

러프집합 기반 GAS 식별 모델 설계

방영근, 조해파, 김남석, 이철희
강원대학교

Design of GAS Identification model based on Rough Sets

Young-Keun Bang, Zhao Haibo, Nam-suk Kim, Chul-Heui Lee
Kangwon National University

Abstract - 인간의 감각 중 후각에 해당하는 가스 센서들에 관한 연구가 현재 상당히 이루어지고 있다. 본 논문에서는 32개의 가스 센서들로부터 측정된 각각의 값들과 GA를 이용하여, 4개의 센서로 구성되는 8개의 센서그룹을 결정한 후 각각의 그룹에서 나타나는 측정값들의 패턴과 러프집합이론을 이용하여 1차 식별 규칙을 생성하였다. 그 다음 8개 가스 그룹의 식별 패턴을 분석하여 다시 러프집합을 통한 2차 식별 규칙을 생성함으로써 보다 효율적이면서도 판단의 정확성을 높일 수 있는 식별 모델을 설계하는 방법을 다룬다.

는 최종 식별 규칙 기반의 생성과정으로 이루어져 있다.

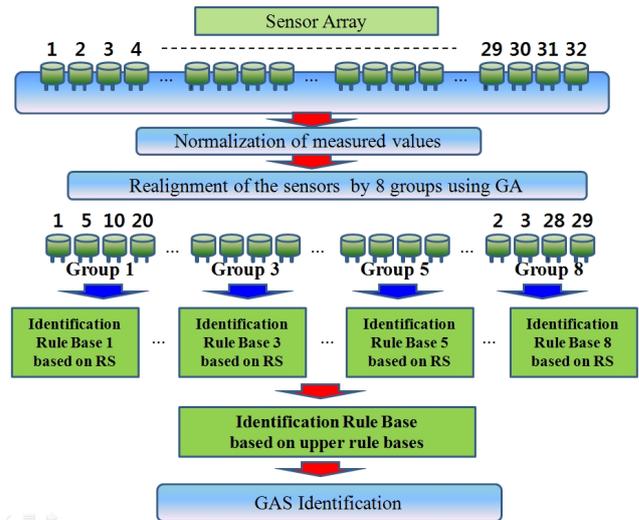
1. 서론

기존의 가스 식별 알고리즘은 패턴인식의 한 분야로 가스에 종류에 따른 전기 신호를 생성해 줄 수 있는 센서 어레이와 이렇게 검출된 전기 신호를 분석하기 위한 다양한 패턴인식 기법들로 구현되어 왔다. 가스 식별 알고리즘의 구현을 위해선 우선, 검출된 신호들은 센서의 유동적 특성으로 인해 한 종류의 가스에 대해서도 검출된 신호의 크기가 다르게 나타날 수 있으므로 이에 적절히 대처할 수 있는 신호 전처리 문제, 또 측정 센서의 개수가 증가할수록 나타나는 특징벡터들의 차원 축소문제, 다음으로 특징 벡터들을 효과적으로 생성하기 위한 분류기법(Classification techniques)이나 군집화기법(Clustering Algorithms)들에 따른 식별능력의 개선 등 다양한 또는 복잡한 구조적 어려움이 존재한다 [1]. 본 논문에서는 32개의 센서를 이용하여 가스의 종류를 식별할 수 있는 식별 모델 설계를 다룬다. 먼저 본 논문에서는 센서로부터 센서들로부터 측정된 신호의 크기들을 일정한 양의 크기로 정규화 한 다음 이렇게 정규화된 신호들을 이용하여 32개의 센서를 4개씩 8개의 그룹으로 분할하는 과정을 거친다. 8개의 그룹 분할을 위해선 실수코딩 유전알고리즘을 적용하여 각 그룹 내의 신호들의 특징들이 잘 드러날 수 있도록 하였으며, 또한 이러한 과정을 통해 32개의 특징 요소를 8개의 큰 특징분류로 표현할 수 있게 되므로 1차 차원축소의 과정으로 이용되게 된다. 다음으로 각각의 그룹은 4개의 특징요소로 이루어져 있으며, 여기서 각각의 그룹에서 가스의 종류를 식별할 수 있는 의사결정 규칙을 생성하기 위해 러프집합 이론의 코어와 리덕트 탐색을 수행하게 된다 [2]. 이러한 과정은 4개의 특징요소 중 불필요한 요소를 제거할 수 있을 뿐만 아니라 적은 수도도 효과적인 의사결정을 내릴 수 있는 규칙기반의 생성이 가능하므로 2차 차원축소의 과정을 수행하게 된다. 이러한 과정을 거치게 되면 하나의 가스에 대해 8개의 식별 규칙이 동작하게 되므로 8개의 의사결정 값이 도출되게 된다. 마지막으로 이렇게 각 그룹에서 도출된 의사결정 값을 이용하여 또 다시 러프집합을 이용한 코어와 리덕트 탐색 과정을 통해 최종 의사결정 규칙을 생성함으로써 센서들과 가스 종류별 특성들을 모두 고려할 수 있는 구조로 모델을 설계한다. 이는 역으로 최종 식별 규칙은 그 양은 적으나 8개 그룹의 식별 패턴을 모두 고려한 형태이고 8개의 그룹은 32개의 센서특성을 모두 고려한 형태가 되는 것이며, 또한 식별 규칙의 생성과정에서 불필요한 특징요소들은 제거가 되며, 초기 방대한 양의 규칙들이 모두 처리되어 소수의 규칙들만으로 구성되므로 효과적인 식별 모델의 역할을 수행할 수 있게 된다.

