*The Korean Journal of Applied Statistics* (2013) DOI: http://dx.doi.org/10.5351/KJAS.2013.26.5.847
**26**(5), 847–856

# A Comparison of the Effects of Optimization Learning Rates using a Modified Learning Process for Generalized Neural Network

Yeochang Yoon[a] · Sungduck Lee[b,1]

[a]Department of Information Security, Woosuk University
[b]Department of Information and Statistics, Chungbuk National University

(Received October 1, 2013; Revised October 16, 2013; Accepted October 16, 2013)

**Abstract**

We propose a modified learning process for generalized neural network using a learning algorithm by Liu *et al.* (2001). We consider the effect of initial weights, training results and learning errors using a modified learning process. We employ an incremental training procedure where training patterns are learned systematically. Our algorithm starts with a single training pattern and a single hidden layer neuron. During the course of neural network training, we try to escape from the local minimum by using a weight scaling technique. We allow the network to grow by adding a hidden layer neuron only after several consecutive failed attempts to escape from a local minimum. Our optimization procedure tends to make the network reach the error tolerance with no or little training after the addition of a hidden layer neuron. Simulation results with suitable initial weights indicate that the present constructive algorithm can obtain neural networks very close to minimal structures and that convergence to a solution in neural network training can be guaranteed. We tested these algorithms extensively with small training sets.

Keywords: Learning algorithm, learning rate, weight scaling, quadratic programming.

## 1. 서론

시계열 자료를 분석하는 일반적인 분석방법은 선형 및 비선형 시계열 모형을 주로 사용해왔다. 최근 신경망을 이용하여 시계열 자료를 분석하고 예측하는 연구가 많은데 특히 단기예측에 있어서 매우 유익한 방법임이 많은 연구의 결과로 인정되어 있다.

Sharda와 Patil (1990)은 여러 가지의 시계열 자료를 사용하여 신경망모형과 Box-Jenkins 방법의 예측을 비교하고 있다. 이들은 신경망방법을 이용한 예측결과가 Box-Jenkins 방법만큼 효율적임을 보였다. 그러나 시계열 자료에 주기성(periodicity)이 있는 경우에는 신경망이 Box-Jenkins 방법보다 별 효과가 없었다. White (1988)은 상관관계가 높은 자료로써 미국 IBM의 일별 주식값 변동자료에 대한 신경망 예측을 하였다. 여기서 한 시점 앞의 예측에서는 신경망방법이 Box-Jenkins 방법보다 평균제곱오

This work was supported by the research grant of the Chungbuk National University in 2011.
[1]Corresponding author: Professor, Department of Information and Statistics, Chungbuk National University, 52 Naesudong-ro, Heungdeok-gu, Cheongju, Chungbuk 361-763, Korea. E-mail: sdlee@chungbuk.ac.kr

차(Mean Squared Error; MSE)가 상대적으로 작게 나타났다. Maasoumi 등 (1994)은 S&P500 주가지수의 예측에서 Box-Jenkins 방법이 아닌 신경망을 이용하여 한 시점 앞의 예측을 하였다. 이들의 연구에서 적절한 신경망의 구조를 선택하기 위해 적용한 방법은 다음과 같다. 입력노드를 1개에서 5개까지 단계별로 변화시킬 때마다 은닉노드 개수를 각각 1개에서 10개까지 변화시키면서, 각각의 경우에 학습오차를 구하여 이 오차가 가장 작아지는 신경망 구조를 택하였다. 예측은 여기서 선택된 구조를 사용하였다. 학습을 위한 오차 판단기준으로는 MSE와 평균절대값오차(Mean Absoluted Error; MAE)를 이용하였다.

신경망 학습과 관련된 알고리즘들이 현재 많이 연구되고 있고 이들 연구들이 성공적으로 다양한 응용 분야에 적용되고 있지만 몇 가지 쟁점 부문에서 완전하게 문제를 해결하지는 못하고 있다. 이들 쟁점들은 초기 가중값과 은닉노드의 개수 결정 문제 그리고 학습의 수렴속도 등과 같은 학습률의 최적 선택의 문제로 요약할 수 있다. 즉, 신경망 학습과정이 주어진 문제의 해로 잘 수렴되도록 하는 방안을 연구하고, 적은 학습량으로 해에 효율적으로 수렴되는 경우의 연구가 많이 있었다 (Diotalevi와 Valle, 2001).

