

퍼지집합과 러프집합을 이용한 계층 구조 가스 식별 시스템의 설계

Design of a Hierarchically Structured Gas Identification System Using Fuzzy Sets and Rough Sets

방 영 근* · 이 철 희†

(Young-Keun Bang · Chul-Heui Lee)

Abstract - An useful and effective design method for the gas identification system is presented in this paper. The proposed gas identification system adopts hierarchical structure with two level rule base combining fuzzy sets with rough sets. At first, a hybrid genetic algorithm is used in grouping the array sensors of which the measured patterns are similar in order to reduce the dimensionality of patterns to be analyzed and to make rule construction easy and simple. Next, for low level identification, fuzzy inference systems for each divided group are designed by using TSK fuzzy rule, which allow handling the drift and the uncertainty of sensor data effectively. Finally, rough set theory is applied to derive the identification rules at high level which reflect the identification characteristics of each divided group. Thus, the proposed method is 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 : Gas identification, Hybrid genetic algorithm, Hierarchical structure, TSK fuzzy rules, Rough set

1. 서 론

대표적 인공 감각기관의 하나인 전자코 시스템은 가스 식별을 비롯한 다양한 분야에서 응용되고 있는데, 작업 환경이나 신체 리듬에 종속적인 인간의 인지 능력에 비해 보다 높은 항상성과 안정성, 그리고 정확성을 제공할 수 있는 장점을 지닌다[1-2]. 그런데 가연성/유독 가스의 발생이나 사용이 크게 늘어나고 그 종류도 다양화되는 추세에 따라 가스 사고의 위험성과 빈도 또한 높아지고 있어서, 유해 가스들을 정확히 검출, 식별할 수 있는 가스 식별 시스템의 개발이 중요한 과제로 대두되고 있다.

대부분의 가스 식별 시스템은 기본적으로 센서를 이용하여 가스 검출을 담당하는 감지부와 취득된 신호(데이터)의 패턴을 분석하여 가스의 종류를 식별하는 신호처리부로 이루어진다[3-4]. 그런데, 동일한 센서일지라도 동작 환경, 장시간 작동에 따른 특성 변화, 노후화에 따른 열화 등 여러 가지 이유로 인해 측정 결과들이 수시로 달라지기 때문에, 개별 센서의 불확실성을 처리하기에 충분한 수의 센서를 조합한 센서 어레이를 사용하는 것이 보통이다. 왜냐 하면, 어레이를 구성하는 각 센서의 데이터를 종합적으로 분석함으로써 단독 센서를 사용할 때보다 측정값의 변동이나 비일관성의 영향을 상당히 감소시켜 가스 감지 및 식별

정확도를 높일 수 있기 때문이다[5].

하지만 센서 어레이를 구성하게 되면 패턴 분석을 위한 데이터의 차원과 양이 커지는 문제가 발생하는데, 다수의 기존 연구들에서는 특징 추출을 통해 차원을 축소한 뒤 확률 통계적 기법들을 활용하여 효율적인 패턴 인식이 가능하도록 접근하였다. 그러나 특징 추출 기법은, 데이터에 내재된 대표적 특성만을 추려 내므로 차원 축소에는 효과적이지만, 복잡하고 난도가 높은 수학적 과정이 요구되는 것이 일반적이며, 경우에 따라선 유동성과 같이 식별 성능에 중요한 특이하고 차별적인 센서 특성이 누락되거나 축소되어 제대로 반영되지 못할 수도 있다. 예를 들어, 추출되는 저차원 특징에 기여도가 큰 센서들 중 하나만 작동 이상이 있거나 고장이 나더라도 특징 패턴의 왜곡이 발생하여 식별 결과에 상당한 오류를 가져오게 된다[6].

이에 대한 대안으로 최근에는 퍼지 이론이나 신경망과 같은 소프트웨어 기법을 이용한 가스 식별 시스템 구현에 관한 연구들이 활발히 이루어지고 있다. 퍼지 이론 기법들에서는 센서 측정 데이터의 퍼지 클러스터링을 통해 차원 축소 문제를, 그리고 데이터에 포함된 유동성과 비일관성 같은 불확실성은 퍼지 분류 규칙에 의해 효과적으로 해결하고 있으며, 신경망 기법들은 병렬 처리 능력에 의해 차원 문제를 극복하는 동시에 측정 데이터의 불확실성은 신경망의 학습을 통해 처리될 수 있도록 하여 식별 성능을 향상시키고 있다[7-9].

본 논문에서는 정확도와 신뢰성이 높은 가스 식별 시스템을 위하여 유전 알고리즘, 퍼지 집합, 러프 집합을 조합하여 그림 1과 같이 센서 어레이의 그룹화와 계층 식별 구조를 갖는 새로운 구현 방법을 제안하였다.

