

EBP 신경망 학습에서의 동적 초기 가중치 선택에 관한 연구

김태훈*, 이일병*

*연세대학교 컴퓨터과학과

e-mail:thkim@csai.yonsei.ac.kr

A Study on Analysis of Dynamic Generation of Initial Weights in EBP Learning

Taehun Kim*, Yillbyung Lee*

*Dept of Computer Science, Yonsei University

요 약

다층 퍼셉트론(MLP) 학습 이론인 오류 역전파 알고리즘은 델타룰과 최급 하강법을 사용하기 때문에 학습시 많은 시간이 소요된다는 단점을 가지고 있다. 때문에 신경망에서의 잘못된 초기 가중치 선택은 오류 역전파 알고리즘을 사용하는 신경망에서의 현격한 학습 성능저하를 발생시키게 된다. 본 논문에서는 학습시 오류 역전파 알고리즘의 수렴시간을 개선하기 위한 신경망의 동적 초기 가중치 선택 알고리즘을 제안한다. 이 알고리즘은 학습전 기존의 선택 가중치와 모든 가중치가 1.0 또는 -1.0 값을 가지는 가중치 집합에서 가중치 변동률을 선추정하여 이들 중 가장 변동률이 큰 경우를 초기 가중치 집합으로 선정하게 된다. 즉, 초기의 가중치 변동률을 차후 성능을 판단하는 지표로 사용하여 잘못된 가중치 선택으로 인한 최악의 학습효율의 가능성을 배제시키고 다층 신경망의 학습특성상 평균 이상의 학습효율을 보장하는 초기 가중치 선택방법이다.

1. 서론

다층 신경회로망은 단일 초평면의 결정영역을 갖는 다층 퍼셉트론에 비해 은닉층을 포함함으로써 보다 영역을 다양화할 수 있게 되며, 학습시의 잡음에 대한 개선된 유연성, 오류처리가 가능하게 된다. 이 같은 오류처리와 유연성 때문에 다층 신경회로망은 문자인식, 음성인식 등의 패턴인식 분야와 클러스터링, 로보틱스 등 다양한 분야에 널리 사용되고 있다. 다층 신경회로망에서는 일반적으로 오류 역전파(EBP : Error BackPropagation) 알고리즘을 기본 학습법으로 사용하고 있다. 이 방법은 목표출력과 실제출력차의 자승합을 목적함수로 최소치를 스텝폭 고정의

최급하강법(steepest descent method)으로 구하는 알고리즘이다. 구해진 값을 이용하여 신경망의 가중치값이 조절되며 가중치값의 변화가 없을 때 비로소 학습은 종료된다. 최급하강법 외에도 목적함수의 최소치를 구하는 알고리즘으로는 Levenberg-Marquardt algorithm, Newton's method, conjugate method, nongradient method과 같은 직접, 간접 탐색법이 사용되고 있다. 하지만 신경망 학습효율의 높이기 위한 다양한 방법들의 개발에도 불구하고 아직도 해결해야 할 많은 단점들이 존재한다.[1][2]

신경망 학습의 문제점 중 하나는 지역해에 빠질 수 있다는 점이다. 만약 지역해를 벗어날 수 있는 알고리즘이 존재하지 않는 한 최적해를 찾아낼 수 없을 뿐만 아니라 벗어난다 하더라도 지역해임을 알아내기 위해서는 많은 시간이 소요된다.[3][4] 또한 초기 가중치값을 선택하는 문제에 있어서도 다층 신

“본 연구는 산자부 뇌신경정보학 사업으로부터 지원을 받아 수행되었음.”