Framling (2004), Zhang와 Jiang (2010) 등의 신경망 학습 알고리즘에 관한 최근의 연구 결과들은 많은 경우에서 수렴 속도를 어떻게 증진시키는가에 주로 초점을 두고 있다. 이들 연구에서 가장 중요한 문제점 중의 하나는 전역 최소값을 찾지 못하고 지역 최소값으로 수렴되는 경우에 이를 해결할 수 있는 방안을 찾는 것이다. 이들이 적용한 알고리즘의 학습과정은 수렴 속도를 개선시키고는 있지만 일반적인 비선형 문제에서 목적한 학습에 도달하지 못하고 지역 최소값에 수렴됨으로써 발생되는 포화(saturation)문제의 개선이 대부분이다.

신경망 학습이 일반적으로 수렴된다는 것은 알고리즘이 문제의 해를 궁극적으로 찾을 수 있다는 것을 의미한다. 다시 말하여 학습 알고리즘은 신경망 학습의 전 과정을 통하여 만날 수도 있는 지역 최소값을 벗어날 수 있게 해야 한다는 것을 의미한다. 이와 같은 문제들의 해결을 위하여 Fukuoka 등 (1998), Parekh 등 (2000)의 관련 연구들은 가중값의 범위 조정(weight scaling)과 평균장 어닐링(mean field annealing)기법 등을 이용하여 지역 최소값 문제를 해결하고 있다.

다층 FNN(Feedforward Neural Network)을 이용한 비선형 문제 해결에 가장 중요한 고려 사항 중의 하나는 은닉층 노드의 구조를 어떻게 결정하는가 하는 문제이다. 네트워크에서 은닉노드의 개수가 불충분하면 문제를 잘 적합시키지 못하며 반대로 노드의 개수가 너무 많으면 네트워크가 과다적합되어 일반화된 모형을 도출시키지 못할 수 있다. 여기서 적절한 은닉노드의 개수 문제는 입력 값 공간의 차원에 따라 다를 수 있다. 은닉노드의 개수가 불충분하면 정확히 추정할 신경망 모형에 필요한 모수의 개수가 충분하지 않아 과소평가된 의사결정을 할 수 있다 (Maghami와 Sparks, 2000). 그러므로 과소 적합문제를 해결하기 위한 방법으로써 네트워크에 필요한 은닉노드의 개수를 충분히 크게 증가시키거나 또는 prune 방법을 이용한 은닉노드의 최적 개수를 설정하면 된다 (Haykin, 2010). 그렇지만 이와 같은 학습 방법은 지역 최소값을 쉽게 벗어나게 해 줄 수는 있지만 학습에 필요한 모수의 개수보다 더 복잡한 네트워크를 학습시킬 수밖에 없기 때문에 수렴 속도가 늦어지고 또한 복잡한 네트워크로 인한 과다적합 문제로 인하여 왜곡된 추정 결론을 초래할 수 있다.

따라서 일반화시킬 수 있는 네트워크 설정 문제를 해결하기 위하여 Anthony와 Bartlett (2009), Roy-Chowdhury 등 (1999) 의 연구에서 일부 알고리즘들은 은닉층이 하나인 단순한 모형에서부터 신경망을 학습하기 시작한다. 그리고 점차적으로 복잡한 네트워크로 확장시켜 학습하는 방법을 적용하기도 한다. 이 경우에 만약 신경망이 수렴되지 않으면서 발산하거나 또는 지역 최소값에 수렴되는 경우에는 은닉층의 노드를 하나씩 추가한 새로운 네트워크를 이용하여 더 높은 차원의 모수공간으로 확대하여 학습할 수 있다. 따라서 학습 속도 뿐만 아니라 전역 최소값으로의 수렴 가능성도 높아진다. 그러나 여기서 발생할 수 있는 또 다른 문제점은 확장된 일반화 네트워크에 필요한 가장 최적의 초기 가중값을 어떻게 다시

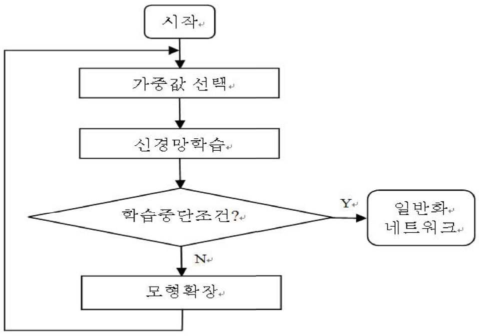


**Figure 2.1.** Modified learning process for generalized neural network

설정하는가에 있으며 또한 가장 적은 회수의 학습량을 이용하여 요구한 허용오차로 어떻게 수렴될 수 있게 하느냐가 중요한 문제가 될 수 있다고 Yam과 Chow (2000) 등에서 연구되었다.

