FMM 신경망 기반의 패턴분류 문제에서 특징의 중요도 판별 기법

박현정, 조일국, 김호준
한동대학교 전산전자공학부
e-mail : hjpark79@empal.com, cik0225@hanmail.net, hjkim@handong.edu

A Feature Saliency Measure in FMM Neural Network-Based Pattern Classification

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

요 약

본 논문에서는 패턴 분류문제에서 특징의 분포와 빈도를 고려하는 FMM(Fuzzy Min-Max) 신경망 구조와 이를 이용한 특징 분석 기법을 소개한다. 이는 기존의 모델에서 균일한 가중치를 고려했을 때 비정상적 학습데이터에 학습 효과가 민감하게 왜곡되는 현상을 방지한다. 또한 학습된 신경망으로부터 각 특징의 중요도를 분석할 수 있게 한다. 본 연구에서는 제안된 모델의 특성을 소개하고 특징 값과 하이퍼박스 간의 관계로부터 특징의 연관도 요소, 중요도 평가 및 특징의 서열화 기법을 제시한다. 이는 패턴 분류 신경망의 노드수를 최적화 함으로써 학습 및 분류 과정에서 연산의 효율성을 증대시킨다.

1. 서 론

특징 선정문제는 n 차원 특징 공간 에서 패턴을 가장 효율적으로 분류할 수 있는 최적의 $n'(n>n')$ 차원 특징 공간을 선별하는 과정이다. 이와 같이 패턴 분류 문제에서 최적의 특징 집합을 선별하는 것은 실행 시간과 분류 성능을 최적화 시키는 매우 중요한 작업으로 다양한 방법론의 연구가 발표된 바 있다[1-7]. Steppe 와 Kenneth 등의 연구에서는 역전파방법을 사용하여 학습된 단일신경망에서 입력 특징의 중요도를 결정하기 위해 통계학적 접근법을 사용한바 있으며[5], Jayanta 등에 의해 제안된 특징 선정에 대한 연구에서는 가중치가 부여된 거리를 소속함수에 부여하여 클러스터를 더욱 잘 구분할 수 있는 평가 지표로 소개하고 있다[7].

FMM 신경망은 1992 년 Simpson 등에 의해 제안된 모델로서 간결하면서도 강력한 학습 기능을 지원한다. 그 후 Gabrys 등에 의해 클러스터링과 분류 과정이 확장된 개념으로 일반화된 모델이 제안되었다. 그러나 특징값과 하이퍼박스 간의 관계를 항상 균일한 가중치로 유지함으로써 특징값의 발생 빈도와 분포정보를 효과적으로 반영할 수 없다는 점이 약점으로 지적된다. 이는 노이즈등 비정상적 데이터에 대하여 바람직하지 않은 학습효과를 보일 수 있게한다. 이에 본 연구에서는 기존의 학습 모델의 동작특성에 빈도요소를 추가하고 이로부터 학습을 통하여 가중치값이 조정되도록 하는 모델을 소개한다. 또한 FMM 신경망의 학습 과정 중 중첩된 영역의 축소 과정에 의해 발생하는 부작용을 최소화 하기 위하여 축소과정을 가중치 값으로 대치하는 학습 기법을 사용하였다.

본 논문에서는 제안된 모델의 활용으로서 특징의 중요도 분석 기법을 제시한다. 새롭게 정의된 활성화 특성에서 가중치 요소로부터 특징과 패턴 클래스간의 4 종류의 연관도 요소를 정의한다. 이들 요소는 주어진 패턴 분류문제에서 특징선정, 분류기 설계 및 지식생성 기법 등에 활용될 수 있다. 제안된 이론은 아이리스 데이터 및 의료진단 데이터를 사용하여 실험적으로 검증하였다.

2. 가중치를 갖는 FMM 신경망

FMM 신경망은 n 차원 하이퍼 공간으로 정의되며 모든 패턴은 패턴 공간 내에서 멤버쉽값을 갖는 하이퍼 박스내에 포함되어 있다. 본 연구에서는 기존의 FMM 신경망 구조에 빈도요소와 가중치 개념을 추가한 형태의 활성화 특성을 갖는 수정된 모델을 고려한다. 이 모델에서 임의의 하이퍼박스 B_j 의 특성은 아래와 같이 정의된다.

