특징의 분포와 가중치를 고려한 FMM 신경망 모델

박현정[O] 조일국 정경훈 김호준
한동대학교 전산전자공학부
hjpark79@empal.com[O], cik0225@hanmail.net, {khjung,hjkim}@handong.edu

An FMM Neural Network Based on Feature Distributions and Weights

Hyun Jung Park[O] Il Gook Cho, Kyeong Hoon Jung, Ho Joon Kim
School of Computer Science and Electronic Engineering, Handong Global University

요 약

본 연구에서는 FMM 신경망을 이용한 패턴 분류 문제에서 학습 패턴에 포함되는 특징의 발생 빈도와 특징값의 분포를 고려하는 네트워크 구조와 학습 방법론을 소개한다. 이를 위하여 하이퍼박스 소속함수의 산출 과정에 세부특징에 대한 가중치 개념이 적용되는 새로운 활성화 특성을 제안한다. 또한 하이퍼박스의 특징 범위와 빈도 및 특징값의 분포를 유지하고 새롭게 정의된 하이퍼박스 생성, 확장, 축소기법을 적용한다. 이는 가중치개념을 통하여 각 특징별 중요도를 서로 다른 값으로 반영할 수 있게 하며, 특징의 분포 정보가 고려되어 기존 FMM 모델에 비하여 노이즈에 의한 영향을 개선하여 학습 효과를 증진시킬 뿐만 아니라 하이퍼박스의 생성 및 확장 과정중에 학습패턴의 순서에 상관없이 동일한 특성을 보일 수 있게 한다.

1. 서 론

퍼지 최대최소 신경망(Fuzzy Min-Max(FMM) Neural Network)은 1992년 Simpson등에 의하여 최초로 제안된 패턴 분류 모델로 매우 간결하면서도 강력한 학습방법을 지원한다[1]. 이후에 내부연산 및 데이터의 표현형태에 대하여 일반화된 모델이 소개된 바 있으며, 그 후 패턴분류 및 인식에 관한 많은 응용분야에 적용하는 연구와 이와 연관된 다양한 신경망 기반 패턴 분류기법에 관한 연구가 활발하게 이루어 지고 있다[2-5].

기존의 FMM 신경망에서는 하이퍼박스의 활성화를 결정하는소속함수가 학습패턴집합에서 관찰되는 특징값의 범위만을 고려할 뿐 그 빈도를 고려하지 않는다. 이는 적은 수의 노이즈 패턴으로 인하여 학습결과가 심하게 왜곡될 가능성이 있고 또한 비정상적인 패턴이 학습패턴에 포함되는 경우 성능저하를 보일 수 있다.

이에 본 연구에서는 학습과정에서 특징값의 빈도와 분포를 고려하는 신경망 모델과 학습방법을 제안한다. 제안된 모델의 동작특성은 가중치요소를 고려한 하이퍼박스 소속함수로서 새롭게 정의된다. 각 세부특징별로 그 발생 빈도와 발생 특징의 범위에 따라 가중치를 서로 다른 값으로 학습되게 함으로써 비정상적인 데이터에 의한 학습 효과의 저하를 방지할 뿐만 아니라, 학습된 신경망으로부터 특징분석기능 등 여러 가지 이점을 얻을 수 있게 한다. 본 논문에서는 새롭게 정의된 하이퍼박스 소속함수에 의한 동작특성과 연관하여 학습과정의 3단계 즉 하이퍼박스 생성, 확장 및 축소과정에서 특징의 빈도, 분포 및 가중치를 결정하는 방법론을 소개한다. 이어서 실제 분류문제를 대상으로 한 실험결과를 분석함으로써 제안된 이론의 타당성을 고찰한다.

2. 퍼지 최대-최소 신경망

FMM신경망은 하이퍼박스 기반의 패턴 분류 모형이다. 하이퍼박스의 개수와 형태는 학습과정에 의해 결정되는데 개별 특징에 대하여 각각 최대값 및 최소값 쌍으로 표현되는 퍼지 구간을 유지한다[1]. 주어진 패턴 A_h 에 대하여 임의의 하이퍼박스 j 에 대한 소속함수 $b_j(A_h)$ 는 식 (1)과 같이 정의된다.

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

식에서 $A_h = (a_{h1}, a_{h2}, \dots, a_{hn})$은 입력패턴으로 총 n개의 특징값으로 이루어진다. 또한 각 특징에 대한 최소점과 최대점은 $U_j = (u_{j1}, u_{j2}, \cdots, u_{jn})$, $V_j = (v_{j1}, v_{j2}, \cdots, v_{jn})$로 표현된다. γ는 특징범위의 가장자리에서 퍼지 소속함수의 기울기를 결정하는 매개변수이다.

