다층 퍼셉트론에서 구조인자 제어 영향의 비교
(Comparison of Factors for Controlling Effects in MLP Networks)

윤 여 창 †
(YeoChang Yoon)

요 약 다층 퍼셉트론(Multi-Layer Perceptron, MLP) 구조는 그의 비선형 적합능력으로 인하여 매우 다양한 실제 문제에 적용되고 있다. 그러나 일반화된 MLP구조의 적합능력은 은닉노드의 개수, 초기 가중값 그리고 학습 회수 또는 학습 오차와 같은 구조인자(factor)들에 크게 영향을 받는다. 만약 이들 구조인자가 부적절하게 선택되면 일반화된 MLP구조의 적합능력이 매우 왜곡될 수 있다. 따라서 MLP구조에 영향을 주는 인자들의 결합 영향을 살펴보는 것은 중요한 문제이다. 이 논문에서는 제어상자(controller box)를 통한 학습결과와 더불어 MLP구조를 일반화할 때 영향을 줄 수 있는 신경망의 일반적인 구조인자들을 실증적으로 살펴보고 이들의 상대효과를 비교한다.

키워드 : 다층 퍼셉트론, 제어상자, 복잡도 제어

Abstract Multi-Layer Perceptron network has been mainly applied to many practical problems because of its nonlinear mapping ability. However the generalization ability of MLP networks may be affected by the number of hidden nodes, the initial values of weights and the training errors. These factors, if improperly chosen, may result in poor generalization ability of MLP networks. It is important to identify these factors and their interaction in order to control effectively the generalization ability of MLP networks. In this paper, we have empirically identified the factors that affect the generalization ability of MLP networks, and compared their relative effects on the generalization performance for the conventional and visualized weight selecting methods using the controller box.

Key words : Multi-Layer Perceptron, controller box, complexity control

1. 서 론

MLP(Multi-Layer Perceptron)구조에 대한 학습은 일반적으로 충분한 개수의 입력과 출력의 쌍으로 이루어진 유한개수의 표본으로 시작된다. 이러한 표본은 전체 모집단을 일반화시키는네트워크 구조를 도출해 내는 학습에 매우 중요하다. 제한된 표본으로부터 추정된 모형을 이용하여 일반화된 네트워크의 능력을 제어하는 것은 복잡도 제어(complexity control)를 통하여 이루어지게 된다[1,2].

일반적으로 모형의 복잡도를 제어하기 위한 방법은 은닉노드의 개수 조절이다. 그러나 MLP구조에서 일반적으로 사용되는 역전파(backpropagation) 알고리즘을 이용한 학습은 은닉노드 개수나 초기 가중값들과 같은 구조인자들을 쉽게 정량화시키기 어렵다.

학습 알고리즘의 일반화된 영향을 이해하기 위해서, 먼저 MLP학습을 위한 최적의 비선형 처리절차가 가중값 공간의 적절한 경로(path)를 따라 규정되어야 한다. 일반적으로 역전파 알고리즘에서는 경사하강추적(gradient descent)방법상의 경로를 따라 경험적인 오류를 감소시키고 있으며, 이를 통하여 가능한 학습 결과가 적절한 경로상에 위치하게 된다. 학습 과정의 경로는 주어진 학습 자료, 비선형 변환함수, 학습 경로상의 초기 가중값 그리고 최종 학습결과를 보여주는 학습회수 또는 학습오차등에 주로 영향을 받게 된다.

Zhong과 Cherkassky[3]는 MLP구조에서 은닉노드의 개수, 초기 가중값 그리고 학습오차등과 같은 구조인자들의 결합 영향을 통한 복잡도 제어에 관하여 논의하고 있다. Yoon[4]의 제어상자[5]를 통한 개선된 학습과정은 초기 가중값의 시각적인 실시간 선택을 통하여 가장 일반적인 네트워크 구조를 유도하고 있다.

