

# 정제된 데이터를 이용한 신경망의 설계 자동화에 관한 연구

An automated neural network design from a well  
organized data set

백 주현\*, 김 홍기\*\*

(Ju-Hyun Back\*, Hong-Gi Kim\*\*)

군장대학 멀티미디어과\*, 충북대학교 전자계산학과\*\*

## ABSTRACT

본 논문에서는 공학적인 체계성을 갖고 초기 연결 가중치 및 임계치를 결정해 주면서, 학습까지 가능한 신경망을 제안한다. 기존의 오류 역전파 신경망을 적용할 때 경험에 의하여 은닉층 노드 수를 결정하거나 임의의 실수 값으로 초기 연결 가중치 및 임계값을 주었을 때 자주 발생하는 학습 마비 현상을 피할 수 있고, Bose 가 제안한 Voronoi 공간 분류에 의한 신경망 구성에서 학습이 불가능하다는 제한적인 단점을 보완하였다. 초기 가중치는 Voronoi 공간 분류가 이루어져 있다고 할 때 Bose 가 제안한 초기 가중치 결정법을 개선하여 사용하고, Bose 의 경우 신경망 노드가 Step function 을 이용하여 정보를 전달하였으나 본 연구에서는 학습이 가능한 함수인 Sigmoid function 을 이용하였다. 제안된 새로운 신경망의 성능 및 효율성을 비교하기 위하여 선형분리가 불가능한 XOR 문제를 실험한 결과, 기존의 학습 가능한 EBP 에서 허용오차 0.05 수준일 때 80 % 정도 학습 마비 현상이 발생하였던 심각한 문제점을 보완할 수 있었고, 또한 학습 속도면에서 8 ~ 9 배 정도 빠른 성능을 나타내었다.

## I. 서론

선형적으로 분리 가능한 패턴 분류의 경우 단층 퍼셉트론 신경망으로 해결할 수 있다.[1] 그러나 일반적으로 선형 분리가 불가능한 경우 다층 퍼셉트론 네트워크(Multilayer Perceptron Network)으로 구성하는데, 이 중에서 일반화된 델타 학습규칙을 활용한 오류 역전파 신경망(Error Backpropagation Network, EBP)은 단순한

구조와 우수한 패턴 분류 능력으로 패턴인식 분야에서 많은 관심을 끌고 있다. 그러나 이 방법은 입력노드의 수와 출력노드의 수를 문제의 특성에 따라 결정하는 것이 가능하지만, 은닉층 노드의 수는 전적으로 설계자의 경험에 의존하고 신경망의 초기 연결 가중치 및 임계값은 일정 영역의 실수에서 임의로 선택하게 되어, 초기 설정에 의해 학습속도가 현저히 떨어지거나 지역 최소치(Local Minima)에 빠질 가

능성으로 인해 학습에 실패할 수 있다는 것은 이미 잘 알려진 사실이다. 이를 해결하기 위하여 경험적 지식에 의한 방법으로 경사치의 재사용, 동적 학습률 적용, 또는 학습 순서 및 입력 자료 표현의 재조정, 학습규칙인 델타규칙의 수정등의 많은 연구가 있어 왔다.[ 2, 3, 4, 5, 6, 7, 8 ]

한편 Bose 는 표준 패턴간의 거리를 유클리드 거리로 가정하여 Voronoi 공간 분류를 이용하여 초기 가중치를 결정하는 방법과 은닉층의 노드 수를 정할 수 있는 새로운 신경망 구조를 제안하였는데 [ 9 ], 여기에서는 정보를 전달하는 방법으로 Step function 을 이용하므로 새로운 데이터에 대한 학습 능력이 전혀 없다는 문제점이 있다. 따라서 본 연구에서는 학습 가능한 기존의 EBP 에서 사용되는 신경망 노드정보 전달함수인 Sigmoid function 을 이용하여 Bose 가 제안한 신경망의 구조를 개선하고자 한다.

본 논문의 구성은 다음과 같다. 2 장에서는 Voronoi Diagram(이하 VoD)을 이용하여 초기 가중치 및 은닉층의 노드 수를 결정하는 Bose 가 제안한 새로운 신경망 설계 방법을 기술하고 제 3 장에서는 이를 기초로 실제 XOR 문제에 있어 본 연구에서 제안된 sigmoid function 을 적용하는 학습 가능한 신경망 구조( 이하 VoD\_EBP)와 EBP 와의 성능 비교를 위한 실험에서 학습 진행에 따른 결과를 그래프를 통해 제시하고 마지막으로 결론에서는 제안된 VoD\_EBP 의 장점 및 향후 연구 방향에 대해서 기술한다.