2. 본론

2.1 제안된 식별 모델의 구조

그림 1은 제안된 가스 식별 모델의 구조를 보여준다. 제안된 가스 식별 모델은 먼저 32개의 가스 센서들로부터 측정된 값들의 모든 합이 일정한 값이 되도록 하는 정규화 과정과, 정규화된 값들을 이용하여 최적 조합의 그룹을 생성하기 위한 GA(genetic algorithms)를 이용한 8개의 가스 센서 그룹 생성과정, 그리고 각각의 그룹들이 효과적으로 가스의 종류를 식별 할 수 있도록 하기 위한 식별규칙기반의 생성 과정, 마지막으로 8개의 그룹에서 내려진 식별 값을 통해 최종 식별을 수행하게 되



<그림 1> 제안된 가스 식별 모델의 구조

2.2 식별 모델의 설계

본 논문에서는 식별 모델 설계를 위해 시스템 최적화에 매우 효과적인 GA와 방대한 정보로부터 효과적이면서도 최소 규칙기반을 생성할 수 있는 러프집합 이론을 사용한다.

2.2.1 GA를 이용한 센서 그룹 생성

32개의 가스 센서들을 8개의 센서 그룹으로 배열하기 위해 본 논문에서는 GA를 이용한다. 먼저 탐색의 방향을 결정하는 GA의 목적함수를 다음과 같이 각 그룹 내의 4개의 가스 센서들이 측정된 측정값의 평균과 특정 값들 사이의 교차상관함수가 큰 쪽으로 탐색할 수 있도록 주어진다.

$$\rho_{MN}^g = \frac{C_{MN}^g}{\sqrt{C_{MM}^g} \sqrt{C_{NN}^g}} \quad (1)$$

여기서, ρ_{MN}^g 는 g 번째 가스 그룹의 교차상관함수를 의미하고, C_{MN}^g 는 g 번째 가스 센서그룹에서 측정값들의 평균과 각각의 측정값들에 대한 교차 공분산을 의미한다. 또한, C_{MM}^g 와 C_{NN}^g 는 g 번째 가스 센서 그룹에서의 평균과 측정 데이터들의 각각의 공분산을 의미한다. 따라서, 이러한 탐색 전략은 측정된 센서 값들의 평균에 대한 분산이 크고, 뿐만 아니라 이에 상응하여 측정된 값들의 교차공분산이 큰 쪽으로 탐색이 진행되므로 가스 종류에 대한 데이터들의 패턴을 보다 명확히 구분 짓게 할 수 있다. 또한, GA의 코딩 기법에는 실수 코딩기법을 적용하였으며, 교배기법으로는 실수코딩의 수정순교배를 이용하고 돌연변이 기법으로는 동적 돌연변이를 채택함으로써 보다 효과적인 가스센서 그룹이 생성될 수 있도록 한다. 마지막으로 엘리트 전략을 사용하여 부모세대의 염색체가 자신의 생성에 기여할 수 있도록 하여 수렴성을 향상시킬 수 있도록 한다. 결국 이러한 탐색전략은 서로 특성이 유사한 센서들끼리 그룹을 형성할 수 있도록 함으로써 8개의 그룹들의 특징들을 보다 효과적으로 표현 할 수 있게 된다.

