

계층별 메트릭 생성을 이용한 계층적 Gaussian ARTMAP의 설계

(A Design of Hierarchical Gaussian ARTMAP
using Different Metric Generation for Each Level)

최 태 훈[†] 임 성 길^{**} 이 현 수^{***}
(TeaHun Choi) (SungKil Lim) (HyonSoo Lee)

요약 본 논문에서는 아날로그 데이터 처리가 가능하고, 온라인 학습, 학습 중 새로운 클래스 추가 등의 특징을 가진 패턴 인식기를 제안하였다. 제안한 패턴 인식기는 계층적 구조를 가지고 있으며, 각 레벨로 서로 다른 메트릭을 적용하여 분류 성능을 향상 시켰다. 제안한 패턴 인식기는 신경망 기반의 패턴 인식 알고리즘인 Gaussian ARTMAP 모델을 기반으로 하고 있다. Gaussian ARTMAP 모델을 계층적으로 구성하고, 계층마다 서로 다른 특징을 학습하도록 하기 위하여 Principal Component Emphasis (P.C.E) 방법을 제안하였으며, 이를 이용하여 새로운 메트릭을 생성하는 방법을 제안하였다. P.C.E는 학습된 입력 데이터들의 분산을 이용하여 클래스 내의 공통 속성을 나타내는 분산이 작은 차원을 제거하고 패턴 간의 서로 다른 속성을 나타내는 분산이 큰 차원만 유지하는 방법이다. 제안한 알고리즘의 학습 과정에서 교사 신호와 다르게 분류된 패턴이 발생하면 잘못 분류된 클래스와 입력된 패턴을 분리하기 위하여 P.C.E를 수행하고 하위 노드에서 학습하게 된다. 실험 결과 제안한 모델은 기존에 제안된 패턴 인식 모델들 보다 높은 분류 성능을 가지고 있음을 확인하였다.

키워드 : ARTMAP, Gaussian ARTMAP, 계층적 분류기, 주성분 강조

Abstract In this paper, we proposed a new pattern classifier which can be incrementally learned, be added new class in learning time, and handle with analog data. Proposed pattern classifier has hierarchical structure and the classification rate is improved by using different metric for each levels. Proposed model is based on the Gaussian ARTMAP which is an artificial neural network model for the pattern classification. We hierarchically constructed the Gaussian ARTMAP and proposed the Principal Component Emphasis(P.C.E) method to be learned different features in each levels. And we defined new metric based on the P.C.E. P.C.E is a method that discards dimensions whose variation are small, that represents common attributes in the class. And remains dimensions whose variation are large. In the learning process, if input pattern is misclassified, P.C.E are performed and the modified pattern is learned in sub network. Experimental results indicate that Hierarchical Gaussian ARTMAP yield better classification result than the other pattern recognition algorithms on variable data set including real applicable problem.

Key words : ARTMAP, Gaussian ARTMAP, Hierarchical Classifier, Principal Component Emphasis

[†] 정 회 원 : 현대엔지니어링(주) 기술혁신개발실 프로젝트IT부
r6master@naver.com

^{**} 학 생 회 원 : 경희대학교 컴퓨터공학과
skan72@khu.ac.kr

^{***} 정 회 원 : 경희대학교 컴퓨터공학과 교수
leehs@khu.ac.kr
(Corresponding author임)

논문접수 : 2009년 3월 6일

심사완료 : 2009년 6월 4일

Copyright©2009 한국정보과학회 : 개인 목적이나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.

정보과학회논문지 : 소프트웨어 및 응용 제36권 제8호(2009.8)

1. 서론

전통적인 패턴 인식기에서는 학습과 리콜(Recall) 과정이 구분되어 있고, 학습 전 클래스의 개수를 결정해야 한다. 하지만 최근 패턴인식 시스템에 관련된 연구가 활발히 진행되고 있는 로봇틱스 분야에서는 모바일 로봇이 주위 환경을 끊임없이 학습하고 인식해야 할 필요성이 강조되고 있다. 따라서 최근 패턴 인식기에서는 학습과 리콜이 어느 시점에서든 새로운 클래스를 자유롭게 추가 및 학습이 가능한 특징이 요구되고 있다[1,2]. 또한 기존에 제안된 패턴 인식기의 분류 성능을 높이기 위한 연구도 활발히 진행되고 있다. 패턴 인식기의 분류 성능을 높이기 위한 방법으로는 효율적인 메트릭을 적용하여 분류 성능을 높이는 연구와[3-5] 패턴 인식기를 계층적으로 구성하여 분류 성능을 높이는 연구들이 진행되었다[6-9].

특히 Hierarchical ART-S 모델[6]과 Gaussian ARTMAP 모델[3]은 신경망 기반의 클러스터링 알고리즘인 Adaptive Resonance Theory를 기반으로 하고 있어 초기에 클러스터의 수를 결정 하지 않아도 학습이 가능하며, 온라인 학습 및 점진적 학습이 가능하다는 장점을 가지고 있다. Hierarchical ART-S 모델은 다수의 ART 모델이 계층적으로 연결된 구조를 가지고 있어 입력 패턴의 계층적 구조를 분석하여 클러스터링 할 수 있는 장점을 가지고 있다. 하지만 이진 데이터만 처리가 가능하고 하다는 단점을 가지고 있고, 패턴 인식에 적용하기 위해서는 교사 신호를 처리 할 수 있도록 수정되어야 한다. Gaussian ARTMAP은 비교사 학습 모델인 ART 모델을 교사 학습이 가능하도록 수정한 ARTMAP 모델을 기반으로 하여 패턴 인식기로 사용할 수 있는 모델이다. 또한 기존 ART의 헤밍 거리에 기반한 메트릭 대신 가우스 분포(Gaussian Distribution)를 이용한 메트릭을 정의하여 적용하였다[1,3]. 하지만 Gaussian ARTMAP은 단일 메트릭을 사용하기 때문에 서로 다른 클래스에 속하는 입력 데이터들의 공통 속성이 많을 경우 분리 성능이 저하되는 단점을 가지고 있다.

