

## 신경망이론에 의한 비중심카이제곱분포의 확률계산

남궁 평<sup>1)</sup>, 구 선 회<sup>2)</sup>

### 요 약

비중심  $\chi^2$ 분포의 누적분포함수의 계산은  $\chi^2$ 검정에서 요구되고 있어 새로운 접근방법으로 신경망 이론을 적용하기 위하여 입력층의 입력노드가 세개, 출력층의 출력노드가 한개 그리고 한개의 은닉층으로 구성된 다층퍼셉트론 네트워크부터 역전파 알고리즘을 개발하여 비중심  $\chi^2$ 분포의 확률계산을 시도하였다. 정확성과 계산속도를 고려하여 기존의 방법과 비교한 결과 효율적임을 알 수 있다.

### 1. 서 론

비중심  $\chi^2$ 분포는 1928년 R. A. Fisher에 의해 다중상관계수 R의 분포를 계산하는 과정에서 유도되었으며 Miller (1958)와 Park (1961)등에 의하면 비중심  $\chi^2$ 분포는 일반화된 Rayleigh분포로서 Rayleigh-Rice 또는 Rice분포라고 불리우고 있다. 이와 같은 비중심  $\chi^2$ 분포의 확률계산을 통하여 범주형자료의 분석에 흔히 사용되는 적합도검정 및 독립성, 동질성검정 등과 같은  $\chi^2$ 검정의 검정력을 구할 수 있다.

비중심  $\chi^2$ 분포에서 분포함수의 값을 계산하는 데에는 일반적으로 많은 시간이 소요될 뿐아니라 정확성이 문제가 되고 있다. 따라서 인공지능(artificial intelligence)분야에서 많은 연구가 이루어진 인공 신경망(artificial neural network : 이하 신경망)이론을 이용하여 비중심  $\chi^2$ 분포의 확률계산을 위한 모형을 설정하고 이 결과를 기존의 계산방법과 비교하고자 한다.

신경망에 관한 연구는 인간의 두뇌와 신경세포 모형에 대한 연구에서 시작되었으며 신경 시스템에서 가장 기본적인 단위는 뉴런(neuron)이며 뇌는 다수의 뉴런이 서로 연결된 신경망으로 구성되어 있다.

### 2. 신경망 이론

#### 2.1 신경망의 구조

신경망은 네트워크의 위상(topology), 유니트(unit), 학습규칙(learning rule)의 세가지 요소로 구

- 1) (110-745) 서울특별시 종로구 명륜동 3가, 성균관대학교 통계학과 교수.
- 2) (560-759) 전라북도 전주시 효자동 3가, 전주대학교 교양전산학 객원교수.

성된다. 네트워크의 위상은 각 유닛들이 연결된 형태로서 기본적으로 입력층(input layer)과 출력층(output layer), 그리고 이 두개의 층을 연결하는 연결강도(connection strength)로 구성되어 있다. 그러나 이러한 네트워크의 위상은 항상 일정한 것이 아니며 경우에 따라 그 위상을 다르게 구성할 수가 있다.

예를 들어, 다층으로 구성된 네트워크인 다층퍼셉트론인 경우에는 가장 하위에 위치한 입력층과 가장 상위에 있는 출력층, 그리고 그 사이에 여러 개의 층으로 이루어진 은닉층(hidden layer)으로 구성되는데, 이 네트워크에서 은닉층의 수 그리고 각 층을 이루고 있는 유닛의 수는 항상 고정되어 있는 것이 아니라 학습 유형에 따라 적절히 조정하여 구성한다.

유닛은 각 층을 이루고 있는 노드이며, 변수의 수 만큼 노드 수가 존재한다. 예를 들어  $p$ 차원의 변수가 입력되어 학습을 하는 경우 이 네트워크는  $p$ 개의 입력노드가 필요하다.

네트워크의 입력층과 출력층, 또는 은닉층을 연결하는 연결강도의 값은 반복이 이루어지면서 계속 조정을 하며 이러한 과정을 학습이라고 하고, 이 학습은 일정한 공식에 의해서 이루어지는 데 이것을 학습규칙이라고 한다.