† Corresponding Author : Dept. of Electrical and Electronic Engineering, Kangwon National University, Korea.
E-mail: chlee@kangwon.ac.kr

* Dept. of Electrical Engineering, Kangwon National University, Korea.

Received : September 20, 2017; Accepted : February 26, 2018

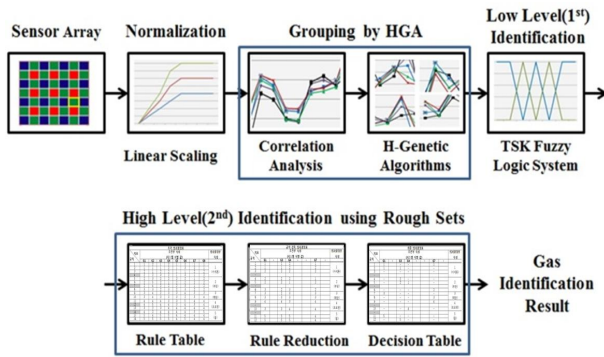


그림 1 제안된 가스 식별 시스템의 순서도

Fig. 1 The flow chart of the proposed gas identification system

우선 센서 어레이의 사이즈와 측정 빈도의 증가에 의한 데이터 차원의 확장에 따라 급격히 높아지는 수치 해석적 어려움을 수반하는 특징 추출에 의한 차원 축소 대신 유전 알고리즘을 이용한 그룹화를 통해 센서 어레이 사용으로 인한 측정 데이터 차원 증가 문제를 해결한다. 그룹화는 최적화 기능이 장점인 유전 알고리즘[10-11]에 의해 어레이 구성 센서들을 유사한 측정 패턴을 가지는 소수의 센서들로 묶인 여러 개의 그룹으로 나누는 과정이다. 이때 진화적 최적화 기법과 통계적 최적화 기법이 결합된 하이브리드 유전 알고리즘(Hybrid Genetic Algorithm : HGA) [12-13]을 사용하여 해의 적합성과 알고리즘의 수렴 속도 모두 양호한 결과를 얻을 수 있도록 하였다. 이렇게 나누어진 각 그룹마다 1차적으로 가스 식별이 이루어지도록 하면, 데이터의 차원이 전체 센서 수 차원에서 각 그룹에 속하는 센서 수 차원으로 대폭 축소되어 특징 추출과 유사한 정도의 패턴 데이터 차원 축소 효과를 얻을 수 있으며, 다음의 설명과 같은 이유로 퍼지 규칙 기반 수립이 좀 더 간단해지고 특징 기술 능력의 정밀도가 높아져서 데이터 분석의 용이성 및 성능 향상을 꾀할 수 있다. 하위 레벨의 1차 식별에는 데이터의 유동이나 비일관성과 같은 불확실성을 잘 다룰 수 있는 퍼지 집합을 이용하여 분류된 그룹 각각에 대해 가스 종류를 판별하는 퍼지 TSK 규칙기반을 구성하여 퍼지 추론을 수행한다. 그런데, 각 그룹은 데이터의 패턴 유사성이 높은 센서들로 묶여 있으므로, 퍼지 집합과 규칙의 수가 줄어드는 등 퍼지 규칙기반 생성이 용이하면서도 양호한 식별 결과를 얻을 수 있게 된다. 1차 식별이 수행되고 나면, 규칙 감축 및 유효 규칙 생성에 강력한 도구인 러프 집합[14-15]을 이용하여 상위 레벨의 식별 규칙을 생성함으로써 각 그룹으로부터 산출된 1차 식별 결과를 조합하여 최종 판단을 내리게 된다. 하위 레벨인 각 그룹별 퍼지 식별 규칙들은 센서 동작 편차나 유동 등에 의한 데이터의 부정확성이나 불확실성에도 잘 대처할 수 있게 해주며, 상위 레벨의 러프 집합에 의한 식별 규칙들은 그룹들 간의 식별 결과의 불일치나 비일관성 등에 잘 대처할 수 있게 해준다. 이와 같은 계층 구조 식별 규칙기반으로 인해 가스 식별이 이중으로 이루어지며, 차원이 축소된 부분적인 추출 특징에 의한 식별이 아니라 센서 어레이에 포함된 모든 센서들의 측

정 패턴이 원형대로 사용된 식별이 이루어지므로 보다 정확하고 신뢰성 있는 식별이 가능해진다. 제안된 가스 식별 시스템은 수학적 모델 구현의 어려움은 비교적 경감하면서도 소프트 컴퓨팅 기법들이 가지는 각각의 장점을 충분히 활용할 수 있고, 또한 서로간의 협력에 의해 각 기법들의 단점과 부담을 완화할 수 있어 가스 식별 시스템의 성능과 설계의 효율성을 동시에 취할 수 있는 구조이다. 5종류의 가스에 대한 센서 측정 데이터들을 이용하여 제안된 방법의 효율성을 검증하였다.

