韓國産業應用學會誌, 제5권, 제1호, 33-38, 2002.2
J. Korean Society of Industrial Application, Vol.5, No.1, 33-38, February, 2002

독립변수의 차원감소에 의한 Polynomial Adaline의 성능개선
Performance Improvement of Polynomial Adaline by Using Dimension Reduction of Independent Variables

조용현*
Yong-Hyun Cho*

< Abstract >

This paper proposes an efficient method for improving the performance of polynomial adaline using the dimension reduction of independent variables. The adaptive principal component analysis is applied for reducing the dimension by extracting efficiently the features of the given independent variables. It can be solved the problems due to high dimensional input data in the polynomial adaline that the principal component analysis converts input data into set of statistically independent features. The proposed polynomial adaline has been applied to classify the patterns. The simulation results shows that the proposed polynomial adaline has better performances of the classification for test patterns, in comparison with those using the conventional polynomial adaline. Also, it is affected less by the scope of the smoothing factor.

Key Words : adaptive principal component analysis, polynomial adaline, feature extraction, pattern classification.

1. 서 론

패턴인식은 임의로 주어지는 패턴을 적당한 클래스에 대응시키는 기능으로 패턴분류와 인식과정으로 구성된다. 그 중에서도 분류과정은 입력패턴의 전처리 과정을 거쳐 특징을 추출하고 군집화를 수행한 다음 표준패턴들을 생성하는 과정이다[1,2]. 이러한 패턴분류를 위해 기존의 통계적 방법이나 신경망을 이용하는 기법들이 많이 연구되었다[1-4]. 최근에는 인간의 두뇌를 모방한 계산방식을 따르는 신경망을 이용한 패턴 분류나 인식기법이 널리 이용되고 있다.

선형적인 입력패턴이 주어질 때 adaptive linear neuron(adaline)을 이용한 효과적인 분류기법이 제안되었다[4-6]. adaline에서 adaptive는 주어진 입력패턴에 대하여 많이 정확한 출력을 나타내도록 허용하는 가중치들을 수정하는 잘 정의된 과정이 존재한다는 의미다. 또한 linear 이란 출력이 입력패턴값의 간단한 선형함수로 주어진다는 것이다. 그러나 adaline은 입력패턴의 선형적인 분류만이 가능하며, 비선형 패턴분류에서는 비효율적이다. 이런 점을 개선하기 위해서

* 정회원, 대구가톨릭대학교 컴퓨터정보통신공학부,교수,工博

712-702 경북 경산시 하양읍 금락리 330

* Professor, School of Computer & Information Comm. Eng., Catholic Univ. of Daegu
E-mail : yhcho@cuth.cataegu.ac.kr

좀 더 복잡한 형태의 polynomial adaline (padaline)이 제안되었다[5,6]. padaline의 기본적인 이론은 adaline의 간단한 선형판별함수를 어떤 경우의 분류도 가능하게 하는 다항식의 판별함수로 대치한 것이다. 여기서 다항식은 어떤 선형함수보다도 더 복잡한 함수를 표현할 수 있다. 따라서 padaline은 adaline처럼 선형적으로 분리가능한 문제들에 제약을 받지 않는다. 특히 padaline은 역전파(backpropagation) 알고리즘의 신경망보다 분류나 인식에서 매우 빠른 특성이 있다[6]. 그러나 입력패턴에 따라 표현되는 다항식은 차원에 따른 계수의 계산이 요구된다. 이는 입력패턴의 차원이 증가하면 그에 따른 계수의 수도 증가하게 되고, 계산 또한 복잡하게 되어 상대적으로 오랜 학습시간이 요구된다. 따라서 입력패턴들 간의 상호관계를 잘 나타내는 주요특징들만 추출하여 계수를 계산하는 데이터로 이용한다면 padaline이 가지는 입력패턴의 고차원에 따른 문제들을 해결할 수 있을 것이다.

본 연구에서는 수치적인 데이터 집합의 통계적 속성을 이용하여 차원을 감소시키는 것으로 널리 알려진 적응적 주요성분분석(principal component analysis : PCA) 기법[3]으로 입력패턴의 특징을 추출한 다음, 추출된 특징을 padaline의 학습패턴으로 이용하는 새로운 기법을 제안한다. 이렇게 하면 PCA가 가지는 대용량의 입력데이터를 통계적으로 독립인 특징들의 집합으로 변환시키는 장점과 padaline이 가지는 우수한 속성을 그대로 살릴 수 있다. 제안된 기법의 padaline을 3개와 5개의 입력패턴을 각각 가지는 분류문제에 적용하여 시뮬레이션하고 그 타당성을 확인하였으며, 기존의 padaline과 그 성능을 비교 고찰하였다.