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

여기서 X 는 $\{X_h, d_h\}$로 학습 패턴 집합이고 U_j 와 V_j 는 하이퍼박스의 최소점과 최대점, F_j 는 하이퍼박스 특징 범위내에 패턴의 빈도수를 나타낸다. 본 논문에서는 이러한 모델을 가중치를 갖는 다는 점에서 WFMM(Weighted Fuzzy Min-Max) 신경망이라 정의한다.

WFMM 신경망의 학습 과정은 구조적 특성상 기존의 FMM 신경망에서 이루어졌던 하이퍼박스 축소과정을 생략할 수 있게 한다. 이는 기존의 하이퍼박스 축소과정에 내재했던 특징범위의 왜곡현상을 보완할 수 있게 한다. 왜냐하면 WFMM 신경망에서는 중첩되는 영역에 대해서 각 하이퍼박스에 대한 소속함수 값이 가중치 적용으로 인하여 서로 다른 값으로 나타나기 때문이다. 따라서 하이퍼박스 중첩 테스트과정과 축소과정이 생략됨으로써 학습시간 단축은 물론 학습 효과도 개선할 수 있게 한다.

따라서 학습과정은 하이퍼박스 생성 및 확장 과정과 가중치 값 조정과정으로 이루어 진다. 하이퍼박스의 생성 또는 확장 여부는 식 (1)의 조건에 의해 결정된다.

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

식 (1)에서 n 은 특징의 수이고 θ 는 확장될 하이퍼박스와의 최대 평균거리로서 [0,1]사이의 값을 갖는다.

하이퍼박스 확장과정은 식(2)와 같이 이루어진다.

$$\begin{cases} f_{ji}^{new} = f_{ji}^{old} + 1 \\ u_{ji}^{new} = \min(u_{ji}^{old}, x_{hi}) \quad \forall i = 1, 2, ..n \\ v_{ji}^{new} = \max(v_{ji}^{old}, x_{hi}) \end{cases} \qquad (2)$$

식에 보인 바와 같이 동일 하이퍼박스가 확장됨에 따라 빈도 요소 값은 1 씩 증가하며 특징 범위의 상하한 값이 재조정된다.

임의의 하이퍼박스의 소속함수는 식(3)과 같이 정의된다.

$$b_j(A_h) = \frac{1}{\sum_{i=1}^{n} w_{ji}} \bullet \sum_{i=1}^{n} w_{ji} [\max(0, 1 - \max(0, \gamma \min(1, a_{hi} - v_{ji}))) + \max(0, 1 - \max(0, \gamma \min(1, u_{ji} - a_{hi}))) - 1.0] \qquad (3)$$

식에서 $A_h = (a_{h1}, a_{h2}, \cdots, a_{hn})$은 입력패턴으로 총 n 개의 특징값이고 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. W_{ji} 는 j 번째 하이퍼박스와 i 번째 특징사이의 연결 가중치를 의미하며 γ 는 특징범위의 가장자리에서 퍼지 소속함수의 기울기를 결정하는 매개변수로 [0,1]사이의 값을 갖는다.

각 학습단계에서 임의의 연결에 대한 가중치 w_{ji} 는 다음과 같이 결정된다.

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

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

이 때 특징범위가 작은 경우 가중치 값이 지나치게 증가하는 현상을 방지하기 위하여 식 (5)와 같이 일정범위 s 내에 대해서는 일정값 이하가 되도록 조정하였다.

3. 세분화된 특징의 중요도 판별 기법

본 연구에서는 그림 1 과 같이 특징 집합과 클래스 집합간의 4 종류의 관계를 고려하여 특징의 연관도 요소를 정의한다. 이를 위하여 다음과 같은 표기를 정의한다.

- X_i : i번째 특징
- x_j : j번째 특징값
- C_i : i번째 클래스
- B_j : j번째 하이퍼박스
- N : 특징의 수
- M : 클래스의 수

