

신경망이론에 의한 비중심t분포 확률계산

구 선 희[†]

요 약

비중심t분포의 누적함수는 두 정규모집단에서 모평균의 동일성 검정에서 검정력 계산 및 모평균에 대한 표준편차의 비에 대하여 신뢰구간을 계산할 때 요구된다. 본 논문에서는 비중심t분포의 누적함수 계산에 신경망 이론을 적용하였다. 신경망은 다층 퍼셉트론이며 학습과정은 역전파 학습알고리즘이다. Fisher가 제시한 확률값과 신경망이론에 의하여 계산한 결과를 비교하였다.

Computation of Noncentral T Probabilities using Neural Network Theory

Gu Son Hee[†]

ABSTRACT

The cumulative function of the noncentral t distribution is desired to calculate power in testing equality of means of two normal populations and confidence intervals for the ratio of population mean to standard deviation. In this paper, the evaluation of the cumulative function of the noncentral t distribution is applied to the neural network theory. The neural network consists of the multi-layer perceptron structure and learning process has the algorithm of the backpropagation. Numerical comparisons are made between the Fisher's values and the results obtained by neural network theory.

1. 서 론

비중심 t분포는 1931년 R. A. Fisher에 의하여 유도되었으며 두 정규모집단에서 모평균의 동일성 검정에서 검정력 계산 및 모평균에 대한 표준편차의 비(ratio)에 대하여 신뢰구간을 계산할 때 적용되고 있다[3, 12].

비중심 t분포에서 분포함수(distribution function)의 값을 계산하는데에는 일반적으로 많은 시간이 소요될 뿐아니라 정확성이 문제가 되고 있다.

이로 인하여 이 분포함수에 대한 연구가 Jennett와

Welch (1939), Johnson과 Welch (1940), Johnson과 Kotz (1970), Deutler (1984), Lenth (1988), Goedhart와 Jansen (1990) 등에 의하여 이루어졌다[5, 7, 12, 13].

본 논문에서는 비중심t분포의 확률계산을 위하여 인공지능(artificial intelligence)분야에서 많은 연구가 이루어진 인공 신경망(artificial neural network: 이하 신경망)이론을 적용하여 먼저 모형을 설정하고 이 계산결과를 Fisher의 누적함수값과 비교하고자 한다.

2. 신경망 이론

2.1 신경망의 구조

신경망은 네트워크의 위상(topology), 유닛(unit), 학습규칙(learning rule)의 세가지 요소로 구성된다

[†] 정 회 원: 전주대학교 전기전자컴퓨터공학부
논문접수: 1996년 5월 16일, 심사완료: 1996년 11월 5일

[1, 2, 6]. 네트워크의 위상은 각 유니트들이 연결된 형태로써 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[8, 9, 16].

유니트는 각 층을 이루고 있는 노드이며, 변수의 수 만큼 노드 수가 존재한다. 예를 들어 p차원의 변수가 입력되어 학습을 하는 경우 이 네트워크는 p개의 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[11].

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 예제(example)를 가지고 적절한 연결강도를 찾는다[14, 18].

먼저 입력노드의 값을 입력받아 은닉층의 연결강도를 곱하고 편(bias)연결강도를 더한 후 시그모이드 함수(sigmoid function)를 통하여 출력층에 전달한다. 출력층은 은닉층의 각 노드의 값을 받아들여 다시 출력층의 연결강도를 곱하고 편연결강도를 더한 후 시그모이드 함수를 통하여 출력한다. 최종에 제시된 출력값을 가지고 사전에 주어진 목표값과 비교하는 데 평균제곱오차를 이용한다. 평균제곱오차가 가장 작은 경우에 목표값에 가장 근사하므로 최종 출력값이 목표값에 가장 근사하게 되는 각 층 각 노드의 연결강도와 편연결강도를 조절하여 학습을 한다[1, 4, 10].

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

$$\omega_m = \omega_{m-1} + C_m \tag{1}$$

여기서 ω_m : 학습횟수(epoch) m 에서의 연결강도 ω 의 값
 C_m : 학습횟수 m 에서의 연결강도 ω 의 변화량이다.

