論文 한국정밀공학회지 제14권 제10호 (1997년 10월)
Journal of the Korean Society of Precision Engineering Vol. 14. No. 10. October. 1997.

분산학습알고리듬의 이론적 분석

조 영 빈*, 권 대 갑*

Theoretical Analysis on the Variance Learning Algorithm

Young-Bin Cho*, Dae-Gab Gweon*

ABSTRACT

분산은 확률 모델을 표현하는 유용한 변수중 하나이다. 입력변수에 대한 함수로 표현되는 조건부 분산을 학습하는 신경회로망에 대한 많은 연구가 있어왔다. VALEAN이라는 신경회로망 역시 이러한 많은 연구중 하나인데 이것은 기본적으로 feedforward 다층 퍼셉트론 구조를 가지며 새롭게 제시된 에너지 함수를 사용하고 있다. 이 논문에서는 이 에너지 모델에 의해 결정되는 피드백에러(델타)가 신경망의 transient, steady state에서 미치는 영향을 다루었다. 과도 상태 분석에서는 델타와 수렴성, 안정성에 관한 내용을 다루고 모의 실험을 하였으며 정상 상태 분석에서는 신경회로망의 정상상태 에러의 크기와 델타의 크기사이의 상관관계에 대하여 다루었다. 학습 알고 리듬이 확률적이므로 정상상태 역시 확률적인 상태를 나타낸다. 따라서 델타의 크기에 따른 정상 상태 에러의 최대치는 확률적인 모델을 가지게 된다. 여기서는 이 확률 관계를 분석적으로 규명하고 이에 따라 원하는 신뢰도로 정상 상태 에러를 제어하기 위해 필요한 델타의 크기를 예측할 수 있는 이론적 배경을 마련하게 된다.

Key Words: Stochastic Process(확률 과정), Stochastic Learning(확률적 학습), Learning Strategy(학습 전략), Error Back-Propagation(오류 역전), Energy Function(에너지 함수)

1. INTRODUCTION

Many researcher showed that feedforward multilayer perceptron can be used to approximate any continuous function.[1-6] Learning algorithms of multilayer perceptron have been studied for this purpose. Among them, the back-propagation algorithm, what is called the generalized delta rule, is one of the most important methods to make multilayer feed-forward neural networks learn a particular function. In most cases, functions approximated by back-propagation algorithm represent deterministic relationships between input and output variables.

When the relations between input and output variables are not only deterministic but also stochastic, variance may be one of the useful parameter to estimate degree of uncertainty caused by

* 한국과학기술원 기계공학과

stochastic relation. Variance estimation is very simple, when its value is constant all over the input domain. It is not simple, however, when the variance is a function of input variables. Donald F. Specht[10] presented a review of two classification method for probability density functions. Weigend[7] developed variance learning algorithm for multilayer perceptron using maximum likelihood concept. William[8] extended its learning algorithm to cover multi-dimensional case for multilayer perceptron.

VALEAN (VAriance LEArning Neural network), which was named and proposed by Young-Bin Cho, is one of the neural networks for this purpose.[9] Its structure and learning algorithm are basically feedforward multilayer perceptron and stochastic version of error back propagation respectively. Instance feedback error and iterative variance estimation technique were newly defined and developed. That paper showed convergence of the learning algorithm on the assumption that instance feedback error is sufficiently small.

Relationships of instance feedback error to transient and steady state of VALEAN are presented in this paper, where the steady state is defined by time duration when any probability of neural network output are independent of time translations and the transient state is defined by time duration that is not the steady state. In transient state analysis, relations of feedback error to convergence and learning stability are studied and computer simulation examples are utilized for the demonstration. In steady state analysis, probabilistic relations of feedback error magnitude to steady state error bound are studied for the evaluation of relative error between true variance and estimated variance made by VALEAN.

