Proceedings of KFIS 2001 Fall Conference, 2001. 12. 1

패턴 인식을 위한 Possibilistic 퍼셉트론 알고리즘

A Possibilistic Perceptron Algorithm for Pattern Recognition

김미경, 이정훈
한양대학교 전자컴퓨터공학부

Kim Mi Kyung, Frank Chung-Hoon Rhee

Computational Vision and Fuzzy Systems Laboratory
Department of Electronic Engineering
Hanyang University
Ansan, KOREA
{mkduck,frhee}@fuzzy.hanyang.ac.kr

ABSTRACT

패턴 인식에서 선형분류가능한 경계면을 찾아 패턴을 분류하는 방법 중 가장 기본적인 방법은 퍼셉트론이라고 볼 수 있다. 하지만 선형분류불가능한 패턴에 대해서는 유용한 결과를 보여주지 못하였다. 먼저 제안된 퍼지 퍼셉트론은 베타영역 설정에 의해 수렴하지 못하는 특성을 보완하였다. 그러나 패턴의 순수한 전형성을 고려해 주지 못하는 단점이 있다. 이에 Crisp의 선형분류 특성과 퍼지의 수렴특성을 합성하고자 Possibilistic 퍼셉트론을 제시한다.

Keywords : crisp perceptron, fuzzy perceptron, possibilistic perceptron, typicality ,membership

Ⅰ. 서 론

경계면 결정함수(decision function)에 의해 패턴을 분류하는 분류자로서, 가장 기본적인 방법은 퍼셉트론이라고 볼 수 있다. 이는 선형분류가능한(lineary separable) 패턴 데이터들에 대해서는 좋은 결과를 보여 왔지만, 클래스간의 경계가 애매한 클래스들에 대해서는 올바른 경계결정함수를 찾지 못하고, 또한 수렴하지 못하는 문제점을 가지고 있다. 이에 발전된 형태인 퍼지 퍼셉트론의 경우에는 패턴에 대해서 소속도 함수를 적용하여, 경계결정함수를 보정할 때 그 역할을 조정하게 된다. 이러한 방법은 알고리즘의 수렴과 함께, 일반적인 퍼셉트론의 방법보다는 바람직한 결과를 보일 수 있다. 그러나 퍼지 퍼셉트론의 특성상 패턴 데이터의 클래스에 대한 전형성을 고려하지 않은 상대적인 거리 개념에 의한 소속도정도 결정과 베타영역의 설정에 의해 그 결과가 달라지는 단점을 보인다. 또한 일반 선형분류가능한 패턴데이터들에 대해서는 일반적인 퍼셉트론보다 수렴결과가 더 좋지 못할 수도 있다. 따라서 본 논문에서는 일반적인 퍼셉트론의 선형분류가능한 데이터에 대한 장점과 퍼지 퍼셉트론의 선형분류불가능한 패턴에 대한 수렴특성을 합성하고자 Possibilistic 퍼셉트론을 제시한다. 기본적으로 본 논문에 제시된 알고리즘은 *PCM (possibilistic C-means)*에 사용된 멤버쉽 결정방법을 사용해 전체 패턴데이터에 대해 해당 클래스의 전형성(typicality)를 고려하도록 하고자한다. 이러한 클래스의 전형성 고려는 선형분류불가능한 데이터들에 대한 경계면 함수 결정에 있어서는 퍼지 퍼셉트론과 유사한 수렴의 특성을 부여하고, 선형분류가능한 패턴들에 대해서는 일반적인 퍼셉트론 방법에 비해 좀더 바람직한 경계면의 위치를 제시할 수 있다.

본 논문에서는 여러 패턴 데이터들을 통해 직선과 곡선으로 이루어진 경계면 함수를 결정해, 그 결과를 비교하여, 본 논문에 제시된 알고리즘의 유용성을 보일 것이며, 마지막으로 영상 데이터를 통해, 영상분할의 예제를 제시할 것이다.

Ⅱ. 본 론

1. Crisp 퍼셉트론

패턴 x_k가 클래스1 또는 클래스2 로 분류

되는 한 쌍의 견본 벡터 { $x_1, \cdots, x_R$ }를 가지고 있다고 가정하자. 만일 이러한 클래스들 중 서로 다른 두개의 클래스를 분류하는 선형 분류함수 g가 존재한다면

$$g(x) = W^t x_k \begin{cases} >0 \text{ if } x_k \in class1 \\ <0 \text{ if } x_k \in class2 \end{cases} \quad (1)$$

$g(x)=0$ 인 지점은 실수평면 R상에서 클래스1에 속하는 벡터들과 클래스2에 속하는 벡터들의 경계면을 정의한다. 이러한 퍼셉트론 알고리즘은 샘플 벡터들을 두개의 클래스로 나누는 경계결정함수를 정의하기 위해 사용되고 있다.

