에러 역전파 학습 성능 향상을 위한 초기 가중치 결정에 관한 연구

김 웅 명, 이 현 수
경희대학교 전자계산공학과

A Study of Initial Weight Determination for Performance Enhancement in Backpropagation

Kim Woong Myung, Lee Hyon Soo
Kyung Hee University

요 약

에러 역전파 신경망에서 학습속도와 수렴률은 초기 가중치의 분포에 따라 크게 영향을 받는다 본 연구에서는 이를 위하여 비교사 학습 신경망(Hebbian learning rule)을 이용한 새로운 초기 가중치 결정 방법을 제안한다 또한 비교사 학습 신경망이 에러 역전파 신경망 학습에 적당하도록 은닉층의 각 뉴런과 연결된 가중치의 norm을 이용하여 학습하였다 시뮬레이션을 통하여 기존 에러역전파 신경망 학습과 그 성능을 비교한 결과 제안한 초기 가중치 표현이 학습속도와 수렴능력에서 우수함을 나타낸다.

1. 서론

현재 에러역전파 신경망은 패턴분류 및 인식, 시스템 제어등 여러 분야에 널리 응용되고 있다 그러나 에러 역전파 신경망은 초기 가중치의 분포가 적절해야만 학습의 수렴 및 속도가 빨라진다. 또한 은닉층 뉴런의 수를 최적으로 결정하기 어렵기 때문에 신경망의 초기 구성시 적당한 은닉층 뉴런의 수와 가중치 분포의 구성이 요구되어 진다[1]. 에러역전파 신경망과 같은 교사학습 신경망은 주어진 입력에 대해 그에 맞는 출력값으로 근사하는 $f(x)$를 구하는 것이다 그리고 주어진 가중치는 입력패턴과을 출력과의 오차를 최소화 하기 위해 가중치를 조정하게 된다 하지만 기울기를 이용하여 가중치를 갱신하므로 국부적인 지역해에 빠질수 있다 Haario과 Jokinen은 이러한 문제점을 입력과 출력패턴들 사이에 선형 회귀방법 이용한 초기 가중치를 제안하였다[1] Denoeux는 입력패턴을 특징 공간(feature space)에 대한 변환후에 가중치를 초기화하는 방법을 제안하였다[2]. 그러나 [1]의 경우, 중간층의 뉴런들이 선형적 출력이 되어야만 하므로 특정 문제에만 적용이 가능하다 [2]의 경우에는 매우 많은 중간층의 뉴런들이 필요하므로 많은 학습시간이 필요하다

본 연구에서는 이러한 접근 방법과는 달리 Hebbian학습을 이용한 초기 가중치 생성 방법을 제안한다

2. Hebbian학습을 이용한 가중치 초기화

에러 역전파 학습시 나타나는 문제점을 간단히 나타내면 다음과 같다 첫번째는 오차 함수의 모양이 미분계수가 0이 되는 지점이 발생할 경우가 생긴다는 것이다 이러할 경우 학습상수 η의 값과 초기 가중치에 의하여 지역해에 빠질수 있다 두 번째는 부적절한 포화현상(Incorrected saturation)이다. 이것은 초기 적당하지 않은 가중치에 의하여 출력함수 포화영역에서 잘 동작하지 않는 것이다[3] 이와 같은 부적절한 포화현상은 출력층과 은닉층 뉴런, 양쪽에서 나타나게 된다 그리고 패턴의 분류의 어려움으로 인한 학습의 진동현상(oscillation)이 발생한다는 것이다 위에서 언급한 문제점은 적절한 초기 가중치 분포에 의하여 문제가 해결될수 있다. 에러역전파 학습에서 적당한 초기 가중치는 학습시 복잡한 입력패턴에 대하여 일어나는 oscillation현상을 빠르게 탈출할 수 있어야 한다[4]. 또한 학습 수렴율면에서 인정되어야 한다 그리고 에러역전파 학습진 입력패턴의 특성이 가중치에 포함되어야 한다.

