

수정된 FMM 을 이용한 특징 선정 기법

박현정, 정경훈, 김호준
한동대학교 전산전자공학부
e-mail : hjkim@handong.edu

A Feature Selection Technique Using a Modified FMM Neural Network

Hyun Jung Park, Kyeong Hoon Jung, Ho Joon Kim
School of Computer Science & Electronic Engineering, Handong Global University

요 약

본 논문에서는 FMM 신경망의 활성화 특성에 가중치 개념을 도입한 패턴 분류 모형을 소개하고 이에 대한 학습 기법을 제안한다. 또한 제안된 모델의 활용으로서 주어진 학습패턴에 대하여 효과적인 특징의 종류와 특징과 패턴 클래스간의 상대적 연관도를 분석하는 방법론을 제시한다. 이를 위하여 새롭게 정의된 하이퍼박스 생성, 확장, 축소의 방법론을 소개하며, 이들 이론에 대하여 의료진단 데이터 등을 사용한 실제 실험을 통하여 유용성을 고찰한다.

1. 서론

패턴 분류 문제에서 특징의 종류에 따라 상대적 중요도를 판별하는 일과 개별 특징값이 특정 클래스에 주는 영향을 분석하는 일은 분류 성능을 향상시키는 측면에서 매우 중요한 의미를 갖는다. 이를 위하여 많은 이론과 구현을 위한 방법론이 소개된 바 있는데 최근의 연구에서는 인공 신경망을 사용하여 패턴 분류의 성능을 향상시키는 연구와 분류에 관한 명시적 지식표현을 위한 규칙생성 기법 등이 연구되고 있다 [1-2]. 퍼지 최대 최소(Fuzzy Min-Max(FMM)) 신경망은 1992 년 Simpson 등에 의해 제안된 패턴분류 모형으로 간결하면서도 강력한 학습기능을 지원한다[3]. 그 후 Gabrys 등은 FMM 신경망의 내부 연산과 데이터의 표현형태를 확장하여 일반화된 모델을 제안하였다[4]. 이들 방법에서는 입력 패턴이 패턴 공간내의 퍼지집합 또는 보통집합(crisp set)의 형태로 정의되는 영역, 이른바 하이퍼박스(hyperbox)로서 정의된다.

본 연구에서는 기본적으로 FMM 신경망에 가중치 개념을 적용하여 확장된 모델을 고려한다. 하이퍼 퍼지 집합 개념을 이용한 FMM 신경망은 내부적으로 최소점과 최대점으로 정의되는 퍼지 구간형태의 멤버쉽 함수를 사용하여 다차원 패턴 공간의 한 영역을 하이퍼박스로 정의한다. 제안된 모델의 활성화 특성은 가중치 개념을 적용함으로써 학습데이터 집합으로부터 특징값의 빈도에 따른 상대적인 학습효과의 변화를

반영할 수 있다. 본 논문에서는 이러한 수정된 FMM 모델의 동작특성을 소개하고 이를 패턴분류를 위한 특징분석 과정에 활용하는 방법론을 제안한다.

2. 배경연구

(1) FMM 신경망

FMM 신경망은 하이퍼박스 퍼지 집합으로 구성되는 패턴분류 모델이다. 하이퍼박스로 모든 특징의 종류에 대하여 완전 클래스 멤버쉽을 가진 패턴들로 이루어진 n 차원 패턴 공간 영역이다. 최소값과 최대값의 범위와 그 가장자리에서 퍼지 멤버쉽을 갖는 퍼지 구간(fuzzy interval)들로서 이루어진 하이퍼박스와 그에 상응하는 멤버쉽 함수는 특징값의 집합과 패턴클래스간의 사상을 지원한다. FMM 신경망 학습은 패턴 공간 내에서 하이퍼박스의 생성(creation), 확장(expansion), 축소(contraction) 등의 과정으로 이루어진다. 임의의 하이퍼박스를 위한 멤버쉽 함수는 아래 수식 (1)과 같다.

$$b_i(A_h) = \frac{1}{2^n} \sum_{i=1}^n [\max(0, 1 - \max(0, \gamma \min(1, a_{hi} - v_{ji}))) + \max(0, 1 - \max(0, \gamma \min(1, u_{ji} - a_{hi})))] \quad (1)$$

식 (1)에서 $A_h = (a_{h1}, a_{h2}, a_{h3}, \dots, a_{hn}) \in I^n$ 은 h 번째 입력 패턴을 의미하며 $u_j = (u_{j1}, u_{j2}, u_{j3}, \dots, u_{jn})$ 은 하이

퍼박스 j 에 대하여 각 특징에 대한 최소값이며, $v_j = (v_{j1}, v_{j2}, v_{j3}, \dots, v_{jn})$ 은 최대값을 의미한다. n 은 특징의 개수이며 γ 는 퍼지 멤버십 함수의 기울기를 결정하는 파라미터로 $[0, 1]$ 사이의 값을 갖는다.