2. LEARNING ALGORITHM of VALEAN

In this section, learning algorithm of VALEAN[9] will be summarized for the introduction. Stochastic relationship $y_s(x)$ may be assumed to be described by the gaussian density function with mean of zero

$$f(y_s) = \frac{1}{\sqrt{2\pi}\cdot\sigma_T(x)}\exp\left(-\frac{y_s^2}{2\sigma_T^2(x)}\right) \tag{1}$$

where $\sigma_T^2(x)$ is the variance of stochastic relationship, $y_s(x)$ and it is the function of input variable x. The iterative variance estimation technique can be used as shown in equation (2).

$$\sigma_N^{t+1}(x) = \sigma_N^t(x) + \Delta\sigma_N^t(x) \tag{2}$$

where $\sigma_N^t(x)$ is the variance estimated by the neural network and $\Delta\sigma_N^t(x)$ is the correction term at the learning time of t respectively. In case of error back propagation, correction term $\Delta\sigma_N^t(x)$ can be regarded as instance feedback error at the learning time t.

$$\lim_{t\to\infty}\sigma_N^t(x) = \sigma_T(x) \tag{3}$$

Equation (3) can be guaranteed when the correction term $\Delta\sigma_N^t(x)$s calculated by the following equations (4.a), (4.b) and (5). Equation (4.a) and (4.b) determine the sign of $\Delta\sigma_N^t(x)$ and Equation (5) determine the magnitude of $\Delta\sigma_N^t(x)$.

$$p(v^- \mid y_s) = \exp\left(-\frac{3y_s^2}{2\sigma_N^2}\right) \tag{4.a}$$

$$p(v^+ \mid y_s) = 1 - \exp\left(-\frac{3y_s^2}{2\sigma_N^2}\right) \tag{4.b}$$

$$\mid\Delta\sigma_N\mid \ll \sigma_N \tag{5}$$

where v^+, v^- mean that the sign of $\Delta\sigma_N$ is positive or negative respectively. Proof is given in reference 9. In the learning process of the neural network, desired output at time $t+1$ is assumed to be $\sigma_N^{t+1}(x)$ instead of $\sigma_T(x)$. So, instance feedback error will be $\Delta\sigma_N^t(x)$ at time t. Although the difference between $\sigma_N^{t+1}(x)$ and $\sigma_T(x)$ is large at the beginning of learning process, it gets small after sufficiently large learning time t.

Figure 1 shows the schematic diagram of VALEAN, where block of "Learning Rule" can be described by Equation (4.a), (4.b), and (5).

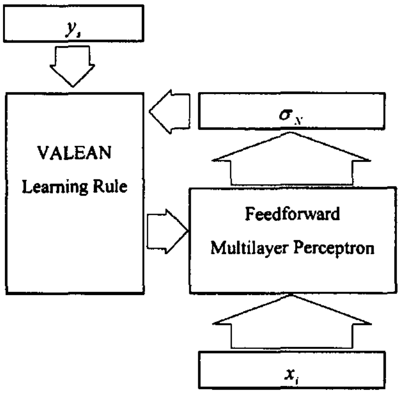

Fig. 1 Schematic diagram of variance learning neural network

$$\delta_i = V_i' \ [\sigma_T(x) - \sigma_N(x)]_i \tag{7}$$

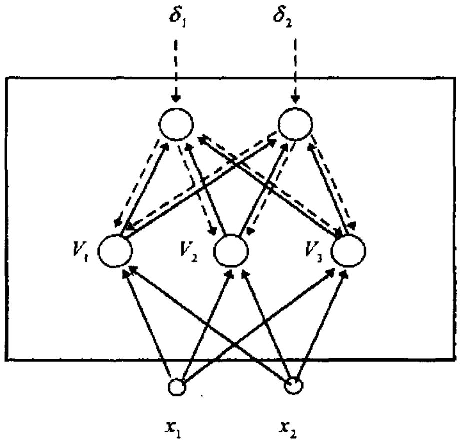

Fig. 2 Back propagation of multilayer perceptron. The solid lines show the forward propagation of signals and the dashed lines show the backward propagation of errors