Hebbian학습은 비교사학습 신경망이다. 이것은 단순히 입력층과 출력층으로 구성되어 있다 가중치 갱신 수식은 다음과 같다

$$\Delta w_{u}(n) = \eta * h_{p} * x_{pi} \qquad (1)$$

$$\text{where.} \quad h = (h_1, h_2, h_3, \cdots, h_m)$$

$$x = (x_1, x_2, x_3, \cdots, x_n)$$

Hebbian학습의 문제점은 학습이 진행됨에 따라서 가중치의 변화량이 계속 증가함에 따라 전체 가중치 크기가 매우 커진다는 것이다 따라서 가중치의 변화량을 적절히 조정하여 가중치의 크기를 제한시키는 것이 필요하다. 수식 2은 이를 위하여 새롭게 제안하는 가중치 변화량을 나타낸다.

$$\Delta w_{p}(n) = \eta * h_{pj} * (1 - \alpha\sqrt{\sum_{h}^{n} w_{hi}^{2}}) * x_{p} \qquad (2)$$

$$\text{where} \quad 0 < \eta < 1 \ , \ 0 < \alpha \leq 1.0$$

위의 수식 1과 비교하여 부가된 항은 수식 3과 같다.

$$(1-\alpha*\sqrt{\sum_{h}^{n} w_{jh}^{2}}) \qquad (3)$$

본 연구에서는 입력층과 중간층 h_{pj}에 연결되어있는 가중치를 하나로 표현할 수 있는 norm 개념을 이용하여 연결강도 갱신크기를 정하였다 norm은 주로 벡터의 크기를 나타내며 일반적으로 다음과 같이 표현할수 있다.

$$\|a\| = \sqrt{a_1^2 + a_2^2 + \cdots, a_m^2} \qquad (4)$$

가중치 w_n의 벡터 크기를 norm으로 표현하면 다음과 같다.

$$\|w_n\| = \sqrt{w_{j1}^2 + w_{j2}^2 + \cdots + w_{jm}^2} \qquad (5)$$

수식 5은 전체 w_n의 전체 크기이며, 이것을 상대적으로 나타낼 만한 크기로 다시 조정하여야 한다 즉, 전체 벡터의 크기에 비례한 기준 벡터의 크기를 1로 정함으로써 학습의 안정화를 꾀하도록 하였다 여기서 η와 α는 1보다 작은 양의 실수이다. h_{pj} , x_{pi}은 -1, 1 사이의 범위이다. 그러므로 $\alpha\sqrt{\sum_{h=0}^{n} w_{jh}^2}$이 1과 유사한 값을 가지면 가중치의 변화가 0이 되므로 학습은 종료하게 된다 여기서 α의 값이 1보다 작아지면 $\alpha\sqrt{\sum_{h=0}^{n} w_{jh}^2}$ 의 항의 값이 증가한다. 따라서 가중치 범위가 증가하게 된다 $h_{pj} \cdot x_{pi}$항은 가중치의 갱신 방향을 결정하게 된다

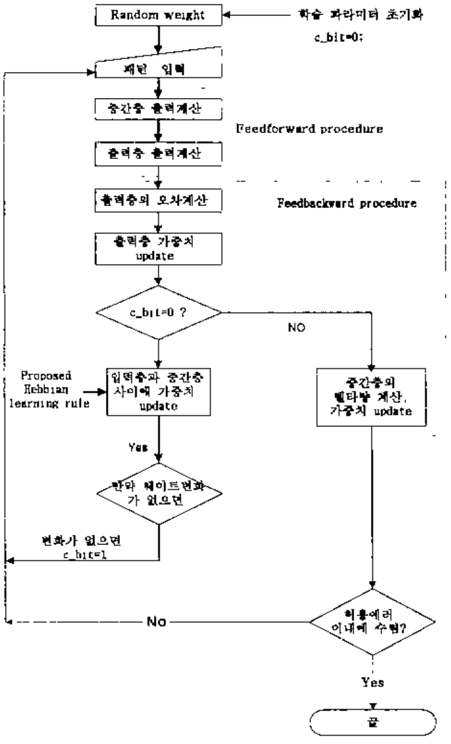

그림 1 제안한 학습 알고리즘의 흐름도

학습의 종결조건은 제안된 Hebbian학습 알고리즘을 이용하여, 모든 은닉 뉴런에 연결된 가중치 변화가 없을 때, 제안한 가중치 변화량에 의한 입력층과 중간층 사이의 가중치 조정을 중단하고 에러역전파 학습을 하게 된다. 이때 입력층과 은닉층 사이의 가중치는 네트워크 구조상에서 입력패턴의 특성을 가지게 된다