다층퍼셉트론의 네트워크에서 입력층의 각 노드는 독립변수의 값을 입력 받으며 은닉층의 노드는 학습과정을 조정하는 역할을 한다. 또한 출력층의 노드들은 종속변수의 값을 계산하고 출력하게 된다. 네트워크는 입력층, 은닉층, 출력층 방향으로 연결되어 있으며 각 층내의 연결과 출력층에서 입력층으로의 직접적인 연결은 존재하지 않는 전방향(feedforward)이다. 다층퍼셉트론은 하나의 조정층(single-adjustable)으로 구성되는 단층퍼셉트론과 유사한 구조를 가지고 있으나 은닉층과 각 유닛의 입출력 특성을 비선형으로 함으로써 네트워크의 능력을 향상시켜 단층퍼셉트론의 여러가지 단점들을 극복하였다.

## 2.2 역전파 알고리즘

역전파 학습 알고리즘은 평균제곱오차(mean square error) 알고리즘의 비선형적인 확장으로 미분의 반복규칙(chain-rule)을 여러번 반복적으로 적용하여 오차의 최저점을 결정하는 최적의 연결강도의 집합을 구하는 데 있다. 그러나 특정영역에 대하여 최적의 연결강도를 결정하는 것이 매우 어려움으로 예제를 가지고 적절한 연결강도를 찾는다.

먼저 입력노드의 값을 입력받아 은닉층의 연결강도를 곱하고 편의연결강도를 더한 후 시그모이드 함수(sigmoid function)를 통하여 출력층에 전달한다. 출력층은 은닉층의 각 노드의 값을 받아 들여 다시 출력층의 연결강도를 곱하고 편의연결강도를 더한 후 시그모이드 함수를 통하여 출력한다. 최종에 제시된 출력값을 가지고 사전에 주어진 목표값과 비교하는 데 평균제곱오차를 이용한다. 평균제곱오차가 가장 작은 경우에 목표값에 가장 근사함으로 최종 출력값이 목표값에 가장 근사하게 되는 각 층 각 노드의 연결강도와 편의연결강도를 조절하여 학습을 한다.

평균제곱오차는 다음과 같다.

$$E = \frac{1}{2} \frac{\sum_{n=1}^N \sum_{k=1}^K (z_{kn} - t_{kn})^2}{NK}, \quad (1)$$

여기서  $N$  : 예제의 수

$K$  : 출력결과 수

$t_{kn}$  :  $n$ 번째 예제에 대한  $k$ 번째 목표결과

$z_{kn}$  :  $n$ 번째 예제에 대한  $k$ 번째 실제결과 이다.

식(1)은 네트워크를 통해 제시된 최종 출력값  $z$ 의 2차 함수이다. 이로부터 오차의 최소값은 식(1)의 1차 미분한 값이 0일때 (즉, 접선의 기울기가 0일때) 존재한다. 여기서 출력값  $z$ 는 연결강도들의 함수로 연결강도 평면상에서 오차표면(error surface)의 최저점을 결정하게 된다. 따라서 오차표면의 기울기를 계산하여 오차표면이 가장 급하강하는 방향으로 연결강도를 변화한다. 오차표면의 기울기가 음이면 연결강도를 조금씩 증가시키며 기울기가 양이면 연결강도를 조금씩 감소하면서  $E$ 의 최저점을 구하게 된다. 또한 기울기의 절대값이 크면 연결강도의 변화도 크게 하고 기울기의 절대값이 작으면 연결강도의 변화도 작게 한다. 여기서 오차표면의 기울기는 연결강도  $\omega$ 에 대하여 오차  $E$ 의 편미분  $\frac{\partial E}{\partial \omega}$ 을 이용한다.