## II. Voronoi Diagram 을 이용한 신경망 설계

기하와 관련된 여러 문제들 예를들어 평면상에 놓여 있는 n 개의 점들에 대하여 각 점에서 가장 가까운 점(nearest neighbor)을 찾는 문제, 점 사이의 거리가 가장 가까운 두 점(two closest point pair)을 찾는 문제, EMST(Euclidean Minimum Spanning Tree)를 찾는 문제, convex hull 을 구한다든지 하는 문제들은 전산학에서

Computational geometry 라는 한 분야로 연구되어 많은 사람들이 관심을 가져왔다.[ 10, 11, 12, 13 ]

VoD 는 일명 Thiessen 다각형 또는 Dirichlet tessellation 이라 불리우며 크게 nearest VoD 와 farthest VoD 으로 구분할 수 있는 데, 일반적으로 말하는 VoD 는 전자의 경우를 말한다. Nearest VoD 란 d-차원에 주어진 n 개의 점에 대해, d-차원의 전 영역을 n 개의 소영역(subdivision)으로 분할한 것으로, 평면상에 주어진 n 개의 점으로부터 VoD 를 구하는 알고리즘은 이미 많은 사람들에 의해 제시되었으며 [ 10, 13 ], 이를 이용하여 앞에서 제시한 여러 문제들을 해결할 수 있을 뿐만 아니라 클러스터링, 패턴인식등에 관련된 문제들에도 응용되고 있다.[ 14, 15 ]

d - 차원 특징공간에서  $\kappa$  개의 서로 다른 클래스에 속하는 패턴집합이 있을 때 입력패턴에 대한 분류는 다음과 같은 6 단계에 의하여 구성된 신경망을 통하여 가능하다.[ 9 ]

단계 1) 입력패턴이 대표값으로 표현되는 특징 공간내에서 패턴들의 i-번째 (  $i = 1, 2, \dots, \kappa$  ) convex hull  $C_i$  를 구한다.

단계 2) 선형 분리 가능한 convex hull 인  $C_i$  들이 서로 겹치는 영역이 존재하지 않는다면 아래 그림과 같이 연결 가중치와 Bias 를 결정하여 처리할 수 있다.

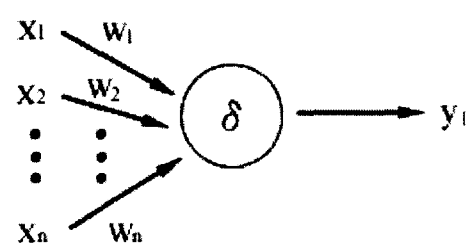


그림 1 초평면  $H_i = \omega_1 x_1 + \omega_2 x_2 + \dots + \omega_n x_n + \delta = 0$  으로 분할하는 경우

단계 3) 각 클래스들이 선형 분리가 불가능한 경우 서로 간에 겹치는 영역이 존재하지 않는다면 두 개의 뉴런 층이 요구되는데 입력층의 노드수는 convex hull 을 이루는 초평면의 수와 같으며, 출력층의 노드 수는 최소  $\kappa - 1$  개

또는 최악의 경우  $\kappa(\kappa-1)/2$  개가 필요하다. 출력층의 노드 수는 구분하고자 하는 클래스들의 갯수와 같다. 만일 임의의 클래스를 이루는 convex hull 에서 적어도 한쌍 이상 서로 간에 겹치는 영역이 존재하면 최소한 3 개 이상의 노드층이 필요하다.

단계 4) 임의의 클래스가 서로 떨어진 클러스터들의 집합으로 구성되고 이들이 convex 하지 않는다면 이들을 구분한 VoD 에서의 유한개 초평면들의 반폐구간(closed half space)들로 표현 가능하다. 여기에서 요구되는 신경망은 3 개층으로 구성되는데 입력층의 노드수는 입력 패턴을 구분하는 초평면의 차원으로 결정하고 둘째층의 노드수는 반폐구간의 갯수이며 출력층의 노드수는  $\kappa$  개의 서로 다른 클래스들의 수이다.