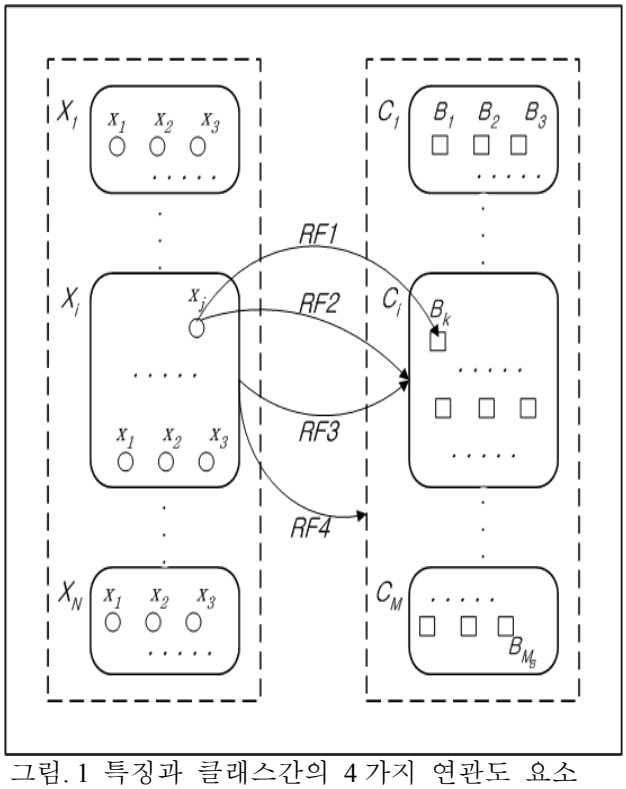

그림. 1 특징과 클래스간의 4 가지 연관도 요소

이들 4 가지 관계로부터 다음과 같은 4 종류의 연관도 요소를 정의한다.

$RF1(x_i, B_j)$: 특징값 x_i와 하이퍼박스 B_j사이의 연관도 요소
$RF2(x_j, C_k)$: 특징값 x_j와 클래스 C_k 사이의연관도 요소
$RF3(X_i, C_k)$: 특징유형 X_i와 클래스 C_k 사이의연관도 요소
$RF4(X_i)$: 주어진문제에서 특징 X_i의 중요도 판별

3.1 특징 값과 패턴 클래스 간의 연관도 요소

학습된 신경망으로부터 특징값과 특정 하이퍼박스 간의 상호 연관도는 가중치값으로서 정의될 수 있다. 즉, 식 (6)과 같이 *i* 번째 특징과 *j* 번째 하이퍼박스 간의 상호 연관도 요소는 이들 간의 가중치 값으로 정의될 수 있다.

$$RF1(x_i, B_j) = w_{ij} \quad (6)$$

이러한 연관도를 사용하여 특정 특징값과 임의의 클래스 간의 연관도 요소를 식 (7)과 같이 정의할 수 있다. 이러한 연관도가 양의 값을 가지면, 특징값과 패턴 클래스 사이에 자극성(*excitatory*) 연관성이 있음을 의미하며, 음의 값을 가지면 그들 사이에는 억제성(*inhibitory*) 연관성이 있음을 의미한다.

$$RF2(x_i, C_k) = \left(\frac{1}{N_k}\sum_{B_j \in C_k} S(x_i, (u_{ji}, v_{ji}))\cdot w_{ij}\right) - \frac{1}{(N_B - N_k)}\sum_{B_j \notin C_k} S(x_i, (u_{ji}, v_{ji}))\cdot w_{ij}\Big) \Big/ \sum_{B_j \in C_k} w_{ij} \quad (7)$$

식에서 N_B 는 총 하이퍼박스의 개수이고 N_k 는 k 클래스에 속하는 하이퍼박스의 수이다. x_i 는 하이퍼박스의 특징 범위로 특징 i 가 속한 하이퍼박스의 최대값을 x_i^L, 최소값을 x_i^U 로 정의한다.

$$S(x_i, (u_i, v_i)) = S((x_i^L, x_i^U), (u_i, v_i)) = \frac{Overlap((x_i^L, x_i^U), (u_i, v_i))}{Max(x_i^U - x_i^L, v_i - u_i)} \quad (8)$$

$$S((x_i^L, x_i^U), (u_i, v_i)) = \begin{cases} 1 & if\ (x_i^L = x_i^U = u_i = v_i) \\ 0 & Otherwise \end{cases} \quad (9)$$

식 (8)은 최종 생성된 하이퍼박스와 임의의 특징 i 의 퍼지 구간 사이의 유사도(*S: similarity*)를 정의하고 있다. 이때 $Max(x_i^U - x_i^L, v_i - u_i)$ 의 두 구간이 한 점으로 나타나 0 이 되는 경우를 방지하기 위하여 식 (9)로써 0 이 되는 경우와 0 보다 큰 경우로 나누고 0 보다 큰 경우 식 (10)를 적용하여 산출한다.