3. TRANSIENT STATE ANALYSIS

3.1 Effect of feedback error magnitude in transient state

In this section, relations of feedback error to convergence and learning stability are studied and computer simulation examples are utilized for the demonstration. Figure 2 shows error back propagation in three layered network. The solid lines show the forward propagation of signals and the dashed lines show the backward propagation of errors ($\delta's$). The back propagation update rule always has the form

$$\Delta w_{pq} = \eta \sum_{\mu} \delta_p^{\mu} \times V_q^{\mu} \tag{6}$$

where subscript p and q refer to the two ends of the input and output connection concerned, superscript μ refers to input pattern, η refers to the learning rate, and V stands for the appropriate input-end activation from a real input. The meaning of δ depends on the layer concerned. For the output layer, it is given by Equation (7).

When the desired output at time $t+1$ is assumed to be $\sigma_N^{t+1}(x)$ instead of $\sigma_T(x)$, Equation (7) is defined by the following equation.

$$\delta_i = V_i' [\Delta\sigma_N]_i \tag{8}$$

where δ has stochastic behavior, since $\Delta\sigma_N$ is stochastic. So learning process is stochastic. Expectation of instance feedback error $\Delta\sigma_N$ at given input pattern x is as follows.

$$\begin{aligned} E[\Delta\sigma_N] &= |\Delta\sigma_N|(p(v^+) - p(v^+)) \\ &= |\Delta\sigma_N| \int_{-\infty}^{\infty} (p(v^+ | y_s) - p(v^- | y_s)) f(y_s) dy_s \\ &= [1 - \frac{2\sigma_N}{\sqrt{3\sigma_T^2 + \sigma_N^2}}] |\Delta\sigma_N| \end{aligned} \tag{9}$$

Equation (9) satisfies the following equations.

$$E[\Delta\sigma_N | \sigma_N << \sigma_T] \approx |\Delta\sigma_N| \tag{10.a}$$

$$E[\Delta\sigma_N | \sigma_N = \sigma_T] = 0 \tag{10.b}$$

$$E[\Delta\sigma_N | \sigma_N >> \sigma_T] \approx -|\Delta\sigma_N| \tag{10.c}$$

As a result, the magnitude of feedback error

$|\Delta\sigma_N|$ means the maximum possible value in the learning process. This is possible when the difference between output value of neural network and true variance is very large as shown in equation (10.a) and (10.c). The following three equations were proposed[9] for the magnitude of feedback error $|\Delta\sigma_N|$

$$|\Delta\sigma_N^t| \equiv C \qquad (11.a)$$

$$|\Delta\sigma_N^t| \equiv r(t)\cdot\sigma_N^t \qquad (11.b)$$

$$|\Delta\sigma_N^t| \equiv r(t)\cdot(\sigma_N^t+\xi) \qquad (11.c)$$

where C is arbitrary constant value, $\gamma(t)$ is decay parameter, which decreases with learning time t, and ξ is additive parameter, small constant value.

3.2 Computer Simulation for Transient Analysis

One hundred sample values with gaussian distribution are used for the computer simulation. Output value of neural network at time t and $t+1$ is assumed to be $\sigma_N^t(x)$ and $\sigma_N^{t+1}(x)$ respectively. σ_N^0 is assumed to be 10^{-8} and two types of γ is used as shown in the following equation and Figure 3.

$$\gamma = \frac{\gamma_0}{1+t_d} \qquad (12.a)$$

$$\gamma = \gamma_0\ \exp(-\frac{t_d^2}{2}) \qquad (12.b)$$

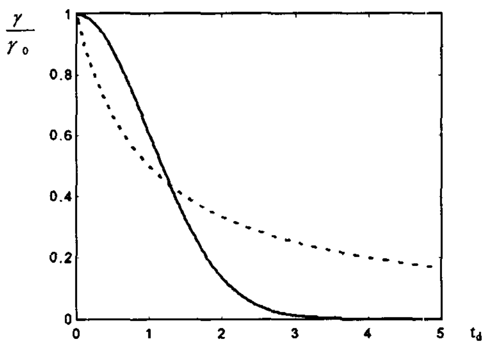

Fig. 3 Two types of decay parameters. The dashed line is geometric decay and the solid line is exponential decay.

where γ_0 is 0.01 and t_d is decay time. Equation (12.a) and (12.b) is defined as geometrical decay and exponential decay, respectively.