그러나 MLP네트워크의 구조인자들을 일반화한 최적 알고리즘의 영향을 분석하기 위하여 본 논문에서는 주어진 학습 자료와 비선형 변환함수는 고정된다고 가정한다. 일반적으로 MLP의 오차평면은 많은 지역 최소값

† 종신회원 : 우석대학교 전산통계학과 교수
yoonyc@core.woosuk.ac.kr
논문접수 : 2003년 4월 10일
심사완료 : 2004년 3월 5일

들을 갖고있기 때문에 특별한 지역최소값은 학습 경로상의 초기 가중값 그리고 최종 학습결과를 보여주는 학습회수를 변화시키면서 찾는다. 예를들어 초기 가중값을 확률난수로부터 발생시킨 작은 값으로 설정할 때, 역전파 알고리즘은 매우 작은 가중값으로 인한 지역최소값에 수렴하는 경향이 있다. 그러나 매우 큰 값을 경사하강추적법 또는 허용오차로 부여하여 학습회수를 제어하면 가중값 경로상에 위치하는 해들을 찾지 못하는 경우도 발생한다. 따라서 초기 가중값의 잘못된 적용은 학습에 매우 큰 영향이 있으므로 학습과정중에 실시간으로 초기값을 재설정함으로써 지역최소값을 벗어날 수 있게 하는 방법을 함께 고려할 수 있다[4].

역전파 알고리즘을 이용한 대부분의 실증연구들은 초기 가중값들의 조건과 변환함수들을 적절히 이용한다[6,7,8,9]. 그러나 이 논문에서는 실제적인 학습의 결과로서 모형 복잡도에 대한 일반적인 초기 가중값과 제어상자를 통한 초기 가중값 그리고 최종 학습결과를 보여주는 학습회수등의 구조인자들에 대한 결합영향을 살펴보고자 한다. 결합영향을 살펴보는 것은 MLP구조를 과다적합하여 추정하는 것을 제어할 수 있다. 또한 통계적 학습이론과 VC이론은 복잡도 제어를 위한 바람직한 대안을 제시할 수 있다[2].

본 논문의 구성은 다음과 같다. 2장에서 우리는 SRM 구조하에서 서로 다른 구조인자들에 대한 일반적인 MLP 복잡도 제어를 논의한다. 3장에서는MLP구조의 일반화를 제어하기 위하여 사용될 수 있는 MLP복잡도 제어를 통하여 각 구조인자들의 결합영향을 살펴본다. 4장은 이에 대한 결론을 다룬다.

2. MLP네트워크의 복잡도 제어

본 논문에서는 Vapnik[2]의 SRM에 의해 제안된 구조와 윤여창[4]의 제어상자를 이용한 구조인자의 실시간 선택방법에 주로 기초하고 있다. SRM(Structural risk minimization)에 따르면 유한 표본으로부터 학습문제를 해결하기 위해서는 근사함수에 대한 구조와 같은 선험적인 규정이 필요하며, MLP에서 구조는 그들의 복잡도에 따라 순서화 된 MLP모형이다. MLP에서 복잡도는 모형의 구조들에 따라 각기 다른 방법들이 이용된다. 따라서 SRM하에서 학습은 주어진 MLP구조 하에서 학습오차의 최소화 그리고 가장 작은 예측오차를 제공하는 모형선택등에 영향을 주로 받는다.

MLP에서 그 구조는 초기 가중값, 은닉노드의 개수 그리고 허용오차에 따라 각각 다음과 같이 정의될 수 있다[2,4].

2.1 초기 가중값

먼저 다음과 같은 구조를 고려하자.

$$S_i = \{A : f(\overrightarrow{x}, \overrightarrow{w}),\ \| w_j^0 \| \le c_i\},\quad j = 1, 2, \ldots, n,\ c_1 < c_2 < c_3 \ldots \tag{1}$$

여기서 $\overrightarrow{w}^0 = \{w_j^0\}$는 최적 알고리즘 A에서 사용되는 초기 가중값 벡터이며 i는 그 구조의 첨자이다. 경사하강추적법은 초기 가중값 주변에서 지역최소값을 찾기 때문에, 전역최소값은 $\| w_j^0 \| \le c_i$을 만족하는 여러 가지 초기조건으로 시작하면서 경험적 위험을 최소화시키면서 가장 잘 적합되는 가중값을 초기 가중값으로 선택한다. 식 (1)의 구조식 S_i는 초기 조건인 $\overrightarrow{w}^0$를 갖는 함수들에 적용되는 모수 추정을 위한 최적 알고리즘 A에 대하여 각각 정의된다. 여기서 경험적 오류는 $\| w_j^0 \| \le c_i$를 만족하는 모든 초기 조건에 대하여 최소화 된다. 비록 신경망의 학습시간을 충분히 함으로써 실제적으로 전역 최소값에 대한 철저한 탐색이 이루어 진다 하여도 초기 가중값은 신경망 학습의 수렴 정도에 중요한 영향을 미칠 수 있다.