변화량 C_m 은 연결강도의 수정이 현재의 수정값 뿐만 아니라 이전 단계에서의 수정값도 고려하여 학습하는 모멘텀(momentum)을 이용한 적응적인 학습규칙 방법(adaptive learning rates method)을 사용하였다. 이것은 일명 delta-bar-delta 라고도 한다.

$$C_m = \mu C_{m-1} - (1 - \mu) e_m d_m \tag{2}$$

여기서 $0 \leq \mu \leq 1$ 로서 $\mu = 1$ 이면 C_m 은 과거의 연결강도의 변화량에는 영향을 받지 않으며 μ 가 클수록 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

$$e_m = \begin{cases} e_{m-1} + \lambda, & d_m f_m > 0 \\ e_{m-1} \times \phi, & d_m f_m \leq 0 \end{cases} \tag{3}$$

이다. 오차가 현재 연결강도에서 감소하는 방향이 최근 감소한 방향과 같을 때는 e 값을 λ 만큼 더해주고 현재 감소하는 방향이 최근 감소한 방향과 반대일 때는 e 값을 ϕ 만큼 곱해준다. 여기서 λ 와 ϕ 의 값은 0과 1 사이의 값을 선택하며 변수 λ, ϕ, μ 는 사용자가 적절히 조절하여 최적해를 찾는다[14, 15, 17].

3. 비중심t분포의 확률계산식

본 논문에서는 비중심t분포의 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 알고리

준을 적용한 결과를 제시하였으며 비중심t분포는 다음과 같이 정의한다.

정의

U 와 X_v 가 독립확률변수로서 U 가 표준 정규분포를 따르고 X_v 가 자유도가 v 인 카이분포이며 δ 가 상수일 때 $\frac{(U + \delta)}{X_v v^{-1/2}}$ 의 분포를 자유도가 v 이고 비중심모수

δ 인 비중심t분포(noncentral t distribution)라 하며 $t'_v(\delta)$ 로 나타낸다[12].

1931년 Fisher가 제시한 비중심t분포의 누적함수는 다음과 같다[12].

$$F(t; v, \delta) = \Pr[t'_v(\delta) \leq t] \\ = \Pr[U + \delta \leq tX_v / \sqrt{v}]$$

$$= \frac{1}{2^{v/2-1} \Gamma(v/2)} \int_0^\infty x^{v-1} e^{-x^2/2} \times$$

$$\frac{1}{\sqrt{2\pi}} \int_0^{tx/\sqrt{v}} \exp[-\frac{1}{2}(u-\delta)^2] du dx$$

여기서 $\Gamma(v) = \int_0^\infty x^{v-1} e^{-x} dx$ 이다.

3.1 모 형

비중심t분포의 누적함수(Fisher,1931)을 이용하여 생성시킨 자료를 다음과 같은 모형으로 분류하였다.

〈표 1〉 모형 1
〈Table 1〉 Model 1

자료	확률(P)	자유도(v)	비중심모수 (δ)
학습 자료	P=0.90 ~ 0.99	v=1	$\delta = 1$
			$\delta = 3$
검정 자료	P=0.90 ~ 0.99	v=1	$\delta = 2$
			$\delta = 4$

〈표 2〉 모형 2
〈Table 2〉 Model 2

자료	확률(P)	자유도(v)	비중심모수 (δ)
학습 자료	P=0.95 ~ 0.99	v=2	$\delta = 1$
			$\delta = 3$
		v=5	$\delta = 3$
			$\delta = 6$

검정 자료	P=0.95 ~ 0.99	v=2	$\delta = 2$ $\delta = 4$
		v=5	$\delta = 5$

〈표 3〉 모형 3
〈Table 3〉 Model 3

자료	확률(P)	자유도(v)	비중심모수 (δ)
학습 자료	P=0.95 ~ 0.99	v=10	$\delta = 10$ $\delta = 30$
		v=15	$\delta = 15$ $\delta = 45$
검정 자료	P=0.95 ~ 0.99	v=10	$\delta = 20$ $\delta = 40$
		v=15	$\delta = 30$ $\delta = 60$

3.2 확률계산

일반적으로 분포함수에서 발생하는 문제로는 서론에서도 제시되었듯이 하나의 함수값을 계산하는데에도 많은 시간이 소요된다는 것이다.

본 논문에서도 비중심t분포의 누적함수값을 Fisher가 제시한 누적함수식으로부터 직접 계산하여 정확한 함수값을 구할 수 있지만 하나의 함수값을 계산하는데에도 많은 시간이 소요됨으로 효율적이라고 할 수 없다.