학습과정에 의해 각 하이퍼박스의 특징범위가 정해지면, 테스트 패턴은 식 (1)을 이용하여 계산된 값 중 가장 높은 값의 하이퍼박스를 찾고 이에 해당되는 클래스로 분류하게 된다. 이 때 각 특징값에 대한 멤버십의 퍼지 특성은 구간 가장자리에서 기울기 파라미터 γ 에 의해 결정된다.

(2) WFMM 신경망

본 연구에서는 가중치 개념이 추가된 FMM 신경망, 소위 WFMM(Weighted Fuzzy Min-Max) 신경망을 제안하고 이의 활용기법을 고찰한다. 즉 식 (2) 와 같은 멤버십 함수로서 하이퍼박스 특성을 새롭게 정의한다.

$$b_j(A_{hi}) = \frac{1}{\sum_{i=1}^n w_{ji}} \cdot \sum_{i=1}^n w_{ji} [\max(0, 1 - \max(0, \gamma \min(1, a_{hi} - v_{ji}))) + \max(0, 1 - \max(0, \gamma \min(1, u_{ji} - a_{hi}))) - 1.0] \quad (2)$$

식에서 w_{ji} 는 j 번째 하이퍼박스와 i 번째 특징사이의 연결 가중치를 의미하며, 나머지 표기는 수식 (1)의 경우와 동일하다. 이는 기존 FMM 신경망의 특성에 비하여 두 가지 요소가 수정된 것이다. 하나는 가중치 요소가 각 특징값에 반영된다는 점이고 두 번째는 각 하이퍼박스의 출력값 산출 방식을 퍼지범위의 최소조건 항과 최대조건 항의 보수(complement) 합계에서 1.0 을 빼줌으로써 출력값의 범위가 $[0, 1]$ 사이가 되도록 조정하였다. 그러므로 평균값 산출을 위한 $(1/2n)$ 계수는 멤버십 값을 표준화 하는 총 가중치합으로 대체하였다.

3. 학습 기법

FMM 신경망의 학습과정은 각 단계에서 하이퍼박스의 생성, 확장, 축소의 과정을 실행해 나가게 된다. 본 논문에서는 기존의 이러한 3 가지 과정에 대하여 가중치 요소의 특성을 반영하도록 변경하였다.

각 학습단계에서, 임의의 하이퍼박스 j 에서 임의의 특징 i 에 대한 가중치 조정은 식 (3)과 (4)의 형태로 이루어진다.

$$w_{ji}^{new} = w_{ji}^{old} + \Delta w_{ji} \quad (3)$$

$$\Delta w_{ji} = \begin{cases} \lambda & \text{if}(v_{-u}^{new} - u_{-u}^{new} \leq s) \\ d \cdot (T - \frac{v_{-u}^{new} - u_{-u}^{new}}{v_{-u}^{old} - u_{-u}^{old}}) & \text{elseif}(\frac{v_{-u}^{old} - u_{-u}^{old}}{v_{-u}^{old} - u_{-u}^{old}} \leq T) \\ MAX_{w_{ji}^{old}} \cdot (\frac{v_{-u}^{old} - u_{-u}^{old}}{v_{-u}^{new} - u_{-u}^{new}} - 1.0), \frac{-w_{ji}^{old}}{2} & \text{Otherwise} \end{cases} \quad (4)$$

다시 말해서 이러한 학습과정은 하나의 하이퍼박스에 속한 두 개 이상의 패턴이 특징 공간의 같은 범위에

에 나타날 때, 가중치가 증가함을 의미하는데 이 때 상대적인 거리와 빈도를 반영한다는 것이다. 식 (3)과 (4)에 의하면, 새롭게 생성된 하이퍼박스의 최소점과 최대점이 일정값 이내의 작은 범위이거나 이전 하이퍼박스에 비해 크기가 일정크기 이내로 확장된 경우 가중치가 높아지고, 하이퍼박스의 크기가 일정크기 이상으로 확장된 경우 가중치가 감소함을 알 수 있다. 따라서 이 수식의 매개변수 s 는 하이퍼박스의 바람직한 특징 범위의 기준 크기이다.