2. 주요성분분석에 의한 입력패턴의 차원감소

2.1 Polynomial Adaline

학습패턴의 저장이 불필요하며 구현을 위한 하드웨어가 간단한 padaline은 분류문제의 해결을 위해서 널리 이용된다[5,6]. 특히 대단히 많은 데이터 베이스나 새로운 학습패턴의 연속적인 흐름이 큰 문제에서는 더욱 효과적이다. 또한 학습속도보다는 시험속도가 좀 더 중요한 문제에서는 우수한 특성을 가진다. padaline을 위한 학습규칙은 지수함수의 합의 항으로 표현된 확률적 신경망(probabilistic neural network)에서 결정한계(decision boundary)의 테일러 급수전개로부터 유도된다[6]. 학습은 학습패턴에 근거한 다항식의 계수를 계산하는 것이며, 그 결과는 다차원 측정공간에서 결정면을 나타내는 일반적인 다항식으로 표현된다. 또한 알고리즘은 반복적이지 않으며, 계수의 최종값은 학습데이터가 한 번 통과하면 결정된다.

한편 2개의 범주 A와 B에 속하는 P 차원의 입력패턴 x_p (i = 1, 2, ⋯, P)에 대해서 표현되는 일반적인 다항식 P(x)는

$$\begin{aligned} P(X) = {} & D_{0\cdots 0} + D_{10\cdots 0}\, X_1 + D_{01\cdots 0}\, X_2 + \cdots + D_{00\cdots 01}\, X_P \\ & + D_{20\cdots 0}\, X_1^2 + \cdots + D_{0\cdots 02}\, X_P^2 + D_{110\cdots 0}\, X_1\, X_2 \\ & + \cdots + D_{z1z2\cdots zP}\, X_1^{z1} X_2^{z2} \cdots X_P^{zP} + \cdots \end{aligned} \tag{1}$$

와 같이 표현된다. 여기서 $z_1, \cdots z_P$ 는 각각 입력패턴의 첫 번째와 P번째 요소들에 대한 일치하는 멱수들이다. 이때 범주 A와 B의 입력패턴들이 각각 m개와 n개로 구성되어 있다면, 식 (1)에서 다항식의 계수 $D_{z1z2\cdots zP}$는 다음과 같이 계산될 수 있다. 즉,

$$D_{z1z2\cdots zP} = \frac{1}{z_1!\, z_2! \cdots z_P!\, s^{2h}} \left[\frac{1}{m} \sum_{i=1}^{m} X_{A_{1i}}^{z_1} X_{A_{2i}}^{z_2} \cdots X_{A_{pi}}^{z_p}\, \mathrm{EXP}_{A_i} - \frac{K}{n} \sum_{i=1}^{m} X_{B_{1i}}^{z_1} X_{B_{2i}}^{z_2} \cdots X_{B_{pi}}^{z_p}\, \mathrm{EXP}_{B_i} \right]$$

$$\mathrm{EXP}_{A_i} = \exp\left(\frac{-L_{Ai}}{2s^2} \right),\ L_{Ai} = X_{Ai}^T X_{Ai}$$