하나의 예제를 가지고 출력층의 노드에서 연결강도에 대한 오차의 1차 편미분은 반복규칙(chain rule)에 의해 다음과 같다.

$$\frac{\partial E}{\partial b} = \frac{\partial E}{\partial z} \frac{\partial z}{\partial v} \frac{\partial v}{\partial b} . \tag{2}$$

그리고 하나의 출력 노드에서의 오차는 식(3)과 같다.

$$E = \frac{1}{2} (z - t)^2 , \tag{3}$$

여기서  $z = \frac{1}{(1 + e^{-v})}$  와  $v = b_0 + \sum_{j=1}^J b_j y_j$  로 부터 다음 식이 유도된다.

$$\frac{\partial E}{\partial z} = z - t \tag{4}$$

$$\frac{\partial z}{\partial v} = z(1 - z) \tag{5}$$

$$\frac{\partial v}{\partial b} = \begin{cases} \frac{\partial v}{\partial b_0} = 1 \\ \frac{\partial v}{\partial b_j} = y_j . \end{cases} \tag{6}$$

또한 연결강도에 대하여 일정한 값을 갖는  $\frac{\partial E}{\partial z} \frac{\partial z}{\partial v}$  을  $p$  라 하면

$$p = (z - t)z(1 - z) \tag{7}$$

이다.

따라서 출력층 노드에서 연결강도에 대한 오차의 편미분은 다음과 같이 정의한다.

$$\frac{\partial E}{\partial b_j} = \begin{cases} p, & \text{편의연결강도} \\ py_j, & \text{은닉연결강도} . \end{cases} \tag{8}$$

마찬가지로 은닉층의 노드에서 연결강도에 대한 1차 편미분은 식(9)와 같다.

$$\frac{\partial E}{\partial a} = \frac{\partial E}{\partial y} \frac{\partial y}{\partial u} \frac{\partial u}{\partial a} . \tag{9}$$

그런데 은닉층의 노드의 오차는 출력층의 노드의 오차와 연결되어 있으므로  $\frac{\partial v_k}{\partial y} = b_k$  와

$u = a_0 + \sum_{i=1}^I a_i x_i$  로 부터 다음과 같이 나타낼 수 있다.

$$\begin{aligned} \frac{\partial E}{\partial y} &= \sum_{k=1}^K \frac{\partial E}{\partial z_k} \frac{\partial z_k}{\partial v_k} \frac{\partial v_k}{\partial y} \\ &= \sum_{k=1}^K p_k \frac{\partial v_k}{\partial y} \\ &= \sum_{k=1}^K p_k b_k \end{aligned} \quad (10)$$

$$\frac{\partial y}{\partial u} = y(1-y) \quad (11)$$

$$\frac{\partial u}{\partial a} = \begin{cases} \frac{\partial u}{\partial a_0} = 1 \\ \frac{\partial u}{\partial a_i} = x_i \end{cases} \quad (12)$$

여기서도 마찬가지로  $\frac{\partial E}{\partial y} \frac{\partial y}{\partial u}$  을  $q$  라 하면

$$q = \left( \sum_{k=1}^K p_k b_k \right) y(1-y) \quad (13)$$

이다.

따라서 은닉층 노드에서 연결강도에 대한 오차의 편미분은 다음과 같이 정의한다.

$$\frac{\partial E}{\partial a_i} = \begin{cases} q, & \text{편의 연결강도} \\ qx_i, & \text{은닉 연결강도} \end{cases} \quad (14)$$

학습은 출력값과 목표값의 차이가 최소가 되도록 연결강도를 계속하여 변화하는 과정인데 연결강도의 변화는 다음과 같다.

$$\omega_m = \omega_{m-1} + c_m, \quad (15)$$

여기서  $\omega_m$  : 학습횟수(epoch)  $m$ 에서의 연결강도  $\omega$ 의 값

$c_m$  : 학습횟수  $m$ 에서의 연결강도  $\omega$ 의 변화량이다.