본 논문에서는 연속적인 데이터에 대해서 온라인 학습이 가능하고 새로운 클래스의 추가가 용이하며 계층적 분류가 가능한 특징을 가지고 있는 패턴 인식 모델을 제안한다. 제안한 모델에서 온라인 학습과 새로운 클래스의 추가를 위해서 신경망 기반의 패턴 인식 모델인 ARTMAP을 기반으로 하였고, 연속적인 데이터 처리를 위해 가우스 분포를 이용한 메트릭을 적용하였다. 분류기의 분류 성능을 높이기 위하여 Gaussian ARTMAP을 계층적으로 구성하는 방법을 제안하였으며, 각 계층마다 서로 다른 메트릭을 적용함으로써 각

층에서 서로 다른 특징이 학습되도록 하였다. 각 층마다 서로 다른 메트릭을 적용하기 위해 각 레벨의 입력 패턴의 공통속성을 제거하고 차이점만을 유지시키는 주 성분 강조(Principal Component Emphasis : P.C.E) 방법을 제안하였다. 입력 패턴에 P.C.E를 적용하면 공통속성을 가지는 차원의 데이터는 제거되고 서로 다른 속성을 가지는 데이터만 남아있기 때문에 서로 유사한 특징을 가지고 있는 패턴들도 쉽게 분류할 수 있다는 장점이 있다.

본 논문은 다음과 같이 구성되어 있다. 2장에서는 제안하는 패턴 분류기의 기반이 되는 Gaussian ARTMAP의 구조 및 동작에 대하여 설명하였으며, 3장에서는 Gaussian ARTMAP의 계층적 구조와 새로운 메트릭을 생성하는 방법에 대하여 설명하였다. 4장에서는 제안하는 모델의 학습 특성을 보여주는 실험과 기존의 모델과 성능 비교를 위한 실험 및 결과에 대하여 기술하였고, 5장에서 본 논문의 결론과 향후 연구 과제에 대하여 기술 하였다.

2. Gaussian ARTMAP의 구조와 동작

Gaussian ARTMAP은 기존의 ARTMAP에 기반하여 병렬 컴퓨팅, 빠른 학습, 클래스의 효과적인 내부 표현, 잡음에 강인한 특성 등을 목적으로 제안된 패턴 인식이 가능한 신경망 모델이다[3]. 클래스의 선택 및 확인 과정에서 fuzzy 연산을 이용하는 대신 가우스 분포를 사용함으로써 클래스의 효과적인 내부 표현, 잡음에 강인한 특성 등 기존의 fuzzy ARTMAP의 단점을 개선하였다. Gaussian ARTMAP은 입력되는 패턴을 서로 유사한 다른 패턴들과 하나의 클러스터로 묶는 역할을 하는 Gaussian ART 부분, 패턴과 함께 입력되는 교사 신호를 이용하여 패턴 인식이 가능하도록 학습하는 ARTMAP 부분으로 구성되어 있다.

2.1 Gaussian ART

2.1.1 네트워크 구조

그림 1은 Gaussian ART의 구조를 나타낸 것이다. F_0 층은 입력을 받는 층으로 입력된 M 차원의 데이터 $I = (I_1, \dots, I_M)$ 를 F_1 층으로 전달하는 역할을 한다. F_1 층은 입력된 데이터를 유지하는 역할을 수행한다. F_2 층은 N 개의 클러스터를 나타내며, 각 클러스터 j 는 M 차원의 평균 μ_j 와 각 차원의 표준편차 σ_{ji} , 그 클러스터로 학습된 학습 패턴의 수 n_j 값으로 정의 된다. 즉 Gaussian ART의 각 클러스터는 $2M+1$ 개의 파라미터로 정의되어 있다. 그림 1에서 $g(\cdot)$, $g'(\cdot)$ 은 각각 클러스터의 판별식과 매칭 함수를 나타내며, ρ 는 경계 파라미터(vigilance parameter)를 나타내고 있다.

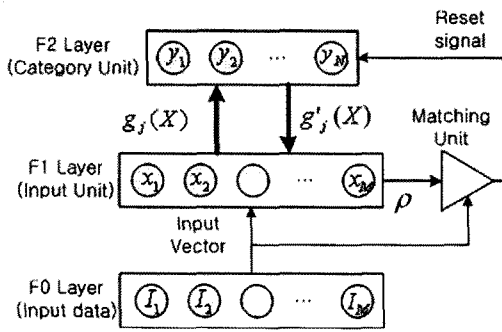


그림 1 Gaussian ART의 구조

2.1.2 클러스터 선택

하나의 패턴 $I = (I_1, \dots, I_M)$ 가 입력되면, 그 패턴이 발생할 가능성이 가장 높은 클러스터 j 를 선택하는데, 이 과정을 클러스터 선택이라고 한다. 클러스터 선택에는 주어진 입력 I 에 대하여 클러스터 j 의 사후 확률이 가장 큰 클러스터를 선택하게 된다. 사후 확률은 식 (1)과 같이 정의된다.

$$P(j|I) = \frac{P(I|j)P(j)}{P(I)} \quad (1)$$

각 클러스터는 가우스 분포로 정의되어 있으므로 주어진 클러스터 j 에 대한 입력 I 의 조건부 확률은 식 (2)와 같이 정의되며, j 에 대한 사전 확률은 식 (3)과 같이 정의된다.

$$P(I|j) = \frac{1}{(2\pi)^{M/2} \prod_{i=1}^M \sigma_{ji}} \exp\left(-\frac{1}{2} \sum_{i=1}^M \left(\frac{\mu_{ji} - I_i}{\sigma_{ji}}\right)^2\right) \quad (2)$$

$$P(j) = \frac{n_j}{\sum_{j=1}^N n_j} \quad (3)$$

식 (1)에서 $P(I)$ 는 모든 클러스터에 대하여 동일하므로 생략하고, 계산의 편의를 위하여 식 (1)에 $(2\pi)^{M/2}$ 를 곱하고 로그를 취하여 각 클러스터 j 의 판별식 $g_j(\cdot)$ 를 식 (4)와 같이 정의한다.

$$g_j(I) = \log((2\pi)^{M/2} P(I|j) P(j)) \\ = -\frac{1}{2} \sum_{i=1}^M \left(\frac{\mu_{ji} - I_i}{\sigma_{ji}}\right)^2 - \log(\prod_{i=1}^M \sigma_{ji}) + \log(P(j)) \quad (4)$$

클러스터 선택은 식 (5)와 같이 판별식이 가장 큰 클러스터 J 로 결정한다.

$$J = \arg \max_j (g_j(I)) \quad (5)$$

2.1.3 공명 및 리셋

입력 패턴에 대한 클러스터가 결정되면 그 클러스터가 입력 패턴의 출처인지를 검증하는 과정을 거치게 되는데 이를 클러스터 공명(resonance)이라고 한다. 만약 선택된 클러스터가 공명 조건을 만족하지 않으면, 그 클

러스터를 제외시키고 나머지 클러스터를 대상으로 다시 클러스터 선택 과정을 반복하게 된다. 클러스터를 선택 과정에서 제외하는 것을 리셋이라고 한다. 한번 리셋된 클러스터는 다음 패턴이 입력될 때 까지 선택 과정에서 제외된다.