또한 학습의 시작단계에서 확률난수를 발생시키는 대신에 학습 자료의 특성에 따라 실시간으로 가장 잘 적합될 수 있는 초기값을 제어상자를 통하여 실시간으로 선택함으로써 지역 최소값을 쉽게 벗어나게 하는 학습과정을 고려할 수 있다. 이때 P차원 가중값 공간을 제어상자의 P차원 가중값 좌표로 일반화 하면 다음과 같이 나타낼 수 있다[4,5].

$$\overrightarrow{w} = \frac{-mk}{m+n}\left(-k \cdot \overrightarrow{l}, -t\right) \tag{2}$$

여기서 가중값 좌표벡터 $\overrightarrow{w}$는 제어상자의 이동변삼각형을 $m : n$으로 내분한 점으로 이루어진 가중값 벡터이다. 제어상자를 통해서 선택된 가중값의 상대좌표는 $(-k,\ t)$이다. $\overrightarrow{l}$은 P차 단위 벡터이다. 여기서 구한 가중값 벡터는 다음 장에서 설명되는 학습 모형의 초기값으로 이용된다.

2.2 은닉노드의 개수

MLP의 사전적 정의는 다음과 같다.

$$f(\overrightarrow{x}, \overrightarrow{w}, V) = \sum_{j=0}^{n} w_j g_j(\overrightarrow{x}, \overrightarrow{v}_j) + w_0 \tag{3}$$

여기서 $g_j(\overrightarrow{x}, \overrightarrow{v}_j)$는 시그모이드형 변환함수이며 모수 $\overrightarrow{v}_j$를 갖는다. 이 구조의 각 요소는 MLP네트워크의 구조인자들이며, n은 은닉노드의 개수다. 따라서 모형선택은 주어진 학습자료에 대한 최적의 은닉노드 개수를 갖는 MLP네트워크를 선택하는 문제이다.

2.3 허용오차

학습의 허용오차를 크게 하는 것은 MLP네트워크의 과다적합을 피하는 일반적인 방법이다. 이 절에서는 수렴과정의 정형화된 영향을 소개하고 과다적합되기 전에 학습을 정지시키는 방법으로 이용될 수 있는 허용오차

를 논의한다. 다음과 같은 구조를 고려하자.

$$S_i = \{A: R_{emp} \le e_i\}, \ e_1 > e_2 > e_3 \cdots \quad (4)$$

여기서 R_{emp}은 경험적인 오류를 줄이기 위한 학습정지 시점을 의미하고, e_i는 최적 알고리즘 A에 대한 최종 학습오차이다.

이와 같은 접근 방법들을 이용한 구조인자들의 영향을 경험적으로 살펴봄으로써 MLP모형의 복잡도를 제어할 수 있다. 다음 장에서 설명되는 MLP복잡도 제어에 대한 분석에서 이들 구조인자들은 경험적으로 하나의 인자가 모형선택에 이용될 때 다른 인자는 고정시킨다. 또한 MLP모형의 예측 효용은 상자그림(boxplot)을 이용하여 비교 분석한다.

3. 실증분석

이 장에서는 제어상자를 통한 초기 가중값의 선택과 더불어 앞 장에서 설명한 세가지 구조인자들이 모형의 일반화에 어떻게 영향을 주는지를 조사한다. 적은 개수의 학습자료를 이용하여 많은 개수의 모수들로 이루어진 MLP네트워크를 의도적으로 학습하기 위하여 다음과 같은 함수를 추정한다고 하자.

$$f(x) = e^{-(x-1)^2} + e^{-(x+1)^2}, \quad x \in [-2.5, 2.5] \quad (5)$$

학습자료: 실험을 단순화하기 위하여 12개의 모의 학습자료 x를 구간 [-2.5, 2.5]의 균등분포에서 발생시키며, 발생된 표본은 평균 0이고 분산 0.005인 정규분포를 따르는 오차를 포함한다고 가정한다.

네트워크 구조: 한 개의 입력값 x와 그에 대응하는 한 개의 출력값 y 그리고 n개의 은닉노드로 이루어진 $1 \times n \times 1$ MLP네트워크를 이용한다. 입력과 출력노드는 선형이며 은닉층은 로지스틱 변환함수를 이용한다.

