

개선된 GFMM 신경망을 이용한 패턴 분류

이조셉, 박진희, 김호준
한동대학교 정보통신공학과

e-mail : joseph.sung.lee@gmail.com, ngu2000@nate.com, hjkim@handong.edu

An Improved General Fuzzy Min-Max Neural Network for Pattern Classification

Joseph S. Lee, Jin-Hee Park, Ho-Joon Kim
Dept. of Information Technology, Handong Global University

요 약

본 논문에서는 일반화된 퍼지 최대-최소 신경망 모델에서 학습 패턴의 빈도요소를 고려하여 개선된 활성화 함수와 학습 방법을 제안한다. 특징공간상에서 하이퍼박스의 활성화를 위한 새로운 기준과 방법을 제시하며, 학습 패턴의 빈도요소가 학습효과에 미치는 영향을 분석한다. 또한 제안된 모델에서 개별 특징값과 하이퍼박스간의 상대적인 연관성을 고려하여 이득치를 계산함으로써, 기존 모델의 하이퍼박스 축소 기법을 대체한 학습효과에 관하여 고찰한다. 실험을 통하여 학습 패턴의 순서 변화와 왜곡된 정보에 안정된 분류기의 성능을 확인한다.

1. 서론

패턴인식 기술은 지난 수십년간 귀납적 기계학습의 분야에서 여러 가지 형태로 발전되어 왔으며, 그 응용은 매우 다양하다. 현재는 음성과 문자를 대상으로 한 패턴인식 시스템이 이미 실용화되고 있으며, 보다 복잡한 패턴 형식을 지닌 화상 데이터간의 분류를 위해서도 수많은 연구가 활발히 진행 중에 있다.

패턴인식에서 사용되는 분류기는 주어진 다차원 특징공간을 결정공간으로 구분 짓기 위해 주어진 데이터에 대해서 학습하는 과정을 거치게 되는데, 이러한 학습 방법에는 크게 두 가지가 있다. 하나는 지도학습에 의한 패턴 분류이고 나머지 하나는 자율학습에 의한 패턴의 군집화이다.

지도학습에 의한 패턴 분류기로서 Simpson 은 하이퍼박스에 기반한 퍼지 최대-최소 (FMM: Fuzzy Min-Max) 신경망을 제안하였으며[1], 이 모델의 활성화 특성과 하이퍼박스의 생성 및 확장 기준을 새로 정의함으로써 자율학습을 위한 또 다른 하나의 FMM 모델을 만들었다[2]. 이 후에, Gabrys 와 Bargiela 는 Simpson 이 제안한 두 가지 모델에서 지도학습과 자율학습의 기능을 통합시켰고, FMM 신경망 모델의 몇 가지 기능을 일반화 시킴으로써 GFMM (General Fuzzy Min-Max) 신경망이라는 모델을 소개하였다[3].

하지만, GFMM 신경망의 기반이 되는 FMM 모델은 비정상적인 패턴이나 학습 패턴의 순서에 의해 왜곡된 하이퍼박스가 만들어질 수 있다는 단점을 지니고 있다. 따라서 FMM 모델을 바탕으로 만들어진 GFMM 신경망도 이와 같은 문제를 상속받고 있는 것이다. FMM 모델의 이러한 문제점을 보완하기 위해 [4]에서는 Simpson 의 모델에 가중치 개념을 적용함으로써

기존의 모델이 갖는 한계점을 개선한 바 있다.

본 연구에서는 [4]에서 사용되었던 빈도요소를 GFMM 신경망에도 적용함으로써 좀더 일반화된 모델을 제시한다. 또한, 학습 패턴의 빈도요소가 반영된 학습 방법의 효과를 분석함으로써, 기존의 모델에서 하이퍼박스 축소기법의 필요성을 대체하는 효과에 대하여 고찰한다. 세부적 실험 및 분석은 실제 데이터인 Iris 패턴 분류 문제를 가지고 데이터 및 특징집합을 선별적으로 적용하여 분석하였다.