선택된 클러스터는 매칭 함수 $g'_j(I)$ 가 주어진 파라미터 ρ 보다 크다면 공명을, 작다면 리셋을 선택한다. 매칭 함수 $g'_j(I)$ 는 식 (6)과 같이 정의된다. 식 (6)은 J 의 분포를 단위 높이로 평균화한 분포이다.

$$g'_j(I) = \log((2\pi)^{M/2} (\prod_{i=1}^M \sigma_{ji}) P(I|j)) \\ = -\frac{1}{2} \sum_{i=1}^M \left(\frac{\mu_{ji} - I_i}{\sigma_{ji}}\right)^2 \\ = g_j(I) - \log(P(j)) + \log(\prod_{i=1}^M \sigma_{ji}) \quad (6)$$

2.1.4 클러스터 학습

선택된 클러스터 J 가 공명 조건을 만족했다면 입력은 클러스터 J 라고 판단하고, 식 (7)과 같이 클러스터 파라미터를 학습시킨다.

$$n_j := n_j + 1 \\ \mu_j := (1 - n_j^{-1}) \mu_j + n_j^{-1} I \\ \sigma_{ji} = \sqrt{(1 - n_j^{-1}) \sigma_{ji}^2 + n_j^{-1} (\mu_{ji} - I_i)^2} \quad (7)$$

만약 모든 클러스터가 리셋된 경우라면 식 (8)과 같이 새로운 클러스터를 추가시킨다.

$$n_{N+1} := 1 \\ \mu_{N+1} := I \\ \sigma_{N+1} := \gamma \quad (8)$$

2.2 Gaussian ARTMAP

클래스의 학습 과정에서 입력 패턴 I 에 대하여 클러스터 J 가 결정되었고, J 가 전체 학습 과정에서 첫 번째로 선택되었다면 J 를 I 의 교사신호로 입력된 클래스 K 에 배정한다.

$$\Omega(J) = K \quad (9)$$

식 (9)는 매핑 함수를 나타낸 것이다. 이 매핑 함수는 다대일 함수로 여러 개의 클러스터가 하나의 클래스에 배정 될 수 있다. 즉 클래스 K 에 해당하는 모든 클러스터는 $\Omega^{-1}(K)$ 로 나타낼 수 있으며, $J \in \Omega^{-1}(K)$ 의 관계가 성립한다.

만약 다른 입력 I' 에 대하여 클러스터 J 가 선택되었고, I' 의 교사신호 K' 이 K 와 같지 않다면, 매치 트래킹 과정을 수행한다. 매치 트래킹 과정은 먼저 경계 파라미터를 $\rho = g'_j(I)$ 로 설정하고 클러스터 J 를 리셋한 후 클러스터 선택 과정을 반복한다. 경계 파라미터를 조정함으로써 입력된 신호를 가장 잘 표현하며, 교사 신호가 동일한 클러스터를 찾거나, 새로운 클러스터를 생성

하게 된다. 조정된 경계 파라미터는 새로운 패턴이 입력 될 때까지 유지하며, 새로운 패턴이 입력되면 원래의 값으로 복원하여 클러스터 선택 과정을 수행한다.

클래스의 판단 과정에서는 가장 적합한 클러스터를 찾는 학습 과정과는 달리 가장 적합한 클래스를 찾는 과정이므로 식 (10)과 같이 입력되는 패턴의 클래스를 결정한다.

$$K = \arg \max_k \left(\sum_{j \in \Omega^{-1}(k)} \exp(g_j(I)) \right) \quad (10)$$

3. 계층적 Gaussian ARTMAP의 제안

본 논문에서는 앞서 기술한 Gaussian ARTMAP을 계층적으로 구성한 계층적 Gaussian ARTMAP을 제안한다. 계층적 구조의 ART는 [7,8]등에서 패턴의 계층적 구조를 분석하기 위한 도구로 사용되었다. 계층적 구조를 패턴 인식 문제에 적용할 경우, 먼저 상위 레벨에서 큰 클래스로 패턴을 분류하고, 하위 레벨에서 세부 클래스로 분류하는 방법을 사용하여, 패턴의 계층적인 구조를 인식 과정에 반영 할 수 있다. 또한 인간이 개와 고양이를 구분 하는 문제와, 개의 각 개체를 구분하는 문제에서 서로 다른 판단 기준을 사용하는 것과 마찬가지로 각 계층마다 서로 다른 메트릭을 적용함으로써 세부 클래스에 대한 인식률을 높일 수 있다는 장점을 가지고 있다. 본 논문에서는 계층 별로 서로 다른 메트릭을 적용하는 방법으로써 주성분 강조(P.C.E) 방법을 제안한다. P.C.E는 한 클래스에서 분산이 작은 입력 차원은 그 클래스의 공통되는 특성이라고 간주하여 제거하고, 분산이 큰 입력 차원만을 추출하여 새로운 입력으로 사용하는 방법을 사용하였다.

3.1 Gaussian ARTMAP의 계층적 구조

그림 2는 제안하는 계층적 Gaussian ARTMAP의 구조를 나타낸 것이다. 제안하는 계층적 구조는 Gaussian ART가 트리 구조로 연결된 것이다. 그림 2에서 가장 상단에 루트 ART가 있고 그 아래에 첫 번째 레벨의 하위 네트워크들이 연결되어 있다. 그림 2는 루트 네트워크를 레벨 0으로 하여 레벨 2까지를 나타낸 것이다.

각 하위 네트워크의 입력은 제안하는 P.C.E 모듈과 연결되어 있다. P.C.E 모듈은 상위 네트워크의 입력 패턴을 입력으로 받아 공통되는 특성을 제거하고, 분산이 큰 차원의 특성만을 추출하여 하위 네트워크의 입력으로 변환하는 역할을 하고 있다.

ART 트리의 우측부분에는 매치 트래킹을 위한 매핑 필드를 나타내었으며, 각 하위 네트워크마다 달리 적용되는 지역적인 경계 파라미터를 각 네트워크의 좌측 혹은 하단에 표현 하였다.