단계 5) 각 초평면의 수식을 통해 입력층의 노드에 대한 연결 가중치를 결정하고 입력층에서 다음 은닉층으로의 연결 가중치는 convex 영역들을 구분한 초평면의 반폐구간에 따라  $H > 0$  인 노드 연결은 1,  $H < 0$  인 노드 연결은 -1 로 결정하고 둘째층 노드의 임계값을  $m - n + 0.5$  의 계산값으로 결정한다. 여기서  $m$  은  $H < 0$  에 속하는 경우의 노드 수이고  $n$  은 전체 연결 노드 수이다.

단계 6) 은닉층에서 출력층으로의 연결 가중치는 단계 5) 에서와 같이 연결 가중치를 결정하고 임계값은  $m - 0.5$  로 계산한다.

### III. 실험 및 토의

이제까지 기술한 VoD 를 이용한 신경망 구성의 6 단계에 따라  $(0,0) \rightarrow 1, (4,0) \rightarrow 0, (0,4) \rightarrow 0, (4,4) \rightarrow 1$  로 구분할 수 있는 XOR 문제에 있어서  $S_1 = \{(0,0), (4,4)\}$  과  $S_2 = \{(4,0), (0,4)\}$  로 이루어지는 convex hull 은 선형 분리가 불가능하므로 두개의 초평면  $H_1 = x_1 - 2, H_2 = x_2 - 2$  으로 각각의 반폐구간 영역을 구분한다.

각각의 반폐구간 영역으로부터 클래스  $S_1 = (H_1^+ \cap H_2^+) \cup (H_1^- \cap H_2^-)$  와  $S_2 = (H_1^- \cap H_2^+) \cup (H_1^+ \cap H_2^-)$  으로 서로 떨어진 영역의 합집합으로 분류할 수 있다.

따라서 단계 4) 에서와 같이 전체 신경망의 입력, 은닉, 출력층의 노드 수는  $2 * 4 * 2$  로, 연결 가중치의 초기 값 및 Bias 는 은 다음과 같이 구성된다.

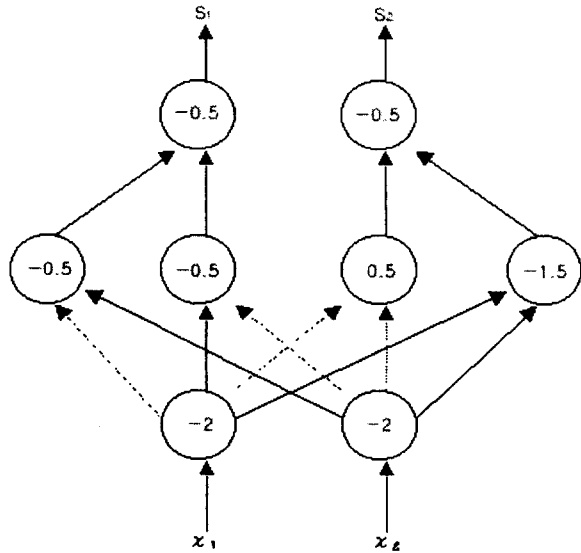


그림 2 제안된 VoD\_EBP 의 전체 구조

제안된 VoD\_EBP 와 기존의 학습 가능한 EBP 와의 성능 비교를 위하여 EBP 에서도 세 개의 층으로 구성하였으며 입력, 은닉, 출력층의 노드수를 경험에 의해 각각  $2*3*2, 2*4*2$  인 경우로 나누어 실험을 하였고 초기 연결 가중치 및 임계값은 Turbo C++ 3.0 에서 제공하는 난수 발생 함수를 이용하여 -1.0 ~ 1.0 사이의 값으로 초기치를 설정하였다.

학습률 0.3 , 허용오차 0.05 수준에서 EBP 의 경우 은닉층의 노드수를 4 개로 하였을 때, 실험 결과 학습횟수에 따른 학습 진행 과정을 그래프로 나타내면 다음 그림과 같다.

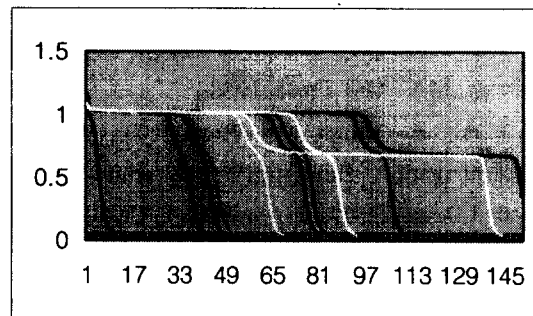


그림 3 제안된 VoD\_EBP 와 은닉층이 4 개인 EBP 에서의 출력층 오차 변화

그림에서와 같이 XOR 문제의 경우 EBP 에서는 초기 학습 진행이 상당히 느리게 진행되어 많은 학습횟수가 요구되었으나 제안된 VoD\_EBP 에서는 초기 가중치의 결정으로 상당히 빠르게 학습이 진행되는 것을 알 수 있다. 참고로 학습률을 0.5, 0.7 인 경우에 대해서도 실험을 한 결과 VoD\_EBP 가 EBP 보다 8 ~ 9 배 정도 빠른 성능을 나타내었고, 항상 학습이 성공적으로 이루어지는 것을 알 수 있었다.