경망은 민감하게 반응하게 되고 잘못된 초기 가중치를 설정하였을 경우 역시 학습효율을 떨어트리게 된다.[5]

오류 역전파 알고리즘의 성능향상을 위해서 학습률 또는 모멘텀의 동적 접근방법이 제시되었고 중복학습을 막아 수렴속도를 올리는 방법도 고려되었다.[6]

본 논문에서는 다층 퍼셉트론 학습 초기 선택된 임의의 가중치와 1.0, -1.0으로 정규화 된 가중치 집합으로부터 각각의 초기 가중치 변동률을 구하고 이를 바탕으로 초기 가중치를 선택하는 동적 초기 가중치 선택 알고리즘을 제안한다. 이 알고리즘을 사용하게 되면 기존 1.0과 -1.0 사이의 범위에서 임의로 선택되던 초기 가중치 방법과 달리 초기 가중치의 잘못된 선택으로 학습효율이 현격히 떨어지는 문제에서 벗어날 수 있다. 오류 역전파 알고리즘을 사용하는 일반 다층 신경망의 수렴 특성상 학습 초기의 가중치 변동이 전체 학습과정 판단의 근거가 될 수 있고, 때문에 주어진 각각의 가중치 집단으로부터 얻어진 일부 초기 가중치 변동률의 총합으로 학습효율을 예측할 수 있게 된다. 물론 초기의 변동률, 한가지 요소가 전체 학습성능을 결정짓는 절대적 지표가 될 수는 없지만 범위내의 무작위 선택이나 임의의 가중치를 선택하는 기존의 방법에 비해 성능, 효율적 측면에서 우위를 갖게 된다.

이를 위하여 오류 역전파 알고리즘을 사용하는 다층 신경망의 학습 과정에 대해 간략히 살펴보고 제안하고 있는 동적 초기 가중치 선택 알고리즘에 제시한다. 다음으로 학습실험을 통해 기존의 방법에 비해 제안된 알고리즘의 학습효율 우위를 입증한다.

3 다층 신경망의 학습방법

오류 역전파 알고리즘을 사용하는 다층 신경망의 학습에서는 델타룰(delta rule)과 최급하강법을 사용하여 임의의 입력에 대해 원하는 출력에 대하여 가중치를 조절하여 수렴하게 하는 방식으로 학습이 이루어진다.

델타룰은 예상 출력과 실제 출력과의 차이를 계산하고 그에 따라 가중치를 변경하는 방법으로 입력 함수가 선형인 네트워크에서 얻어진 모든 입력 패턴으로부터 실제출력과 목표출력과의 오차 제곱이 최소가 되도록 신경망의 가중치를 조정하게 된다. 이를 위해서는 패턴쌍에서의 오차의 제곱을 각각 연결 가중치로 미분한 것이 델타룰에서의 연결 가중치

변화량에 비례한다는 것을 보여주면 된다.

연결 가중치는 델타룰에 의하여 연결 가중치 공간상에 주어지는 오차의 제곱을 높이로 하는 곡면에 대하여 오차의 제곱이 가장 많이 감소하는 방향으로 가중치를 변화시킨다. 즉, 주어진 결과에 대하여 최급하강법을 사용하게 된다. 이러한 일련의 학습과정을 수식적 측면에서 접근해 보면 다음과 같다.

패턴 P 에서의 목표출력 t 와 실제출력 o 에 대하여 오차는 아래 식1과 같이 나타낼 수 있고 이러한 모든 패턴들의 합을 구하면 전체 네트워크상에서의 출력 오차를 구할 수 있다.

$$E_P = \frac{1}{2} \sum_j (t_{pj} - o_{pj})^2 \quad (1)$$

연결 가중치의 변화는 해당 연결의 종점에서의 오차 δ 와 연결 시점의 출력 o 의 곱에 비례하게 되며 이때 학습률(learning rate) η 가 가중치 변동량을 조절해 학습 수렴속도를 향상시킬 수 있다.

$$\Delta_p W_{ji} = \eta \delta_{pj} o_{ri} \quad (2)$$

오차신호는 출력에 대한 오차를 초기값으로 재귀적 수행을 함으로써 구해질 수 있으며 이 통해 델타룰과의 유사성을 확인할 수 있다. 따라서 네트워크 출력과 목표출력과의 차이로부터 출력층에서의 오차를 구할 수 있게 된다.