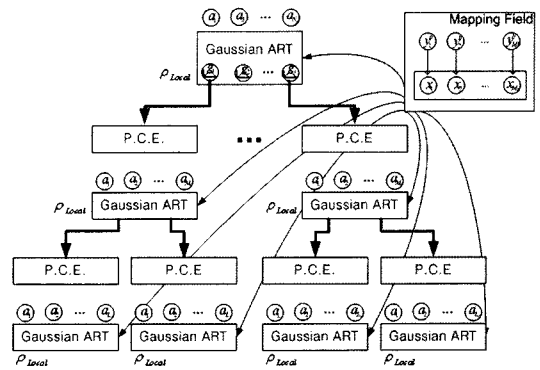


그림 2 계층적 Gaussian ARTMAP의 구조

3.2 클러스터 선택 및 공명/리셋

클래스 정보를 교사 신호로 갖는 입력 I 가 입력되면 루트 네트워크에서 식 (4)와 식 (5)를 이용하여 클러스터 선택 과정을 수행한다. 선택된 클러스터 J 에 대하여 식 (6)을 경계 파라미터 ρ 와 비교하여 클러스터 공명 및 리셋을 결정하고, 식 (7)이나 식 (8)을 이용하여 클러스터 학습을 수행한다. 클러스터의 선택, 클러스터의 공명/리셋 결정 및 클러스터 학습의 동작은 Gaussian ART의 동작과 동일하다.

3.3 매치 트래킹 / 하위 네트워크 생성 판단

교사 신호 K 와 함께 입력된 패턴 I 에 대하여 클러스터 J 가 선택되었고, J 가 첫 번째 선택된 경우라면, 식 (9)와 같이 클러스터 J 를 클래스 K 에 배정한다. 만약 교사 신호 K' 와 함께 입력된 패턴 I' 에 대하여 선택된 클래스가 J 이고 $K \neq K'$ 이라면 매치 트래킹 과정을 수행하거나 하위 네트워크를 생성한다. 매치 트래킹을 수행 할 것인가 혹은 하위 네트워크를 생성 할 것인가의 판단은 입력 패턴의 차원에 의해 결정된다. 입력 패턴의 차원이 2보다 작은 경우는 더 이상 차원 감소를 수행 할 수 없으므로 기존의 Gaussian ARTMAP의 매치 트래킹을 수행하여 현재 레벨의 네트워크에서 클러스터의 수를 증가시킨다.

입력 패턴의 차원이 2이상인 경우는 하위 네트워크를 생성한다. 하위 네트워크를 생성하는 경우는 P.C.E를 거쳐 새로운 입력 패턴을 생성하고, 지역적인 경계 파라미터를 결정된 후에 하위 네트워크에서 클러스터 선택, 클러스터 공명/리셋 판단, 클러스터 학습 과정을 거쳐 지역적인 매핑함수 $\Omega_{(t,r)}(\cdot)$ 를 정의하는 과정을 거친다. 이후 매치 트래킹 수행 및 하위 네트워크의 생성 여부는 루트 네트워크와 동일하게 적용된다.

3.4 주성분확장 방법 및 지역적 경계 파라미터 결정

상위 네트워크에서 하위 네트워크 생성이 결정된 경우 상위 네트워크 입력을 하위 네트워크의 입력으로 변

환하는 P.C.E 과정을 수행한다. P.C.E 과정은 클러스터 내에서 비슷한 값을 가지는 입력 차원, 즉 분산이 작은 입력 차원을 제거하고, 분산이 큰 차원만을 사용함으로써 하위 클러스터의 분포를 보다 클러스터링 하기 좋은 형태로 변환한다. 레벨 l 에서 네트워크 번호 r 의 입력은 식 (11)과 같이 정의 된다.

$$I_i^{(l,r)} = \begin{cases} I_i^p & \text{if } \sigma_i^2 > \frac{1}{M_p} \sum_{s=1}^{M_p} \sigma_s^2 \\ NULL & \text{otherwise} \end{cases} \quad (11)$$

식 (14)에서 I_i^p 는 현재 하위 네트워크의 부모 네트워크로 입력되는 i 번째 차원의 입력값, M_p 는 부모 네트워크의 입력 차원수를 나타낸다.

하위 네트워크에서 사용되는 지역적인 경계 파라미터는 매칭 트래킹에서 사용되는 임시의 경계 파라미터와 유사하게 식 (12)와 같이 정의한다. 지역 경계 파라미터 ρ_i^l 에서 l 은 하위 네트워크의 레벨을, r 은 그 레벨에서의 네트워크 번호를 나타낸다.

$$\rho_i^l = g'(I_i^{(l,r)}) \quad (12)$$

하위 네트워크의 각 클러스터는 사용되지 않는 차원을 NULL로 표시한 M 차원의 평균과 분산, 각 클러스터에 학습된 패턴의 수를 포함하여 $2M+1$ 개의 파라미터로 정의되며, 각 하위 네트워크 마다 고유의 경계 파라미터를 저장하고 있다. 지역적인 매핑함수 $\Omega_{(l,r)}(\cdot)$ 는 루트 네트워크에서 정의하는 방법과 동일한 방법으로 결정한다.

3.5 클래스 판단 과정

계층적 구조의 Gaussian ARTMAP에서 클래스 판단 과정은 기본적으로 기존의 Gaussian ARTMAP의 클래스 판단 과정과 동일하다. 그러나 계층적 구조에서는 하위 네트워크 마다 지역적인 매핑 함수가 존재하므로 어떤 레벨의 매핑 함수까지 클래스 판단 계산에 포함할 것인가를 결정할 수 있고, 더 높은 레벨을 선택할수록 분류율은 높아지게 되며 상대적으로 계산의 복잡도는 증가하게 된다. 입력 패턴 I 에 대하여 식 (13)과 같이 클래스를 판단한다.

$$K = \arg \max_k \left(\sum_{j \in \Omega^{-1}(k)} \exp(g_j(I)) + \sum_{j \in \Omega_{(l,r)}^{-1}(k)} \exp(g_j(I)) + \dots \right) \quad (13)$$

4. 실험 및 결과 분석

제안한 알고리즘의 분류 성능을 평가하기 위해 두 가지 측면에서 실험을 수행하였다. 첫 번째, 알고리즘을 제안하기 위해서 고려했던 효율적인 분류기로서의 특징을 확인하기 위한 실험을 하였다. 분류기의 특징을 확인하기 위해 아날로그 데이터에 대한 온라인 학습, 학습

중 클래스 추가의 용이성, 계층적 분류에 따른 분류 성능 변화에 대한 실험을 수행하였다.