2.2.2 러프집합을 이용한 식별규칙 생성

GA에 의해 결정된 8개의 센서그룹은 각각에 상응하는 식별규칙을 생성하게 되며 방법은 동일하므로 하나의 그룹에 대한 식별규칙의 생성과정을 설명한다. 먼저 4개의 센서 각각의 측정값들을 이용하여 5개의 기준 값을 정의하게 된다. 이러한 방법에는 다양한 군집화 과정을 통할 수 있지만, 본 논문은 단순히 최대값, 최소값, 평균을 이용하여 정의하였다. 이렇게 다섯 개의 기준 값이 정의되면 측정데이터들과의 거리를 계산하여 측정된 데이터가 가까운 쪽의 기준 값 번호를 의사결정표의 조건부 속성값으로 사용한다. 또한, 이 의사결정표의 의사결정부 속성 값에는 5개 가스의 식별번호로 1에서 5까지 적용한다. 이러한 과정을 거치면 150개의 초기 의사결정 규칙이 생성되게 된다. 아래의 표는 초기 의사결정 표의 한 예를 보여준다.

<표 1> 초기의 의사결정 표

규칙	조건부 속성 (측정값에 상응하는 기준 값 번호로 정의)				의사결정부 속성 가스 번호
	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	
5	2	2	2	1	2(부탄)
6	2	1	2	2	
7	2	1	3	1	
8	3	2	3	2	
9	3	3	4	3	3(메탄)
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	5(물)
15	5	4	4	4	
16	5	3	5	4	

의사결정 규칙을 감축하기 위해선 먼저 비일관적인 규칙을 처리하여야 한다. 표 1에서 규칙 4와 규칙 7, 그리고 규칙 12와 규칙 14는 똑같은 조건부 속성값에 의해 다른 가스의 종류를 지시하고 있으므로 유용한 규칙으로 간주될 수 없다. 따라서 규칙 4, 7, 12, 14는 삭제된다. 다음으로는 불필요한 속성을 제거하여야 한다. 센서 1의 모든 측정값들을 제거하고 센서 2, 3, 4의 속성값들에 따른 비일관적 규칙의 생성 여부를 살펴보면 규칙 11과 규칙 16은 같은 조건부 속성값에 대하여 서로 다른 가스를 식별하게 된다. 따라서 센서 1은 규칙을 특징화하는 부분에 있어 없어서는 안 되는 속성이 되므로 삭제 될 수 없다. 반면에 센서 2의 경우에는 모든 속성값들을 배제하더라도 비일관적인 규칙이 생성되지 않는다. 따라서 센서 2는 가스의 종류를 식별함에 있어 불필요한 요소가 되므로 제거될 수 있다. 표 2는 일련의 과정을 통해 차원이 축소된 형태의 의사결정 표를 보여준다.

<표 2> 비일관적 규칙과 불필요한 속성이 제거된 형태

규칙	조건부 속성 (측정값에 상응하는 클러스터 번호로 정의)			의사결정부 속성 가스 번호
	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	
7	3	4	3	3(메탄)
8	3	2	3	
9	4	5	4	4(프로판)
10	3	4	4	
11	5	4	4	
12	5	5	4	5(물)

마지막으로 최종 규칙기반의 감축을 위해 러프집합 이론의 코어와 리

덕트 속성을 탐색하면 된다. 먼저, 규칙 1에 대한 코어와 리덕트를 탐색하면, 센서 1의 속성과 1이 포함된 규칙의 상한근사는 다음과 같이 정의된다.

$$S11 = \{1,2\} \quad (2)$$

다음 센서 3의 속성값 2를 포함하는 규칙의 상한근사는

$$S12 = \{1,4,5,8\} \quad (3)$$

마지막으로 센서 4의 속성값 1에 대한 규칙의 상한근사는

$$S13 = \{1,4\} \quad (4)$$

로 정의되어진다. 그리고 이 규칙 1은 의사결정부 속성의 1을 식별하는 규칙이므로 이에 대한 상한근사는 다음과 같이 정의된다.