본 논문의 구성은 다음과 같다. 2 절에서는 GFMM 신경망에 대한 배경 이론을 간략하게 설명하고, 3 절에서는 하이퍼박스 축소 과정이 분류기의 성능에 미치는 영향에 대해 알아본다. 4 절을 통해 제안된 개선점과 학습 기법을 설명하고, 5 절은 이를 바탕으로 한 학습 효과의 분석을 보여준다. 6 절에서 실험을 통해 분류 성능을 평가 한 후, 마지막 절에서는 본 연구에 대한 결론을 내린다.

2. GFMM 신경망

GFMM 신경망이 기존의 FMM 모델과 구별되는 특징은 네 가지로 요약 될 수 있다. 첫째, 신경망의 입력이 단일 패턴에서 퍼지 하이퍼박스도 받아들일 수 있다는 점이며, 둘째로는 활성화 함수가 퍼지집합 연산을 사용하는 형태로 변경되고 하이퍼박스의 확장 조건이 바뀌었다는 점이다. 셋째로, 이 신경망은 레이블된 데이터와 레이블 되지 않은 데이터로, 분류와 군집화 두 가지 기능을 동시에, 또는 개별적으로 수행할 수 있다는 점이다. 마지막 특징은 하이퍼박스의 최대 크기를 결정하는 매개변수가 적응적으로 변할 수 있다는 것이다. 여기서 분류기 성능에 가장 큰 영

향을 미치는 것은 결정 공간에 영향을 주는 두 번째 특징인 하이퍼박스 활성화 함수의 특성이 될 것이다. Simpson 이 정의한 하이퍼박스의 퍼지 집합은 GFMM 에서 다음과 같이 정의된다.

$$B_j = \{X_h, U_j, V_j, b_j(X_h, U_j, V_j)\} \quad (1)$$

위에서 $X_h = [X_h^l, X_h^u]$ 는 최소점과 최대점을 나타내는 벡터로 구성되며, 이는 h 번째 입력 패턴 집합을 의미한다. n 차원에 대해 $U_j = (u_{j1}, u_{j2}, \dots, u_{jn})$ 는 하이퍼박스 B_j 의 최소점을, $V_j = (v_{j1}, v_{j2}, \dots, v_{jn})$ 는 최대점을 나타낸다. 함수 b_j 는 GFMM 신경망의 활성화 특성으로서 아래의 식으로 표현된다.

$$b_j(X_h) = \min_{i=1 \dots n} (\min([1 - f(x_{hi}^u - v_{ji}, \gamma_i)], [1 - f(u_{ji} - x_{hi}^l, \gamma_i)]) \quad (2)$$

위의 소속 함수에서 두 개의 매개변수를 가진 함수 f 는 다음과 같이 퍼지구간의 범위를 결정한다.

$$f(r, \gamma) = \begin{cases} 1 & \text{if } r\gamma > 1 \\ r\gamma & \text{if } 0 \leq r\gamma \leq 1 \\ 0 & \text{if } r\gamma < 0 \end{cases} \quad (3)$$

여기서 매개변수 γ 는 하이퍼박스 영역의 가장자리에서 퍼지소속 함수의 기울기를 의미한다. GFMM 에서 하이퍼박스의 확장을 제한하는 조건 식은 다음과 같다.

$$\forall_{i=1 \dots n} (\max(v_{ji}, x_{hi}) - \min(u_{ji}, x_{hi})) \leq \theta$$

이 조건에서 θ 는 하이퍼박스의 최대 크기를 결정하는 매개변수로, 입력된 특징 값 x_{hi} 가 하이퍼박스와의 최소거리 θ 내에 있다면 해당 하이퍼박스를 확장시키고, 최소거리 밖에 있다면 새로운 하이퍼박스를 생성하게 된다.

이러한 GFMM 모델의 활성화 함수의 특징은 특징과 하이퍼박스 간의 상호 작용은 모두 대등한 비중으로만 고려된다는 점이다. 이러한 특징은 응용 문제에서 다음 절에서 설명할 두 가지의 경우, 바람직하지 않은 성능 저하를 야기 할 수 있다.