두 번째, 분류기로서의 분류 성능 향상에 대한 실험을 하였다. 분류 성능 향상을 비교하기 위해서 제안한 알고리즘과 Gaussian ARTMAP 알고리즘을 이용하여 계층적 분류에 의한 유사한 데이터들의 분류에 대한 실험과 두 알고리즘의 분류 성능을 비교하는 실험을 하였다. 분류기의 특성 및 분류 성능 실험을 위한 데이터는 각각의 실험에 적합한 데이터를 가우시안 분포를 이용하여 2차원, 4차원, 10차원에 해당하는 데이터를 생성하여 사용하였다. 제안한 모델에서 클래스의 개수와 클러스터링 성능을 결정하는 파라미터인 ρ 와 γ 값은 각각 $\rho=0.5$, $\gamma=0.5$ 로 하였다.

그리고 제안한 모델이 실제의 데이터에도 응용 가능함을 보이기 위한 실험을 진행하였다. 실험 데이터는 기존에 패턴인식 알고리즘에서 성능평가로 사용되었던 데이터 중 Glass Data, Breast Cancer Data, Letter Recognition Data를 선택하였다. 제안한 모델과 다른 패턴 인식기의 분류 성능 비교를 위해서 각 모델의 학습과 리콜 데이터의 개수를 동일하게 조정하였다.

4.1 제안한 계층적 Gaussian ARTMAP의 동작 특성

4.1.1 온라인 학습 실험

제안한 알고리즘의 특성 중에서 온라인 학습은 데이터를 입력 데이터가 교차신호를 포함하고 있는 경우 기존의 네트워크에 입력 데이터를 학습하는 기능을 하고, 입력 데이터가 교차신호를 포함하지 않는 경우 기존의 네트워크에 학습된 클래스 정보를 이용하여 분류를 수행할 수 있는 기능이다. 제안한 알고리즘의 온라인 학습 특성을 나타내기 위해 3개의 클래스를 가지고 있는 데이터를 지속적으로 학습하였으며, 실험 결과의 시각화를 위해 데이터의 차원은 2차원으로 하였다. 실험 데이터의 2차원 평면에 의한 분포는 그림 3과 같다.

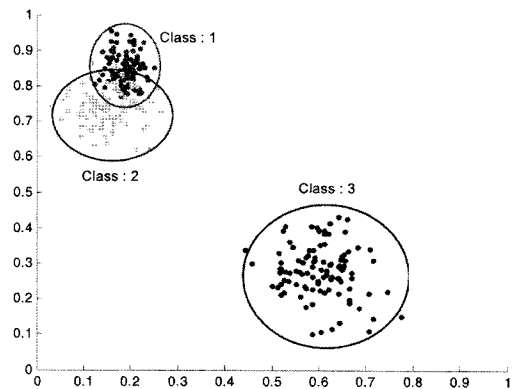
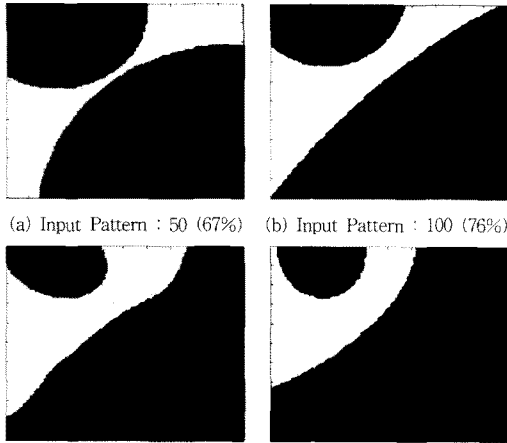


그림 3 2차원 3개 클래스 데이터의 분포



(a) Input Pattern : 50 (67%) (b) Input Pattern : 100 (76%)
(c) Input Pattern : 200 (89.7%) (d) Input Pattern : 300 (92%)
그림 4 On-Line 학습 실험 결과(팔호 안은 분류율)

실험은 주어진 데이터 집합 중에서 초기 훈련 데이터 50 개를 이용하여 학습하였으며, 이후 훈련 데이터를 100, 200, 300개로 증가 시켜가며 학습에 따라서 클래스 경계가 변하는 모습을 확인하였다. 실험 결과는 그림 4와 같다.

실험 결과 제안한 알고리즘은 입력 데이터를 학습함에 따라서 결정 경계가 입력 데이터에 적합하게 학습되고 있음을 확인할 수 있었다. 또한 학습 데이터의 숫자가 증가함에 따라 분류 성능이 향상됨을 확인할 수 있었다. 실험을 통해 제안한 알고리즘이 온라인 학습 특성을 가지고 있음을 확인할 수 있다.

4.1.2 학습 중 클래스 추가 실험

제안한 알고리즘이 새로운 클래스에 해당하는 데이터를 입력 받는 경우 새로운 클래스를 추가하고 기존의 클래스를 업데이트 하는 과정을 보여주기 위한 실험은 총 5개의 클래스를 가지는 2차원 데이터를 이용하였다. 입력 데이터의 분포는 그림 5와 같다.