변화량  $c_m$ 은 연결강도의 수정이 현재의 수정값 뿐만 아니라 이전 단계에서의 수정값도 고려하여 학습하는 모멘텀(momentum)을 이용한 적합한 학습규칙방법(adaptive learning rules method)을 사용하였다. 이것은 일명 delta-bar-delta 라고도 한다.

$$c_m = \mu c_{m-1} - (1-\mu) e_m d_m, \quad (16)$$

여기서  $0 \leq \mu \leq 1$  로서  $\mu = 0$  이면  $c_m$ 은 과거의 연결강도의 변화량에는 영향을 받지 않으며  $\mu$ 가 클 수록 과거의 연결강도의 변화량에 크게 영향을 받는다.

또한  $d_m = \sum_{n=1}^N \left( \frac{\partial E}{\partial \omega_m} \right)_n$  은 학습횟수  $m$ 에서의 모든 예제에 대한 연결강도에 대한 오차의 편미분의 합이다.  $d_m$ 의 부호가 양의 방향이면 연결강도가 내려감에 따라 오차가 감소하지만  $d_m$ 의 부호가 음의 방향이면 연결강도가 올라감에 따라 오차가 감소하게 된다.  $e_m$ 은 학습횟수  $m$ 에서의  $d_m$ 의 방향과 과거오차의 편미분의 방향  $f_m$ 에 영향을 받는다. 즉,

$$e_m = \begin{cases} e_{m-1} + x, & d_m f_m > 0 \\ e_{m-1} \times \phi, & d_m f_m \leq 0 \end{cases} \quad (17)$$

이다. 오차가 현재 연결강도에서 감소하는 방향이 최근 감소한 방향과 같을 때는  $e$ 값을  $x$ 만큼 더해주고 현재 감소하는 방향이 최근 감소한 방향과 반대일 때는  $e$ 값을  $\phi$ 만큼 곱해준다. 여기서  $x$ 와  $\phi$ 의 값은 0과 1사이의 값을 선택하며 변수  $x, \phi, \theta, \mu$ 는 사용자가 적절히 조절하여 최적해를 찾는다.

### 3. 비중심카이제곱 확률계산식

비중심  $\chi^2$ 분포의 확률계산을 하기 위하여 입력층의 입력노드(독립변수)가 3개 (자유도, 비중심모수, 백분위수) 이며 출력층의 출력노드(종속변수)가 1개 (확률) 그리고 1개의 은닉층으로 구성된 다층퍼셉트론 네트워크로 부터 역전파 학습 알고리즘을 적용한 결과를 제시하였으며 비중심  $\chi^2$ 분포는 다음과 같이 정의한다.

#### 정의

확률변수  $U_1, U_2, \dots, U_v$  가 서로 독립이고 각 각 표준정규분포를 따르며  $\delta_1, \delta_2, \dots, \delta_v$  가 상수일 때  $\sum_{j=1}^v (U_j + \delta_j)^2$  의 분포를 자유도가  $v$ 이고 비중심모수가  $\lambda = \sum_{j=1}^v \delta_j^2$  인 비중심  $\chi^2$ 분포(non-central  $\chi^2$  distribution)라 하며  $\chi_v^2(\lambda)$ 로 나타낸다.

1928년 Fisher가 제시한 비중심  $\chi^2$ 분포의 누적함수는 다음과 같다.

$$F(x; v, \lambda) = \Pr[\chi_v^2(\lambda) \leq x] \\ = \int_0^x \frac{1}{2} \left( \frac{y}{\lambda} \right)^{\frac{1}{4}(v-2)} I_{\frac{1}{2}(v-2)}(\sqrt{\lambda} \sqrt{y}) \exp \left[ -\frac{1}{2}(\lambda + y) \right] dy, \quad (x > 0)$$

여기서  $I_k(x) = \left( \frac{1}{2} x \right)^k \sum_{j=0}^{\infty} \frac{\left( \frac{1}{4} x^2 \right)^j}{j! \Gamma(k+j+1)}$  는 차수가  $k$ 인 제1종의 수정 Bessel함수이다.

#### 1 모형

비중심  $\chi^2$ 분포의 누적함수 (Fisher,1928)를 이용하여 생성시킨 자료를 다음과 같은 모형으로 분류하였다.