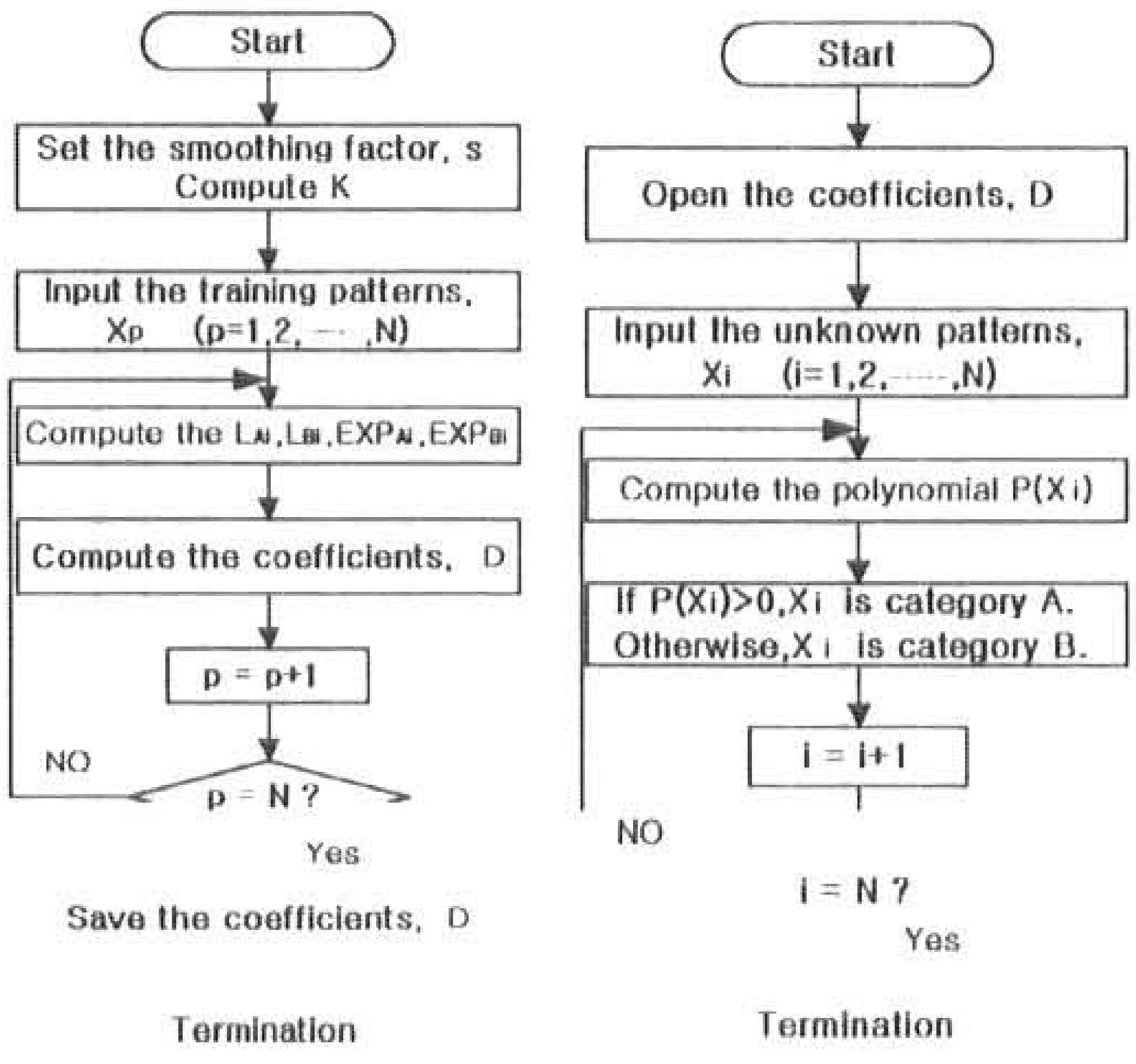
$$\mathrm{EXP}_{B_i} = \exp\left(\frac{-L_{Bi}}{2s^2} \right),\ L_{Bi} = X_{Bi}^T X_{Bi}$$

$$h = \sum_{j=1}^{P} z_j \tag{2}$$

이다. 여기서 T는 전치를 나타낸다. 식에서 $K = \frac{n\, l_B}{m\, l_A}$로 결정되며, 이때 l_A와 l_B는 각각 범주 A를 범주 B로와 범주 B를 범주 A로 결정할 손실이다. 또한 s는 평활요소(smoothing factor)로 그 값이 클수록 좀 더 복잡한 결정면을 형성하는 다항식이 되며, 작은 값일수록 좀 더 평활된 면을 형성한다. 특히 P(X)=0는 범주 A와 B 사이를 결정하는 결정면의 수학적 표현식이다. 따라서 입력패턴 X로 표현되는 다항식 P(X)를 계산한 다음, P(X)>0이면 입력패턴은 범주 A에 속하고, 그렇지 않고 P(X)<0이면 범주 B에 속하게 된다.

이상에서 서술된 식으로 표현되는 padaline의 학습과정과 분류과정의 흐름을 각각 도시하면

다음의 Fig. 1과 같이 나타낼 수 있다. 그림에



(a) 학습과정 (b) 분류과정

Fig. 1 Padaline의 학습과 분류과정의 흐름도

서 학습은 알려진 패턴으로부터 다항식의 계수를 구하는 과정이고, 분류는 모르는 입력패턴에 대한 분류과정이다. 한편 padaline에서 보면 계수는 일반 신경망에서의 가변되는 연결가중치(weights)와 동일하며, 최종적으로 양자화기를 이용하여 출력을 결정하게 된다.