학습알고리즘: 일반적인 LM(Levenberg-Marquardt) 알고리즘을 이용한다. LM 알고리즘은 비선형함수의 오차제곱합(MSE)을 최소화시키기 위하여 고안되었으며 뉴턴 방법의 일종이다. 이 방법은 효율성을 측정하는 척도가 오차제곱합인 신경망 학습에 잘 적용되며, 적정 개수의 네트워크 모수에 대한 가장 빠른 신경망 학습알고리즘으로 알려져 있다[10]. 최대 학습 회수는 200회까지로 한정하면서 오차 최소화를 검토한다.

초기 가중값: 초기값의 경험적인 범위 설정[4]은 0<c<10이다. 각각의 c값에 대하여 구간 [$-c$, c]로부터 뽑은 확률난수를 이용하여 30회 학습하고, 가장 작은 학습오차를 보이는 초기값을 네트워크의 구조인자로 선택한다. 제어상자를 이용한 실시간 초기 가중값은 같은 c값 구간에서 선택된다. 확률난수로부터 발생시킨 초기값의 적용 결과를 살펴보면서, 네트워크에 가장 잘 적합될 수 있는 가중값들을 실시간으로 다시 초기 가중값으로 선택한 후 학습하게 된다. 선택된 초기값은 식 (2)의 가중값 좌표로부터 구할 수 있다. 그러므로 선택된 네트워크는 확률난수를 이용한 초기화에 의한 가장 근접한 최종 예측모형으로 진행될 수 있다.

학습의 중단시점: 초기 가중값의 영향과 은닉노드의 개수를 조사할 때 0으로 먼저 설정되며, 학습중단 규칙은 서로 다른 충분히 작은 값을 설정한 최종 학습오차 e = {0, 1e-6, 0.001, 0.01}에 의해서 제어된다.

예측효율: 진실함수와 학습된 네트워크에 의해 추정된 모형으로부터 구한 MSE를 이용하여 측정한다. 이에 대한 시각적인 비교를 위하여 네트워크에 의해 추정한 실제 출력값을 그래픽으로 표현한다. 추정량들은 같은 크기의 모의 추출된 학습표본에서 구한 네트워크에 의해 구해지며, 각 실험은 같은 크기의 모의 추출된 서로 다른 모의 학습표본을 이용하여 200회 반복한다. 그리고 예측된 MSE의 경험 분포는 그 분포의 95, 75, 50, 25 그리고 5퍼센타일을 표시할 수 있는 상자그림으로 나타낸다.

분석 결과를 시각적으로 나타내기 위하여 먼저 적합그림에서 각 점(t)들은 학습표본이고 출력값(z)들은 각각 실선으로 연결한다. 그리고 상자그림에서는 주어진 은닉노드의 개수(n)와 가중값의 범위(c)에 대한 MSE를 나타낸다.

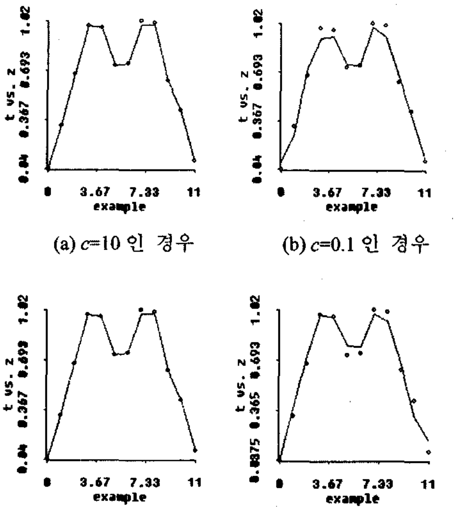
3.1 초기 가중값의 효과[11,12]

초기 가중값의 효과를 파악하기 위하여 먼저 n=9, e=0 그리고 초기값의 범위를 c = {10, 0.1}로 설정한다. 최대 학습 회수는 200회이다.