<표 1> 모형 1

자료	확률값(P)	자유도(v)	비중심모수 ( $\lambda$ )
학습자료	P=0.95 ~ 0.999	v=1	$\lambda = 1$ $\lambda = 3$
		v=2	$\lambda = 2$ $\lambda = 6$
		v=5	$\lambda = 5$
검정자료	P=0.95 ~ 0.999	v=1	$\lambda = 2$
		v=2	$\lambda = 4$
		v=5	$\lambda = 10$

<표 2> 모형 2

자료	확률값(P)	자유도(v)	비중심모수 ( $\lambda$ )
학습자료	P=0.95 ~ 0.999	v=10	$\lambda = 10$ $\lambda = 30$
		v=15	$\lambda = 15$ $\lambda = 45$
		v=20	$\lambda = 20$
검정자료	P=0.95 ~ 0.999	v=10	$\lambda = 20$
		v=15	$\lambda = 30$
		v=20	$\lambda = 40$

### 3.2 확률계산

#### 3.2.1 모형 1의 확률계산

네트워크에서 학습자료를 가지고 학습한 결과 다음과 같은 모수에서 목표결과인 Fisher의 참값과 신경망을 이용한 근사값과의 차이가 작게 됨을 알 수 있다. 8개의 은닉노드 그리고 모수가  $x = 0.009$ ,  $\phi = 0.3$ ,  $\mu = 0.2$ ,  $\theta = 0.3$  에서 최적의 값을 갖는다. 그리고 이 모수로 부터 최종 학습

결과의 연결강도를 가지고 검정자료로부터 비중심  $\chi^2$  분포의 확률계산을 하며 최종 학습결과의 연결강도는 다음과 같다.

<표 3> 입력노드에서 은닉노드로의 연결강도 (모형 1)

	입 력 노 드				
	j	$a_{0j}$	$a_{1j}$	$a_{2j}$	$a_{3j}$
은닉노드	1	0.458563	0.206361	0.081557	-0.093273
	2	-1.873473	-0.170342	-0.259860	0.120981
	3	0.309210	0.104493	0.102144	-0.087077
	4	-0.450607	-0.064258	-0.087153	0.076711
	5	0.195063	0.309282	0.259293	-0.297252
	6	-0.734100	-0.398016	-0.810924	0.091696
	7	0.351816	0.011089	0.124230	-0.083472
	8	-0.198235	-0.172584	0.389876	0.217612

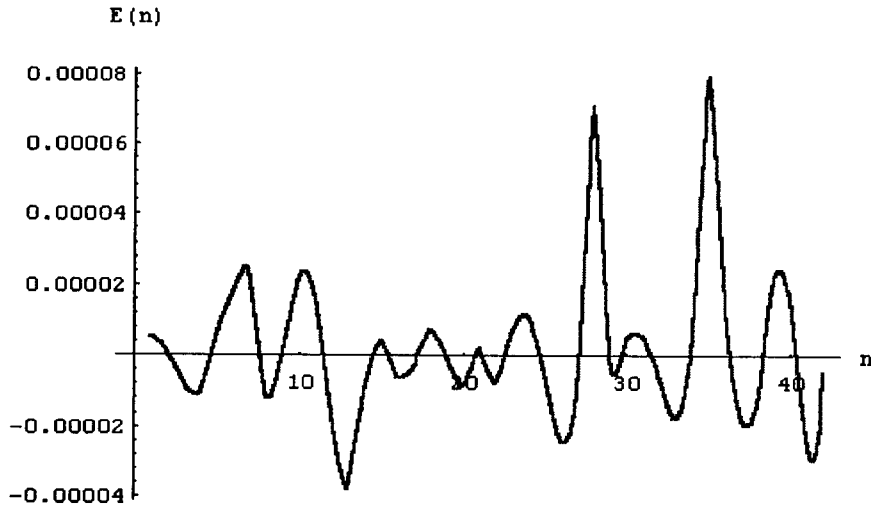
여기서  $a_{0j}$  는 은닉노드  $j$ 의 편의연결강도,  $a_{ij}$  는 입력노드  $i$ 로 부터 은닉노드  $j$ 로의 연결강도이다.