$$D1 = \{1,2,3\} \quad (5)$$

이렇게 각각의 속성값들에 대한 상한근사가 정의 되면 조건부 속성값들의 하한근사가 의사결정부의 범주에 완전히 포함되는지를 판단하여 코어와 리덕트를 찾게 되며, 다음과 같다. 먼저 식 (2)와 (3)의 하한 근사는 규칙 1이 되며, 이는 의사결정부의 상한근사에 완전히 포함되므로 필요한 성분이 된다. 다음 식(3)과 (4)에 대한 하한근사는 규칙 1,4가 되며, 여기서 규칙 4는 의사결정부 상한근사에 포함되지 않으므로 이는 불필요한 성분이 된다. 마지막으로 식 (2)와 (4)에 대한 하한근사는 규칙 1이 되고, 이는 의사결정부 상한근사에 완전히 포함되므로 필요한 성분이 된다. 따라서 필요한 성분들에 대한 식으로 정리하면 다음과 같다.

$$S11 \cap S12 \subseteq D1 \quad (6)$$

$$S11 \cap S13 \subseteq D1 \quad (7)$$

그러므로 S11은 1번 가스 식별을 위한 공통 속성이 되며, 따라서 꼭 필요한 요소인 코어속성이 된다. 또한 S12와 S13는 규칙 1에 대한 리덕트 속성이 된다. 표 2에서 진하게 표시된 부분이 코어 속성을 의미하며, 리덕트 속성들은 위의 방법들을 통해 정의 할 수 있게 된다. 이러한 형태로 8개 그룹에 대한 식별 규칙을 만들고 그들의 결정 값 또한 1에서 5사이의 가스 번호로 나타나므로 8개의 속성을 갖는 초기의 최종 의사결정 표를 작성할 수 있으며, 이를 같은 방법으로 코어와 리덕트를 탐색하면 적은 수의 최종 식별 규칙기반을 생성할 수 있게 된다.

3. 시뮬레이션

아래의 표는 학습에 사용된 150개의 측정된 값들을 재입력으로 사용하였을 때의 결과를 보여준다.

<표 3> 식별 시스템의 성능

식별 단계	가스 종류	그룹 번호(성능: 백분율)							
		1	2	3	4	5	6	7	8
그룹별 성능	아세톤	100	100	100	63.3	100	100	100	100
	부탄	100	36.6	100	100	100	100	100	56.6
	메탄	23.3	100	90	36.6	100	96.6	10	100
	프로판	13.3	30	93.3	100	93.3	100	100	100
	물	90	96.6	70	100	100	90	40	100
그룹별 평균		65.3	72.6	90.6	80	98.6	97.3	70	91.3
시스템의 최종 성능	아세톤	부탄		메탄		프로판		물	
	100	100	100	100	100	100			

표 3을 살펴보면 8개의 각 그룹의 가스 식별율은 평균 83.24%로 나타났다. 이들에 대한 2차 식별 규칙을 이용한 관별 후에는 모든 가스에 대하여 제안된 모델이 정확히 식별하였음을 알 수 있다. 이는 32개의 센서들이 가지는 측정 패턴들을 1차로 그룹화하여 식별하고 이를 토대로 2차 식별 규칙을 생성하여 각 그룹의 식별 패턴을 반영함으로써 정확한 식별을 할 수 있음을 보여주는 결과로 간주 할 수 있다.

3. 결 론

본 논문은 GA와 러프집합을 이용한 계층구조의 식별 규칙을 갖는 가스 식별모델의 설계에 관한 방법을 다루었다. 제안된 식별 모델은 센서들의 모든 특성을 고려하면서도 차원이 축소된 형태의 효과적인 식별 규칙을 생성하여 식별하는 구조로 식별 규칙생성을 위한 복잡한 군집화 과정 없이도 결과에서 나타나듯이 학습에 이용된 모든 가스 데이터들을 정확히 식별할 수 있음을 보여 주었다.

[참 고 문 헌]

[1] Ricardo Gutierrez-Osuna, "Pattern Analysis for Machine Olfaction: A Review", IEEE SENSORS JOURNAL, vol. 2, no. 3, pp. 189-202, 2002.
 [2] B. Walczak, D. L. Massart, "(Tutorial) Rough sets theory", Chemometrics and Intelligent Laboratory Systems, vol. 47, pp. 1-16, 1999.