만일 위와 같이 두개의 클래스를 나눌 수 있는 선형 경계면 함수가 존재한다면, 퍼셉트론 알고리즘은 몇 번의 반복 수행을 거쳐 유용한 경계면을 찾아내게 된다[1]. 그러나 선형분류불가능한 패턴들의 경우에는 이러한 경계면 함수가 존재하지 않아 무한횟수의 반복 연산을 거쳐서도 수렴하지 못하는 문제점이 있다. 따라서 퍼셉트론 알고리즘 종료조건을 수정하는 것에 의해 아주 좋은 결과는 아니지만, 선형분류가능한 경우와 선형분류불가능한 경우 모두 비교적 좋은 결과를 얻을 수 있도록 할 수 있다. 퍼지 집합 이론을 퍼셉트론 알고리즘에 포함하는 것은 이러한 문제에 대한 또 다른 방법이다.

Crisp 퍼셉트론 알고리즘의 주된 약점은 두개의 클래스가 선형분류불가능할때 이 알고리즘은 하나의 적절한 경계면 함수를 찾아내지 못하고 무한 반복연산을 수행할 수 있다. 만약 이러한 무한 반복연산을 방지하기 위해 임의의 지점에서 반복연산을 종료시키면 그 결과 경계면 함수는 해당 패턴에 최후의 경계면 임을 보장할 수 없게 된다. 경계면에서 겹쳐진 패턴들은 퍼셉트론 알고리즘이 수렴하지 못하고 잘못 동작하게되는 주된 원인이 된다. 이 같은 패턴 벡터들은 자기 클래스가 아닌 상관이 없는 클래스들에 대해 상대적으로 상관이 없어야 하는데 기존의 Crisp 퍼셉트론 알고리즘에서는 이들에 대해서도 완전한 weight(1을 갖는다)를 갖게된다.

2. Fuzzy 퍼셉트론

퍼지 퍼셉트론에서는 위에서 설명한 Crisp 퍼셉트론의 단점을 보완하기 위해 소속이 불확실한 패턴 벡터들이 weight vector 수정에 적게 영향을 미치게 하도록 퍼셉트론 알고리즘을 수정한다. 이를 위해 weight벡터를 수정하는 과정을 보완한다[2].

Crisp 퍼셉트론에서, 이러한 weight 수정 단계는 $W \leftarrow W + c x_j$로 한다. Crisp 퍼셉트론의 weight 수정단계에 멤버쉽 함수 값을 적용하기 위해 아래와 같이 수정한다.

$$W \leftarrow W + | u_{1j} - u_{2j} |^m c x_j \quad (2)$$

x_j가 모두 그 소속이 애매한 벡터라면 weight를 보정하는 값은 0이 된다. 정의된 조건에 따라 퍼지 멤버쉽을 할당하면[2].

FOR all pattern x_k

IF x_k **is class1**

$$u_{1k} = \frac{\exp(f(d_2 - d_1)/d) - \exp(-f)}{2(\exp(f) - \exp(-f))}$$

$$u_{2k} = 1 - u_{1k}$$

Else

$$u_{1k} = 1 - u_{2k}$$

$$u_{2k} = \frac{\exp(f(d_1 - d_2)/d) - \exp(-f)}{2(\exp(f) - \exp(-f))}$$

END IF

END FOR

여기서, $d_1(d_2)$은 벡터로부터 클래스1(2)의 평균까지의 거리이다. 그리고 d는 두 클래스의 평균값 사이의 거리이다. 상수 f는 멤버쉽값이 0.5를 향해 감소하는 비율을 제어하는데 이 값은 반드시 양수이어야 한다. 퍼지 멤버쉽 분할이 생성되고 난 후, 베타영역을 설정하여 선형분류 불가능한 부분을 제외하고 weight를 보정하게된다.

$$\text{BETA} = \frac{1 - e^{-f}}{2(e^f - e^{-f})} + \epsilon \text{ where } \epsilon \geq 0. \quad (3)$$