<표 4> 은닉노드에서 출력노드로의 연결강도(모형1)

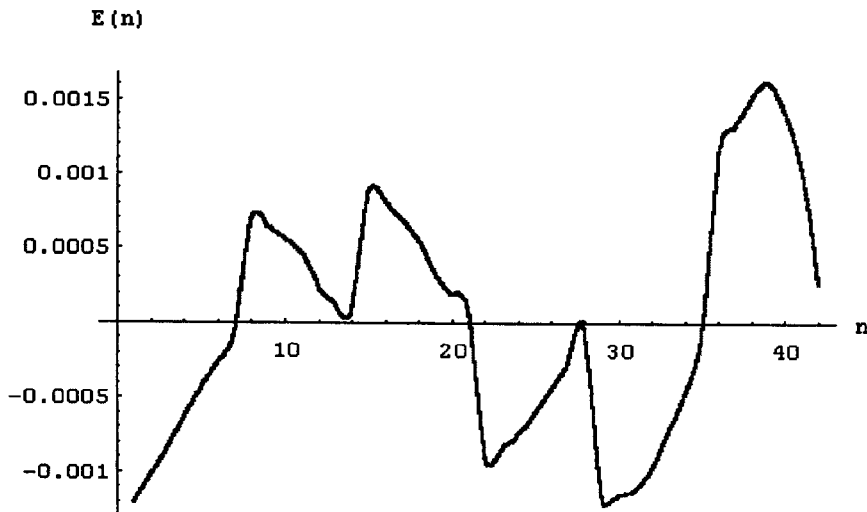
	출 력 노 드	
	j	$b_j$
은닉노드	0	1.247051
	1	-1.667231
	2	5.807258
	3	-1.428062
	4	2.429413
	5	-1.031085
	6	4.268556
	7	-1.495991
	8	1.206108

여기서  $b_0$  는 출력노드의 편의연결강도,  $b_j$  는 은닉노드  $j$ 로 부터 출력노드의 연결강도이다.

다음은 모형 1의 Fisher의 참값과 신경망을 이용한 근사값과 비교하여 학습결과와 검정결과의 오차들을 제시하였다. 학습자료의 근사값은 대략 소수점 네째자리까지 Fisher의 참값과 동일하며 Fisher의 참값과의 최대오차는  $79 \times 10^{-6}$  이다. 또한 검정자료의 근사값은 대략 소수점 세째자리까지 Fisher의 참값과 동일하며 Fisher의 참값과의 최대오차는  $150.4 \times 10^{-5}$  임을 알 수 있다.



<그림 1> 학습결과의 오차 1 : 오차  $E(n)$ 은 Fisher의 참값과 신경망 방법을 이용한 근사값과의 차이이며,  $E(35) = 79 \times 10^{-6}$ 의 최대오차를 나타내고 있다.



<그림 2> 검정결과의 오차 1 : 오차  $E(n)$ 은 Fisher의 참값과 신경망 방법을 이용한 근사값과의 차이이며,  $E(38) = 150.4 \times 10^{-5}$ 의 최대오차를 나타내고 있다.



3.2.2 모형 2의 확률계산

네트워크는 7개의 은닉노드 그리고 모수가  $x = 0.009$ ,  $\phi = 0.3$ ,  $\mu = 0.4$ ,  $\theta = 0.3$  에서 최적의 값을 가짐을 알 수 있다. 이 모수로 부터 최종 학습결과의 연결강도를 가지고 검정자료를 이용하여 비중심  $\chi^2$ 분포의 확률계산을 한 결과는 다음과 같다.

<표 5> 입력노드에서 은닉노드로의 연결강도 (모형 2)

	입 력 노 드				
	j	$a_{0j}$	$a_{1j}$	$a_{2j}$	$a_{3j}$
은닉노드	1	0.397121	-0.029338	0.043027	-0.028580
	2	-0.334947	-0.059440	-0.039361	0.041638
	3	0.195075	0.261684	0.008101	-0.178951
	4	-0.548756	-0.040949	-0.041362	0.040328
	5	0.294556	0.111210	0.057747	-0.060380
	6	-1.336726	-0.105710	-0.144325	0.091683
	7	0.557329	0.293616	0.359229	-0.166078

여기서  $a_{0j}$  는 은닉노드  $j$ 의 편이연결강도,  $a_{ij}$  는 입력노드  $i$ 로 부터 은닉노드  $j$ 로의 연결강도이다.