그러나 신경망이론을 적용하는 경우 비록 최적의 모수를 찾는 데 있어 많은 시간과 노력이 필요하지만 한 번 학습이 끝난 후에는 비중심t분포의 함수값을 소요되는 시간없이 구할 수 있게 된다.

신경망을 적용하기 위하여 먼저 Fisher가 제시한 비중심t분포의 누적함수식으로부터 직접 계산한 Fisher의 함수값을 학습자료로 사용하였다.

다음은 각 모형에 대한 학습자료이다.

〈표 4〉 모형 1의 학습자료
〈Table 4〉 Training data of Model 1

자유도	비중심 모수	백분위수	확률
1	1	8.6	0.90034
1	1	9.5	0.90964

1	1	10.7	0.91966
1	1	12.3	0.93002
1	1	14.4	0.94016
1	1	17	0.94927
1	1	21.1	0.95909
1	1	28	0.96915
1	1	43	0.97991
1	1	86	0.98995
1	3	23.9	0.90018
1	3	26.5	0.90992
1	3	30	0.92038
1	3	34.1	0.92992
1	3	39.8	0.93993
1	3	47	0.94911
1	3	59	0.95945
1	3	79	0.96971
1	3	120	0.98005
1	3	240	0.99003

〈표 6〉 모형 3의 학습자료
 〈Table 6〉 Training data of Model 3

자유도	비중심 모수	백분위수	확률
10	10	16.3	0.949677
10	10	16.8	0.958948
10	10	17.5	0.969073
10	10	18.5	0.979259
10	10	20.3	0.989703
15	15	21.8	0.948543
15	15	22.4	0.959500
15	15	23.1	0.969415
15	15	24.1	0.979547
15	15	25.8	0.989664
10	30	47.9	0.949814
10	30	49.4	0.959499
10	30	51.5	0.969927
10	30	54.4	0.979947
10	30	59.5	0.989967
15	45	64.7	0.949403
15	45	66.3	0.959449
15	45	68.4	0.969710
15	45	71.3	0.979772
15	45	76.3	0.989891

〈표 5〉 모형 2의 학습자료
 〈Table 5〉 Training data of Model 2

자유도	비중심 모수	백분위수	확률
2	1	6.0	0.95009
2	1	6.7	0.95944
2	1	7.8	0.96964
2	1	9.6	0.97968
2	1	13.7	0.98988
2	3	13.9	0.95004
2	3	15.6	0.96005
2	3	18	0.96978
2	3	22.2	0.97999
2	3	31.5	0.98999
5	3	7.1	0.95000
5	3	7.5	0.95969
5	3	8.1	0.97047
5	3	8.9	0.98007
5	3	10.4	0.98986
5	6	13	0.94989
5	6	13.7	0.95967
5	6	14.7	0.97006
5	6	16.1	0.97982
5	6	18.8	0.98992

자료를 앞에서 제시한 모형에 따라 분류하여 학습한 결과 다음과 같은 모수에서 학습오차가 최저가 됨을 알 수 있었다.

모형 1, 2, 3 모두 7개의 은닉노드와 모수가 $X=0.009$, $\phi=0.3$, $\mu=0.2$, $\theta=0.3$ 에서 최적의 값을 갖는다.

다음은 모형에 대한 학습과 검정결과의 오차를 제시하고 있다.

3.2.1 모형에 대한 학습과 검정결과의 오차

자유도와 비중심모수에 대하여 학습자료(60개)와 검정자료(55개)중에서 일부분을 비교하고 있으며 각각의 전체자료에 대한 오차는 그림을 이용하여 보여주고 있다.

Fisher의 함수값과 비교하여 신경망을 이용한 학습과 검정결과는 다음과 같다.