2. 데이터 전처리-정규화 및 그룹화

2.1 측정 데이터의 정규화

센서 어레이에서 감지/측정되는 가스 데이터들은 각 센서의 위치 및 물리적 성능 편차, 온-습도를 비롯한 측정 환경의 변화, 반복 측정에 의한 유동 등으로 인해 매 측정마다 그리고 센서 간에도 값이 달라진다. 하지만 가스가 동일하다면 센서 어레이 집합의 데이터 패턴은 매우 유사하기 때문에, 가스 종류 식별에는 절대적인 측정값보다는 데이터 패턴이 더 효과적인 요소로 작용한다. 그러므로 매 측정마다 생길 수 있는 측정 편차를 보정하여 데이터 패턴의 상대적인 비교가 가능하도록 데이터를 정규화하는 것이 필요하다. 그런데 가장 일반적인 정규화 기법과 같이 매 측정 회차 측정값들 중 최대값으로 그 측정 회차의 측정값들을 정규화하는 것은 어레이 내의 각 센서들의 측정값에 대한 상대적 보정에 불과할 뿐으로, 측정 회차별 변화 경향을 상대 비교하기 위한 정규화로는 적합하지 않다. 따라서 본 논문에서는 다음과 같은 방법으로 정규화를 수행하였다.

$$NV = \alpha(i) \sum_{j=1}^n MV(i,j) \tag{1}$$

여기서, NV 는 정규화 목표값, $\alpha(i)$ 는 i 번째 측정 회차 측정값들에 대한 가중치, 그리고 $MV(i,j)$ 는 i 번째 측정 회차의 j 번째 센서의 측정값이다. 식 (1)의 정규화에 의해 측정 회차에 상관없이 센서 어레이에 의한 측정값들의 합이 항상 같게 되므로, 반복 측정된 데이터들의 패턴을 상호 비교하기가 용이해진다.

2.2 HGA를 이용한 센서 그룹화

센서 어레이에 의한 측정 데이터는 센서 수만큼의 차원을 가지며, 어레이를 구성하는 센서 수가 많을수록 차원이 더 커질 것이다. 이러한 데이터의 고차원 특성은 분석의 어려움을 야기한다. 그룹화는 측정 패턴이 유사한 센서들끼리 조합을 이루도록 어레이 센서들을 적절한 크기의 여러 개의 작은 그룹으로 분할함으로써 고차원 패턴 데이터를 저차원으로 감축하여 효과적인 패턴 분석과 식별 규칙이 생성될 수 있도록 만드는 과정이다. 그룹화에는 센서 조합의 효율적인 최적화를 위해 하이브리드 유전 알고리즘을 이용한다. 유전형의 진화와 자연 선택을 모사하여 집단 기반 탐색 전략을 추구하는 유전 알고리즘은 전역 최적해가 존재하

는 영역을 빠르게 탐색할 수 있지만 수렴 영역에서 정확한 최적해를 찾아내는 데에는 상대적으로 긴 시간을 필요로 하므로, 지역 최적해를 보다 효과적으로 탐색할 수 있는 국부 탐색 기법을 유전 알고리즘에 조합시켜 전역 최적해로의 수렴성을 개선한 것이 하이브리드 유전 알고리즘이다[13].

유전 알고리즘의 진화 방향을 결정하는 적합도 함수 Fit 는 다음과 같이 교차 상관함수를 사용하여 정의한다.

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

여기서, ρ_{MN}^g 는 g 번째 센서 그룹의 교차 상관 함수, 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 공분산이며, C_b 는 Fit 가 항상 양의 값을 갖도록 보정해주는 바이어스이다.

세대를 진화해가며 그룹화를 진행하게 되면, 교차 상관함수가 갖는 특성으로 인해 초기 그룹들은 강상관성의 특성을 갖는 센서들의 조합이 될 것이며, 중간쯤에서는 무상관적 특성을 보이는 센서들의 조합, 그리고 후속 세대로 갈수록 역상관성이 강한 센서들의 조합으로 구성되어진다. 그러므로 종료 조건의 선택에 따라 다양한 형태의 상관성으로 조합된 센서 그룹들을 생성할 수 있으며, 이는 식별 규칙 생성을 위한 정보의 편중 현상을 극복함과 동시에 보다 정확한 식별 규칙 생성을 유도할 수 있게 해준다.