2.2 적응적 특징추출을 위한 PCA

PCA는 n차원 입력공간의 데이터를 m차원 출력공간의 데이터로 투영시키는 것이다[3]. 여기서 m ≪n 이며, 이는 입력데이터 벡터의 대부분 내부정보를 유지하도록 차원의 감소를 얻는 것이다. 자기상관행렬(auto-correlation matrix) $\mathbf{R_{xx}} = \langle \mathbf{xx}^T \rangle$를 가진 평균이 영인 입력벡터 $\mathbf{x} = [\, x_1, x_2, \cdots, x_n \,]^T$에 대해서 생각해 보자. 여기서 $\langle \mathbf{x} \rangle$는 기대치를 나타낸다. 또한 $\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_m$ 이 $\mathbf{R_{xx}}$ 의 고유벡터와 직교되는 연결가중치 벡터라 할 때, $\mathbf{w}_1 = [\, w_{11}, w_{12}, \cdots, w_{1n} \,]^T$는 가장 큰 고유치 λ_1과 일치하며, $\mathbf{w}_2 = [\, w_{21}, w_{22}, \cdots, w_{2n} \,]^T$는 두 번째로 큰 고유치 λ_2와 $\mathbf{w}_n = [\, w_{n1}, w_{n2}, \cdots, w_{nn} \,]^T$는 가장 작은 고유치 λ_n과 각각 일치한다. 이상의 관계를 행렬방정식으로 나타내면 다음과 같다. 즉,

$$\mathbf{R_{xx}}\, \mathbf{w}_j = \lambda_j\, \mathbf{w}_j \quad (j = 1,2, \cdots, n) \tag{3}$$

이다.

여기서 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$ 이다.

주어진 입력벡터 **x**를 위한 첫 번째 m개의 주요 특징벡터 **y**는 다음의 선형변환식으로 나타낼 수 있다. 즉,

$$\mathbf{y} = \mathbf{W}\, \mathbf{x} \tag{4}$$

이다. 여기서 $\mathbf{W} = [\, \mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_m \,]^T \in \mathbb{R}^{m \times n}$ 이며, T는 전치를 나타낸다. 이 식은 행렬 **W**의 행이 가장 큰 고유치와 일치하는 상관행렬 $\mathbf{R_{xx}}$의 고유벡터임을 의미한다. 다시 말하면, 입력 데이터 공간에서 m차원의 주요특징을 나타내는 부공간은 $\mathbf{R_{xx}}$의 m개 주요 고유벡터에 의해 구성된 부공간으로 정의된다. 결국 PCA는 $\langle \| \mathbf{w}_j^T \mathbf{x} \|_2^2 \rangle$가 최대인 고유벡터 $\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_m$의 방향을 찾는 것이다.

Sanger 등[3]은 일반화된(generalized) 헤비안 규칙을 이용함으로써 정상과정의 m개의 가장 중요한 주요특징들을 계산하기 위한 다중 뉴런 모델을 제안하였다. 한편, Foldiak[3]은 망의 입력과 출력사이의 연결가중치 경신에는 정규화된 헤비안규칙을 이용하고, 망의 출력사이의 측면연결 가중치 경신에는 반 헤비안규칙(anti-Hebbian rule)을 함께 이용한 학습알고리즘을 제안하였다. 일반적으로 Sanger 등에 의해 제안된 학습규칙보다는 Foldiak 에 의해 제안된 학습규칙이 수렴속도 면에서 더 우수한 것으로 알려 있다[3].

Fig. 2는 n개의 입력뉴런과 m개의 출력뉴런으로 구성된 입력과 출력뉴런간 및 출력뉴런 상호간의 측면연결을 가진 단층신경망의 구조이다.