학습표본의 개수가 12개이고 은닉노드의 개수가 9개인 경우에 그림 1의 (a)와 (c)는 과다 적합인 경우로서 이는 가중값의 초기화가 MLP네트워크의 일반적인 적합능력에 영향을 주고 있음을 보여준다. 은닉노드의 개수를 9로 고정하였을 때, 초기 가중값이 크면 주어진 학습자료를 너무 복잡한 모형으로 추정함으로써 과다적합을 일으킨다. 이와는 반대로 초기 가중값이 작으면 더 좋은 MLP모형을 유도하는 경향이 있다. 은닉노드의 개수가 각각 4, 6인 경우에는 서로 다른 작은 값 c의 결과간의 차이는 매우 작다. 이 결과는 그림 2의 (b)와 (c)에서 알 수 있다. 그러나 (d)는 모형의 모수가 과소 적용되어 어떠한 초기 가중값에도 영향을 받지 않음을 알 수 있다.

그림 1의 (c)와 (d)는 제어상자를 통한 초기값의 설정을 통한 학습 결과를 나타내고 있다. 여기서 우리는 경험적으로 유의한 차이는 아니지만 과다 적합을 피하고 있음을 확인할 수 있다.

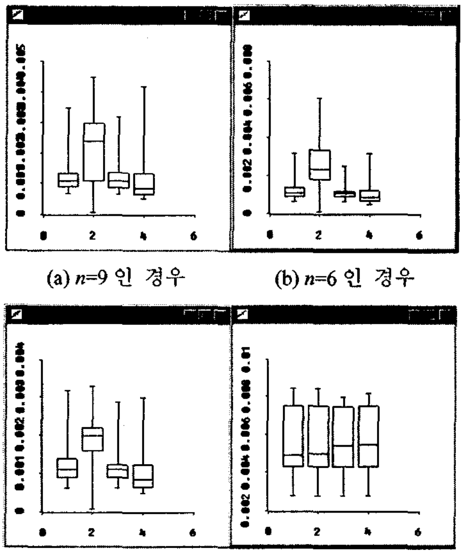
e = {0, 1e-6}과 같이 학습 회수가 아주 충분한 경우에는 예측의 정도가 c≤5인 서로 다른 작은 값 c에 민감하게 반응하지 않는 것을 알 수 있다. 이러한 결과로 좋은 추정 효율은 초기화 범위값 c를 충분히 작게 함으



(a) c=10 인 경우 (b) c=0.1 인 경우

(c) c=10 에서 제어상자 (d) c=0.1 에서 제어상자

그림 1 초기 가중값의 영향: c와 과다적합 관계



(a) n=9 인 경우 (b) n=6 인 경우

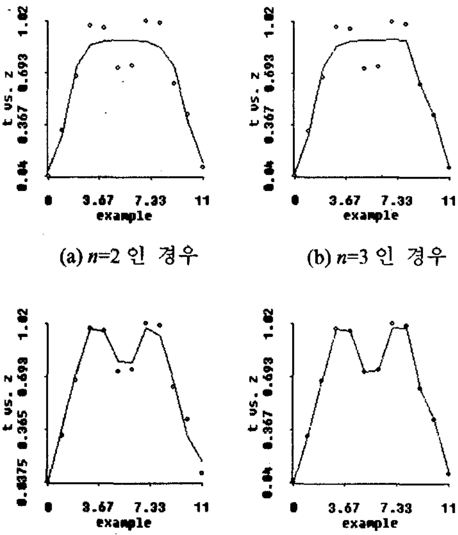
(c) n=4 인 경우 (d) n=3 인 경우

그림 2 은닉노드의 개수에 따른 초기 가중값의 영향

로써 제어할 수 있다.