식 (4)는 가중치 변화분을 나타낸 식으로, 세가지의 경우로 구분하여 조정됨을 보인다. 우선 새롭게 구성되는 특징범위가 일정범위 이내, 즉 가장 이상적인 작은 범위 이내이면 일정 상수값을 증가시킨다. 반면 일정범위 보다 크지만 새롭게 생성된 범위가 이전보다 작거나, 일정비율 이내에서 증가한 경우라면 증가분에 반비례하여 결정한다. 그러므로 매개변수 T 는 1 보다 큰 1 근처의 값이다. 세 번째로 생성된 범위가 지나치게 커진 경우이다. 이는 식에 보인 바와 같이 커진 비율에 비례하여 가중치 값을 감소시킨다. 파라미터 λ 와 d 는 학습률 상수인데 특징영역이 확장함에 따라 해당 특징에 대한 가중치가 얼마나 급격히 증가시킬 것인가를 결정하는 요소이다.

각 가중치 값은 하이퍼박스가 생성되는 시점에 초기값 1.0 으로 설정된다. 식 (2)에서 가중치 값이 항상 1.0 이라면 이는 기존 FMM 의 특성과 유사해 지는데, 이는 제안된 모델이 기존 FMM 모델의 특성을 좀 더 일반화한 개념임을 의미한다.

그림 1 은 FMM 모델과 제안된 모델의 학습과정에서 하이퍼박스에 대한 확장과정을 비교하여 보인 것이다. 제안된 모델의 하이퍼박스 확장과정은 다음과 같은 식으로서 이루어진다.

$$n\theta \geq \sum_{i=1}^n (\max(v_{ji}, x_{hi}) - \min(u_{ji}, x_{hi}))$$

이면

$(x_{hi} < u^{old})$ 의 경우

$$u_{ji}^{new} = u^{old} - \frac{1}{w_{ji}}(u^{old} - x_{hi})$$

$(x_{hi} > v^{old})$ 의 경우

$$v_{ji}^{new} = v^{old} + \frac{1}{w_{ji}}(x_{hi} - v^{old})$$

제안된 모델의 핵심은 기존의 FMM 모델에 비해 다음과 같이 두 가지 장점을 기대할 수 있게 한다는 것이다.

- 학습패턴에 포함될 수 있는 노이즈나 비정상적인 데이터에 의한 영향을 줄일 수 있다. 데이터 집합내에서 상대적으로 빈도가 적은 패턴에 대해 하이퍼박스 영역의 확장을 점진적으로 이루어지게 하고, 가중치 값을 재조정하여 이러한 비정상 패턴에 민감하게 학습되는 현상을 보완한다.
- 두 번째로 특징에 대한 분석기능을 지원한다는 점이다. 가중치 값으로부터 각 특징과 클래스간의 관계를 해석해 낼 수 있으며, 이는 특징종류에 대한

유용성과 특징값에 대하여 하이퍼박스에 대한 상대적인 연관도를 서로 다른 값으로 판별해 낼 수 있게 한다.