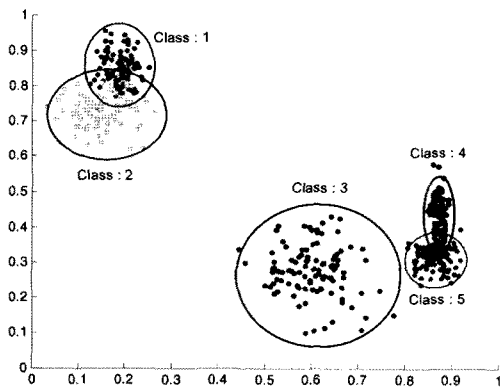


그림 5 2차원 5개 클래스 데이터의 분포

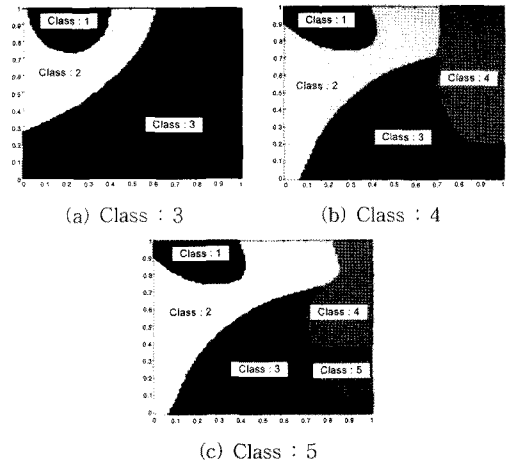


그림 6 클래스 추가에 의한 클래스 경계의 변화

실험은 주어진 데이터 집합 중에서 클래스 1, 2, 3에 포함되는 훈련 데이터 300개를 이용하여 학습을 하여 초기 클러스터를 형성했다. 이후 클래스 4에 해당하는 데이터를 포함한 입력 데이터 100개를 추가로 학습하였고, 클래스 5에 해당하는 데이터를 포함한 입력 데이터 100개를 최종적으로 학습에 사용하였다. 실험 결과는 그림 6과 같다. 실험 결과 학습 중 새로운 클래스에 해당하는 데이터가 입력되는 경우 새로운 클래스가 추가되고 기존의 클래스에 의한 판단경계도 새로운 클래스를 반영하여 이동되는 것을 확인할 수 있다.

4.1.3 계층적 분류에 따른 분류 성능의 변화

제안한 모델의 계층적 분류에 따른 분류 성능 향상을 보이기 위한 실험은 다음과 같다. 가우시안 분포를 가지는 15개 클래스로 구성된 10차원 데이터를 생성하여 계층적 Gaussian ATRTMAP을 이용하여 각 클래스별로 1,000개의 입력 데이터를 학습시킨 후, 계층적으로 구성된 네트워크를 이용하여 트리 구조의 레벨 1의 노드부터 최대 트리 높이(Height)에 해당하는 노드까지 리콜 과정을 단계별로 실시하여 분류 성능을 측정하였다.

15개의 클래스로 구성된 데이터로 학습한 후 생성된 트리 구조의 예는 그림 7과 같다. 그림 7에서 원 안의 숫자는 클래스 인덱스, 팔호 안의 숫자는 P.C.E에 의해 선택된 차원의 인덱스, ρ_{local} 은 P.C.E 과정 이후 하위 클래스 생성에 사용된 지역적 ρ 값을 의미한다.

15개의 클래스에 대해 임의의 순서로 입력 데이터 총 1,000개를 학습한 이후 학습에 사용되지 않은 데이터 5,000개를 이용하여 리콜 과정을 거쳤다. 리콜은 트리 구조의 레벨 1에 해당하는 노드들에서부터 트리의 높이(Height)인 레벨 4에 해당하는 노드로 순차적으로 실험하여 분류 성능을 측정 하였다. 또한 리콜 과정에서 각

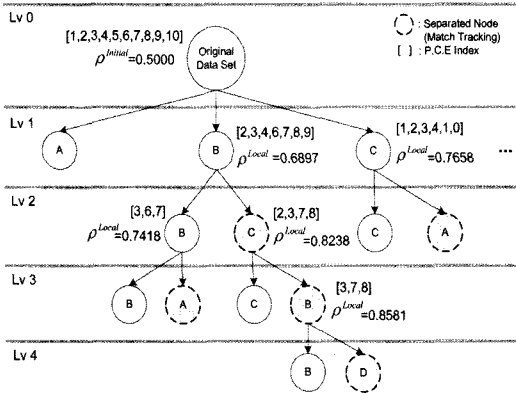


그림 7 계층적 Gaussian ARTMAP의 학습 결과

계층적 레벨별로 입력 데이터에 대한 승자 노드를 선택하기 위해 형제노드에 접근하는 평균 횟수(Node Hit)를 측정하였다. Node Hit는 식 (14)와 같다. 식 (14)에서 N 은 입력 데이터의 수, L 은 트리 접근 레벨, J 는 레벨 $l-1$ 에서 선택된 승자 노드의 연결된 자식 노드(Child Node)의 수를 의미한다.

$$node\ hit = \frac{1}{N} \sum_{i=1}^L J_{i-1} \quad (14)$$

실험 결과는 표 1과 같다. 표에서 Hit Rate는 제안한 모델에 리콜 데이터를 입력하여 출력된 클래스 결과와 데이터의 교사 신호가 동일한 패턴의 수를 비율로 나타낸 것이다. Node Hit는 식 (14)을 이용해 계산한 각 레벨별 평균 노드 접근 횟수를 의미한다. 그리고 Time은 각 레벨에서 총 리콜에 걸린 시간을 측정한 것이다.

표 1 제안한 모델의 레벨별 분류 성능

	Hit Rate	Node Hit	Time
Level 1	57.75%	135	7.6s
Level 2	89.52%	140.07	39.49s
Level 3	90.42%	140.27	41.73s
Level 4	90.42%	140.33	41.92s

실험 결과 리콜 과정에서 접근 레벨이 증가 할수록 분류 성능이 좋아지는 것을 확인할 수 있다. 레벨1과 레벨 2사이에 분류성능이 급격히 증가하는 것은 레벨2로 하위 클래스를 형성하는 과정에서 많은 노드들이 생성되었기 때문이다. 그리고 레벨3에서 레벨4로 진행되면서 새롭게 생성된 클래스가 4개만 존재하기 때문에 분류율이 향상되지는 않았다. Node Hit는 접근 레벨이 증가함에도 불구하고 크게 증가하지 않고 있다. 이는 계층적 구성에 의한 분류를 실행할 때 필요한 노드에 접근하는 시간보다 해당 레벨의 노드 정보를 바탕으로 새로운 네트워크를

구성하는데 많은 시간이 걸리는 것을 의미한다.

4.2 기존 패턴 인식기와의 비교 실험

본 논문에서 제안한 알고리즘의 분류 성능을 비교하기 위해서 학습 데이터 개수 증가별 분류 성능 비교를 위한 실험과 실제 패턴 인식 알고리즘의 응용 가능성을 측정하기 위해 [10]의 데이터를 이용하여 다른 패턴 인식 알고리즘과 분류성능을 비교하였다.