$$\delta_{pj} = (t_{pj} - o_{pj}) f_j'(net_{pj}) \quad (3)$$

준선형 입출력 함수의 미분계수와 출력개체 k 의 오차 δ_{pk} 그리고 그 사이의 연결 가중치 w_{kj} 를 이용하면 은닉층에 있는 개체에 대한 오차를 구할 수 있고 그 식은 다음과 같이 표현된다.

$$\delta_{pj} = f_j'(net_{pj}) \sum_k \delta_{pk} w_{kj} \quad (4)$$

이렇게 구해진 네트워크 각 층에서의 오차와 출력결과 그리고 연결 가중치 변동량을 통해 다층 신경망 전체의 차후 연결 가중치 변동량을 구할 수 있게 되며, 이때 오차 진동을 줄이기 위한 모멘텀(momentum)까지 고려하면 아래 식5와 같은 일반 연결 가중치 변화의 최종 관련식을 도출해 낼 수 있다[7].

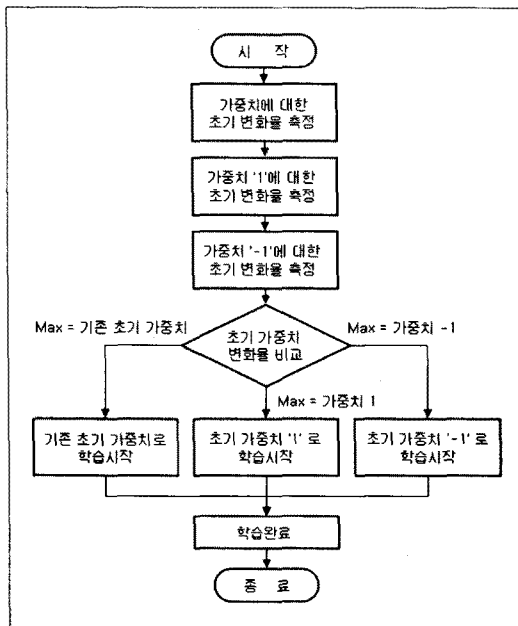
$$\Delta_p W_{ji} (n+1) = \eta \delta_{pj} o_{ri} + \alpha \Delta_p W_{ji} (n) \quad (5)$$

4. 동적 초기 가중치 선택 알고리즘

이 장에서는 오류 역전파 알고리즘을 사용하는 다층 신경망에서의 학습시간을 개선시키기 위해 초기 임의 선택된 가중치를 동적으로 수정해주는 알고리즘을 제안한다. 이 알고리즘은 다층 신경망에서 사용되는 임의 초기 가중치 선택방법 사용시 발생할 수 있는 학습시간 증가에 따르는 학습효율 저하를 방지할 수 있다. 기존 사용되었던 임의 초기 가중치 선택방법은 사전 정확히 예측할 수 없는 다층 신경망의 특성상 좋은 학습효율을 가질 수 있지만 잘못된 초기 가중치 선택으로 발생할 수 있는 현격한 학습시간 상승과 이로 인한 학습효율 저하를 막을 수 없었다. 하지만 제시한 동적 초기 가중치 선택 알고리즘을 사용하면 이러한 최악의 경우를 사전 차단할 수 있다.

언급 했던 바와 같이 각각의 초기 가중치만으로는 그에 따르는 정확한 학습효율 결과값을 구해낼 수 없다. 하지만 초기 가중치 변동률을 고려한다면 통계상 경험적(heuristic) 학습효율 예측이 가능하다. 즉, 지역 최소점에 도달하여 다시 학습을 시작하는 경우와 같이 학습의 재시작이 이루어지지 않는 경우 초기 연결 가중치 변동률은 전체적인 학습을 판단의 근거가 될 수 있다.