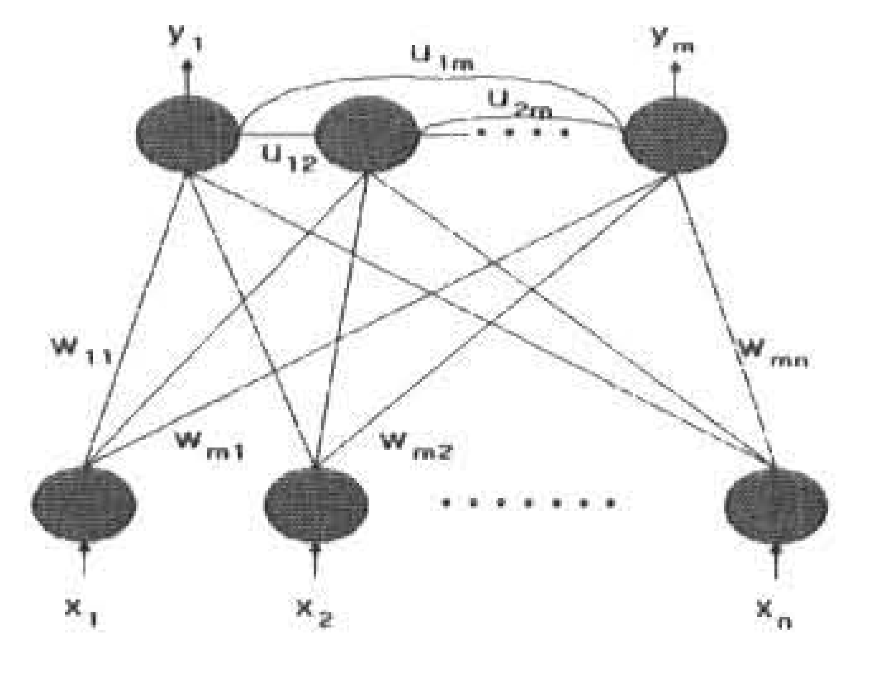


Fig. 2 m개의 첫 번째 주요특징 추출을 위한 측면연결의 단층신경망

그림에서 입력과 출력의 관계를 나타내면 다음과 같다. 즉,

$$y_i = \sum_{j=1}^{n} w_{ij}\, x_j + \sum_{q=1}^{i-1} u_{iq}\, y_q \;,(i = 1,2,\cdots,m) \qquad (5)$$

이다. 여기서 w_{ij}는 입력뉴런과 출력뉴런을 연결하는 연결가중치이고, u_{iq}는 출력뉴런 상호간의 측면연결 가중치이다. 이때 출력뉴런 간의 상호연결을 보면 뉴런 i는 다만 q<i인 뉴런에만 연결된다. 따라서 신경망을 학습시켜 입출력 뉴런간 및 출력뉴런 상호간의 연결가중치를 각각 구함으로써 주어진 n개의 입력벡터 **x**로부터 m개의 주요특징벡터 즉, 출력벡터 **y**를 구할 수 있다. 입력과 출력뉴런 간의 연결가중치 경신에는 정규화된 헤비안규칙을 이용하고 출력뉴런 간의 측면 연결가중치 경신에는 반 헤비안규칙을 이용하는 Foldiak의 학습알고리즘은 다음과 같다. 먼저 입출력 뉴런간의 연결가중치 w_{ij}의 경신규칙을 살펴보면 다음과 같다. 즉,

$$w_{ij}(t+1) = w_{ij}(t) + \eta\ [\ y_i(t)\ x_j(t) - w_{ij}(t)\ y_i(t)^2],$$
$$(i=1,2,\cdots,m,\ j=1,2,\cdots,n) \qquad (6)$$

이다. 또한 출력뉴런 간의 측면 연결가중치 u_{iq}의 경신규칙은 다음과 같다. 즉,

$$u_{iq}(t+1) = u_{iq}(t) + \rho\ \ y_i(t)\ y_q(t),\ (i > q) \qquad (7)$$

이다. 식 (6)의 η와 (7)의 ρ는 학습율이다. 따라서 식 (6)과 (7)을 이용하여 입력과 출력뉴런 간의 연결가중치 및 출력뉴런 상호간의 측면 연결가중치를 각각 경신시켜 식 (5)에 대입하면 m개의 주요특징들을 추출할 수 있다.

결국 PCA는 높은 차원의 입력공간을 더 낮은 차원의 표현공간으로 사상시켜 입력데이터가 가지는 두드러진 특징들을 추출하는 기법이다. 이는 입력 데이터 내에 존재하는 주요특징들을 추출함으로써 입력의 개수를 감소시키는데 이용될 수 있다. 따라서 학습데이터의 입력패턴 각각에 적응적 PCA를 이용하여 특징을 구한 다음 이를 padaline의 입력으로 이용한다. 이렇게 하면 입력패턴의 고차원에 따른 다항식의 계수 계산시간이 줄어들어 padaline이 가지는 우수한 속성을 더욱 더 잘 살릴 수 있을 것이다.

3. 시뮬레이션 결과 및 분석

제안된 기법의 성능을 평가하기 위하여 Foldiak 학습알고리즘의 단층신경망과 padaline을 각각 구성하였다. 특징추출을 위한 단층신경망의 입력뉴런과 출력 뉴런사이의 초기 연결가중치는 각각 랜덤시드(random seed)를 이용하여 -1에서 +1 사이의 임의 값으로 설정하였다. 학습은 전체 반복회수가 10,000회 이상이거나 전체 오차값이 허용치 10^{-4} 이하일 때 종료되도록 하였다.