염색체의 코딩을 위해 본 논문에서는 부호화와 복호화 과정이 요구되지 않으면서도 최적 탐색성능과 수렴성을 강화한 하이브리드 유전 알고리즘의 다양한 연산자들에 적합한 실수 코딩 기법을 적용하였다[12-13]. 다시 말해, 센서 어레이를 구성하는 각각의 센서들에 부여된 번호가 염색체를 구성하는 유전 정보로 적용되며, 그룹화를 위해 정의된 센서의 개수가 염색체의 길이가 된다. 만약 그룹에 포함된 센서의 수를 n_{gs} 개라고 하면, 초기 집단 생성에 코딩된 모든 염색체의 길이는 n_{gs} 개로 정의되게 된다. 따라서 어레이는 구성하는 총 센서의 수가 n_s 개라면, 분류되는 그룹의 개수는 n_s/n_{gs} 개로 정의되게 된다. 위와 같은 방법으로 초기 모든 집단의 생성이 끝나면 다음 세대를 선택하기 위한 재생산 연산자로는 다음과 같이 구배(gradient)와 유사한 형태의 연산자를 사용한다.

$$\begin{aligned} \bar{s}_{ic}(G+1) = & round[s_{ic}(G) + \eta_i \frac{[Fit_{bt}(G) - Fit_i(G)]}{Fit_{best}(G)}] \\ & \times [s_{bt(c)}(G) - s_{ic}(G)] \end{aligned} \quad (3)$$

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

여기서, $s_{ic}(G)$ 는 G 세대의 i 번째 염색체의 c 번째 요소이고, $Fit_{bt}(G)$ 는 G 세대의 가장 우수한 적합도이며, $Fit_i(G)$ 는 G 세대 개체들의 각각의 적합도를 의미한다. 또한, η 는 재생산 연산자 변수로 각 염색체별로 독립적이거나 공통적으로 정의하여 사용할 수 있다.

식 (3)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. 수정 단순교배의 교배점은 $cp \in [1, n_{gs} - 1]$ 인 구간에서 무작위로 선택된다. 한 쌍의 염색체는 선택된 교배점을 기준으로 교배점 이후의 유전 정보들은 서로 교환하여 교배를 수행하며, 교배점에서의 유전정보는 다음과 같은 방법으로 수행되며 이를 통해 교배로 인한 유전 정보의 불연속성을 완화할 수 있도록 수행된다.

$$\begin{aligned} \tilde{s}_{cp}^a &= round(\lambda \bar{s}_{cp}^b + (1-\lambda) \bar{s}_{cp}^a) \\ \tilde{s}_{cp}^b &= round(\lambda \bar{s}_{cp}^a + (1-\lambda) \bar{s}_{cp}^b) \end{aligned} \quad (1 \leq cp \leq n_{gs} - 1) \quad (4)$$

여기서, $\lambda \in [0, 1]$ 의 값을 가지는 곱인수를 의미하며, \tilde{s}_{cp} 는 교배를 위해 선택된 두 부모 염색체인 \bar{s}^a 와 \bar{s}^b 의 유전 정보들을 교배점 cp 로부터 곱인수와와의 조합을 통해 교배점에서 생성되는 각각의 자손에 대한 유전정보를 의미한다.

이러한 식 (4)는 교배는 교배로 인한 유전 정보의 불연속성을 완화함으로써, 생성되는 자손들이 항상 적합한 형태로 생성될 수 있게 변환하는 과정이다. 돌연변이 연산자에는 동적 돌연변이 [11-13] 연산자를 사용하여 초기에는 균등확률 탐색 후 세대수 증가에 따라 지역적 탐색을 수행하게 하여 알고리즘의 수렴성을 개선하였으며, 마지막으로 우수한 유전정보를 포함한 부모가 자손의 생성에 항상 기여할 수 있도록 엘리트 전략을 사용하였다 [12-13].

유전알고리즘을 통해 하나의 센서 그룹이 결정되면, 그 센서 그룹에 포함된 센서들을 제외한 나머지 센서들에 대해 그룹화를 반복 수행하여 원하는 개수($N = n_s/n_{gs}$)의 그룹을 형성하게 된다.

3. 계층 구조 식별 시스템의 설계

유전 알고리즘에 의해 센서들의 그룹화가 완료되면, 1차적으로 각 그룹별로 독립적으로 가스 식별을 수행하는 하위 식별 시스템과 이들의 식별 패턴을 종합하여 2차적으로 최종 판별 결과를 만들어내는 상위 식별 시스템으로 구성된 계층 구조 형태로 가스 식별 시스템을 구현한다. 계층 구조에 의한 정보 표현은 지식의 연계성이 높아 효율적으로 정보를 이용할 수 있고, 복잡성을 최소화함으로써 분석의 어려움을 완화할 수 있는 장점을 갖는다.