학습과정은 각 하이퍼박스에서 각 특징에 대한 최대 및 최소점을 결정하는 과정으로 하이퍼박스의 생성, 확장, 축소의 3가지 과정으로 이루어 진다. 본 연구에서 제안하는 모델에서는 식 (1)의 특성에 추가로 가중치 요소 및 특징의 무게중심을 유지하여 빈도 및 분포를 고려할 수 있게 한다. 또한 특징 구간의 양쪽 끝에서 퍼지 소속함수를 결정하는 기울기는 각 특징별로 학습될 수 있도록 하였다.

3. 특징의 분포와 가중치를 고려한 FMM 신경망 모델

3.1 하이퍼박스 멤버쉽 함수

본 연구에서는 기존의 FMM모델의 활성화 특성에서 빈

도와 분포를 고려할 수 있도록 수정된 모델을 제안한다. 이를 위하여 새로운 형태의 하이퍼박스 멤버쉽 함수를 정의한다. 우선 하이퍼박스 B_j를 다음과 같이 최대, 최소점 뿐만아니라 각 특징범위의 무게중심과 특징의 발생빈도를 포함하여 새롭게 정의한다.

$$B_j = \{X, U_j, V_j, C_j, F_j, f(X, U_j, V_j, C_j, F_j)\} \quad \forall X \in I^n$$

여기서 U_j는 하이퍼박스의 최소점, V_j는 최대점, C_j는 무게중심점, F_j는 발생 빈도수이다. 이로부터 식 (2)와 같이 하이퍼박스 소속함수가 적용된다. 식에서 w_{ji}는 특징의 발생 빈도로 결정되는 i번째 특징과 j번째 하이퍼박스 간의 연결 가중치이고 γ_{jiU}와 γ_{jiV}는 j번째 하이퍼박스의 i번째 특징의 최소점과 최대점에서의 퍼지구간을 결정하는 매개변수이다. 나머지 표기는 식 (1)과 동일하다.

$$b_j(A_h) = \frac{1}{\sum_{i=1}^{n} w_{ji}} \bullet \sum_{i=1}^{n} w_{ji}[\max(0, 1 - \max(0, \gamma_{jiv} \min(1, a_{hi} - v_{ji}))) + \max(0, 1 - \max(0, \gamma_{jiu} \min(1, u_{ji} - a_{hi}))) - 1.0] \quad (2)$$

$$\gamma_{jiU} = \frac{\gamma}{R_U} \qquad R_U = \max(s, u_{ji}^{new} - u_{ji}^{old})$$
$$\gamma_{jiV} = \frac{\gamma}{R_V} \qquad R_V = \max(s, v_{ji}^{old} - v_{ji}^{new}) \quad (3)$$

위 식에서 보인 바와 같이 제안된 모델에서는 가중치를 추가한 것 외에도 함수의 형식을 변형하였다. 식에서 가중치 요소는 특정 하이퍼박스에 대하여 각 특징값의 상대적인 중요도를 서로 다른 값으로서 표현할 수 있게 한다. 기존 FMM 모델에서는 각 하이퍼박스가 0.5와 1사이의 값을 출력하게 되는데, 이를 좀 더 일반적인 모델의 특성을 갖도록 하기 위하여 0와 1사이의 값을 갖도록 식의 형태를 변형하였다.

학습과정에서 가중치 변화는 생성되는 하이퍼박스에 나타나는 특징별 발생 빈도수와 특징의 범위에 따라 결정된다. 즉 각 학습단계에서 임의의 연결에 대한 가중치 w_{ji}는 식(4)에 의해서 조정된다.

$$w_{ji} = \frac{\alpha f_{ji}}{R} \quad (4)$$

$$R = \max(s, v_{ji} - u_{ji}) \quad (\text{단}, s > 0) \quad (5)$$

식에서 보인 바와 같이 가중치 값은 특징범위에 나타난 빈도에 비례해서 결정된다. 단 특징범위가 일정량 이하인 경우 가중치가 지나치게 증가하는 현상을 보완하기 위해 식 (5)와 같이 일정범위 s 이내에 대해서는 일정값 이하로 조정되도록 하였다.

α는 학습률 상수인데 특징영역이 확장함에 따라 해당 특징에 대한 가중치가 얼마나 증가시킬 것인가를 결정하는 매개변수이다.

3.2 학습 알고리즘

제안된 모델에서 각 특징들에 대한 가중치 값은 하이퍼박스 생성, 축소 및 확장의 3단계의 과정을 반복하면서 재조정된다.