제안된 기법을 이용한 padaline을 3개의 입력변수를 가지는 10개의 학습패턴과 5개의 입력변수를 가지는 100개의 학습패턴의 2 범주(A,B)를 가진 분류문제를 대상으로 펜티엄Ⅲ-700 컴퓨터를 이용하여 시뮬레이션한 후 그 타당성을 확인하였으며, 기존의 입력패턴을 그대로 이용하는 padaline과 그 성능을 비교 고찰하였다. 또한 실험에서는 10개의 학습패턴에서 범주 A와 B에 속하는 패턴의 수는 각각 6개와 4개로 하였으며, 그 중에서 4개와 2개의 총 6개의 패턴은 학습패턴으로 나머지 4개의 패턴은 시험패턴으로 이용하였다. 한편 100개의 학습패턴에서도 범주 A와 B에 속하는 패턴의 수는 각각 60과 40개로 하였으며, 그 중에서 40개와 20개의 총 60개의 패턴은 학습패턴으로 나머지 40개의 패턴은 시험패턴으로 이용하였다.

Table 1. 평활요소 s의 변화에 따른 분류율 비교

평활요소,s	기존 padaline (입력변수 3개)	제안된 padaline (입력변수 2개)
0.1	60	60
0.5	60	70
1.0	70	80
1.5	80	90
2.0	80	90
2.5	90	90
3.0	80	90
5.0	50	60
10.0	40	40

Table 1은 평활요소 s와 3개 입력변수를 가지는 기존 padaline과 2개의 입력변수를 가지는 제안된 padaline에 의한 분류성능을 나타낸 것이다. 표에서 보면 평활요소의 값에 기존의 padaline과 제안된 padaline의 분류성능이 의존됨을 알 수 있다. 특히, 기존의 padaline에서는 평활요소의 범위가 2.0에서 3.0사이의 값일 때와 제안된 padaline에서는 1.0에서 3.0사이일 때가 상대적으로 우수한 성능이 있다. 이는 제안

된 padaline이 기존 padaline보다 평활요소의 설정에서도 용이함을 확인할 수 있다. 특히 평활요소의 값이 2.5일 때는 서로 동일한 분류성능을 보이나 다른 값에서는 제안된 padaline이 기존의 padaline보다 전체적으로 우수한 분류성능이 있음을 알 수 있다.

Fig. 3은 평활요소 s의 값을 2.5로 고정하고 입력변수 3개와 PCA를 적용한 입력변수 2개 및 1개로 할 때의 분류율을 나타낸 것이다. 여기서 입력변수 3개는 기존의 padaline을 의미한다. 그림에서 보면 입력변수가 3개와 2개일 때는 동일한 성능을 보이나 하지만 입력변수의 수를 줄이면 상대적으로 padaline을 학습시키는 시간은 줄어든다. 한편 PCA를 적용한 입력변수 1개일 때는 오히려 성능이 감소됨을 확인할 수 있다. 이는 3개의 입력변수들의 속성이 1개의 변수에 충분히 반영되지 못하기 때문으로 추측된다.