본 연구에서는 일반화 네트워크를 찾기 위하여, 은닉노드의 개수와 학습패턴을 각각 하나씩 점차적으로 증가시키면서 학습주기마다 확인되는 허용오차의 변화를 이용하여 각 단계에서 비교된 최적의 모형을 선택할 수 있도록 하는 일반화 신경망의 개선된 학습과정을 제안한다. 그리고 이 개선된 학습 알고리즘을 신경망에 적용할 때 나타나는 신경망 학습의 효과 즉, 초기 가중값, 학습량 그리고 학습오차 등과 같은 주요 학습률들의 변화에 대하여 살펴본다. 개선된 학습과정은 초기 가중값의 범위설정에 관한 Wu와 Zhang (2002)의 연구 결과와 Liu 등 (2001)의 신경망 학습과정을 이용하고, Al-Shareef와 Abbod (2010)의 최적화 초기 가중값을 이용하 여 일반화 네트워크를 구할 수 있도록 한다.

본 연구의 구성은 다음과 같다. 2장에서는 학습의 각 단계에서 비교된 최적의 네트워크를 일반화 네트워크로 선택하기 위하여 제안된 일반화 신경망의 개선된 학습과정에 대하여 논의한다. 3장에서는 개선된 학습과정을 이용한 일반화 신경망으로 전형적인 비선형 함수를 학습 추정하고, 기존의 학습 알고리즘을 통한 학습 효율성의 비교를 제시한다. 4장에서는 이에 대한 결론과 향후 연구방향을 다룬다.

## 2. 일반화 신경망의 개선된 학습과정

본 연구에서 제시하는 일반화 신경망의 개선된 학습과정의 전체적인 흐름도는 Figure 2.1과 같다. Figure 2.1에서 가중값은 설정된 각 구간에서 발생시킨다. 발생된 가중값을 이용하여 네트워크를 학습시키고, 주어진 허용오차를 만족할 때까지 네트워크를 단순 모형에서부터 점차로 확장시키는 학습과정을 통하여 일반화 네트워크를 구축해 나간다.

개선된 신경망 학습과정은 일반화 네트워크의 최적 설정을 중심으로 한다. 개선된 알고리즘에서 학습패턴들은 학습과정에서 허용오차를 만족하지 않을 때마다 하나씩 차례로 증가시키며, 하나의 패턴은 모든 패턴이 학습될 때까지 먼저 학습된 패턴들과 결합하여 학습되도록 선택한다. 학습 초기에는 한 개의 패턴으로 시작하고 여기서의 허용오차를 만족할 수 있도록 가중값을 선택한다. 이때 적용하는 가중값은 Gunaseeli와 Karthikeyan (2007), 그리고 Wu와 Zhang (2002)의 범위를 이용한다. 학습주기마다 결정되는 모형은 학습자료로부터 추가적인 패턴으로 결정하게 된다. 일반적으로 새롭게 추가된 패턴을 이용하면 허용오차보다 더 큰 오차를 발생시킬 수 있다. 그러므로 만약 오차가 주어진 허용오차를 만족하면서 감소하고 있다면, 또 다른 패턴이 같은 방법으로 결합되면서 학습할 수 있다. 만약 그렇지 않다면 네트워크 모수의 개수를 증가 시킨다. 이러한 방법으로 증가된 은닉층 노드와 이와 연결된 새로운 초기 가중값들을 결정할 수 있는 방법을 개발할 필요가 있다. 여기서 시스템 오차를 제로에 가깝게 감소시킬 수

있는 가중값들의 선택은 이차계획법(quadratic programming)과 같은 최적화 기법들을 이용한다. 이와 같은 방법을 이용한 개선된 학습 알고리즘을 수식화하여 그 학습 결과를 전형적인 FNN 학습 알고리즘을 이용한 결과와 비교해 본다.

본 연구에서 신경망 학습을 위해 고려되는 네트워크는 입력값 $N_i$ 은닉층 노드 $N_h$ 출력층 노드 $N_o$개로 구성된 FNN이다. 먼저 $W_{jk}$는 $k$번째 입력노드와 $j$번째 은닉층 노드 사이의 연결 가중값이고, $v_{ij}$는 $i$번째 출력층 노드와 $j$번째 은닉층 노드 사이의 연결 가중값이라 하면, 은닉층과 출력층의 절편항은 각각 다음과 같다.

$$W = [w_{jk}] \in R^{N_h \times (N_i+1)},$$
$$V = [v_{ij}] \in R^{N_0 \times (N_h+1)}.$$

학습 자료인 입력패턴과 출력패턴은 각각 다음과 같다.

$$x = [x_0, x_1, \dots, x_{N_i}]^T,$$
$$d = [d_0, d_1, \dots, d_{N_0}]^T,$$

여기서 $x_0 = 1$은 가중값 $w_{j0}$이고 $j = 1, 2, \dots, N_h$이다. 따라서 네트워크에서 은닉층의 출력값과 출력층의 출력값은 각각 다음과 같다.