이같은 이론적 배경을 바탕으로 학습이 시작하기



(그림1) 제안된 초기 가중치 선택 알고리즘 순서도

전 기존 선택되었던 초기 가중치와 통계적 초기 가중치 상한선인 1.0로 이루어진 연결 가중치 집합 그리고 하한선인 -1.0로 이루어진 집합 이렇게 세가지 경우에 대해 초기 가중치 변동률을 측정한다. 그 결과값을 근거로 이들 중 가중치 변동률이 가장 큰 가중치 집합을 선택하고 선택된 집합을 초기 가중치로 설정, 다시 학습을 시작한다.

이러한 일련의 학습 전처리 과정을 그림1과 같이 흐름도(flow chart)로 표시하였다.

5. 실험결과

5.1 실험환경

본 연구에 사용된 시스템 사양은 P4-2.8GHz, Memory 512M이며 다층 신경망 구현과 해당 응용프로그램으로는 C++를 이용하였다. 오류 역전파 알고리즘을 사용한 다층 퍼셉트론 환경에서의 순수한 초기 가중치 변화가 미치는 영향만을 판단하기 위해 표1과 같이 간단한 XOR문제를 학습 데이터로 사용하였고, 다층 신경망의 구성에 있어도 은닉층이 1개로 이루어진 간단한 신경망 패턴을 실험에 사용했다.

<표1> 학습자료로 사용될 XOR문제 입/출력값

| | | | | |
|--------|--------|--------|--------|--------|
| 입력값 | (0, 0) | (0, 1) | (1, 0) | (1, 1) |
| 목표 출력값 | 0 | 1 | 1 | 0 |

5.2 실험결과

학습시 허용오차 범위는 소수점 5자리 이하로 설정하고 초기 가중치에 따른 영향을 확인하기 위해 학습율과 초기 가중치를 변경해 보았다.

<표2> 초기 가중치 1인 집합에서의 학습결과

| | | |
|---------------|----------------------|---------------------|
| 구 분 | 학습율(η) : 0.01 | 학습율(η) : 0.1 |
| epoch | 229,658 | 86,557 |
| run time(sec) | 12.375 | 4.671 |

<표3> 초기 가중치가 -1인 집합에서의 학습결과

| | | |
|---------------|----------------------|---------------------|
| 구 분 | 학습율(η) : 0.01 | 학습율(η) : 0.1 |
| epoch | 409,766 | 420,486 |
| run time(sec) | 22.078 | 22.703 |

표2, 표3의 결과와 같이 동일 학습 데이터와 신경망 패턴으로 학습한 경우에도 학습효율이 0.01인 경우는 1.78배, 0.1인 경우는 4.58배의 차이를 보였다.

즉, XOR문제 학습데이터 경우에 있어서는 초기 가중치 -1 집합에 비해 가중치 1의 집합이 적절한 경우임을 알 수 있다.

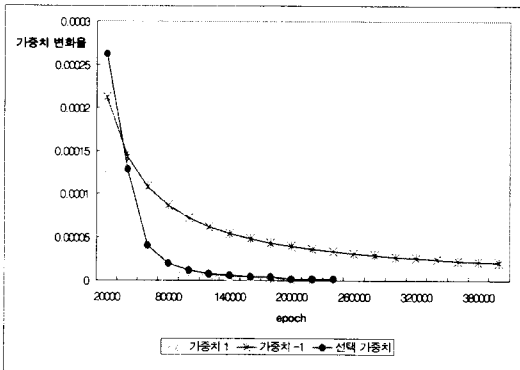
<표4> 임의의 초기 가중치 집합에서의 학습결과

| 구 분 | 학습율(η) : 0.01 | 학습율(η) : 0.1 |
|---------------|----------------------|---------------------|
| epoch | 457,135 | 151,130 |
| run time(sec) | 26.250 | 8.672 |

기존 임의의 초기 가중치 선택의 경우, 표4 나타난 바와 같이 특정 학습율에 한하여 위 세가지 가중치 집합의 경우 중 가중치 -1 집합은 선택했을 때보다 좋은 성능을 나타내기도 하지만 가중치 1 집합의 경우보다는 우수하지 않았다. 따라서 이와 같은 경우 기존의 방법의 사용한다면 초기 가중치를 모두 1로 지정한 경우보다 최적의 학습효율의 보장할 수 없게 된다.