3.1 퍼지 집합을 이용한 하위(1차) 식별 시스템

N 개의 그룹이 각자 독립적으로 가스 식별을 수행하는 하위 식별 시스템의 구축에는 TSK 퍼지 모델을 이용한다. TSK 퍼지 모델은 언어적 규칙기반과 수학적 모형을 조합하여 각각의 이점을 동시에 취할 수 있도록 한 것으로, 규칙의 전건부에는 퍼지 집합을 이용하여 불확실성을 포함한 데이터의 특성을 묘사할 수

모두 동일함에도 불구하고 결론부 속성이 다른 비일관적인 규칙들을 제거한 다음, 러프 집합의 식별 불가능성을 이용하여 규칙 감축을 위한 작업을 진행하게 된다. 그 첫 단계로 규칙 생성에 영향을 주지 않는 불필요한 속성을 제거하는데, 식 (10)에서 조건부 속성 $G1, G2, \dots, GN$ 중 어떤 하나를 제거하여도 식별을 위한 규칙기반이 비일관적으로 바뀌지 않는다면 그 속성은 불필요한 속성이므로 삭제하여도 지장이 없다. 예를 들어, 다음과 같은 의사 결정 규칙에서

$$\begin{aligned} R1 : G1 = gas1 \ \& \ G2 = gas1 \ \& \ G3 = gas1 \ \rightarrow \ D = gas1 \\ R2 : G1 = gas1 \ \& \ G2 = gas2 \ \& \ G3 = gas1 \ \rightarrow \ D = gas1 \\ R3 : G1 = gas1 \ \& \ G2 = gas2 \ \& \ G3 = gas3 \ \rightarrow \ D = gas2 \end{aligned} \quad (11)$$

속성 $G3$ 를 삭제하면 규칙 $R2, R3$ 가 같은 조건부 속성 값을 가지고도 다른 결론을 얻게 되므로 $G3$ 는 삭제해서는 안 된다. 반면에 속성 $G1$ 과 $G2$ 는 하나를 삭제하여도 비일관적인 규칙이 생성되지 않으므로 둘 중 하나는 삭제해도 무방하다. 모든 조건부 속성에 대해 이와 같이 비일관성을 확인하여 불필요한 속성을 삭제함으로써 규칙들이 보다 단순화된다. 식 (12)은 이렇게 단순화된 규칙의 예이다.

$$\begin{aligned} R1 : G1 = gas1 \ \& \ G5 = gas2 \ \& \ G7 = gas1 \ \rightarrow \ D = gas1 \\ R2 : G1 = gas1 \ \& \ G5 = gas3 \ \& \ G7 = gas2 \ \rightarrow \ D = gas1 \\ R3 : G1 = gas2 \ \& \ G5 = gas1 \ \& \ G7 = gas2 \ \rightarrow \ D = gas1 \\ R4 : G1 = gas2 \ \& \ G5 = gas2 \ \& \ G7 = gas1 \ \rightarrow \ D = gas2 \\ R5 : G1 = gas2 \ \& \ G5 = gas1 \ \& \ G7 = gas3 \ \rightarrow \ D = gas2 \\ R6 : G1 = gas3 \ \& \ G5 = gas3 \ \& \ G7 = gas2 \ \rightarrow \ D = gas2 \\ R7 : G1 = gas3 \ \& \ G5 = gas2 \ \& \ G7 = gas3 \ \rightarrow \ D = gas3 \\ R8 : G1 = gas2 \ \& \ G5 = gas3 \ \& \ G7 = gas2 \ \rightarrow \ D = gas3 \end{aligned} \quad (12)$$

최소 의사 결정 규칙을 얻기 위한 규칙 감축의 다음 단계는 식 (12)와 같이 비일관적인 규칙과 불필요한 속성들이 제거된 규칙기반에 대해 코어와 리덕트를 탐색하여 규칙의 표현을 위한 꼭 필요한 최소한의 속성 값들만 남기게 된다. 식 (12)의 규칙 $R1$ 에 대하여 먼저 조건부 속성의 상한 근사와 의사결정부 속성 범주를 구하면 다음과 같고,

$$\begin{aligned} \overline{G1} = \{R1, R2\}, \quad \overline{G5} = \{R1, R4, R7\} \\ \overline{G7} = \{R1, R4\}, \quad D = \{R1, R2, R3\} \end{aligned} \quad (13)$$

이를 이용하여 조건부 속성들의 하한 근사를 구하면

$$\begin{aligned} \underline{A1} = \overline{G1} \cap \overline{G5} = \{R1\}, \quad \underline{A2} = \overline{G1} \cap \overline{G7} = \{R1\} \\ \underline{A3} = \overline{G5} \cap \overline{G7} = \{R1, R4\} \end{aligned} \quad (14)$$