$$y_i = f\left(\sum_{k=0}^{N_i} w_{jk} x_k\right), \quad j = 1, 2, \dots, N_h, \tag{2.1}$$

$$z_i = f\left(\sum_{j=0}^{N_i} v_{ij} y_j\right), \quad i = 1, 2, \dots, N_o, \tag{2.2}$$

여기서 $y_j$는 $j$번째 은닉노드의 출력값이고 $z_i$는 $i$번째 출력노드의 출력값이다. 그리고 $f(\cdot)$는 시그모이드형 변환함수로서

$$f(u) = \frac{1}{(1 + e^{-u})}$$

이다. 또한 $f^{-1}(\cdot)$은 존재하고, 모든 $u \in R$에 대하여 $f(0) = 0$, $f(u) = -f(-u)$이다. $i = 1, 2, \dots, N_o$에서 $y_0 = 1$과 $v_{i0}$는 출력층의 절편항이다. 그리고 주어진 패턴에 대한 오차함수 $E_p$는 다음과 같다.

$$E_p = \frac{1}{2}\sum_{i=1}^{N_o}(z_i - d_i)^2.$$

신경망 학습에서 $N_p$패턴은 $p = 1, 2, \dots, N_p$일 때 $x^p = [x_0^p, x_1^p, \dots, x_{N_i}^p]^T$와 $d^p = [d_0^p, d_1^p, \dots, d_{N_0}^p]^T$이다. 또한 $x_0^p = 1$이다. 따라서 신경망은 전체 오차 $E(N_p)$가 아주 작은 허용오차 $\varepsilon > 0$를 만족시키는 네트워크를 찾는 것이다. 여기서 $E(N_p)$는 다음과 같다.

$$E(N_p) = \frac{1}{N_p}\sum_{p=1}^{N_p} E_p.$$

학습할 문제로 새로 결합될 패턴이 있다고 하고 그 패턴번호가 $L$이라고 하자. 이때 $L - 1 < N_p$패턴에 대하여 학습된 신경망이 허용오차 $\varepsilon$를 만족하고 있다고 가정하자. 그러면 오차 $E(L) < \varepsilon$를 만족

하는 신경망을 찾는 것이 새로운 학습 목표가 된다. 시스템에서 발생시킬 가중값으로는 Wu와 Zhang (2002)의 초기 가중값 구간인 $[-c, c]$의 균등분포로부터 랜덤하게 발생시켜 학습을 시작하면서 은닉노드가 $N_h$개인 신경망을 두 가지 방법으로 연속적으로 학습한다. 신경망 구조는 Al-Shareef와 Abbod (2010)의 최적화 초기 가중값을 이용한다. 개선된 가중값의 발생방법은 다차원 자료에 확장 적용될 수 있다. 첫번째는 새로운 문제에서 $L$패턴에 대한 특정 오차를 구할 수 있도록 학습하는 경우이고, 두번째는 지역 최소값에 빠지는 경우로서 큰 오차로 인한 시스템의 학습 종료가 발생되는 경우이다. 첫번째 경우에서는, 같은 과정을 순차적으로 새로운 패턴에 반복 적용함으로써 신경망이 목적한 바를 구할 수 있도록 학습할 수 있다. 두번째 경우에서는, 단지 은닉노드의 개수가 $N_h$인 신경망 구조로는 오차를 $\varepsilon$ 보다 작게 줄일 수 없다는 것을 의미한다. 다시 말하여 신경망은 오차를 더욱 감소시키기 위하여 최소한 은닉노드의 개수를$N_h + 1$로 확장하게 되고, 따라서 이와 같은 경우의 학습문제는 다시 $N_h + 1$개의 은닉노드와 $L$개의 패턴을 갖는 새로운 신경망으로 재구성 된다.

$L - 1$패턴에 대한 원래의 신경망 구조가 최소 개수의 은닉노드를 갖는다면 $L$패턴에 대한 신경망 학습 결과도 거의 유사하게 최소 개수의 노드를 갖는다. 또한 어떤 하나의 패턴에 대하여, 한 개의 은닉노드 즉 $N_h + 1$인 신경망에 대하여 오차가 제로가 되는 가중값 $W \in R^{1\times(N_i+1)}$와 $V \in R^{(N_o\times 2)}$는 항상 존재함을 증명할 수 있다. 그러면 시스템의 허용오차를 만족시키기 위하여 두번째 패턴을 첫번째 패턴과 함께 학습할 수 있다. 학습과 모형 확장이라는 두 가지 단계의 학습과정은 모든 패턴이 함께 학습될 때까지 반복한다. 이와 같이 일반화 네트워크를 위한 은닉노드의 개수가 최소 개수로 확장될 수 있고, 궁극적으로 해에 수렴될 수 있는 신경망을 구할 수 있도록 개선된 학습과정의 개요는 다음과 같다.

