@hanmail.net



변 형 기 (邊 衡 起)

1961년 8월 1일
1990년 Manchester Univ. 전자전기(석사)
1995년 Manchester Univ. 계측공학(박사)
현재 강원대 정보통신공학과 교수
E-mail : byun@kangwon.ac.kr



이 철 희 (李 哲 熙)

1959년 7월 4일
1985년 서울대학교 전기공학과(석사)
1989년 서울대학교 전기공학과(박사)
현재 강원대 전기전자공학과 교수
E-mail : chlee@kangwon.ac.kr