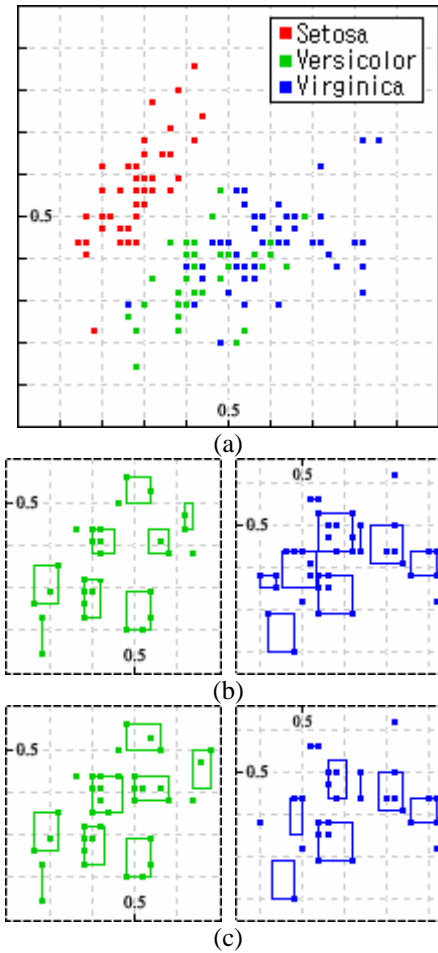
3. 하이퍼박스 축소의 영향

기존 모델은 하이퍼박스 생성과 확장 후 중첩영역이 발생하면 패턴에 대한 분류결과가 모호해지므로 축소 연산을 수행한다. 이 과정에서 응답의 모호성은 해결되나, 하이퍼박스 형태는 학습되는 패턴의 순서와 비정상적인 패턴에 의해 정보왜곡이 발생 한다.

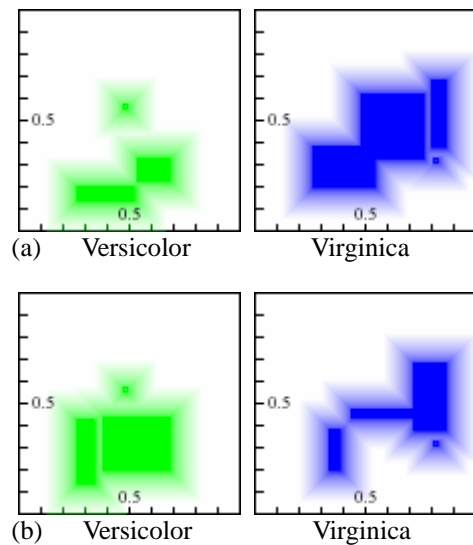
우선 축소 과정의 수행에 따른 결정공간은 학습 패턴의 입력 순서에 따라 크게 변할 수 있다. 이는 나중에 학습되는 패턴이 기존에 있던 다른 클래스의 하이퍼박스를 침범하는 현상을 가져오기 때문이다.

그림 1(a)와 같이 분포한 학습 패턴에서 Versicolor, Virginica 순으로 패턴을 학습시키면 각 클래스는 그

림 1(b)와 같은 하이퍼박스를 형성하게 된다. 반대로 Virginica, Versicolor 순으로 학습 시켰을 경우 결정공간은 그림 1(c)와 같이 된다.



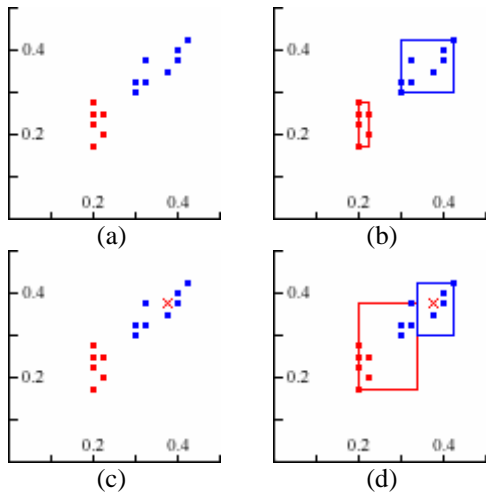
(그림 1) 학습 순서 변화에 의한 하이퍼박스 영역 ($\theta = 0.09$)