**알고리즘 2.1**: 개선된 신경망 학습과정: 입력 : 출력 패턴은 각각 입력 패턴 $N_p$와 연결된다. 출력 : 주어진 허용오차 $\varepsilon$을 만족하고 가능한 최소 개수의 은닉노드로 학습된 네트워크.

단계 1: 먼저 $l = 1$로 설정한다. 학습 자료로 부터 한 개의 패턴을 선택한다. 허용 오차 $\varepsilon$를 만족하기 위해 선택된 패턴을 이용하여 은닉노드 하나로 신경망을 학습한다. 초기 가중값은 주어진 설정 범위 내에서 발생된다.

단계 2: 만약 $l < N_p$가 만족되면 특정 범주에 맞는 다음 번 패턴을 선택한다. 그리고 $l = l + 1$로 설정하고 단계 3으로 간다. 그렇지 않으면 끝.

단계 3: 만약 학습 알고리즘이 $\varepsilon$을 만족하는 범위 안에서 $E(l)$을 감소시킬 수 있으면 단계 2로 간다. 그렇지 않으면 단계 4로 간다.

단계 4: 최종적으로 변화된 신경망의 가중값을 저장한다. 은닉노드의 개수를 하나 증가시키고, 새롭게 변화된 모형에 대한 초기 가중값을 주어진 설정범위 내에서 다시 설정한다. 단계 3으로 간다.

알고리즘 2.1의 단계 1과 단계 4에서 초기 가중값은 Gunaseeli와 Karthikeyan (2007)의 구간에서 발생시키고, Wu와 Zhang (2002)의 초기 가중값 구간인 $[-c, c]$까지 균등분포로부터 확장하면서 랜덤하게 발생시킨다. 단계 4에서 초기 가중값의 결정은 다음과 같이 이차계획법을 이용한다. 먼저 알고리즘의 단계 3에서 실행된 가중값 행렬을 $W$와 $V$라고 하자. 또한 네트워크의 입출력 관계식을 행렬식으로 표현하기 위하여 $x^p$를 주어진 입력 패턴이라 하고 $z^p_\alpha$는 목표값 $z^p$에 대한 실제 출력값이라고 하면 다음과 같이 수식화 된다.

$$\mathbf{y^p} = \begin{pmatrix} 1 \\ \cdots\cdots \\ f(Wx^p) \end{pmatrix} = [1, f(W_1x^p), \ldots, f(W_{N_h}x^p)]^T$$

$$Z^p_\alpha \cong f^{-1}(z^p) = Vy^p. \tag{2.3}$$

여기서 $W_1, W_2, \ldots, W_{N_h}$는 행렬 $W$의 행 벡터이다. 은닉노드가 추가된 후의 새로운 가중값 행렬은 다음과 같다.

$$\overline{\mathbf{W}} = \begin{pmatrix} W \\ \cdots \\ W_r, \end{pmatrix}$$

$$\overline{V} = [V \vdots V_c],$$

여기서 $W_r$는 추가된 가중값에 대한 행 벡터이고 $V_c$는 열 벡터이다. 입력패턴 $x^p$을 확장된 신경망 모형으로 표현하면 다음과 같다.

$$\overline{y}^p = [1, f(W_1 x^p), \ldots, f(W_{N_h} x^p), f(W_r x^p)]^T, \tag{2.4}$$

$$\overline{z_\alpha}^p = [V \vdots V_c]\overline{y}^p = Vy^p + V_c f(W_r x^p) = z_\alpha^p + V_c f(W_r x^p). \tag{2.5}$$

따라서 모든 적용 가능한 패턴 $p$에 대해서 $\overline{z_a}^p \approx d_a^p$가 된다. 여기서 $d_a^p \cong f^{-1}(d^p)$와 $d^p$는 $x^p$에 대응되는 목표 출력값이다. 식 (2.4)와 (2.5)로부터, 만약 패턴 $x^p$가 지역 최소값으로 수렴시키게 하는 새로운 패턴이라면 $x^L$로 표시한다. 따라서 지금까지의 $x^p$에서 $p = 1, 2, \ldots, L-1$이 된다. 즉 $W_r$과 $V_c$를 결정할 때 $\overline{z_a}^L = d_a^L$을 고려할 수 있고 따라서 다음과 같다.

$$V_c f\left(W_r x^L\right) = d_a^L - Vy^L. \tag{2.6}$$

식 (2.6)은 $x^L$과 $d_a^L - Vy^L \neq 0$로 나타내는 한 개 패턴에 대한 식이다. 또한 만약 패턴 $x^P$가 사전에 학습된 패턴이라면 $W_r$과 $V_c$의 결정은 다음 식에 의한다.