Decay time t_d is defined as follows.

$$t_d = \frac{t}{N^\mu} \qquad (13)$$

where N^μ is the number of pattern.

Figure 4. (a) shows the learning process of VALEAN using Equation (11.a). The value of C is $0.01\sigma_T$, $0.001\sigma_T$, $0.0001\sigma_T$, respectively. The

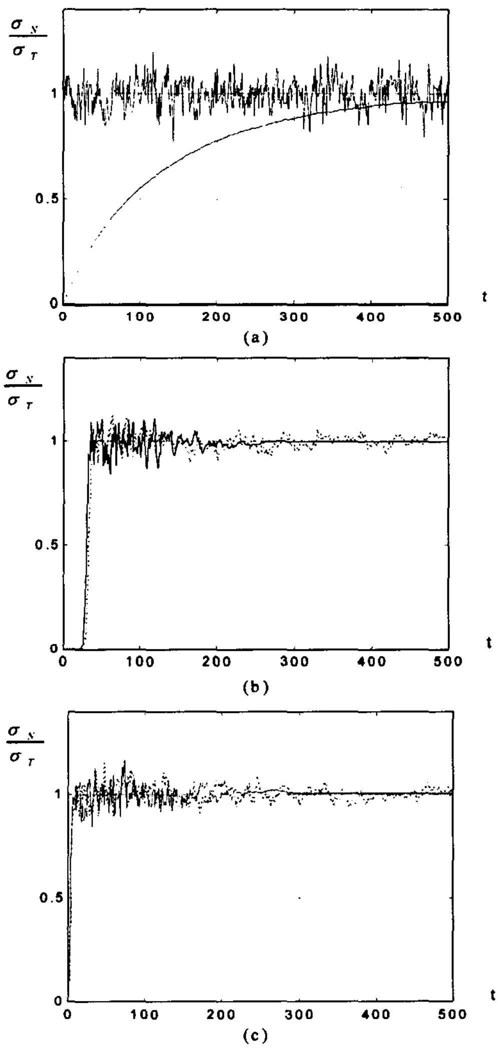

Fig. 4 Transient State of Variance Learning Process

large value of C makes the time of transient state to be short and the learning process to be unstable. It is not adequate to use larger value than 0.01 σ_T in practical situation, because neural network is plausible to diverge at large feedback error. So, the value of C is desired to have large value at the beginning and small value at the end of the learning process. It is impossible to make C have the desired value, however, since true value of variance σ_T is unknown.

Figure 4. (b) shows the learning process of VALEAN using Equation (11.b). σ_N is used for determination of feedback error magnitude instead of unknown σ_T.

Shortcoming of this method is that the update speed may be inadequately slow when the output of neural network has smaller value than the true value. This situation can occur in the beginning of learning process. Solid line is exponential decay and dashed line is geometric decay in this figure.

Figure 4. (c) shows the learning process of VALEAN using Equation (11.c). To overcome the shortcoming of Equation (11.b), additive term ε is used and its value is 0.01 in this simulation. In transient state, decay term γ and additive term ε make good performance in estimating the true variance σ_T. Solid line is exponential decay and dashed line is geometric decay in this figure.