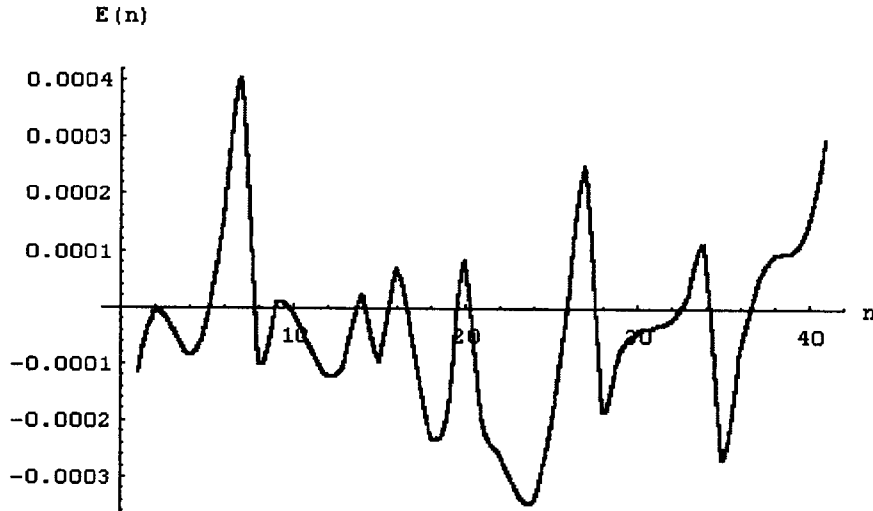
<표 6> 은닉노드에서 출력노드로의 연결강도 (모형 2)

	출 력 노 드	
	j	$b_j$
은닉노드	0	1.539630
	1	-1.554436
	2	2.340432
	3	-1.486466
	4	3.036736
	5	-1.290002
	6	3.412711
	7	-1.458087

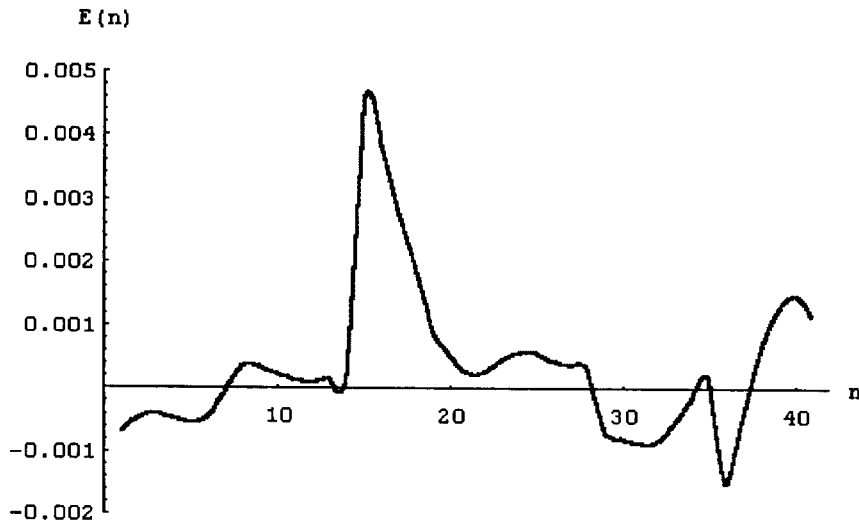
여기서  $b_0$  는 출력노드의 편이연결강도,  $b_j$  는 은닉노드  $j$ 로 부터 출력노드의 연결강도이다.