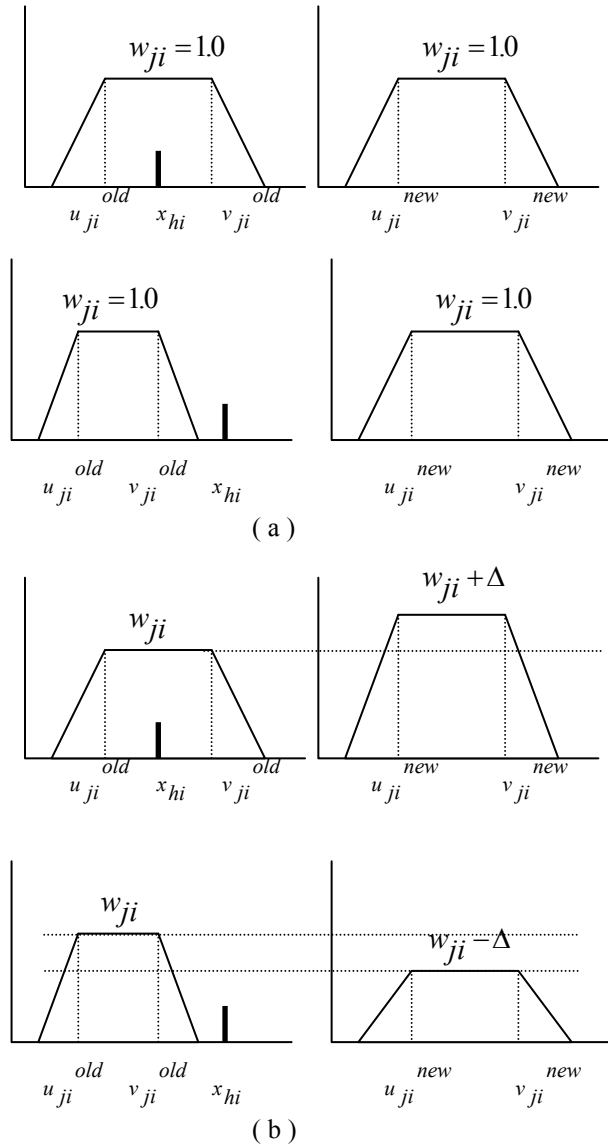


그림 1) (a) FMM 모델, (b) WFMM 모델

4. 특징 분석 기법

본 절에서는 학습된 신경망으로부터 특정 패턴 클래스의 각 특징의 상호 연관성 요소를 판별하는 방법을 소개한다. 패턴 분류 응용의 측면에서 볼 때, 이는 두 가지 점에서 유용한 기법이다. 우선 주어진 문제에서 어떤 특징 종류가 전체적인 분류 과정에 효과적으로 활용될 수 있는가 하는 점을 지원한다. 두 번째로 어떤 특징값이 특정 클래스에 가장 중요한 값으로 연관되어 있는가를 판별할 수 있다는 점이다.

이를 위하여 본 연구에서는 각 특징과 패턴 클래스에 대하여 연관도(RF : Relevance Factor)를 다음과 같이 정의한다.

각 하이퍼박스 퍼지집합 $B_j = \{X, U_j, V_j, f(X, U_j, V_j)\}$, $\forall X \in I^n$ 이라고 정의하자. 그러면 하이퍼박스 퍼지집합의 정의를 사용하여 C 번째 패턴 클래스 C_k 를 다

음과 같이 정의할 수 있다.

$$C_k = \bigcup_{j \in K} B_j$$

K 는 클래스 k 와 관련된 하이퍼박스의 색인 집합이다. $U_j = (u_{j1}, u_{j2}, u_{j3}, \dots, u_{jn})$ 는 B_j 의 최소점의 집합이고 $V_j = (v_{j1}, v_{j2}, v_{j3}, \dots, v_{jn})$ 은 B_j 의 최대점의 집합이다.

학습된 신경망으로부터 주어진 클래스의 연관성 특징 목록을 아래와 같이 산출한다. 즉 클래스 K 와 관련된 특징표시 f_i 의 연관도(RF)는 식 (8)과 같이 정의된다.

$$RF(f_i, k) = \left(\frac{1}{N_k} \sum_{B_j \in C_k} S(f_i, (u_{ji}, v_{ji})) \cdot w_{ji} \right) - \frac{1}{(N_B - N_k) \sum_{B_j \notin C_k} S(f_i, (u_{ji}, v_{ji})) \cdot w_{ji}} / \sum_{B_j \in C_k} w_{ji} \quad (8)$$

식에서 상수 N_B 는 총 하이퍼박스의 개수이며 N_k 는 클래스 k 에 속한 하이퍼박스의 개수이다. 식 (8)에서 특징 f_i 는 최종적으로 생성된 하이퍼박스에 나타나는 특징범위이다. 그러므로 이는 최소값 및 최대값 쌍으로 이루어진 퍼지 구간의 형식으로 나타난다. 임의의 특징 f_i 의 최대값을 f_i^L , 최소값을 f_i^U 로 정의하고, 두 퍼지 구간의 유사도(similarity) S 를 식 (9)와 같이 정의한다.

$$S(f_i, (u_i, v_i)) = S((f_i^L, f_i^U), (u_i, v_i)) = \frac{\text{Overlap}((f_i^L, f_i^U), (u_i, v_i))}{\text{Max}(f_i^U - f_i^L, v_i - u_i)} \quad (9)$$