(그림 2) 학습 순서의 변화에 따른 결정공간의 변화 ($\gamma = 7.0, \theta = 0.30$)

같은 패턴에 대해서 나중에 학습 된 경우가 먼저 학

습 되었을 때보다 하이퍼박스가 차지하는 영역이 커진다는 것을 확인할 수 있다. 이 현상은 그림 1(a)처럼 클래스간의 경계가 명확하지 않거나 θ 의 값이 클 때 그림 2처럼 더 심각해진다. 동일한 학습 패턴에 대해 일정한 분류 결과가 나오는 것이 바람직하겠지만 하이퍼박스의 축소과정 때문에 기존 모델의 분류 신뢰도는 떨어지게 된다. 두 번째로는 학습 패턴의 빈도요소에 대한 문제이다. 주어진 학습 패턴 집합에서 대부분의 데이터가 신뢰할 수 있는 데이터로 잘 표본화 되어있다 하더라도 단 한 개의 왜곡된 데이터를 포함하고 있다면, 학습의 형태를 극단적으로 왜곡시킬 수 있다.



(그림 3) 왜곡된 학습 데이터에 의한 패턴 분류 ($\theta = 0.20$)

그림 3(a)와 같은 학습 패턴에 대해 하이퍼박스는 그림 3(b)와 같이 생성될 수 있다. 하지만, 만약 이 학습 패턴에 그림 3(c)와 같은 왜곡된 특징(x 표시)이 포함되어 있을 때 θ 가 0.20 이상인 경우 그림 3(d)와 같은 하이퍼박스가 생성될 것이다. 이 때 원래 하이퍼박스에 속해 있던 패턴의 반이 제외된 것을 볼 수 있다. 즉, 많은 수의 데이터에 의하여 잘 학습된 하이퍼박스의 형태가, 단 한 개의 비정상적인 데이터로 인하여 극단적으로 변형될 수 있는 것이다. 본 논문에서는 이러한 GFMM 모델이 갖는 특징의 영향을 개선할 수 있는 방법에 대하여 다음 절에서 고찰한다.

4. 빈도요소를 고려한 학습

앞 절에서 설명한 문제점을 개선하기 위해 빈도요소를 고려한 학습 방법과 이득치를 계산하여 문제를 해결하는 방법론을 고찰한다. 우선 기존의 식 (1)에서와는 달리 다음과 같이 정의된 하이퍼박스 B_j 를 고려한다.

$$B_j = \{X_h, U_j, V_j, f_j, b_j(X_h, U_j, V_j, f_j)\}$$

여기서 추가된 속성은 f_j 로서 이 값은 j 번째 하이퍼박

스에 대하여 학습에 사용된 패턴의 빈도를 의미한다. f_j 는 처음에 0으로 초기화 되어 있고 아래의 식에 의해 하이퍼박스가 생성되거나 확장 될 때, 또는 학습 패턴을 포함한 경우 해당 하이퍼박스에 대해서 계산된다.

$$f_j^{new} = f_j^{old} + 1$$

각 하이퍼박스의 이득치 값은 특징공간의 각 차원에 대하여 빈도요소를 고려하여 다음 식과 같이 산출된다.

$$g_{ji} = \frac{f_j}{N} \times (1 - R_{ji})$$

여기서 N 은 학습 패턴의 총 개수이며 $R_{ji} = v_{ji} - u_{ji}$ 로 i 번째 차원에서의 하이퍼박스의 폭이다. 즉, g_{ji} 는 j 번째 하이퍼박스의 차원 i 에 대한 밀도를 의미한다. 이러한 이득치는 특징의 빈도에 비례하고 발생 범위의 폭에 반비례 한 값으로 결정된다. 이득치는 하이퍼박스의 멤버십을 결정하는 식 (2)에 다음과 같이 적용 된다.