$$V_c f(W_r x^p) = \overline{z_a}^p - \overline{z_a}^p \approx 0, \quad p = 1, 2, \ldots, L-1. \tag{2.7}$$

새로운 노드에 대응되는 가중값인 $W_r$과 $V_c$를 결정할 수 있다면 은닉노드를 새롭게 추가한 후에도 $\overline{W}$와 $\overline{V}$로 주어지는 가중값을 갖는 확장된 신경망은 식 (2.6)과 (2.7)을 만족한다. 먼저 $V_c$는 상수라고 가정하자. 즉 네트워크는 단지 한 개의 노드를 갖는다. 여기서 설명되는 접근방법은 출력노드가 한 개 이상의 네트워크에도 쉽게 적용될 수 있지만, 본 연구에서는 출력노드를 한 개로 단순화 시킨 경우만 다룬다. 식 (2.6)으로부터 $V_c$가 다음 식처럼 선택될 수 있음을 알 수 있다.

$$V_c = \alpha\left(d_a^L - Vy^L\right), \tag{2.8}$$

여기서 $\alpha$는 $|\alpha| > 1$을 만족하는 상수이다. 또한 시그모이드형 변환함수는 $|f(\cdot)| < 1$를 만족한다. 따라서 식 (2.6)은

$$f\left(W_r x^L\right) = \frac{1}{\alpha}$$

또는

$$W_r x^L = f^{-1}\left(\frac{1}{\alpha}\right) \tag{2.9}$$

와 같다.

최적 처리절차를 이용하여 $W_r$를 계산하기 위하여 식 (2.7)은 다음과 같은 범위를 만족한다.

$$-\delta \le V_c f\left(W_r x^p\right) \le \delta, \quad p = 1, 2, \ldots, L-1, \tag{2.10}$$

여기서 $\delta > 0$는 아주 작은 상수이다. 따라서 식 (2.10)은 다음과 같이 정리될 수 있다.

$$-f^{-1}\left(\frac{\delta}{|V_c|}\right) \le W_r x^p \le f^{-1}\left(\frac{\delta}{|V_c|}\right), \quad -\lambda \le W_r x^p \le \lambda,\ p = 1, 2, \ldots, L-1, \tag{2.11}$$

여기서 $\lambda > 0$은 아주 작은 값이다. 그러므로 식 (2.9)와 식 (2.11)의 조건 하에서 다음 식을 최소화 시키는 이차계획법을 사용할 수 있다.

$$W_r W_r^T + k\lambda^2. \tag{2.12}$$

만약 다음과 같이 부등식

$$-\lambda + f^{-1}\left(\frac{1}{\alpha}\right) \le W_r x^L \le f^{-1}\left(\frac{1}{\alpha}\right) + \lambda \tag{2.13}$$

를 이용하여 식 (2.9)를 부등식으로 표현하는 경우에도 식 (2.11)과 식 (2.13)의 조건 하에서 $\lambda > 0$을 최소화한다면 같은 방법으로 최적처리절차를 이용하여 추가된 가중값 벡터 $W_r$을 계산할 수 있다.

## 3. 시뮬레이션 연구

일반화 신경망의 학습과정을 개선시키기 위한 네트워크를 학습 추정하기 위하여 다음 식 (3.1)과 같은 비선형 함수를 이용한다.

$$f(x) = e^{-(x-1)^2} + e^{-(x+1)^2}, \quad x \in [-2.6, 2.6]. \tag{3.1}$$

실험을 단순화하기 위하여 12개의 모의 학습자료 $x$를 구간 $[-2.6, 2.6]$의 균등분포에서 발생시킨다. 발생된 표본은 평균 0이고 분산 0.005인 정규분포를 따르는 오차를 포함하고 있다. 적용된 학습 알고리즘의 단계 3은 LM(Levenberg-Marquardt) 알고리즘을 이용한다. 학습의 중단시점으로 설정한 최종 학습 허용오차는 $\varepsilon = 0.01$이고, 초기 가중값의 범위는 $|c| < w$이다. Yam과 Chow (2000)의 연구와 이를 이용한 실증연구인 Gunaseeli와 Karthikeyan(2007)의 초기 가중값은 $|w| < 4.36$이다. 이 초기 가중값과 함께 여기서의 가중값의 초기 발생 범위는 $w = 0.001 \sim 10$이다. 또한 오차 판단기준은 MSE이다.