<표5> 제안한 알고리즘 사용 학습결과

| 구 분 | 학습율(η) : 0.01 | 학습율(η) : 0.1 |
|---------------|----------------------|---------------------|
| epoch | 229,958 | 86,857 |
| run time(sec) | 12.500 | 4.828 |



(그림2) epoch에 따른 가중치 변화율($\eta=0.01$)

제안된 동적 초기 가중치 선택 알고리즘의 초기 가중치 변동율은 초기 전처리 학습과정 중 매 50 epoch당 10회 동안의 변동율 총합을 통해 구하고 이를 바탕으로 경우의 최적 초기 가중치를 선택하게 된다.

이 알고리즘을 사용한다면 표5와 그림2에서처럼 세가지 가중치 집합에서는 최고의 성능을 보장하는 가중치 1 집합의 학습효율보다는 낮지만 그와 근접한 학습효율을 보여줄 수 있고 최악의 경우인 가중치 -1 집합은 선택하지 않을 수 있다.

6. 결론 및 향후과제

본 논문에서는 오류 역전파 알고리즘을 사용하는 다층 신경망의 성능을 개선시키기 위하여 초기 가중치 선택에 있어 동적으로 접근하는 방법을 제안하였다. 제안된 알고리즘의 경우, 임의의 선택된 초기 가중치와 범위 말단의 값들로만 이루어진 가중치 1.0 / -1.0 집합과의 학습초기 가중치 변동량을 구해 이를 바탕으로 초기 가중치를 선택하게 된다. 최선의 학습효율에 근접하는 성능을 기대할 수 있고 무엇보다 좋지 못한 학습율을 보일 가능성이 있는 초기가중치 선택의 경우를 미리 제외시킬 수 있다는 데 이 알고리즘의 이점이 있다.

하지만 다양한 형태를 갖는 학습데이터, 네트워크 패턴에 대한 적응력이 확인되지 않았고 비일반적인 진행형태의 학습이 계속될 경우 기존 임의의 가중치 선택 방법보다도 좋지 못한 성능을 보일 수 있다.

따라서 특이한 패턴을 갖는 학습데이터, 네트워크의 분석과 제어에 관한 연구가 진행되어야 하며, 이러한 경우에도 적용할 수 있는 알고리즘을 찾는다면 보다 보편적인 학습효율 기대할 수 있다.

참고문헌

- [1] C.Charalambous, "Conjugate gradient algorithm for efficient training of artificial neural networks", IEEE Proceedings, Vol. 139, No.3, pp.301~310
- [2] M.T. Hagan and M. Menhaj, "Training Feed forward Networks with the Marquardt algorithm", IEEE Tran. on Neural Networks, Vol.5, No.6, 1994
- [3] D.Nguyen, B.widrow, "Improving the Learning speed of 2-layer neural networks by choosing initial values of the adaptive weights", IJCNN, San Diego CA, 21-16(1990)
- [4] Cheung, et al, "Relative Effectiveness of training set patterns for backpropagation", IJCNN, Vol.1, pp.673~678, 1990
- [5] D.E. Rumelhart, G.E. Hinton and R.J. Williams, "Learning internal representations by error propagation", Explorations in the Micro Structure of Cognition, vol.1 : Foundations. MIT Press pp.318~362, 1986
- [6] Ulf D. Schiller, Jochen J. and Steil, "Analyzing the weight dynamics of recurrent learning algorithms", Neurocomputing 63(2005) 5~23
- [7] Simon Haykin, "Neural Networks : a comprehensive foundation", Prentice Hall