학습자료의 함수값은 대략 소수점 네째자리까지 Fisher의 함수값과 동일하며 Fisher의 함수값과의 최대오차는 1.59×10^{-4} 이다. 또한 검정자료의 함수값은 대략 소수점 세째자리까지 Fisher의 함수값과 동

표 7) Fisher의 함수값과 신경망이론에 의한 함수값 비교 (학습자료)

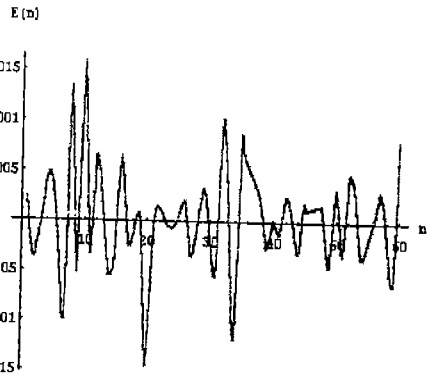
Table 7) Comparisons of the Fisher's values and the results obtained neural network theory(training data)

자유도 (v)	비중심 모수 (δ)	백분 위수 (x)	Fisher의 확률 (P_1)	신경망의 확률 (P_2)	오차 (E)
1	1	9.5	0.909640	0.909676	0.000036
1	1	43	0.979910	0.979964	0.000054
1	3	39.8	0.939930	0.939954	0.000024
1	3	240	0.990030	0.990178	0.000148
2	1	7.8	0.969640	0.969640	0.000000
2	3	13.9	0.950040	0.950020	0.000020
5	3	10.4	0.989860	0.989773	0.000087
5	6	14.7	0.970060	0.970030	0.000030
10	10	17.5	0.969073	0.969070	0.000003
15	15	21.8	0.948543	0.948563	0.000020
10	30	59.5	0.989967	0.989989	0.000022
15	45	76.3	0.989891	0.989810	0.000081

표 8) Fisher의 함수값과 신경망이론에 의한 함수값 비교 (검정자료)

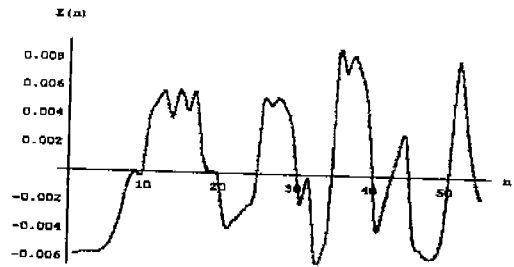
Table 8) Comparisons of the Fisher's values and the results obtained neural network theory(validation data)

자유도 (v)	비중심 모수 (δ)	백분 위수 (x)	Fisher의 확률 (P_1)	신경망의 확률 (P_2)	오차 (E)
1	2	16	0.900290	0.906155	0.005865
1	2	80	0.979970	0.979870	0.000100
1	4	35.4	0.910070	0.905247	0.004823
1	4	319	0.990000	0.990180	0.000180
2	2	11	0.960140	0.963535	0.003395
2	4	18.1	0.949720	0.943642	0.006078
5	5	11.6	0.959770	0.969934	0.010164
5	5	15.9	0.989790	0.992207	0.002417
10	20	36.4	0.979893	0.972726	0.007167
15	30	43.2	0.949178	0.952846	0.003668
10	40	72.4	0.979884	0.985285	0.005401
15	60	95.0	0.979803	0.979308	0.000495



(그림 1) 학습결과의 오차 : 오차 $E(n)$ 은 Fisher의 함수값과 신경망 방법을 이용한 함수값과의 차이이며, $E(10) = 1.59 \times 10^{-4}$ 의 최대오차를 나타내고 있다.

(Fig. 1) Errors for training data



(그림 2) 검정결과의 오차 : 오차 $E(n)$ 은 Fisher의 함수값과 신경망 방법을 이용한 함수값과의 차이이며, $E(36) = 8.637 \times 10^{-3}$ 의 최대오차를 나타내고 있다.

(Fig. 2) Errors for validation data

일하며 Fisher의 함수값과의 최대오차는 8.637×10^{-3} 임을 알 수 있다.

Fisher의 함수값과 신경망에 의한 함수값과의 차이인 검정자료 오차로부터 상당히 좋은 결과임을 알 수 있다. 또한, 최적의 모수로부터 구한 최종 학습결과의 연결강도를 가지고 우리가 구하고자 하는 비중심t분포의 누적함수값 여러개를 동시에 소요되는 시간 없이 구할 수 있다.

4. 결 론

본 논문은 새로운 접근방법으로 신경망 이론을 적용하여 비중심t분포의 확률 계산을 하였다. 이 신경망 이론은 다층의 구조를 갖는 복잡한 신경망 역전파 학습 알고리즘으로 연결강도 공간에서 오차표면의 기울기 감소를 기본으로 하였다.