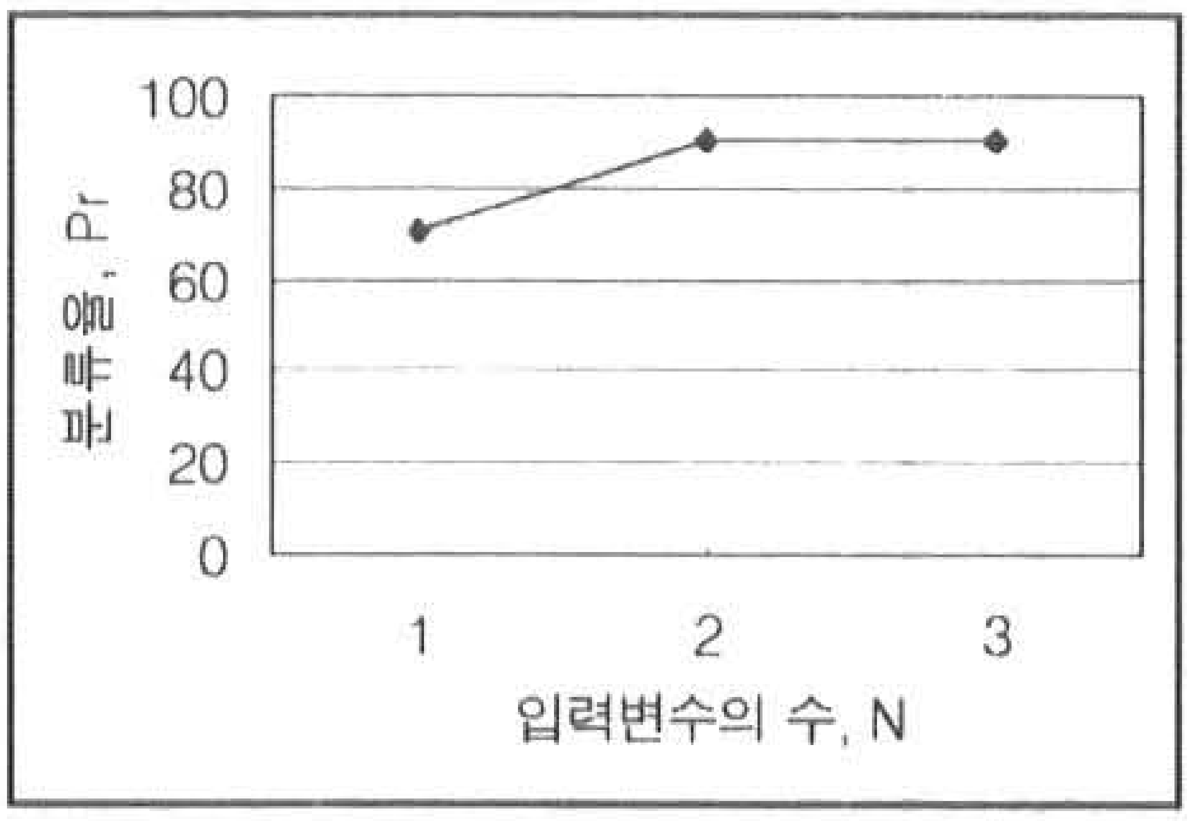


Fig. 3. 입력변수의 수에 따른 분류성능

Table 1. 평활요소 s의 변화에 따른 분류율 비교

평활요소,s	기존 padaline (입력변수 5개)	제안된 padaline (입력변수 4개)
0.1	41.8	55.2
0.5	51.3	65.7
1.0	58.4	74.3
1.5	73.1	81.6
2.0	84.6	89.3
2.5	81.5	87.2
3.0	61.4	75.9
5.0	50.9	62.5
10.0	21.2	20.7

Table 2는 평활요소 s와 5개 입력변수를 가지는 기존 padaline과 4개의 입력변수를 가지는 제안된 padaline에 의한 분류성능을 나타낸 것이다. 표에서도 보면 평활요소의 값에 기존의 padaline과 제안된 padaline의 분류성능이 의존됨을 알 수 있다. 특히, 기존의 padaline에서는 평활요소의 범위가 1.5에서 2.5 사이의 값일 때와 제안된 padaline에서는 1.0에서 3.0사이일 때가 상대적으로 우수한 성능이 있다. 이는 제안된 padaline이 기존 padaline보다 평활요소의 설정도 용이함을 다시 한번 보여준다. 여기서도 제안된 padaline의 경우 기존의 padaline보다 전체적으로 우수한 분류성능이 있음을 알 수 있다.

Fig. 4는 평활요소 s의 값을 2로 고정하고 제안된 padaline에서 입력변수를 5개, 4개, 3개, 2개, 그리고 1개로 할 때의 분류율을 나타낸 것이다. 여기에서도 입력변수 5개는 기존의 padaline을 의미한다. 그림에서 보면 입력변수가 4개일 때 가장 우수한 분류성능을 보이며, 입력변수를 줄일수록 상대적인 성능도 감소함을 확인할 수 있다. 이도 줄인 입력변수들이 문제의 학습패턴에 대한 각 입력변수의 속성을 충분히 반영하지 못하기 때문이다.