3.2 은닉노드 개수의 효과

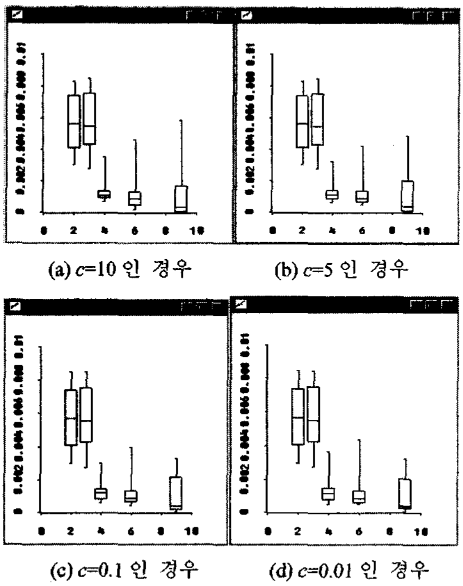
은닉노드 개수의 선택 효과를 논의하기 위하여 먼저 c=10, e=0 그리고 은닉노드의 개수변화는 n={2,3,4,6,9}을 설정한다. 그림 3은 그 결과를 나타내고 있다. 여기서 은닉노드의 개수가 n=2,3과 같이 작은 경우는 과소적합을 보이고 충분한 학습이 이루어지지 않았음을 알 수 있다. 왜냐하면 앞에서 설명한 바와 같이 주어진 학습 표본자료가 $n \leq 3$인 은닉노드로는 충분히 적합시킬 수 없기 때문이다. 그러나 n=9와 같이 너무 많은 경우에는 과다적합을 보인다. 그림 4의 상자그림은 은닉노드가 4개인 경우에서 상대적으로 최적의 결과를 보인다. 따라서 주어진 학습자료에서 은닉노드의 개수는 학습 결과에 민감한 반응을 보인다.



(a) n=2 인 경우 (b) n=3 인 경우

(c) n=4 인 경우 (d) n=9 인 경우

그림 3 은닉노드의 영향: n과 과다적합 관계



(a) c=10 인 경우 (b) c=5 인 경우

(c) c=0.1 인 경우 (d) c=0.01 인 경우

그림 4 초기 가중값에 따른 은닉노드 개수의 영향

3.3 허용오차의 효과

먼저 n=9, c=10 그리고 학습중지 시점을 위한 허용오

차의 변화는 e = {0, 1e-6, 0.001, 0.01}이라 하자. n과 c는 빠른 수렴결과를 얻기 위하여 큰 값으로 설정한다. 이 경우에 $e=0$은 다소 과다적합을 보이고 $e=0.01$는 과소적합을 보인다. 이는 허용오차가 모형 복잡도를 제어하는 구조인자로써 영향을 주고 있음을 보여준다. 또한 $e=0$에 대한 예측 효용은 e의 다른 값에 대한 것들보다 유의적으로 더 나쁘지 않음을 보여준다. 그러므로 이 인자는 주어진 학습자료에 대한 초기 가중값과 은닉노드의 개수와 같은 모형 복잡도 제어에 유의하지는 않음을 알 수 있다.

3.4 구조인자들의 상대효과

그림 2는 은닉노드의 개수 n={3,4,6,9}의 서로 다른값에 대한 초기 가중값의 상자그림 결과이다. 여기서 각 창의 세 번째와 네 번째 상자그림은 서로 다른 c에 대하여 제어상자를 통한 초기 가중값을 선택한 후 학습한 결과이다. 은닉노드의 개수가 n=4,6인 경우에, 서로 다른 c에 대하여 거의 유사한 예측효용을 갖는 모형을 추정할 수 있다.

n=3인 경우에는 과소 적합으로 인한 충분한 학습이 이루어지지 않고 있음을 알 수 있다. 다시 말하면 예측효용은 최적의 은닉노드 개수를 선택하였을 경우에만 c 값의 영향을 줄일 수 있다. 즉 모형의 복잡도 제어는 c 값 보다는 은닉노드의 개수에 따라 더 영향을 받을 수 있음을 의미한다.

그림 4는 c={10, 1, 0.1, 0.01}에 대하여 은닉노드 개수의 영향을 보여주는 상자그림이다. 여기서 c가 작은 경우에 서로 다른 n은 유사한 예측정도를 나타냄을 알 수 있다. 그러므로 예측효용 측면에서 살펴볼 때 적절하게 작은 c값을 선택하면 n의 영향을 줄일 수 있게 된다. 즉 복잡도 제어는 n의 값보다는 초기 가중값이 더 중요한 인자임을 보여준다.

그림 2와 그림 4에서는 초기 가중값과 은닉노드 개수의 결합 영향을 나타내고 있다. 주어진 네트워크에서 한 개의 구조인자를 잘 선택하면 다른 구조인자에 관계없이 좋은 모형을 도출할 수 있음을 보았다. 이는 응용문제에서 복잡도 제어가 어떤 이상적인 고정값으로 한가지를 정한 후 다른 두 가지 구조인자중의 하나에 대하여 실행될 수 있다는 것을 의미한다.