과 같은데, 식 (14)에서 $A1$ 과 $A2$ 의 속성 값들은 의사결정부 속성 범주에 완전히 포함되지만 $A3$ 는 완전히 포함되지 않는다. 따라서 규칙 $R1$ 의 표현을 위한 코어와 리덕트 속성은 $A1$ 과 $A2$ 의 조합으로 다음과 같이 탐색될 수 있다.

$$R1 : A1 \cup A2 = (G1 \cap G5) \cup (G1 \cap G7) = G1 \cap (G5 \cup G7) \quad (15)$$

따라서 규칙 $R1$ 에 꼭 필요한 코어 속성은 $G1$ 이고, $G5$ 와 $G7$

은 세부 표현을 위한 리덕트 속성이다. 두 개의 리덕트 성분 중 하나만을 선택하여 규칙을 생성하게 되면 규칙 $R1$ 은 다음의 둘 중 하나로 간략화 할 수 있다.

$$\begin{aligned} R1 : G1 = gas1 \ \& \ G5 = gas2 \ \rightarrow \ D = gas1 \\ R1 : G1 = gas1 \ \& \ G7 = gas1 \ \rightarrow \ D = gas1 \end{aligned} \quad (13)$$

이러한 방법으로 모든 규칙들에 대하여 코어 속성과 리덕트 속성을 찾으면 감축된 최소의 규칙기반을 얻게 되며, 이 규칙기반에 의해 최종적인 가스 식별 결과를 산출하게 된다.

4. 시뮬레이션 및 검증

제안된 식별 시스템의 효용성을 입증하기 위하여 아세톤, 부탄, 메탄, 프로판, 물 5종류의 가스에 대한 식별을 수행하였다. 시뮬레이션에 사용된 데이터들은 32개의 센서로 구성된 센서 어레이를 이용하여 5종류의 가스들에 대해 각 가스별로 25회씩 반복 측정된 것이며, 그 중 15회의 측정 데이터가 하위 및 상위 식별 규칙기반 구축을 위한 학습 데이터로 사용되었고, 10회의 측정 데이터는 검증 데이터로 활용하였다.

먼저 센서 그룹화를 위해 각 그룹에 포함되는 센서의 수(염색체 길이)는 하위 TSK 퍼지 논리에 적용하는 부분과 그룹 분할의 정확성을 위해 4개의 센서들로 구성될 수 있도록 하였다. 따라서 HGA를 이용한 센서 그룹화에 의해 결정되는 그룹 수는 32개의 센서들에 대하여 4개씩 그룹을 진행하므로 총 8개의 센서 그룹에 의한 1차 판별이 수행되게 되며, 아래의 표는 이를 위해 사용된 HGA의 매개변수들을 보여주고 있다.

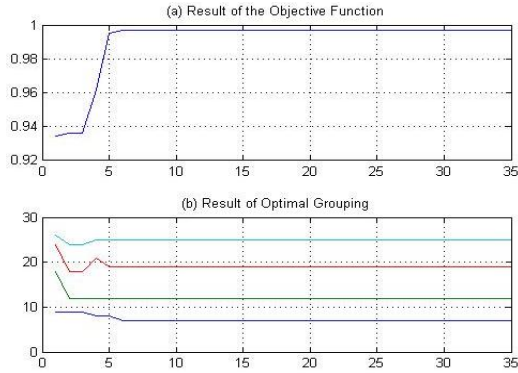
표 1 HGA에 사용된 매개변수

Table 1 Parameters of HGA

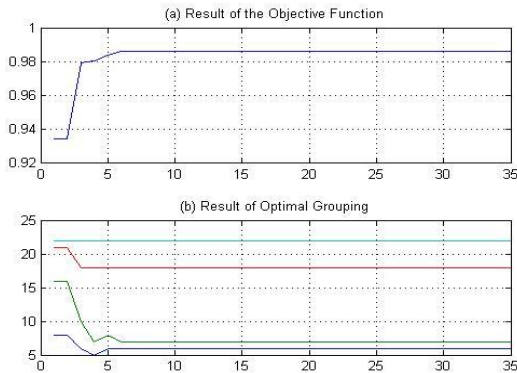
Maxgen	Maximum repetition time	35
Popsiz	Population size(G_{max})	30
Lchrom	Chromosome Length(n_{gs})	4
Etha	Reproduction parameter(η)	1.7
Pcross	Crossover probability(p_c)	0.9
Pmutat	Mutation probability(p_m)	0.1

그림 2는 첫 번째 그룹화 과정과 두 번째 그룹화 과정에 대한 HGA의 목적함수의 변화 및 센서 조합에 대한 결과를 보여준다.