비선형 함수 식 (3.1)에 대한 Wu와 Zhang (2002)의 학습 결과와 본 연구에서 개선한 학습과정을 이용한 결과의 비교는 Table 3.1과 같다. Table 3.1에서 첫번째 열은 두 가지 학습결과를 비교하기 위하여 초기 가중값의 발생 범위의 설정을 나타내고 있다. 개선된 학습 과정은 설정된 각 구간에서 발생시킨 초기값을 이용하여 네트워크를 학습한다. 이때 주어진 허용오차를 만족할 때까지 네트워크를 확장시키면서 최종적으로 선택된 일반화 네트워크를 이용하여 비선형 함수를 추정하고, 그때의 학습량과 MSE를 각각 계산한 결과이다. 이러한 개선된 학습결과를 Wu와 Zhang (2002)의 연구 결과와 비교하였다. 여기서 초기 가중값의 범위는 $w = 0.01$ 주변에서 두 방법간의 최선의 결과를 보여주고 있으며, $w = 5$와 $w = 10$인 경우에 Wu와 Zhang (2002)의 연구 결과에서는 학습이 수렴되지 않고 발산되고 있는 경우도 있지만 개선된 알고리즘에서는 수렴으로 진행되고 있음을 알 수 있다. Gunaseeli와 Karthikeyan (2007)의 초기 가중값 발생 범위 $|w| < 4.36$에서는 두 연구 결과가 모두 수렴되고 있음을 알 수 있다. 이 경우에도 일반화 신경망의 개선된 학습과정의 학습량과 학습오차가 좀 더 효율적임을 알 수 있다.

**Table 3.1.** Learning result depend in initial values.

| $c$ | Wu와 Zhang (2002)의 결과 | | 개선된 알고리즘 | |
|---|---|---|---|---|
| | 학습량 | MSE | 학습량 | MSE |
| 0.001 | 13172 | 0.013661 | 1227 | 0.012571 |
| 0.01 | 4644 | 0.010080 | 452 | 0.010102 |
| 0.1 | 18726 | 0.011074 | 684 | 0.010874 |
| 1 | 5863 | 0.011755 | 576 | 0.011545 |
| 2 | 6963 | 0.01227 | 635 | 0.012221 |
| 5 | $\infty$ | | 701 | 0.012567 |
| 10 | $\infty$ | | 817 | 0.012983 |

## 4. 결론

본 연구에서는 FNN의 학습 초기에 설정한 각 구간에서 발생시킨 가중값을 이용하여 네트워크를 학습시키고, 주어진 허용오차를 만족할 때까지 네트워크를 단순 모형에서부터 점차로 확장시키면서 일반화 네트워크를 구하는 개선된 학습과정을 제안하였다. 또한 이 학습과정을 이용한 신경망 학습을 통하여, 각 초기 가중값에 따른 학습량과 학습오차를 비교하였다. 제안된 개선된 학습과정은 Liu 등 (2001)의 학습 알고리즘과 Wu와 Zhang (2002)의 초기 가중값, Gunaseeli와 Karthikeyan (2007)의 초기 가중값의 범위 설정에 관한 연구 결과를 이용하여 일반화 네트워크를 구하고 이들의 학습효율을 살펴보았다. 일반화 네트워크를 구하여 이를 통한 신경망 학습을 하도록 제안된 개선된 알고리즘은 Wu와 Zhang (2002)의 초기 가중값의 범위 설정에 관한 연구 결과와 비교할 때 거의 유사하게 허용오차 내에서 시스템의 오차를 만족했다. 그리고 학습의 발산이 나타나는 경우에도, 제안된 개선된 학습과정은 수렴으로 진행되고 있음을 알 수 있다. 그리고 Gunaseeli와 Karthikeyan (2007)의 초기 가중값 발생 범위를 적용한 경우에서도 두 연구 결과가 비슷한 결과를 보였다. 이 연구에서는 일반화 네트워크를 찾는 과정의 복잡성 때문에 출력노드를 한 개로 제한하였지만, 보다 복잡한 네트워크에 적용하는 문제는 향후 과제로 남긴다.

## References

Al-Shareef, A. J. and Abbod, M. F. (2010). Neural Network Initial Weights Optimisation, *12th International Conference on Computer Modelling and Simulation*, 57–61.

Anthony, M. and Bartlett, P. L. (2009). *Neural Network Learning: Theoretical Foundations*, Cambridge University Press.

Diotalevi, F. and Valle, M. (2001). Weight Perturbation Learning Algorithm with Local Learning Rate Adaptation for the Classification of Remote-Sensing Images, *Proceedings of European Symposium on Artificial Neural Networks*, 217–222.

Framling, K. (2004). Scaled Gradient Descent Learning Rate: Reinforcement Learning with Light-seeking Robot, *Proceedings of International Conference on Informatics in Control, Automation and Robotics*, 3–11.

Fukuoka, Y., Matsuki, H., Minamitani, H. and Ishida, A. (1998). A modified back-propagation method to avoid false local minima, *Neural Network*, **11**, 1059–1072.