학습의 각 단계에서 주어진 패턴에 대하여 같은 클래스에 속하는 하이퍼박스들 중 가장 크게 반응하는 하이퍼박스로부터 식(6)에 따라 확장여부를 결정하여 해당 특징에 대한 빈도, 퍼지구간범위 및 무게중심은 식(7)과 같이 수정한다.

$$n\theta \ge \sum_{i=1}^{n} (\max(v_{ji}, x_{hi}) - \min(u_{ji}, x_{hi})) \quad (6)$$

$$\begin{cases} f_{ji}^{new} = f_{ji}^{old} + 1 \\ u_{ji}^{new} = \min(u_{ji}^{old}, x_{hi}) \\ v_{ji}^{new} = \max(v_{ji}^{old}, x_{hi}) \\ c_{ji}^{new} = (c_{ji} * f_{ji}^{old} + x_{hi}) / f_{ji}^{new} \end{cases} \quad (7)$$

이렇게 확장된 하이퍼박스가 특징공간상에서 다른 클래스의 하이퍼박스와 중첩이 되는 경우 이를 축소한다. 이 경우 정보의 소실을 최소화 하기 위하여 다차원 특징공간에서 가장 작은 범위가 중첩되는 차원(dimension)에서 축소한다. 하이퍼박스 중첩 테스트는 아래와 같이 하며 $\delta^{old} - \delta^{new} > 0$인 경우 $\Delta = i$, $\delta^{old} = \delta^{new}$를 넣고 중첩되는 차원이 없는 경우 $\Delta = -1$을 넣어준다.

$$case1: u_{ji} < u_{ki} < v_{ji} < v_{ki}$$
$$\delta^{new} = \min(v_{ji} - u_{ki}, \delta^{old})$$
$$case2: u_{ki} < u_{ji} < v_{ki} < v_{ji}$$
$$\delta^{new} = \min(v_{ki} - u_{ji}, \delta^{old})$$
$$case3: u_{ji} < u_{ki} < v_{ki} < v_{ji}$$
$$\delta^{new} = \min(\min(v_{ki} - u_{ji}, v_{ji} - u_{ki}), \delta^{old})$$
$$case4: u_{ki} < u_{ji} < v_{ji} < v_{ki}$$
$$\delta^{new} = \min(\min(v_{ji} - u_{ki}, v_{ki} - u_{ji}), \delta^{old})$$

$\Delta > 0$인 경우 중첩 부분 삭제를 위해 하이퍼박스 영역 축소를 하게 된다. 제안된 학습 방법에서의 축소는 각 하이퍼박스의 축소 특징의 발생 빈도를 고려한 축소를 하게 된다. 축소방법은 4가지 경우에 대하여 서로 다른 방법으로 이루어 지는데 식 (8)은 첫 번째 경우에 대한 예이다.

$$case1: u_{j\Delta} < u_{k\Delta} < v_{j\Delta} < v_{k\Delta}$$

$$\begin{cases} v_{j\Delta}^{new} = v_{j\Delta}^{old} - \frac{f_{k\Delta}}{f_{j\Delta} + f_{k\Delta}}(v_{j\Delta}^{old} - u_{k\Delta}^{old}) \\ u_{k\Delta}^{new} = u_{k\Delta}^{old} + \frac{f_{j\Delta}}{f_{j\Delta} + f_{k\Delta}}(v_{j\Delta}^{old} - u_{k\Delta}^{old}) \\ f_{j\Delta}^{new} = f_{j\Delta}^{old} * (\frac{v_{j\Delta}^{new} - u_{j\Delta}^{new}}{v_{j\Delta}^{old} - u_{j\Delta}^{old}}) \\ c_{j\Delta}^{new} = u_{j\Delta}^{new} + (c_{j\Delta}^{old} - u_{j\Delta}^{old}) * (\frac{v_{j\Delta}^{new} - u_{j\Delta}^{new}}{v_{j\Delta}^{old} - u_{j\Delta}^{old}}) \end{cases} \quad (8)$$

이어서 최종단계로 하이퍼박스 퍼지구간의 재조정이 식 (9)와 같이 이루어진다. 앞 절에서도 설명한 바와 같이 본 논문에서는 학습데이터의 노이즈패턴을 최소화 하는 방법으로 학습패턴 집합을 모두 학습 시킨 후에 하이퍼박스 B_j의 특징별 c_{ji}를 고려하여 하이퍼박스 최소점(u_{ji})과 최대점(v_{ji}) 그리고 멤버쉽 함수에 영향을 주는 γ_{jiU}와 γ_{jiV}를 재조정한다.

$$u_{ji}^{new} = c_{ji} - \min(c_{ji} - u_{ji}, v_{ji} - c_{ji})$$

$$v_{ji}^{new} = c_{ji} + \min(c_{ji} - u_{ji}, v_{ji} - c_{ji})$$

$$\gamma_{jiU} = \frac{\gamma}{R_U} \qquad R_U = \max(s, u_{ji}^{new} - u_{ji}^{old}) \qquad (9)$$

$$\gamma_{jiV} = \frac{\gamma}{R_V} \qquad R_V = \max(s, v_{ji}^{old} - v_{ji}^{new})$$