이러한 연관도 요소는 주어진 패턴 분류 문제에서 특정 특징값이 임의의 클래스를 분류하는 과정에서 상대적인 중요도 값을 반영한다. 이는 학습패턴으로부터 패턴 분류를 위한 명시적인 지식표현으로서 추출 할 수 있는 가능성을 시사한다. 다시 말해서 특정 클래스에 대한 분류 기준에 가장 중요한 특징 값을 발견함으로써 규칙 생성기법등에 활용 될 수 있다.

$$Overlap((d_i^L, d_i^U), (u_i, v_i)) = \begin{cases} d_i^U - u_i & if(d_i^L \le u_i \le d_i^U \le v_i) \\ v_i - u_i & if(d_i^L \le u_i \le v_i \le d_i^U) \\ d_i^U - d_i^L & if(u_i \le d_i^L \le d_i^U \le v_i) \\ v_i - d_i^L & if(u_i \le d_i^L \le v_i \le d_i^U) \\ 0 & Otherwise \end{cases} \quad (10)$$

3.2 특징의 유형과 패턴 클래스간의 연관도 산출

본 절에서 제안하고 있는 클래스 k 와 관련된 특징 x 의 연관도(*RF3*)는 식 (11)과 같이 *RF2* 를 사용하여 정의될 수 있다.

$$RF3(X_i, C_k) = \frac{1}{L_i}\sum_{x_l \in X_i} RF2(x_l, C_k) \quad (11)$$

식에서 L_i 는 i 번째 특징에 속하는 특징값의 개수를 의미한다. 다시 말해서, *RF3* 는 특정 클래스에 속한 특징 집합들의 상대적 연관도에 대한 평균치를 의미한다. 이는 패턴 분류문제에서 학습이 완료된 신경망으로부터 일종의 지식을 추출할 수 있게 한다. 예를 들어 특정 클래스를 분류할 때 가장 중요하게 반영되는 특징의 종류를 서열화 또는 선별 선택할 수 있게 한다.

이들 연관도를 활용하여 특정 특징이 주어진 문제의 분류 과정에서 상대적 중요도를 평가할 수 있다. 이는 다음과 같이 정의될 수 있다.

$$RF4(X_i) = \frac{1}{M}\sum_{j=1}^{M} RF3(X_i, C_j) \quad (12)$$

특정 특징의 상대적 중요도는 각 개별 클래스에 대하여 평균적인 중요도를 산출 함으로써 평가된다. 이는 주어진 문제에서 가장 효과적인 특징을 선별할 수 있게 한다. 따라서 패턴 분류 신경망을 설계하는 과정에서 최적의 입력노드 수를 결정할 수 있게 함으로써 성능과 계산량 및 신경망의 규모를 최적화 할 수 있게 한다. 또한 실세계로부터 추출한 원시 데이터로부터 학습 패턴 집합을 효과적으로 선별 할 수 있게 할 것이다.

4. 실험 결과 및 고찰

제안된 모델의 성능을 평가하기 위하여, 본 실험에서는 패턴 분류 문제에서 잘 알려진 아이리스 데이터와 클리브랜드 의료 진단 데이터를 사용하여 특징의 중요도 판별 기법을 적용하였다. 아이리스 데이터는 4 개의 특징을 가진 3 개의 클래스가 각 50 개의 패턴으로 구성되어 있으며 클리브랜드 의료 진단 데이터는 총 297 개의 패턴에 대하여 13 개의 특징으로부터 5 개의 클래스로 분류하는 문제이다.

표 1 은 패턴 분류 성능 평가에 대한 실험 결과이다. 클리브랜드 의료 데이터에 대하여 기존의 FMM 신경망과 제안된 모델을 사용한 결과를 상호 비교하였다. 표에서 보인 바와 같이 학습 패턴의 수가 증가함에 따라 인식률이 점진적으로 상승하였고 기존의

FMM 신경망 보다 우수한 성능을 보였다.