Gunaseeli, N. and Karthikeyan, M. (2007). A Constructive Approach of Modified Standard Backpropagation Algorithm with Optimum Initialization for Feedforward Neural Networks, *International Conference on Computational Intelligence and Multimedia Application*, 325–331.

Haykin, S. (2010). *Neural Networks and Learning Machines*, 3rd Ed., PHI Learning Private Limited.

Liu, D., Chang, T. S. and Zhang, Y. (2001). A New Learning Algorithm for Feedforward Neural Networks, *Proceedings of IEEE, International Symposium on Intelligent Control,* 39–44.

Maasoumi, E., Khotanzad, A. and Abaye, A. (1994). Artificial neural networks for some macroeconomic series : A first report, *Econometric Reviews*, **13**, 105–122.

Maghami, P. G. and Sparks, D. W. (2000). Design of neural networks for fast convergence and accuracy: Dynamics and control, *IEEE Trans. Neural Networks*, **11**, 113–123.

Parekh, R., Yang, J. and Honavar, V. (2000). Constructive neural-network learning algorithms for pattern classification, *IEEE Trans. Neural Networks*, **11**, 436–451.

RoyChowdhury, P., Singh, Y. P. and Chansarkar, R. (1999). Dynamic tunneling technique for efficient training of multilayer perceptrons, *IEEE Trans, Neural Networks*, **10**, 48–55.

Sharda, R. and Patil, R. B. (1990). Neural Networks as Forecasting Experts : An Empirical Test,*Proceeding of the IJCNN Meeting*, 491–494.

White, H. (1988). Economic Prediction Using Neural Networks : The Case of IBM Stock Prices, *Proceedings of the Second Annual IEEE Confernece in Neural Networks*, **2**, 451–458.

Wu, Y. and Zhang, L. (2002). The Effect of Initial Weight, Learning Rate and Regularization on Generalization Performance and Efficiency, *Proceedings on ICSP*, 1191–1194.

Yam, J. Y. F. and Chow, T. W. S. (2000). A Weight Initialization Method for Improving Training Speed in Feedforward Neural Network, *IEEE Trans. Neural Networks*, **30**, 219–232.

Zhang, Y. and Jiang, Q. (2010). An Improved Initial Method for Clustering High-Dimensional Data, *2nd International Workshop on Database Technology and Applications*, 1-4.

# 일반화 신경망의 개선된 학습 과정을 위한 최적화 신경망 학습률들의 효율성 비교

윤여창[a] · 이성덕[b,1]

[a]우석대학교 정보보안학과, [b]충북대학교 정보통계학과

(2013년 10월 1일 접수, 2013년 10월 16일 수정, 2013년 10월 16일 채택)

**요 약**

본 연구에서는 Liu 등의 학습 알고리즘과 Wu와 Zhang의 초기 가중값의 범위 설정, 그리고 Gunaseeli와 Karthikeyan의 초기 가중값에 관한 연구 결과를 이용하여 일반화 네트워크를 구할 수 있는 개선된 학습을 제안하고, 최적화된 신경망 학습률들을 이용하여 개선된 학습 과정의 학습효율등을 비교해 본다. 제시된 알고리즘을 이용한 학습에서 학습 초기에는 가장 단순한 학습 패턴과 은닉층으로부터 학습을 시작한다. 신경망 학습과정 중에 지역 최소값에 수렴되는 경우에는 가중값 범위 조정을 통하여 지역 최소값 문제를 해결하고, 지역 최소값으로부터 탈출이 용이하지 않으면 은닉노드를 점차적으로 하나씩 추가하면서 학습한다. 각 단계에서 새롭게 추가된 노드에 대한 초기 가중값의 선택은 이차계획법을 이용한 최적 처리절차를 이용한다. 최적 처리절차는 은닉층의 노드가 추가된 후의 새로운 네트워크에서 학습회수를 단순히 증가시키지 않아도 주어진 학습 허용오차를 만족시킬 수 있다. 본 연구에서 적용한 개선된 알고리즘을 이용하면서 초기 가중값들에 관한 기존 연구들을 적용하면 신경망 학습시의 수렴 정도를 높여주고 최소한의 단순 구조를 갖는 일반화 네트워크로 추정할 수 있게 된다. 이러한 학습률들을 변화시키는 모의실험을 통하여 기존의 연구 결과와의 학습 효율을 비교하고 향후 연구 방향을 제시하고자 한다.

주요용어: 학습 알고리즘, 학습률, 가중값 범위 조정, 이차계획법.

---

이 논문은 2011 년 충북대학교 학술지원사업의 연구비 지원에 의해 연구되었음.
[1]교신저자: (361-763) 충북 청주시 내수동로 52, 충북대학교 정보통계학과, 교수.
E-mail: sdlee@chungbuk.ac.kr