비중심t분포의 확률 계산을 Fisher의 함수값과 신경망을 이용한 함수값과 비교하여 학습결과와 검정 결과 및 그에 따른 오차들을 그림을 이용하여 제시하였다. 학습하고 검정한 결과를 Fisher의 함수값과 비교한 결과는 다음과 같다.

첫째, 학습자료의 함수값은 대략 소수점 네째자리까지 Fisher의 함수값과 정확하며 검정자료의 함수값은 소수점 세째자리까지 Fisher의 함수값과 정확하게 나타남을 알 수 있다.

둘째, 학습하는데 소요되는 시간, 즉 신경망의 최적의 모수를 찾는 데 있어 많은 시간과 노력이 필요하지만 한 번 학습이 끝난 후에는 최종의 연결강도를 가지고 소요되는 시간 없이 비중심t분포의 확률계산을 수행한다.

셋째, 정확성과 계산속도를 고려하여 볼 때 Fisher의 누적함수식에 의한 방법보다 효율적임을 알 수 있다.

참 고 문 헌

[1] 김대수, '신경망 이론과 응용 (I), (II)', 하이테크 정보, 1994.
 [2] 이성환, '패턴인식의 원리 (I), (II)', 홍릉과학출판사, 1994.
 [3] P. J. Bickel and K. A. Doksum, 'Mathematical Statistics: Basic ideas and selected topics,' San Francisco: Holden-Day, 1977.
 [4] B. Cheng and D. M. Titterton, "Neural Networks: A review from a statistical perspective," Statistical Science, Vol. 9, pp. 2-54, 1994.
 [5] T. Deutler, "A series expansion for the cumulants of the X-distribution and a Cornish-fisher-expansion for the noncentrality parameter of the noncentral T-distribution," Communications in Statistics, Simulation, Vol. 13(4), pp. 507-513,

1984.
 [6] J. H. Friedman and W. Stuetzle, "Projection pursuit regression," Journal of the American Statistical Association, Vol. 76, pp. 817-823, 1981.
 [7] P. W. Goedhart and M. J. W. Jansen, "An integral useful in calculating non-central t and bivariate normal probabilities," Applied Statistics, Statistical Algorithms, pp. 496-497, 1990.
 [8] J. N. Hwang and H. Li. M. Maechler and D. Martin and J. Schimert, "A comparison of projection pursuit and neural network regression modeling," Neural Information Processing System, pp. 1159-1166, Nov. 1991.
 [9] J. N. Hwang and H. Li. M. Maechler and D. Martin and J. Schimert, "Regression modeling in back-propagation and projection pursuit learning," IEEE transactions on neural networks, Vol. 5, pp. 342-353, 1994.
 [10] L. K. Jones, "A simple lemma on greedy approximation in Hilbert space and convergence rates for projection pursuit regression and neural network training," The Annals of Statistics, Vol. 20, pp. 608-613, 1992.
 [11] H. John and K. Anders and G. P. Richard, 'Introduction to the Theory Neural Computation,' Addison-Wesley Publishing Co. Inc, 1991.
 [12] N. L. Johnson and S. Kotz, 'Continuous Univariate Distributions,' Boston: Houghton Mifflin, 1970.
 [13] R. V. Lenth, "Cumulative distribution function of the non-central t distribution," Applied Statistics, Statistical Algorithms, pp. 185-189, 1988.
 [14] D. E. Rumelhart, E. H. Geoffrey, and J. W. Ronald, "Learning representations by backpropagating errors," Nature, Vol. 323, pp. 533-536, 1986.
 [15] M. Smith, 'Neural Networks for Statistical Modeling,' New York: Van Nostrand Reinhold, 1993.
 [16] D. F. Specht, "A general regression neural network," IEEE transactions on neural networks,

Vol. 2, pp. 568-576, 1991.

[17] H. White, "Learning in artificial networks: A Statistical perspective," *Neural Computation*, Vol. 1(4), pp. 425-469, 1989.

[18] H. White, "Connectionist nonparametric regression: multilayer feedforward networks can learn arbitrary mappings," *Neural Networks*, Vol. 3, pp. 535-549, 1990.



구 선 희

1986년 숙명여자대학교 수학과 졸업(이학사)

1989년 숙명여자대학교 대학원 수학과 졸업(이학 석사)

1996년 성균관대학교 대학원 통계학과 졸업(경제학박사)

1996년~현재 전주대학교 전기전

자컴퓨터 공학부 객원교수

관심분야: 신경망이론, 전산수학