그림 2를 살펴보면, 두 경우 모두 7세대 이전에 수렴하는 성능을 보여주고 있어 HGA의 탐색 성능이 우수함을 알 수 있으며, 목적함수의 변화와 센서 번호의 변화를 살펴보면 제안된 HGA에 의해 상관성이 높은 센서들로 그룹화가 진행됨을 알 수 있으며, 이러한 특성은 식별을 위한 좋은 조건으로 작용할 수 있을 것이다. 같은 방법으로 나머지 센서들에 대하여도 그룹화를 진행하게 되며, 8개의 모든 그룹에 대하여 센서 그룹화가 완료되면 각 그룹별로 TSK 퍼지 모델을 이용하여 각 센서데이터에 대한 가스 식별을 수행하게 된다.



(a) Process of the first sensor grouping



(b) Process of the second sensor grouping

그림 2 HGA에 의한 센서 그룹화 결과

Fig. 2 The result of the sensor grouping by HGA

표 2 그룹들의 식별 결과로 표현된 초기 의사결정 표

Table 2 Initial decision table expressed by identification results of all the groups

Initial decision table					
Index rule	Conditional attribute			Decisional attribute	
	G1	...	G7, G8		
1	1	...	2	1	1 (Acetone)
2	1		3	2	
3	2		1	2	
4	2		3	1	2 (Butane)
5	2		2	1	
6	2		2	2	
7	2		3	1	3 (Methane)
8	3		3	2	
9	3		4	3	4 (Propane)
10	3		2	3	
11	4		5	4	5 (Water)
12	4		4	3	
13	3		4	4	
14	4		4	3	
15	5		4	4	
16	5		5	4	

표 2에서 각 그룹별로 수행된 1차 식별 결과는 다소 부정확함을 볼 수 있다. 그러므로 높은 식별 정확도를 얻기 위해서는 추가로 이 결과를 이용하여 2차 식별이 수행되어야 한다. 표 2에서 규칙 4와 7, 12와 14는 비일관적인 규칙이므로 제외하고, 나머지 12개의 규칙에 대해 불필요한 속성을 제거하면 G1, G7, G8의 3개 속성만 남게 된다. 이에 대해 코어와 리덕트를 구하면 표 3의 의사결정표가 얻어진다. 표 3에서 음영으로 나타낸 부분이 코어이다.

표 4는 설계된 가스 식별 시스템을 이용하여 학습 데이터에 대해 식별한 결과이다. 표 4에서 보면, 각 그룹별 퍼지 식별 시스템의 성능은 평균 93.2%의 식별률을 나타내고 있다. 이는 센서 측정 데이터의 불확실성과 더불어 1차 식별 시스템의 규칙기반을 최소화하기 위해 단순하게 2개의 퍼지 집합을 사용한 결과로 보인다. 하지만 상위 레벨의 러프 집합에 의한 최종 식별 규칙을 적용하면 모든 가스 데이터에 대하여 100% 식별하는 결과를 보여준다.

이러한 결과가 설계된 식별 시스템이 학습 데이터에 대해 최적화되어 있기 때문에 당연한 것 아니냐고 생각할 수 있으므로, 시스템 설계에 전혀 사용되지 않은 50(=5×10)개의 검증 데이터를 이용하여 가스 식별을 수행하도록 하였다. 검증 데이터의 가스 센서별 측정 데이터 패턴은 그림 3과 같다. 그림에서 보듯이

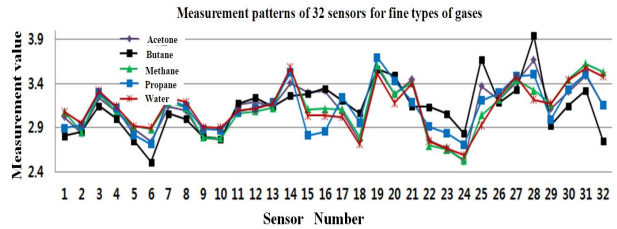


그림 3 센서별 검증 데이터의 측정 패턴

Fig. 3 Measurement patterns of each sensor for test data

표 3 코어와 리덕트로 표현된 최종 식별 의사결정표

Table 3 The final decision table expressed by Core and Reduct

Final decision table				
Index rule	Conditional attribute			Decisional attribute
	G1	G7	G8	
1	1	2		1 (Acetone)
2	1		2	
3		1	2	
4	2		1	2 (Butane)
5		2		
6	3	3		3 (Methane)
7		4	3	
8	3	2	3	4 (Propane)
9	4		4	
10	3		4	5 (Water)
11	5		4	
12	5	5		

표 4 학습데이터에 대한 식별 결과

Table 4 Identification results for the learning data

Identification stage	Gas	Fuzzy system (performance : %)								
		G1	G2	G3	G4	G5	G6	G7	G8	
Performance (Fuzzy system)	Acetone	80	88	76	96	92	76	84	100	
	Butane	84	100	100	100	68	88	92	100	
	Methane	96	88	80	100	100	84	100	96	
	Propane	100	100	96	100	100	100	96	100	
	Water	100	100	96	96	96	80	100	100	
Average of each group		92	95.2	89.6	98.4	91.2	85.6	94.4	99.2	
Performance after secondary identification		Acetone		Butane		Methane		Propane		Water
		100		100		100		100		100

같은 가스에 대해서도 각 센서의 측정값이 상당한 차이를 보이고 있음을 알 수 있다.