BETA영역의 설정으로 인해 선형분류가 불가능한 벡터들을 weight변화시킬 때 제외시킴으로써 언제나 수렴하는 결과를 얻을 수 있다[2]. 하지만 퍼지 퍼셉트론에서 멤버쉽을 구하는 식이 프로토타입과 패턴사이의 거리에 의존하므로 패턴의 분포가 원형이 아닌 경우에서는 만족스럽지 못한 결과를 얻게된다. 또 BETA영역의 설정으로 최적의 경계면 함수를 찾지 못하더라도 수렴하는 결과를 나타낸다. 이는 각 패턴의 소속도 측정에 있어, 상대적인 클래스 평균과의 거리에 의해 할당받은 멤버쉽 값을 사용하기 때문에 각 패턴의 weight 보정과정에 대한 기여도가 실제 패턴의 클래스에 전형성과는 차이가 있기 때문이다.

3. 제안한 Possibilistic 퍼셉트론

퍼지 분류는 각 반복연산과정에서 하나의 주어진 클래스에 하나의 패턴벡터가 전적으로 기여할 필요가 없다는 관점에서 전통적인 Crisp방법에 비해 장점이 있는 것으로 알려져 왔다. 하지만 fuzzy는 한 패턴의 각 클래스에 해당되는 멤버쉽의 합이 1이 되게 확률적인 제약을 사용하였다. 이러한 제약에 의해 생성되는 상대적인 수로 멤버쉽값을 할당하기 때문에 패턴의 전형성을 고려할 수 없다. 멤버쉽의

Possibilistic한 해석을 용이하게 하기 위해 멤버쉽의 sum에 대한 제한을 두지 않는 방법이다 [3].

기본적인 과정은 fuzzy 퍼셉트론과 동일하지만, weight를 보정할 때 사용하게 되는 멤버쉽 값을 할당하는데 있어서 패턴의 전형성을 고려한다는 것이 다르다. 이러한 요구를 만족하는 *PCM*의 objective function은 다음과 같이 공식화된다.

$$J_m(L,U)=\sum_{i=1}^{C}\sum_{j=1}^{N}(u_{ij})^m d_{ij}^2 + \sum_{i=1}^{C}\eta_i\sum_{j=1}^{N}(1-u_{ij})^m \quad (4)$$

여기서 η_i 는 적절한 양수이다. 첫 번째 항은 패턴 벡터에서 프로토타입까지의 거리는 가능한 한 가까울 것을 요구한다. 이에 반해 두 번째 항에서는 거리가 가까울수록 멤버쉽 u_{ij} 에게 가능한 한 큰 값을 갖도록 강요한다. 이와 같이하여 패턴 데이터의 전형성을 고려할 수 있다.

$$u_{ij}=\frac{1}{1+\left(\frac{d_{ij}^2}{\eta_i}\right)^{\frac{1}{m-1}}} \quad (5)$$

이와같이, 각 iteration에서 멤버쉽값 u_{ij}의 업데이트된 값은 x_j와 β_i의 거리에 의존한다. 프로토타입의 적합성의 관점에서, 클래스의 한 점의 멤버쉽은 그 클래스의 프로토타입으로 부터 거리에 의해 결정되어야만 하고 다른 클래스들과 겹치지 않아야 한다.

식 (6)은 모든 feature 포인트들 x_j들로 구성됨을 설명하는 도메인에 대한 β_i 클래스를 위한 Possibility distribution function를 정의한다. m 값은 퍼지화 정도와 멤버쉽 함수(Possibility distribution)의 모양을 결정한다. m이 1에 가까워질수록 패턴의 퍼지화정도가 낮아지게 되고, m이 무한대에 접근할 때 퍼지화 정도가 가장 커진다. 실제 각종 패턴인식 문제에 대해서 m=2 로 설정했을때 비교적 좋은 결과를 나타낸다. η_i 의 값은 클래스에서 멤버쉽값이 0.5가되는 점의 거리를 결정한다. 따라서, 각 클래스를 위한 멤버쉽 함수의 바람직한 대역폭에 의존하여 선택되는 것이 필요하다.

$$\eta_i=K\frac{\sum_{j=1}^{N}u_{ij}^m d_{ij}^2}{\sum_{j=1}^{N}u_{ij}^m} \quad (6)$$

이러한 선택은 η_i가 클래스 β_i의 평균 퍼지 클래스간 거리에 적절하게 만든다.