#### IV. 결론

본 연구에서는 Bose 가 제안한 Voronoi 공간 분류에 의한 신경망 구성을 토대로 학습까지 가능한 새로운 신경망 VoD\_EBP 의 구성을 제안하였다. 또한 기존의 신경망에서 발생하는 문제점인 지역 최소치에 빠져 학습 실패가 이루어지는 단점이 보완되었고, 초기 연결 가중치 및 임계값을 공학적인 체계에 의해 미리 근사한 값으로 결정하여 줌으로써 초기 학습 진행부터 상당히 빠른 속도로 학습이 진행되었다. 그러나 아직 Voronoi Diagram 에 의한 convex hull 의 구성이 일반적인 패턴인식에 있어 학습 패턴으로 주어지는 모든 입력벡터에 대하여 적용될 수 있는지에 대한 의문은 해결되지 않은 상황이며, VoD\_EBP 가 기존의 EBP 보다 항상 효율적일 것이라는 이론적인 연구가 이루어지지 않아 향후 이에 대한 지속적인 연구가 필요하다.

#### V. 참고문헌

[ 1 ] M. L.Minsky and S. A. Papert, *Perceptrons* Cambridge, MA : MIT Press, 1988  
 [ 2 ] Ali A. Minai and Ronald D. Williams, "Backpropagation Heuristics : A study of the Extended Delta-Bar-Delta Algorithm", *IJCNN*, Vol 1, pp. 595-600, 90.  
 [ 3 ] Hedong Yang and Clark C. Guest, "Linear Discriminants, Logic Functions, Backpropagation, and Improved Convergence", *IJCNN*, Vol 3, pp. 287-292, 90. Jun  
 [ 4 ] Norio Bada, "A Hybrid Algorithm for Finding

the Global Minimum of Error function of Neural Networks", *IJCNN*, Vol 1, pp. 585-588, 90. Jan  
 [ 5 ] P. Burrascano and P. Lucci, "Smoothing backpropagation cost function by Delta Constraining", *IJCNN*, Vol 3, pp. 75-80, 90. Jun  
 [ 6 ] Tariq Samad, "Backpropagation Improvements Based on Heuristic Arguments", *IJCNN*, Vol 1, pp. 565-568, 90. Jan  
 [ 7 ] 조성배, 김진형, "오류 역전파 학습 알고리즘의 속도 개선방법: 분류 및 비교", *한국 정보과학회 논문지*. 91.11 Vol.18, No.6 November, pp. 649-660.  
 [ 8 ] 권회용 외 4, "바람직하지 않은 포화상태 방지에 의한 역전파 학습의 개선", *한국 정보과학회 논문지*. 93.5 Vol.20, No.5 May, pp. 656-661.  
 [ 9 ] N. K. Bose "Neural Network Design Using Voronoi Diagrams", *IEEE Transaction on Neural Networks* Vol. 4, No 5, September 1993.  
 [ 10 ] D.T.Lee and F.P.preparata, "The All Nearest Neighbor Problems for convex polygon," *Information Processing Lett.*, vol.7, pp.189-192, June 1978  
 [ 11 ] Christos H.Paradimitriou and M.Yannakakis, "The complexity of Restricted Spanning Tree Problems," *JACM*, vol.29, pp.285-309, Apr.1982.  
 [ 12 ] D.Mccallum and D.Avis, "A Linear Algorithm for finding the convex hull of a Simple polygon," *Information Processing lett.*, vol 9, pp.201-206, Dec.1979.  
 [ 13 ] M.I. Shamos. *Computational Geometry*, Ph.D. thesis, Yale Univ., 1978.  
 [ 14 ] Godfried T.Toussaint, "Pattern Recognition and Geometrical Complexity," *IEEE*, 5-th International Conference on Pattern recognition, pp.1324-1347, 1980 .  
 [ 15 ] J.J.A.Hartigan, *Clustering Algorithms*, John Wiley & Sons, 1975.