식 (9)에서 두 퍼지구간이 모두 한 점으로 이루어진 데이터라면 분모 $\text{Max}(f_i^U - f_i^L, v_i - u_i)$ 는 0 이 되는데 이 경우는 별도로 고려하여 식 (10)을 적용한다. 즉 두 구간이 동일한 한 점일 경우 유사도 값은 1.0 이며, 서로 다른 두 점일 경우 0 이 된다.

$$S((f_i^L, f_i^U), (u_i, v_i)) = \begin{cases} 1 & \text{if } (f_i^L = f_i^U = u_i = v_i) \\ 0 & \text{Otherwise} \end{cases} \quad (10)$$

만약 $\text{Max}(f_i^U - f_i^L, v_i - u_i)$ 가 0 보다 크다면 값은 식 (11)를 적용하여 산출된다.

$$\text{Overlap}((f_i^L, f_i^U), (u_i, v_i)) = \begin{cases} f_i^U - u_i & \text{if } (f_i^L \leq u_i \leq f_i^U \leq v_i) \\ v_i - u_i & \text{if } (f_i^L \leq u_i \leq v_i \leq f_i^U) \\ f_i^U - f_i^L & \text{if } (u_i \leq f_i^L \leq f_i^U \leq v_i) \\ v_i - f_i^L & \text{if } (u_i \leq f_i^L \leq v_i \leq f_i^U) \\ 0 & \text{Otherwise} \end{cases} \quad (11)$$

만약 $RF(f_i, k)$ 가 양의 값을 가지면 이것은 클래스 k 와 특징 f_i 사이의 자극성(excitatory) 연관성이 있음을 의미한다. 그러나 $RF(f_i, k)$ 가 음의 값을 가지면 그들 사이는 억제성(inhibitory) 연관성을 가진다는 것

을 의미한다. 또한 주어진 클래스에 대하여 중요한 역할을 하는 흥미있는 특징들의 목록을, 계산된 연관성 요소 값을 사용하여 선정할 수 있다. 이는 각 특징의 종류에 대하여 모든 값과 모든 하이퍼박스에서 연관성 요소의 값을 구분하여 합산하고 이를 정렬하여 출력함으로써 이루어 질 수 있다.

5. 실험 결과 및 고찰

실험은 패턴 분류 문제에서 잘 알려진 아이리스 데이터와 의료진단 데이터[5]를 사용하여 결과를 고찰하였다. 아이리스 데이터는 총 3 개의 클래스에 각각 50 개의 패턴으로 구성되며 각 패턴은 4 종류의 특징으로 이루어진다. 제안된 모델을 사용한 특징 분석 실험결과를 표 1에 보였다. 이 실험에서 파라미터 값은 여러 번의 튜닝을 통하여 $\theta = 0.3$, $\gamma = 0.5$, $T = 1.2$, $d = 0.1$, $s = 0.05$ 로 조정하였다. 표 1에 보인 바와 같이 각 패턴 클래스별로 각 특징들의 RF 값을 산출할 수 있다.

표 1 아이리스 데이터를 사용한 특징 분석 실험 결과

패턴클래스	RF와 가중치를 가진 연관성특징 목록
Setosa	F4:(0, 0.125), RF=0.038, W=1.810, F3:(0.016, 0.118), RF=0.033, W=1.909, F1:(0.027, 0.222), RF=0.019, W=1.683, F2:(0.375, 0.833), RF=0.006, W=1.468
Versicolor	F3:(0.508, 0.652), RF=0.023, W=1.465, F4:(0.5, 0.687), RF=0.020, W=1.516, F2:(0.125, 0.541), RF=0.005, W=2.128, F1:(0.333, 0.75), RF=0.003, W=1.31
Virginica	F3:(0.658, 0.779), RF=0.030, W=1.981, F4:(0.75, 0.916), RF=0.028, W=1.473, F2:(0.208, 0.666), RF=0.004, W=1.619, F1:(0.416, 0.722), RF=0.006, W=1.667

표 2는 클리블랜드 의료진단 데이터에 대한 실험 결과이다. 학습 데이터는 13개의 특징을 가진 다섯개의 클래스로 총 패턴 수는 279개이다. 여기에 $\theta = 0.93$, $\gamma = 0.5$, $T = 1.01$, $d = 30.0$, $s = 0.05$ 을 사용하여 학습되었다.