$$b_j(X_h) = \min_{i=1..n} (\min([1 - f(x_{hi}^u - v_{ji}, \gamma_i, g_{ji})], [1 - f(u_{ji} - x_{hi}^l, \gamma_i, g_{ji})]))$$

여기서 하이퍼박스 내의 영역에서는 각 차원에 대한 이득치를 반영할 수 있도록 식 (3)을 다음과 같이 수정한다.

$$f(r, \gamma, g) = \begin{cases} 1 & \text{if } r\gamma > 1 \\ r\gamma & \text{if } 0 \leq r\gamma \leq 1 \\ -g & \text{if } r\gamma < 0 \end{cases}$$

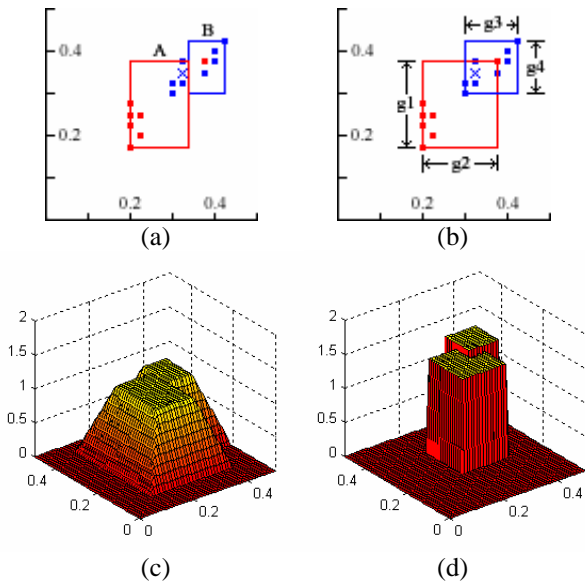
기존의 모델은 하이퍼박스내의 학습패턴 빈도와 관계 없이 일정한 값으로 반응하기 때문에 하이퍼박스의 중첩 영역 내에서는 분류결과의 모호성이 존재한다. 본 연구에서 제안하는 모델에서는 각 개별 특징별로 서로 다른 이득치 값이 반영되므로 하이퍼박스의 중첩영역을 제거하지 않고서도 모호한 응답특성의 문제를 개선할 수 있다.

다음 절에서는 하이퍼박스 축소기법을 대체한 학습 방법의 효과에 대해서 분석한다.

5. 학습효과 분석

그림 3(c)와 같은 15 개의 학습 패턴에 대해 GFMM의 결정공간은 그림 4(a)와 같을 수 있다. 이 경우, 클래스 B의 패턴이 군집 되어있던 영역에 분류를 위한 특징(x 표시)이 입력되었을 때, 그림 4(c)와 같은 소속도로는 입력 특징이 클래스 A로 판단될 것이다. 그림 4(b)는 제안된 학습 방법에 의한 하이퍼박스의 형태이다. 하이퍼박스의 중첩영역이 존재하지만, 그림 4(d)와 같이 각 하이퍼박스가 가지고 있는 이득치 값을 비교함으로써 중첩영역에 의한 모호성을 해결할

수 있다. 그림 4(b)에서의 각 이득치는 표 1 과 같다.



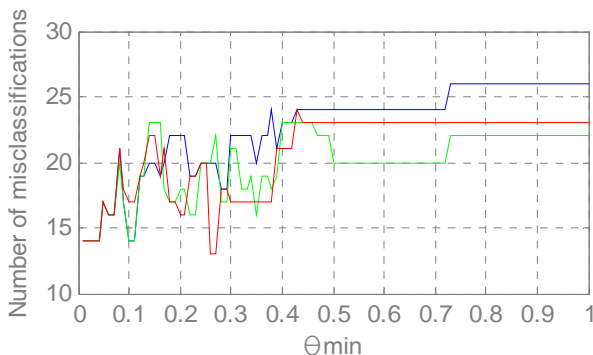
(그림 4) 빈도요소가 중첩영역에서 미치는 영향

<표 1> 그림 4(b)에서 두 하이퍼박스의 중첩영역에 대한 각 차원의 이득치

	빈도수	하이퍼박스 폭	이득치
g1	7	0.200	0.373
g2	7	0.175	0.385
g3	8	0.125	0.467
g4	8	0.125	0.467