제안된 식별 시스템의 최종 예측 결과는 총 50개의 검증 데이터 중에서 아세톤에 대해서만 1회 식별 오류가 발생하였으며, 나머지 데이터에 대해서는 모두 올바르게 식별하여 높은 정확도 (98%)를 나타내었다. 이 1회의 오류는 그 데이터가 아세톤에 대한 다른 측정 데이터들의 패턴과 차이가 많이 나는 악성(ill conditioned)의 경우인 것으로 판단된다.

5. 결 론

본 논문은 하이브리드 유전 알고리즘과 퍼지 집합, 그리고 러프집합 이론을 결합하여 각각의 장점을 살려 구현이 용이하고 성능이 좋은 효율적인 계층 구조 가스 식별 시스템을 설계하였다. 최적화 기능이 뛰어난 유전 알고리즘을 이용하여 센서 어레이를 유사성을 갖는 소수의 센서들로 묶인 여러 개의 그룹으로 나누고 TSK 퍼지 모델을 이용하여 각 그룹별로 1차 가스 식별 시스템을 최소 규칙을 가지도록 구축함으로써 패턴 분석을 위한 데이터의 차원 축소와 가스 식별을 위한 규칙기반 생성이 보다 편리하게 이루어질 수 있도록 하였다. 그리고 상위 레벨 최종 식별 규칙을 러프 집합을 이용하여 생성함으로써 규칙기반을 최소화하면서도 식별 시스템의 성능을 효과적으로 개선할 수 있는 방법을 제시하였다. 제안된 계층 구조식별 시스템은 하위 레벨의 퍼지 식별 규칙기반에 의해 센서 동작 편차나 유동 등으로 인한 데이터의 부정확성이나 불확실성을 효과적으로 처리할 수 있으며, 상위 레벨의 러프 집합 식별 규칙기반에 의해 그룹들 간의 식별 결과의 불일치나 비일관성 등에 잘 대처하여 식별 정확도를 높일 수 있게 해준다. 그리고 부분적인 추출 특징에 의한 식별이 아니라 센서 어레이에 포함된 모든 센서들의 측정 패턴이 원형대로 사용된 식별이 이루어지므로 보다 정확하고 신뢰성 있는 식별이 가능해진다. 제안된 시스템의 이러한 우수성은 시뮬레이션 결과에서도 확인할 수 있다. 따라서 제안된 설계 기법은 가스 식별뿐만 아니라 다양한 식별 시스템의 설계에 활용될 수 있으리라 사료된다.

감사의 글

2013년도 강원대학교 학술연구조성비(120130080)로 연구 하였습니다.

References

- [1] R. Gutierrez-Osuna, "Pattern Analysis for Machine Olfaction : A Review", IEEE Sensors Journal, vol. 2, no. 3, pp. 189-202, 2002.
- [2] 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.
- [3] 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.
- [4] D. Vlachos, J. Avaritsiotis, "Fuzzy neural networks for Gas Sensing", Sensors and Actuators B, vol. 3, pp. 77-82, 1996
- [5] R. Gutierrez-Osuna and H. T. Nagle, "A Method for Evaluating Data-Preprocessing Techniques for Odor Classification with an Array of Gas Sensors", IEEE Trans. on Systems, Man, and Cybernetics-part B: Cybernetics, vol. 29, no. 5, pp. 626-632, 1999.
- [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. 33, pp. 77-82, 1996.
- [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, 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] J. M. Mendel, "Uncertainty, Fuzzy Logic and Signal Processing", Signal Processing, vol. 80, pp. 913-933, 2000.

저 자 소 개



방 영 근 (Young-Keun Bang)

1974년 4월 7일생. 2003년 강원대(삼척) 전기공학과 졸업(석사). 2010년 강원대(춘천) 전기전자공학과 졸업(박사)

E-mail: b2y2c1@hanmail.net



이 철 희 (Chul-Heui Lee)

1959년 7월 4일생. 1985년 서울대학교 전기공학과 졸업(석사). 1989년 서울대학교 전기공학과 졸업(박사). 현재 강원대학교 전기전자공학과 교수

E-mail: chlee@kangwon.ac.kr