다음은 모형 2의 Fisher의 참값과 신경망을 이용한 근사값과 비교하여 학습결과와 검정결과의 오차들을 플롯을 이용하여 제시하였다. 학습자료의 근사값은 대략 소수점 네째자리까지 Fisher의 참값과 동일하며 Fisher의 참값과의 최대오차는  $40.3 \times 10^{-5}$  이다. 또한 검정자료의 근사값은 대

략 소수점 세째자리까지 Fisher의 참값과 동일하며 Fisher의 참값과의 최대오차는  $45.69 \times 10^{-4}$  임을 알 수 있다.



<그림 3> 학습결과의 오차 2 : 오차  $E(n)$ 은 Fisher의 참값과 신경망 방법을 이용한 근사값과의 차이이며,  $E(7) = 40.3 \times 10^{-5}$ 의 최대오차를 나타내고 있다.



<그림 4> 검증결과의 오차 2 : 오차  $E(n)$ 은 Fisher의 참값과 신경망 방법을 이용한 근사값과의 차이이며,  $E(15) = 45.69 \times 10^{-4}$ 의 최대오차를 나타내고 있다.

## 4. 결 론

새로운 접근방법으로 신경망 이론을 적용하여 비중심  $\chi^2$ 분포의 확률을 계산하였다. 이 신경망 이론은 다층의 구조를 갖는 복잡한 신경망 역전파 학습알고리즘으로 연결강도 공간에서 오차표면의 기울기 감소를 기본으로 하였다.

비중심  $\chi^2$ 분포의 확률계산을 Fisher의 참값과 신경망을 이용한 근사값과 비교하여 학습결과와 검정결과의 오차들을 플롯을 이용하여 제시하였다. 학습하고 검정한 결과를 Fisher의 참값과 비교한 결과는 다음과 같다.

첫째, 학습자료의 근사값은 대략 소수점 네째자리까지 Fisher의 참값과 정확하며 검정자료의 근사값은 소수점 세째자리까지 Fisher의 참값과 정확하게 나타남을 알 수 있다. 둘째, 학습하는데 소요되는 시간 즉, 신경망의 최적의 모수를 찾는 데 있어 많은 시간과 노력이 필요하지만 한번 학습이 끝난 후에는 최종의 연결강도를 가지고 검정자료로 부터 소요되는 시간이 없이 비중심  $\chi^2$ 분포의 확률계산을 할 수 있다. 셋째, 계산결과를 정확성과 계산속도를 고려하여 기존의 방법과 비교할 때 효율적임을 알 수 있다.

## 참 고 문 헌

- [1] 김대수 (1994). 「신경망 이론과 응용 (I),(II)」, 하이테크정보.
- [2] 이성환 (1994). 「패턴인식의 원리 (I),(II)」, 홍릉과학출판사.
- [3] Abdel-Samad, A. I. and Ashour, S. K. (1990). On the computation of non-central chi-square distribution function, *Communications in Statistics, Simulation*, 19(4), 1279-1291.
- [4] Bickel, P. J. and Doksum, K. A. (1977). *Mathematical Statistics : Basic ideas and selected topics*, San Francisco : Holden-Day.
- [5] John, H. and Anders, K. and Richard, G. P. (1991). *Introduction to the Theory Neural Computation*, Addison-Wesley Publishing Co. Inc.
- [6] Johnson, N. L. and Kotz, S. (1970). *Continuous Univariate Distributions*, Boston : Houghton Mifflin.
- [7] Pao, Y. H. (1989). *Adaptive Pattern Recognition and Neural Network*, Addison-Wesley Publishing Co. Inc.
- [8] Park, J. H. (1961). Moments of the generalized Rayleigh distribution, *Quarterly of Applied Mathematics*, Vol. 19, 45-49.
- [9] Posten, H. O. (1989). An effective Algorithm for the noncentral chi-squared distribution function, *The American Statistician*, Vol. 43, 261-263.
- [10] Rumelhart, D. E., Geoffrey, E. H. and Ronald, J. W. (1986). Learning representations by backpropagating errors, *Nature*, Vol. 323, 533-536.
- [11] Smith, M. (1993). *Neural Networks for Statistical Modeling*, New York.