표 1. 학습 패턴의 수에 따른 성능 비교
(클리버랜드 의료 데이터)

학습패턴의수	기존 FMMNN	제안된 모델
	에러율	에러율
10	35	32
20	29	27
30	21	19
40	14	12
50	7	5

표 2 는 특징 분석 기법에 대한 실험 결과이다. 아이리스 데이터 분류 문제에서 각 패턴 클래스와 관련된 특징값에 대하여 상호 연관도 *RF2* 를 평가하였다. 표에서는 각 클래스 별로 가장 높은 연관도를 갖는 특징 2 개씩을 보였다.

표 2. 특징값과 특정 클래스 간의 연관도 요소
(아이리스 데이터)

특징값	클래스	RF2
F4 : (0.0, 0.13)	Setosa	0.312
F1: (0.03, 0.22)	Setosa	0.190
F3: (0.51, 0.65)	Versicolor	0.443
F2: (0.13, 0.54)	Versicolor	0.158
F3: (0.65, 0.78)	Virginica	0.272
F2: (0.21, 0.67)	Virginica	0.133

이어서 특징 유형과 클래스와의 관계 *RF3* 에 대한 실험을 클리브랜드 의료데이타에 대하여 적용하고 그 결과를 표 3 에 보였다.

표 3. 특징 유형과 특정 클래스간의 연관도 요소 (RF3)
(클리버랜드 의료데이터)

특징유형	클래스	RF3
F12	0	0,327
F3	0	0.240
F2	1	0.387
F12	1	0.311
F3	2	0.382
F8	2	0.133
F9	3	0.402
F2	3	0.176
F2	4	0.377
F5	4	0.266

각 클래스에 대하여 연관도가 높은 특징 2 개씩을 선택하여 표에 나타내었다. 이는 패턴 분류문제에서 특정 클래스의 분류에 가장 영향을 주는 특징의 종류로서 문제에 관한 중요한 정보로 활용될 수 있을 것이다.

5. 결 론

본 논문에서는 패턴 분류문제에서 특징의 분포와 빈도를 고려하는 퍼지 최대-최소 신경망 구조를 소개하고, 이를 이용하여 특징과 클래스 간의 상호연관도를 분석하기 위한 방법론을 고찰하였다. WFMM 신경망은 가중치를 고려한 동작특성을 가지고 있으며 기존의 신경망의 하이퍼박스 중첩 부분 제거 과정의 부작용을 최소화 하기 위해 축소과정을 가중치로 대체하여 학습한다.

특징 분석 기법에서 4 종류의 연관도 요소는 패턴 분류 문제의 분석과정에 활용될 수 있으며 알려지지 않은 문제에 대한 해결책을 설계하는데 도움을 줄 수 있다. 또한 제안된 특징의 연관도 요소는 특정 특징값이 임의의 클래스를 분류하는 과정에서 상대적인 중요도 값을 반영함으로써 학습패턴으로부터 패턴 분류를 위한 명시적인 지식표현으로서 추출 할 수 있는 가능성을 시사하였다. 또한 제안된 특징별 중요도 평가와 특징의 서열화 기법이 특정클래스에 대한 분류 기준에 가장 중요한 특징 요소를 발견함으로써 규칙 생성기법등에 활용 될 수 있다.

* 이 연구는 과학기술부 뇌과학 연구개발사업으로 수행되었음

참고문헌

[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, A. Bargiela, “General Fuzzy Min-Max Neural Network for Clustering and Classification,”IEEE Transactions on Neural Networks, Vol.11, No.3, 2000.

[3] Ho J. Kim, Tae W. Ryu, Thai T. Nguyen, Joon S. Lim, S. Gupta, “A Weighted Fuzzy Min-Max Neural Network for Pattern Classificatin and Feature Extraction,” Lecture Notes in Computer Science, LNCS 3046, pp.791-798, 2004.

[3] S. Mitra, 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] Hung-Pin Chiu, Din-Chang Tseung, “Invariant Handwritten Chinese Character Recognition using Fuzzy Min-Max Neural Networks,” Pattern Recognition Letters, Vol.18, pp.481-491, 1997.

[5] Jean M. Steppe, Kenneth W. Bauer, Jr. “Improved feature screening in feedforward neural networks,” Neurocomputing , 13, pp.47-58, 1996

[6] Mahn M. Lee, Kuo H. Chen, I. F. Jiang, "A Neural network Classifier with Disjunctive Fuzzy Information," Neural Networks, Vol.11, No.1, pp.1113-1125, 1998.

[7] 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.