그림 1은 제안된 학습 알고리즘의 전체 흐름도를 나타낸 것이다. 제안한 학습 알고리즘은 Batch학습이 아닌 On-line학습방법으로 수행하며 입력 패턴마다 가중치 갱신이 일어나게 된다

3. 시뮬레이션 및 고찰 분석

본 연구에서 제안한 학습 알고리즘의 우수성을 검증하기 위해 Sun Ultra-1상에서 C언어로 구현된 Simulator로 수행하였으며, 뉴런의 활성함수는 tanh를 사용하였다. 입력패턴의 복잡도가 높은 4parity문제와 2spiral문제를 대상으로 시뮬레이션 하였다.

4parity문제

4parity문제의 경우 입력층 뉴런 4개, 중간층 뉴런 8개, 출력층 뉴런 1개로 구성하였다 그리고 초기 임의의 가중치 범위는 -0.5~0.5이다. η에 따른 학습횟수를 보이기 위해, α값을 0.5, 최대반복횟수를 30000번, 허용에러를 0.001로 시뮬레이션을 수행하였다

그림 2의 초기화 iteration은 제안된 가중치 학습 알고리즘을 이용해 초기 가중치를 생성하는 학습 횟수이다 제안된 알고리즘의 전체 학습횟수는 초기화 가중치 생성 학습 횟수와 에러 역전파 신경망 학습 횟수를 합하여 나타낸 것이다

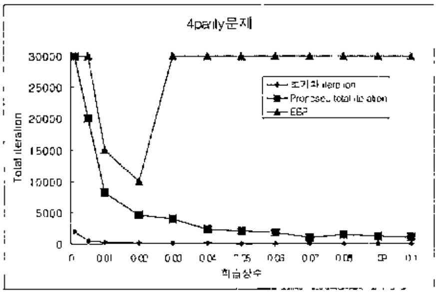

그림 2 4parity문제에서 학습상수 η값에 따른 학습 회수

시뮬레이션결과 에러 역전파 신경망은 학습상수의 범위가 0.01과 0.02의 범위에서 최대 반복횟수내에 수렴됨을 알수 있다. 하지만 제안한 신경망 학습은 학습상수가 0.001이외의 η값에서 수렴됨을 알수 있고, 전체 학습의 속도가 빠름을 알수 있다 이는 4parity문제에서 제안한 신경망 학습 알고리즘이 학습초기 osillation현상을 회피하여 학습하기 때문이다

2-spiral 문제

2-spiral문제는 그림 3에 나타난 바와 같이 출력 패턴이 2개이면서 입력패턴이 50개인 2개의 나선형 모양이다 입력 패턴 25개는 목표출력이 -1이고, 나머지 25개는 -1이다

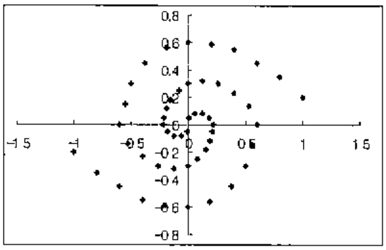

그림 3 2spiral 문제

네트워크 크기는 입력층의 뉴런 2개, 은닉층의 뉴런 10개, 출력층의 뉴런 1개로 시뮬레이션을 수행하였다. 시뮬레이션 조건은 초기 가중치의 범위가 -0.5~0.5이고, α값이 1.0, 허용에러는 0.01, 최대반복회수는 100000번이다. 그림 4에서 에러 역전파 학습의 경우는 학습상수 η에 따라서 작은 구간에서 수렴을 하지만 제안된 가중치 생성 알고리즘의 경우 전체 η구간에 빠른 학습이 진행되었음을 알 수 있다.