4.2.1 Gaussian ARTMAP과 성능비교

제안한 알고리즘과 Gaussian ARTMAP 모델의 성능을 비교하였다. Gaussian ARTMAP 모델의 파라미터는 제안한 모델의 최종 레벨에서 평균 Node Hit와 유사한 Node Hit를 가지도록 $\rho = 0.8$ 로 설정하였다. 실험에 사용한 데이터는 계층적 분류 특성을 보이기 위해 생성했던 15개의 클래스 부류를 가지는 10차원 데이터를 사용하였다. 실험은 각 패턴 인식 모델에 대해서 학습 데이터의 개수를 100개에서부터 2,000개 까지 100개씩 증가시키며 진행하였다. 실험 결과는 그림 8과 같다.

실험 결과 제안한 모델은 학습 패턴의 개수가 증가함에 따라 분류율이 비슷한 Node Hit를 가지는 Gaussian

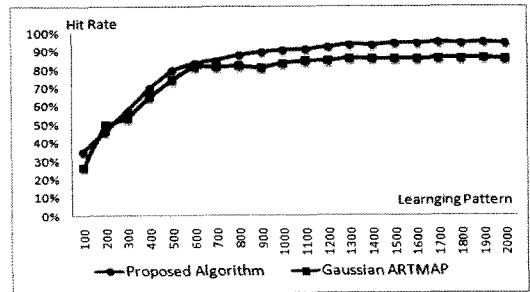


그림 8 제안한 모델과 Gaussian ARTMAP의 분류율

ARTMAP모델 보다 안정적으로 증가하였다. 학습 데이터 개수의 변화에 따른 평균 분류 성능은 제안한 모델이 Gaussian ARTMAP 모델보다 약 6.45% 향상되었다.

4.2.2 기존 패턴 인식 모델과 분류 성능 비교

제안한 알고리즘과 다른 패턴 인식 모델과 분류 성능 비교를 위해 실제 패턴 인식에 사용되는 데이터 중에서 UC Irvine Machine Learning Repository에 기증된 Glass Data, Breast Cancer Data, Letter Recognition Data[10]를 선택하였다. 각 데이터 집합의 특성은 표 2와 같다. 표 2에서 차원별 평균 분산은 각 데이터 집합의 차원별 데이터를 정규화 하여 평균 분산을 계산하였다.

Glass Data는 1987년 B. German에 의해서 생성된 데이터 집합으로 7종류의 유리에 대해서 주요 9개 원소의 함량 비율을 특징으로 사용하고 있다. 데이터에 포함되어 있는 7종류의 유리는 평면 처리 빌딩용 유리, 일반

표 2 실험에 사용된 데이터

데이터특성	Data Set		
	Glass Data	Breast Cancer	Letter
클래스	7	2	26
차원	9	30	16
평균 분산	0.00012	0.00083	0.02386
학습 데이터	214	463	4,500
리콜 데이터	214	569	9,000

빌딩용 유리, 평면 처리 자동차 유리, 일반 자동차 유리, 그릇용기, 식탁 용기, 헤드 램프로 구성되어 있다. 데이터 집합은 총 214개의 훈련 데이터로 구성 되어있다.

Breast Cancer Data는 Wisconsin 대학교의 Cancer Research Center에서 이미지 프로세싱 패키지의 실험을 위해 생성되었으며 데이터는 유방 세포핵의 영상 사진을 전처리 과정을 거친 후 특징을 추출하여 생성하였다. 데이터의 클래스는 각 세포핵에 대해서 양성과 음성 2가지로 구분 되어 있다. 데이터의 특징은 30차원으로 반지름, 조직, 둘레, 면적, 곡률, 밀집도, 함몰도, 함몰점, 대칭성, 프랙탈 차원에 대한 평균, 오차, 최대값으로 구성되어있다. 훈련 데이터는 569개로 이루어져 있다.

Letter Data는 바이너리 이미지 기반의 20개의 다른 영어 폰트에 대해서 특징을 추출하여 생성된 데이터이다. 데이터 구성은 알파벳 A부터 Z까지 총 26개의 클래스로 구성되어 있다. 각 알파벳에 대해서 수평 위치, 수직 위치, 가로 길이, 세로 길이, On 픽셀의 개수, 모든 On 픽셀의 평균 가로 길이, 모든 On 픽셀의 평균 세로길이, 세로 픽셀의 제곱 평균(mean squared), 가로 픽셀의 제곱 평균, 가로 픽셀과 세로 픽셀의 평균 곱(mean product), 가로 픽셀 평균의 제곱(squared mean), 세로 픽셀 평균의 제곱, 경계 픽셀의 평균 값, 수직 경계의 평균 값, 가로 경계의 평균 값, 상하 대칭되는 픽셀의 평균값의 총 16 차원의 특징 벡터로 구성되어 있다.

표 3은 제안한 모델과 기존의 패턴 인식 모델의 분류 성능을 비교한 것이다. 실험에 사용한 학습 데이터와 테스트 데이터의 수는 각각의 모델을 제안한 논문 [11], [12], [13]과 동일하게 사용하였으며, Boosted NN, Neive k-nn, BP#1, BP#2, Ada Boost, Assemble 모델의 분류 성능은 각각의 논문에서 제시된 분류 성능이다. Gaussian ARTMAP과 제안한 모델의 경계 파라미터는 4.2.1의 실험과 동일한 방법으로 결정하였다.

Glass data의 경우, 제안한 알고리즘이 Boosted NN 이나 Neive k-nn 알고리즘보다 약 9%이상 높은 분류 성능을 가지고 있었고, Gaussian ARTMAP 보다 2.34% 높은 분류 성능을 가지는 것을 확인하였다.

Breast Cancer Data의 경우, 제안한 알고리즘이

표 3 실제 데이터에 대한 모델별 분류율

모델명	Glass Data	Breast Cancer	Letter
Boosted NN	75.60%	N/A	N/A
Neive k-nn	73.20%	N/A	N/A
BP#1	N/A	93.40%	N/A
BP#2	N/A	94.00%	N/A
Ada Boost	N/A	N/A	89.89%
Assemble	N/A	N/A	90.50%
G-ARTMAP	80.84%	82.95%	88.83%
Proposed	83.18%	94.02%	91.04%

* N/A (not available) : 참고 논문상에 비교 데이터 없음

Back Propagation 알고리즘보다 약 0.3% 높은 분류 성능을 보였다. Glass Data의 분류 성능 증가에 비해 낮은 증가율을 보이는 이유는 Breast Cancer Data Set의 경우 각 차원별 분산이 작기 때문이다. 제안한 알고리즘에서는 하위 노드를 생성하고 오분류된 데이터를 분리하여 학습 과정에서 공통 속성을 제거하게 된다. 하지만 Breast Cancer Data의 경우 차원별 분산이 작으므로 공통 속성이 많지 않기 때문에 분류 성능의 증가폭이 작게 나타난다.