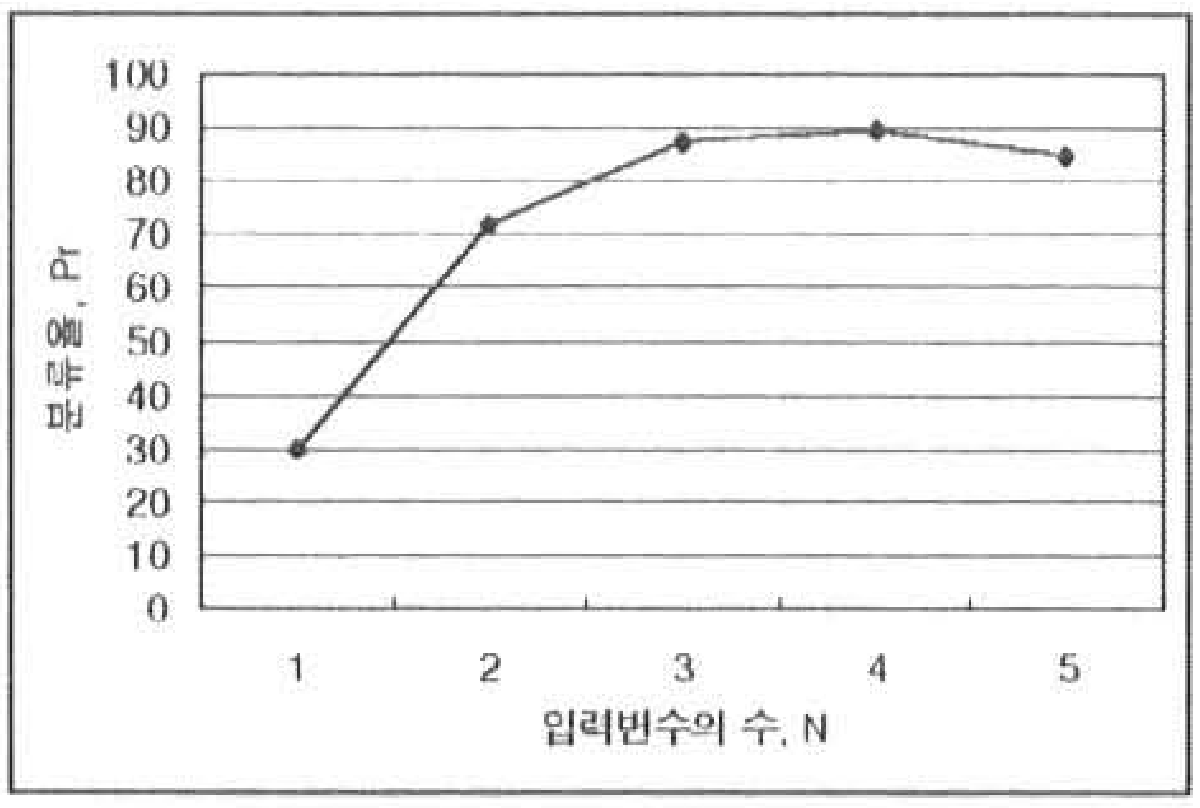


Fig. 4 입력변수의 수에 따른 분류성능

4. 결 론

본 논문에서는 적응적 학습알고리즘의 PCA 기법으로 polynomial adaline의 성능을 개선하는 방법을 제안하였다. 제안된 방법에서는 적응적 학습알고리즘의 PCA 기법을 이용하여 입력변수의 특징을 추출하고 이를 polynomial adaline의 학습데이터로 이용하였다. 이는 주요성분분석 기법이 가지는 대용량의 입력 데이터를 통계적으로 독립인 특징들의 집합으로 변환시키

는 속성을 살려 입력데이터의 차원을 감소시킴으로서 고차원의 데이터에 따른 polynomial adaline이 가지는 제약을 해결하기 위함이다.

제안된 기법의 polynomial adaline을 3개의 입력변수를 가진 10개 패턴 및 5개의 입력변수를 가진 100개의 패턴들을 분류하는 문제에 적용하여 시뮬레이션한 결과, 기존의 polynomial adaline에 의한 결과보다 더 우수한 분류성능이 있음을 확인할 수 있었다. 그리고 커널함수의 평활요소 설정 면에서도 우수한 특성이 있음을 확인할 수 있었다.

향후 제안된 기법의 padaline을 좀 다양한 문제에 적용하며, 효율적인 응용을 위하여 제안된 padaline에서 성능결정 요소들에 대한 연구도 뒤따라야 할 것이다.

참고문헌

1) E. Gose, R. Johnsonbaugh, and S. Jost, : Pattern Recognition and Image Analysis, Prentice-Hall, London, (1996)

2) 김상운 : 패턴인식 입문, 홍릉과학출판사,(1997)

3) S. Haykin, : Neural Networks : A Comprehensive Foundation, IEEE Press, New York, (1994)

4) J. A. Freeman and D.M. Skapura, Neural Networks : Algorithms, Applications, and Programming Techniques, Addison-Wesley, (1991)

5) M. Caudill and C. Butler, : Understanding Neural Networks, MIT express, Cambridge, (1992)

6) D. F. Specht, : Probabilistic Neural Networks and the Polynomial Adaline as Complementary Techniques for Classification, IEEE Trans. on Neural Networks,Vol.1, No.1, pp.111-121, Mar. (1990)

(2001년 9월 6일 접수, 2002년 2월 20일 채택)