Possibilistic 퍼셉트론의 결과는 다른 클래스 분류 기법들처럼 초기화에 따라 결과가 크게 달라질 수 있다. Possibilistic 퍼셉트론에서, 클래스들은 많은 이동성을 갖지 못한다. 그 이유는 각 데이터들이 단지 하나의 클래스에만 소속되기 때문이다. 그래서, Possibilistic 퍼셉트론에서도 전체의 최소값으로 수렴되기 위해 합리적인 초기화방법이 요구된다. 퍼지 퍼셉트론은 Possibilistic 퍼셉트론과 결과된 퍼지 분할에 사용된 η_i를 측정하는데 훌륭한 초기화 요소들을 공급할 수 있고 새로 얻어진 멤버쉽 값들에 의해 weight 보정값이 패턴의 전형성을 고려한 값으로 바꿔게 되는 것이다. 이후의 실험에서 제안한 Possibilistic 퍼셉트론 알고리즘의 유용성을 보여주고 있다.

III. 실험결과 및 분석

본 논문에 제시된 방법의 효과를 측정하기 위해, Possibilistic과 퍼지, 그리고 Crisp 퍼셉트론을 여러 종류의 패턴 데이타에 대해 수행하고, 그 결과를 비교한다. 패턴집합은 아이리스, two class, ellipse 데이타들에 대해 각각 퍼셉트론을 수행하였다. 아이리스 데이터는 클래스가 3개인 데이터이므로 클래스 2,3만이 선형분류가불가능하기 때문에 이들 클래스만을 사용하여 실험하였다. Crisp과 Possibilistic 퍼셉트론은 수렴하지 않으므로 종료조건을 weight의 변화가 없을 때가 아닌 최대 반복 횟수를 정해놓고 그 중에서 에러가 가장 적을 때의 weight를 찾아 1-Jack knife를 수행하여 confusion matrix를 비교하였다. 각 표에서는, fuzzifier f의 값은 3.0으로, m은 2.0으로, c는 2, E는 0.02로 각각 설정되었다.

표 1. Twoclass 데이터의 confusion matrix비교

Two class	Crisp		Fuzzy		Possibilistic	
	class1	class2	class1	class2	class1	class2
class1	86	35	114	7	96	25
class2	0	121	20	101	3	118

표 2. Iris 데이터의 confusion matrix비교

Iris	Crisp		Fuzzy		Possibilistic	
	class1	class2	class1	class2	class1	class2
class1	43	7	49	1	46	4
class2	2	48	5	45	0	50

위의 실험 결과를 보면 crisp에 비해 fuzzy와 Possibilistic이 더 나은 결과를 나타내는 것을 볼 수. 있다. 패턴 데이터의 클래스 모양이 선형으로 구분이 불가능한 경우의 비교는 다른 결과를 보여준다. .아래 결과는 ellipse데이타의 confusion matrix이다.

표 3. ellipse 데이터의 confusion matrix비교

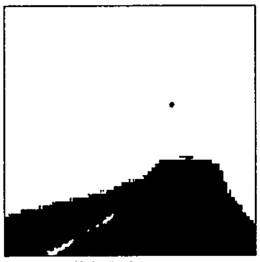
Ellipse	Crisp		Fuzzy		Possibilistic	
	class1	class2	class1	class2	class1	class2
class1	37	13	31	19	44	6
class2	1	49	27	23	1	49

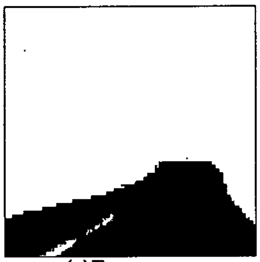
퍼지의 경우는 약 50%밖에 인식하지 못하는 결과를 나타내었다. 이는 평균까지의 거리함수에 기초한 멤버쉽 할당으로 인해 선으로 분류가 불가능한 패턴 클래스에 대해서는 올바른 구분을 할 수 없음을 나타낸다.

모두 최대 반복 횟수를 1000번으로 하고 전혀, 수렴하지 못하는 crisp 퍼셉트론의 종료를 위한 충분한 반복연산에 대한 제한이다. 이러한 제한내에서의 Possibilistic 퍼셉트론은 퍼지 퍼셉트론의 경우처럼 하나의 경계면 함수를 완벽히 찾지는 못한다. 그러나 충분한 반복연산후의 weight의 보정범위는 충분히 매우 작기 때문에 crisp에 비해서는 좋은 성능을 보이게 된다. 또한 퍼지 퍼셉트론으로서 경계면이 최적화되지 못해 완벽하게 수렴하지는 않았지만 Possibilistic 퍼셉트론의 결과 경계면이 더 좋은 성능을 보였다.