4. STEADY STATE ANALYSIS

Steady state can be defined as the time duration $[t_1, t_2]$, in which the probability function $p(\sigma'_N)$ is identical for any time t, where $t_1 < t < t_2$ and T is sufficiently long time to obtain $p(\sigma'_N)$. Vibratory behavior of the estimated output is one of the characteristics of stochastic learning algorithm as shown in Figure 4. Magnitude of feedback error, $|\Delta\sigma_N|$ is closely related to the probability function $p(\sigma_N)$ in steady state, where σ_N is the output of VALEAN. Relations between magnitude of feedback error and the probability function of the neural network output is studied in this section. From the equation (4.a), the following equation can be driven.

$$p(v^-) = \int_{-\infty}^{\infty} p(v^- | y_s) \cdot f(y_s) dy_s = \frac{\sigma_N}{\sqrt{3\sigma_T^2 + \sigma_N^2}} \quad (14)$$

Let the magnitude of feedback error $|\Delta\sigma_N|$ be one N' th of true variance σ_T, where N is a positive integer. Variance σ_N estimated by VALEAN is assumed to be described by n times $|\Delta\sigma_N|$, where n is zero or a positive integer.

$$\sigma_T = N \cdot |\Delta\sigma| \quad (15)$$

$$\sigma_N = n \cdot |\Delta\sigma| \quad (16)$$

Using equation (15), (16) and a right triangle whose base side, vertical side and interior angle are $\sqrt{3}\sigma_T$, σ_N and θ, respectively, as shown in figure 5. Equation (14) can be described by the following equation.

$$p(v^- | n) = p(v^- | \sigma_N = n|\Delta\sigma|) = \frac{n}{\sqrt{3N^2 + n^2}} = \sin(\theta)$$

$$p(v^- | n+1) = \sin(\theta + \delta) \quad (17)$$

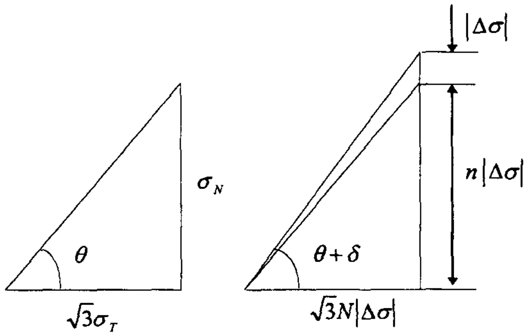

Fig. 5 A right triangle with base side, vertical side and interior angle are $\sqrt{3}\sigma_T$, σ_N and θ, respectively.

Let interior angle be $\theta + \delta$, when σ_N is $(n+1)|\Delta\sigma|$.

$$\tan\,\delta = \frac{\tan(\theta+\delta)-\tan(\theta)}{1+\tan(\theta+\delta)\ \tan\theta} \tag{18}$$
$$= \frac{\sqrt{3}N}{3N^2+n(n+1)}$$
$$\lim_{\delta\to 0}\ \tan\delta = \lim_{\delta\to 0}\ \cos\ \delta = \delta$$

Using the theorem 6 in appendix, probability function $p(\sigma_N)$ satisfies the following equation.

$$p(v^+,\ n) = p(v^-, n+1) \tag{19}$$

or

$$p(n)\ \ p(v^+ \mid n) = p(n+1)\ \ p(v^- \mid n+1) \tag{20}$$

Let $p(n+1)$ be described by the following equation.