Letter Data의 경우, Ada Boost보다 약 1.15%, Assemble보다 약 0.54% 분류 성능이 향상된 것을 확인할 수 있다. 또한 Gaussian ARTMAP의 분류 성능 보다 2.22% 높은 분류 성능을 가지고 있음을 확인할 수 있다.

5. 결론

본 논문에서는 효율적인 패턴 인식기의 특성 중에서 온라인 학습, 학습 중 클래스 추가의 용이성, 계층적 분류가 가능한 특성을 가지는 패턴 인식기를 제안하였다. 제안한 패턴 인식기는 기존의 신경망 모델 기반의 패턴 인식기인 Gaussian ARTMAP을 계층적으로 구성하였다. Gaussian ARTMAP을 계층적으로 구성함으로써 기존의 모델에서 서로 다른 클래스에 속하는 데이터들의 공통 속성이 많은 경우 쉽게 분류하지 못하던 데이터들에 대해서도 분류를 할 수 있다는 장점이 있다. 계층적 Gaussian ARTMAP 모델의 하위 클래스 생성을 위해서 입력 데이터의 차원 중에서 중요한 차원만 선택하는 P.C.E 방법을 제안하였다.

제안한 모델의 동작 특징을 확인하기 위한 실험을 하였다. 실험결과 제안한 모델이 온라인 학습, 학습 중 클래스 추가, 계층적 분류 등의 특징을 가지고 있음을 확인하였다. 또한 계층적 Gaussian ARTMAP 모델과 Gaussian ARTMAP과의 분류 성능을 비교하기 위한 실험을 하였다. 실험 결과 제안한 모델의 분류 성능이 동일한 Node Hit를 가지는 Gaussian ARTMAP 모델

보다 약 6.45% 향상됨을 확인하였다. 또한 실제 데이터에 대한 다른 패턴인식 알고리즘과의 분류 성능 비교를 위한 실험에서도 제안한 알고리즘이 높은 분류율을 가지는 것을 확인하였다.

향후 연구로는 PCA(주성분분석)에 기반한 P.C.E를 이용한 입력 데이터의 차원을 감소하는 방법 대신 Fisher Discriminant Analysis에 기반하여 입력 차원을 감소하고 그에 적합한 매트릭을 생성하여 분류 성능을 향상시키는 연구가 필요하다.

참고 문헌

- [1] Carpenter, G.A, Grossberg S., "ARTMAP: A self organizing neural network architecture for fast supervised learning and pattern recognition," *Proceedings of the International Joint Conference on Neural Networks (IJCNN 91)*, pp.565-588, 1991.
- [2] Carpenter, G.A., Grossberg, S., and Rosen, D.B. "Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system," *Neural Networks*, vol.4, pp.759-771, 1991.
- [3] Williamson, J. R., "Gaussian ARTMAP: A Neural Network for Fast Incremental Learning of Noisy Multidimensional Maps," *neural networks*, vol.9, no.5, pp.881-897, 1996.
- [4] Xu, H., Vuskovic, M., "Mahalanobis distance-based ARTMAP network," *Neural Networks Proceedings 2004*, vol.3, pp.2353-2359, 2004.
- [5] Dagher, I., "Art networks with geometrical distances," *Journal of Discrete Algorithms archive*, vol.4, Issue.4, pp.538-553, 2006.
- [6] Bartfai, G., "An ART-based Modular Architecture for Learning Hierarchical Clusterings," *Journal of Neurocomputing* on volume 13, pp.31-45, 1996.
- [7] Fraley, C., "Algorithms for model-based Gaussian Hierarchical Clustering," *Society for industrial and applied mathematics, SIAM J. SCI. COMPUT.* vol.20, no.1. pp.270-281, 1998.
- [8] Dasgupta, S., Long, P.M., "Performance guarantees for hierarchical clustering," *Journal of Computer and System Sciences* archive Volume 70, Issue 4, pp.555-569, 2005.
- [9] 김승석, 광근창, "계층적 클러스터링과 Gaussian Mixture model을 이용한 뉴로 퍼지 모델링", *퍼지 및 지능 시스템* vol.13, no.5, pp.512-519, 2003.
- [10] UC Irvine Machine Learning Repository, <http://archive.ics.uci.edu/ml/index.html>, Website
- [11] Athitsos, V., Sclaroff, S., "Boosting nearest neighbor classifiers for multiclass recognition," *Computer Vision and Pattern Recognition, 2005 IEEE Computer Society Conference on 25-25 June 2005* pp.45-45, 2005.

- [12] Liu, H., Motoda, H., Dash, M., "A monotonic measure for optimal feature selection," In *Proceedings of European Conference on Machine Learning* Volume Year 1398/1998, pp.101-106, 1998.
- [13] Bennett, K. P., Demiriz, A., Maclin, R., "Exploiting unlabeled data in ensemble methods," *International Conference on Knowledge Discovery and Data Mining archive Proceedings of the eighth ACM SIGKDD*, pp.289-296, 2002.



최 대 훈

2007년 경희대학교 컴퓨터공학과(공학사). 2009년 경희대학교 컴퓨터공학과(공학석사). 2009년 현재 현대엔지니어링(주) 기술혁신 개발실 프로젝트IT부. 관심분야는 신경망, 패턴인식



임 성 길

1997년 경희대학교 수학과(이학사). 1999년 경희대학교 전자계산공학과(공학석사). 1999년~현재 경희대학교 컴퓨터공학과 박사과정. 관심분야는 패턴인식, 신경망, ASA(Auditory Scene Analysis)



이 현 수

1979년 경희대학교 전자공학과(공학사). 1982년 일본 게이오대학원 전기공학과(공학석사). 1985년 일본 게이오대학원 전기공학과(공학박사). 1999년~2000년 미국 오레곤 주립대학교 전기 및 컴퓨터공학과 방문연구원, 미국 캘리포니아대학교(U.C.I) 전기 및 컴퓨터공학과 방문연구원. 2005년~2008년 경희대학교 전자정보대학 학장 겸 정보통신대학원 원장. 1985년~현재 경희대학교 컴퓨터공학과 교수. 관심분야는 컴퓨터구조 및 VLSI, 병렬처리, 패턴인식, 신경망, 음성처리