예를 들어 본 논문에서 고려한 학습표본의 분포특성을 고려하여 n은 3보다 큰 특정값인 4이상의 값으로 설정될 수 있고 c의 최적값은 교차타당성 검토에 의하여 선택될 수 있다. 또는 다른 경우로서 c를 10으로 설정하고 최적의 n값을 선택할 수 있다. 이는 MLP네트워크의 모형 복잡도를 제어하기 위한 c를 이용할 때 과다한 적합의 n값을 이용할 수 있게 된다. 즉 다음과 같이 설명할 수 있다. $n \leq 3$인 아주 작은 값으로 인하여 c에 관계없이 과소 적합된 MLP모형을 초래하는 경우가 있으며 이러한 값들은 실제 응용문제에서 일반적으로 알려지지 않고 있다. 그러나 충분히 큰 n값에 의한 과다적합은 최적화된 c값에 의하여 피할 수 있다.

4. 결 론

본 논문에서는 상자그림과 같은 잘 정의된 통계적 방법을 이용하여 구조인자들의 영향을 일반화시킬 수 있는 MLP네트워크들을 비교하였다. 적용한 통계 그래픽스는 구조인자들의 영향에 대하여 직관적인 평가를 제시할 수 있었다. 즉 허용오차는 다른 두 가지 인자인 초기 가중값과 은닉노드의 개수만큼 영향을 주지 않음을 확인하였고 이는 MLP네트워크의 일반화를 제어하는 더 좋은 척도가 됨을 보았다. 제어상자를 통한 초기 가중값의 적용 결과는 다소 과다 적합을 피할 수 있는 대안이 될 수도 있음을 확인하였지만 그 차이가 유의하지는 않았다. 우리의 경험적인 분석으로 요인들의 상호 작용이 MLP네트워크의 복잡도 제어에 상대적으로 적은 영향만이 있음을 보았다. 이러한 논의는 실제적인 모형 제어전략의 한 방법으로 제시될 수 있겠다. 학습표본에 급격한 변화 즉 특이값이 존재하는 자료에 대한 일반적인 학습방법은 향후 연구과제로 남긴다.

참 고 문 헌

[1] Cherkassky, V. and Mulier, F., "Learning from data-Concepts, Theory and Methods," Wiley, New York, 1998.
[2] Vapnik, V., "The Nature of Statistical Learning Theory," Wiley, New York, 1995.
[3] Zhong, S. and Cherkassky, V., "Factors Controlling Generalization Ability od MLP Networks," Proceedings of International Joint Conference On Neural Networks, 1999.
[4] 윤여창, "제어상자를 이용한 단순 신경망의 개선된 학습과정", 정보과학회논문지 : 소프트웨어 및 응용, 제28권 제4호, pp.338-345, 2001.
[5] Easton, G.S., "A Simple Dynamic Graphical Diagnostic Method for Almost Any Model," Journal of the American Statistical Association, 89, pp.201-207, 1994.
[6] Kim, Y.K. and Ra, J.B., "Weight value initialization for improving training speed in the backpropagation network," Proceedings of International Joint Conference On Neural Networks, Vol.3, pp.2396-2401, 1991.
[7] Espinosa, C.H. and Redondo, M.F., "Multilayer feedforward weight initialization," Proceedings of International Joint Conference On Neural Networks, pp.166-170, 2001.
[8] Thimm, G. and Fiesler, E., "High-order and multi-

layer perceptron initialization," IEEE Transactions on Neural Networks, Vol.8, pp.349-359, 1997.
[9] Smith, M., "Neural Networks for Statistical Modeling," Van Nostrand Reinhold, New York, 1993.
[10] Hagan, M.T., Demuth, H.B. and Beale, M., "Neural Network Design," PWS, Boston, 1995.
[11] Atiya, A. and Ji, C., "How Initial Conditions Affect Generalization Performance in Large Networks," IEEE Transaction on Neural Networks, Vol.8, No.2, March, pp.448-451, 1997.
[12] Cherkassky, V. and Shepherd, R., "Regularization Effect of Weight Initialization in Back Propagation Networks," Proceedings of International Joint Conference On Neural Networks, pp.2258-2261, 1998.

윤 여 창
1986년 성균관대학교 통계학과(학사) 1989년 성균관대학교 대학원 통계학과(석사). 1996년 성균관대학교 대학원 통계학과(박사). 1995년~현재 우석대학교 전산통계학과 부교수. 관심분야는 신경망, 인공지능, 시계열자료분석