6. 분류 성능의 실험적 평가

본 논문의 실험적 평가를 위해 150 개의 Iris 패턴에서 각 클래스별로 25 개의 학습 패턴을 무작위로 뽑아 학습 패턴으로 선정하였다. 그림 1(a)의 차원에 대해 기존의 방법과 제안된 방법에 따라 분류기를 학습 시킨 후, 나머지 75 개의 패턴으로 분류를 실시했을 때 나타난 오류의 개수로 인식 성능을 측정하였다.



(그림 5) 기존의 GFMM 과 제안된 모델로 분류를 실시했을 때 측정된 오류 개수

실험의 결과는 그림 5 의 그래프와 같다. 이 그래프에서 파란색 선은 기존의 모델을 Setosa, Versicolor, Virginica 순으로 학습 시킨 경우이며, 초록색 선은

Versicolor 와 Virginica 의 순서를 바꿔서 학습 시킨 경우를 표시한다. 그리고 빨간색 선은 제안된 모델의 오인식 개수를 나타낸다. 이 모델은 학습 패턴의 클래스 순서에 영향을 받지 않으므로 하나의 학습 순서에 대해서만 기록하였다.

그림 5 의 그래프에서 나타난 바와 같이 축소 과정을 대체한 학습 방법이 기존 모델의 성능에 비해 크게 떨어지지 않음을 확인하였다.

7. 결론

본 논문에서는 GFMM 신경망에 빈도요소를 고려한 활성화 특성을 적용하여 분류기의 성능을 개선시킨다. 하이퍼박스의 이득치는 학습과정 중에 하이퍼박스에 속해 있던 패턴의 빈도수와 하이퍼박스를 이루는 특정 범위에 의해서 산출된다. 이로써 서로 다른 클래스 간의 하이퍼박스 중첩 영역에 대한 모호성을 해결하고 하이퍼박스의 축소 과정을 대체한다. 하이퍼박스의 축소 과정이 없는 분류기의 결과는 학습 패턴의 입력 순서에 크게 달라지지 않고 동시에 노이즈의 영향에도 덜 민감하게 반응한다. 실험 데이터로는 Iris 데이터를 사용하였고, 하이퍼박스의 최대 크기와 퍼지 구간의 기울기가 동일할 때 기존의 GFMM 모델에 비해 개선된 성능을 보였다.

향후 연구로서 퍼지 구간에 빈도요소를 적용하여 확장 시킬 예정이며, 특징 분석 기법에 대한 연구도 병행 할 예정이다.

* 본 연구는 21 세기 프론티어 연구개발사업의 일환으로 추진되고 있는 정보통신부의 유비쿼터스컴퓨팅및네트워크원천기반기술개발사업의 지원에 의한 것임

참고문헌

- [1] P. K. Simpson, "Fuzzy Min-Max Neural Networks-Part 1: Classification," IEEE Transactions on Neural Networks, vol. 3, no. 5, pp. 776-786, 1992.
- [2] P. K. Simpson, "Fuzzy Min-Max Neural Networks-Part 2: Clustering," IEEE Transactions on Fuzzy Systems, vol. 1, no. 1, pp. 32-45, 1993.
- [3] Bogdan Gabrys, Andrzej Bargiela, "General Fuzzy Min-Max Neural Network for Clustering and Classification," IEEE Transactions on Neural Networks, vol. 11, no. 3, pp. 769-783, 2000.
- [4] Ho-Joon Kim, Hyun-Seung Yang, "A Weighted Fuzzy Min-Max Neural Network and Its Application to Feature Analysis," International Conference on Natural Computation, vol. 3612, pp. 1178-1181, 2005.