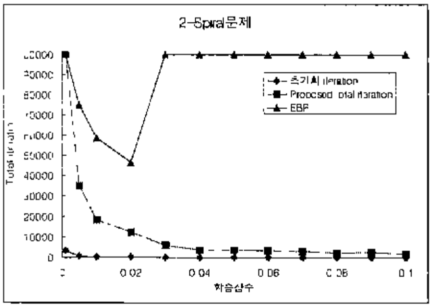

그림 4 2spiral문제에서 학습상수 η에 따른 학습 횟수

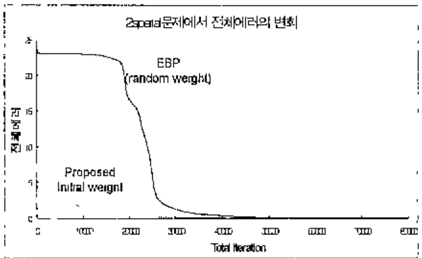

그림 5 에러 역전파 학습과 제안한 초기 가중치에 의한 학습 에러의 변화
(허용에러 : 0.001, η=0.009, α=1.0)

그림 5에서, 에러 역전파 학습이 긴 oscillation구간이 나타나서 학습이 안됨을 알수 있다. 에러 역전파 학습과 제안된 초기화 가중치 생성으로 학습한 결과를 학습 횟수마다 전체 에러 변화를 보인 것이다. 에러 역전파 학습은 잘못된 초기 가중치로 인하여 학습 초기 수렴이 안되고 있는 것을 알수 있다. 하지만 제안된 알고리즘은 문제에 적당한 가중치를 생성함으로써 적은 학습횟수만으로 수렴함을 알 수 있다.

본 논문에서 제안한 알고리즘은 입력패턴과 은닉층 출력간에 효율적으로 패턴들을 분포시키고, 출력층에서의 학습을 용이하게 함을 알 수 있다. 이것은 입력 패턴에 맞게 초기의 가중치를 적당하게 확장시킴으로써 에러역전파 학습을 용이하게 한다. 제안하는 신경망 학습 알고리즘은 초기에 주어진 가중치를 norm을 이용하여 일정하게 확장시킴으로써 입력패턴의 특성을 잃지 않으면서 적당한 가중치 공간의 확장을 의미한다.

4. 결론

본 연구에서는 에러 역전파 학습에 적당한 초기 가중치 표현 알고리즘을 제안하였으며 시뮬레이션을 통하여 학습속도와 수렴율에서 성능을 비교하였다. 또한 초기 가중치 표현 알고리즘이 에러 역전파 신경망의 입력층과 은닉층 사이에 적당한 가중치를 생성하기 위해 제안된 Hebbian학습방법을 이용하는 것이 임의의 초기 가중치로 학습하는것보다 효과적임을 알 수 있었다. 이를 위하여 학습이 안정화 될 수 있도록 입력층과 은닉층의 가중치 변화에 norm을 이용하였다.

향후 보다 효과적인 초기 가중치 결정방법을 위해 적당한 α값과 입력패턴 값에 대한 관련 연구가 필요로 되어진다.

[참고문헌]

[1] H.Haario and P.Jokinen,"Increasing the learning speed of backpropagation algorithm by linearization", in Artificial Neural Networks, T.Kohonen, Ed. Amsterdam, The Netherlands:Elsevier, pp629-634,1991,

[2] T.Denoeux, R. Lengelle, and S.Canu, "Initialization of weights in a feedforward neural network using prototypes", in Artficial Neural Networks, T. Kohonen, K.Makisara, O. Simuba, and J.Kangas, Eds. Amsterdam, The Netherlands:Elsevier, pp623-628,1991,

[3] Lee Y. S.Oh and M.Kim, "The effect of initial weights on premature saturaion in backpropagation learning", IJCNN'1991 Vol1, pp.765-770, 1991

[4] Nicolaos B.Karayiannis, "Accelerating the Training of Feedforward Neural Networks Using Generalized Hebbian Rules for Initializing the Internal Representations", Tran on Neural Networks Vol7 No2, pp419-426,1996