이러한 패턴분류자들을 숲과 길의 두 부분으로 분류가 가능한 자연영상에 대해 영상분할 실험에 적용해 보았다. 메디안필터, excess green, homogeneity, entropy의 특징 영상을 이용하여 각 클래스 별로 100개의 패턴을 추출하여 weight 함수를 찾아 영상 데이타를 분할하였다. 아래그림은 영상 분할의 결과이다.


(a)원래영상


(b)Crisp


(c)Fuzzy

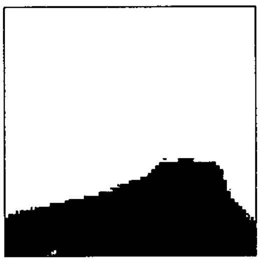

(d)Possibilistic

그림 1. 영상 분할의 예

원래 영상을 보면 길과 숲 이렇게 2개로 분할 가능한 이미지이다. 이를 분할 알고리즘에 적용하면 그림1과 같은 결과를 얻을 수 있다. 3개의 퍼셉트론 모두가 어느 정도 숲의 영상은 깨끗하게 찾아내었다. 하지만 길 영상분할 결과를 자세히 보면 crisp과 퍼지에서는 길의 중앙선부분이 숲 영상으로 분할된 것을 알 수 있다. 이에 반해 Possibilistic 퍼셉트론은 중앙선부분도 더욱더 확실한 길로 인식되는 결과를 볼 수 있다.

Ⅳ. 결 론

퍼지 퍼셉트론은 선형분류불가능한 경우에 적절한 경계면을 찾지 못하는 crisp 퍼셉트론의 주된 문제점을 해결하는 방법으로서 제시되었다. 퍼지 기술들은 경계면 결정에 큰 역할을 하지 못하는 패턴들을 찾아내고 이들을 경계면 계산에서 다른 패턴들보다 작은 weight를 부여하게 된다. 퍼지 퍼셉트론에서의 weight는 클래스 평균과의 상대적인 거리에 의한 값이다. 따라서 실제 패턴의 클래스에 따른 전형성이 고려된 값이라 볼 수 없다. 본 논문에서는 Possibilistic의 기술을 사용하여 멤버쉽 할당단계에서부터 해당클래스에 대한 샘플 벡터의 전형성에 직접 의존해 해당 벡터의 weight보정단계에 직접 영향을 미치도록 하였다. 퍼지 퍼셉트론은 평균까지의 거리함수에 기초한 멤버쉽 할당으로 인해 선으로 분류가 불가능한 패턴 클래스에 대해서는 만족스럽지 않은 결과를 나타내었다. 수렴이라는 측면에서 보면 퍼지 퍼셉트론이 가장 우수하지만 데이터의 클래스가 섞여 있는 경우 crisp과 퍼지에 비해 Possibilistic 퍼셉트론이 우수하다는 것을 알 수 있었다. 이는 nosie 패턴이 첨가된 데이터에서 더 유용할 것으로 생각된다. 그러므로, Possibilistic 퍼셉트론은 두개의 클래스가 서로 선형분류가능한 여부에 상관없이 두개의 클래스 사이에 바람직한 경계면을 생산하는 신뢰성 있는 방법이다. 본 논문에 제시된 알고리즘은 향후 뉴럴 네트웍을 사용하여 여러 클래스에 대해서도 동작할 수 있는 클래스 분류자로 확장 될 것이다.

Ⅳ. 참고문헌

[1] J. T. Tou and R. C. Gonzalez. *Pattern Recognition Principles. Reading.* MA : Addison-Wesley. 1974.

[2] J. M. Keller and D. J. Hunt "*Incorporating Fuzzy Membership Functions into the Perceptron Algorithm,*" IEEE Trans. PAMI, vol. 7, no. 6, pp. 693-699, 1985

[3] R. Krishnapuram and J. M. Keller. "*A Possibilistlic Approach to Clustering,*" IEEE Trans. Fuzzy systems, vol 1. no. 2, pp. 98-110, 1993

[4] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms.* NewYork: Plenum, 1981