표 2 클리블랜드 의료 데이터를 사용한 특징 분석 예

패턴클래스	RF와 가중치를 가진 연관성특징 목록
0	F12:(0, 0), RF=0.004, W=1.8, F9:(0, 0), RF=0.003, W=1.8, F3:(0.666, 0.666), RF=0.003, W=1.5, F2:(1.0, 1.0), RF=0.003, W=1.8
1	F2:(1.0, 1.0), RF=0.001, W=1.6, F1:(0.229, 0.604), RF=0.001, W=2.819, F12:(0, 0), RF=0.001, W=1.3, F6:(0, 0), RF=0.001, W=1.6
2	F3:(1.0, 1.0), RF=0.003, W=1.2, F6:(0, 0), RF=0.002, W=1.2, F8:(0.244, 0.541), RF=0.002, W=2.950, F10:(0.225, 0.451), RF=0.001, W=2.399

3	F9:(1.0, 1.0), RF=0.003, W=1.3, F6:(0, 0), RF=0.002, W=1.3, F2:(1.0, 1.0), RF=0.002, W=1.3, F12:(0, 1.0), RF=0.001, W=1.22
4	F2:(1.0, 1.0), RF=0.002, W=1.2, F3:(1.0, 1.0), RF=0.002, W=1.2, F5:(0.182, 0.641), RF=0.001, W=2.057, F13:(1.0, 1.0), RF=0.001, W=1.2

표 1과 2에서 보인 바와 같이 제안된 모델은 패턴 분류 문제에서 특정 클래스와 특징간의 상호 연관도를 분석해 낼 수 있게 한다. 이는 실제 응용에서 지식 생성 기법이나, 신경망 분류기의 학습효과 분석 등에 사용될 수 있다.

6. 결론

기존의 FMM 신경망은 신경망의 학습능력과 퍼지집합이론의 장점을 조화시킨 일종의 퍼지신경망 모델이다. 제안된 모델은 이러한 퍼지 신경망모델의 장점을 활용하면서 보다 확장된 기능을 지원하는 패턴 분류 기법이다. 그러므로 학습을 통한 지식 획득이 가능할 뿐만 아니라, 기존의 순수 신경망에 비해 인간과 같은 논리적인 능력과 전문지식을 결합할 수 있는 가능성을 지원한다. 제안된 신경망의 가중치 요소는 학습패턴에서 특징값의 빈도의 차이에서 오는 영향을 반영한다. 즉 잘 구성된 학습데이터는 각 패턴 클래스에 대하여 균등한 빈도로 나타남이 적당함을 전제로 한다면, 상대적으로 빈도가 적은 비정상 패턴이 분류 시스템의 성능에 영향을 주는 것은 바람직하지 못하다. 이에 본 연구에서 제안하는 모델은 가중치변화 과정에서 이러한 빈도 요소를 반영하게 함으로써 이러한 문제점을 개선하였다. 본 연구에서 제안된 가중치 요소와, 특징과 클래스의 연관도 정의 등은 다른 패턴 분류 방법론과 결합하여 지식분류 및 분석기능과, 분류과정에 대한 시스템 설명기능 등을 구현하는데 활용될 수 있다.

* 이 연구는 과학기술부 뇌과학 연구개발사업으로 수행되었음

참고문헌

[1] Raphael Feraud, Fabrice Clerot, "A Methodology to Explain Neural Network Classification," Neural Network, Vol.15, pp.237-246, 2002.
 [2] Jayanta Basak, Raat K. De, Sankar K. Pal, "Unsupervised Feature Selection using a Neuro-Fuzzy Approach," Pattern Recognition Letters, Vol.19, pp.997-1006, 1998.
 [3] Patrick K. Simpson, "Fuzzy Min-Max Neural Networks - Part 1: Classification," IEEE Transactions on Neural Networks, Vol.3, No.5, pp.776-786, 1992.
 [4] B. Gabrys, A. Bargiela, "General Fuzzy Min-Max Neural Network for Clustering and Classification," IEEE Transaction on Neural Networks, Vol.11, No.3, 2000.
 [5] C.L. Blake, C.J. Merz, UCI Repository of machine learning databases [http://www.ics.uci.edu/~mllearn/MLRepository.html]. University of California, Irvine, 1998.