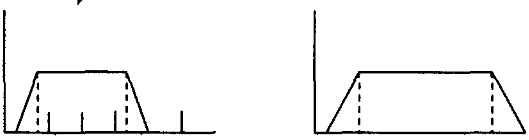

그림1. FMM 신경망

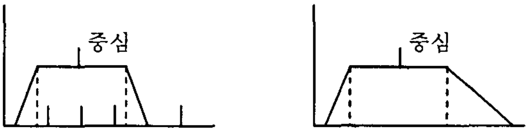

그림2. 가중치를 고려한 FMM 신경망

위의 그림 1,2에서 보는 바와 같이 무게중심을 이용하여 재조정 함으로 기존의 여러 번의 학습과정을 통해서 조정된 결과가 한 개의 잘못된 비정상적 데이터로 인하여 크게 왜곡되는 것을 방지할 수 있게 한다.

4. 실험 결과 및 고찰

제안된 모델의 성능평가를 위하여, 제안 모델을 아이리스 데이터의 분류문제에 적용하였다. 아이리스 데이터는 총 150개의 패턴을 가지며, 4개의 특징과 3개의 클래스로 구성된다. 성능평가는 두가지로 수행하였다. 첫번째로는 학습패턴수를 바꾸어가며 적은 학습패턴수에 대한 분류성능을 확인하였으며, 두번째로는 무작위로 1개의 학습 패턴을 왜곡하였을때의 분류결과를 통해 노이즈에 대한 영향을 확인하였다. 표 1은 학습패턴 수의 증가에 따른 분류결과이며, 표 2는 노이즈의 영향개선에 대한 실험 결과이다.

표 1. 학습패턴 수에 대한 분류결과

학습패턴수	FMM신경망	제안된FMM신경망
18	144	144
27	144	146
36	145	146
45	146	147

표 2. 왜곡된 학습패턴에 대한 분류결과

학습패턴수	FMM신경망	제안된FMM신경망
9	111	132
27	141	142
54	143	145
72	144	147

결과를 보면, 적은수의 학습패턴에 대한 분류성능과 왜곡된 학습패턴의 삽입에 대한 분류성능이 보다 제안모델의 분류성능이 효율적임을 알 수 있다.

5. 결론

본 연구에서 제안된 모델은 기존의 FMM 신경망의 학습 능력과 퍼지집합 특성에 특징 범위와 발생 빈도를 활용할 수 있도록 하는 패턴 분류 모형이다. 학습시 특징값의 범위와 발생 빈도를 이용하여 특징의 중요도를 평가함으로써 일정비율의 학습패턴에 대하여 기존 모델보다 효율적인 패턴 분류 성능을 가짐을 보여 주었다. 또한 제안된 모델은 하이퍼박스 재조정시에 특징값의 분포, 즉 통계적인 정보를 활용하여 노이즈 패턴을 가지고 있는 학습데이터가 패턴 분류 성능에 주는 영향을 최소화 하였을 뿐만 아니라, 학습패턴의 순서에 상관없이 동일한 특성을 가지게 하여 안정된 분류성능을 유지한다. 향후 연구로는, FMM신경망 학습과정 중 중첩 축소시 발생하는 부작용을 보완하는 방법으로, 축소 과정을 제안된 모델의 특징 발생 빈도를 이용한 분류로 대치하여 부작용을 최소화하여 패턴분류 성능을 향상시키는 연구를 고려하고 있다.

* 이 연구는 과학기술부 뇌과학 연구개발사업으로 수행되었음

6. 참고 문헌

[1] P. Simpson, " Fuzzy Min-Max Neural Networks-Part 1:Classification," IEEE Transactions on Neural Networks, Vol.3, No.5,pp.776-786,1992.

[2] B. Gabrys and A. Bargiela, " General Fuzzy Min-Max Neural Network for Clustering and Classification," IEEE Transactions on Neural Networks, Vol.11, No.3, 2000.

[3] S. Mitra and Y. Hayashi," Neuro-Fuzzy Rule Generation: Survey in Soft Computing Framework," IEEE Transactions on Neural Networks, Vol.11, No.3, pp.748-768, 2000.

[4] C.L. Blake, and C. J. Merz, " UCI Repository of machine learning databases [http://www.ics.uci.edu/~mlearn/MLRepository.html]," University of California Irvine, Department of Information and Computer Science,1998.

[5] Jayanta Basak, Rajat K. De, Sankar K. Pal," Unsupervised Feature Selection using a Neuro-Fuzzy Approach," Pattern Recognition Letters, Vol.19, pp.997-1006, 1998.