$$p(n+1) = p(n) + \Delta p(n) \tag{21}$$

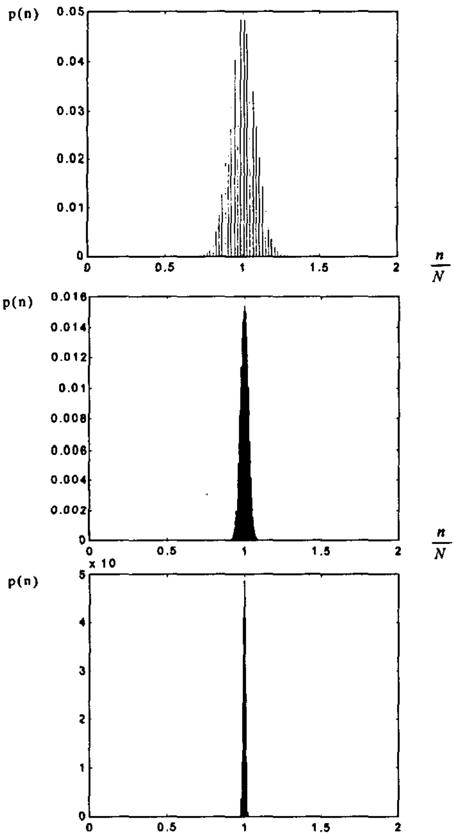

Fig. 6 Probability function of estimated variance for N=100, 1000, 10000 respectively

Using the equation (17), (19) and (21), equation (20) can be described by

$$p(n)\cdot[1-2\sin\theta-\delta\cos\theta] = \Delta p(n)\cdot(\sin\theta+\delta\cos\theta) \tag{22}$$

or

$$\Delta p(n) = r(n,N)\cdot p(n) \tag{23}$$

where $r(n,N)$ is defined by the following function using equation (18).

$$r(n,N) = \frac{(3N^2+n^2+n)\cdot(\sqrt{3N^2+n^2}-2n)-3N^2}{n(3N^2+n^2+n)+3N^2} \tag{24}$$

As a result, probability function $p(n)$ can be calculated by the following equation

$$p(n) = p(0)\cdot\prod_{i=0}^{n-1}(1+r(i,N)) \tag{25}$$

where $p(0)$ can be calculated by the following condition.

$$\sum_{n=0}^{\infty} p(n) = 1$$

Figure 6. shows the shape of the probability function $p(n)$ with respect to N Large value of N, which means small value of $\Delta\sigma$, makes the shape of probability function $p(n)$ be sharp as shown in this figure. In practice, $p(0)$ has too small value to deal with. So, maximum value of probability distribution function $P(0,n_1)$, $P(n_2,\infty)$ which are the left or right tail area of probability density function $p(n)$ and explained in theorem 6 and 7 of appendix, should be used.

Let the confidence level α be defined with respect to maximum relative error as follows.

$$\alpha = \sum_{N-M}^{N+M} p(n) \tag{26}$$

where M defines maximum relative error, max (E_r), as shown in equation (27)

$$\max\,(E_r) = \max\left|\frac{\sigma_T-\sigma_N}{\sigma_T}\right| = \left|\frac{N-M}{N}\right| \tag{27}$$

Figure 7. shows the confidence level α with respect to maximum relative error max (E_r) and magnitued of instance feedback error, which can be described by N.

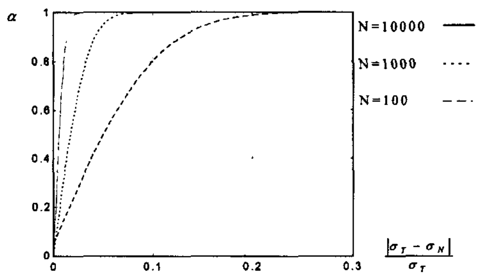

Fig. 7 Confidence level α with respect to maximum relative error

The following empirical formula can be used to describe the relation between relative error and magnitude of feedback error.

$$E = \frac{|\sigma_T - \sigma_N|}{\sigma_T / \sqrt{N}} \tag{28}$$

$$P[E \le k_\alpha] = \alpha \tag{29}$$

Table 1 shows the value of k_α with respect to α . For example, when the relative error is desired to be smaller than 0.01 with 0.99 confidence level, minimum number of N is as follows.

$$k_\alpha = k_{0.99} = 2.0982$$
$$\frac{|\sigma_T - \sigma_N|}{\sigma_T} \le \frac{k_\alpha}{\sqrt{N}} \le 0.01 \tag{30}$$
$$N \ge (k_\alpha / 0.01)^2 = 44{,}521$$

As a result, when the value of γ is set to 1 over N, maximum relative error is expected to be smaller than 0.01 with 0.99 confidence level, where γ is the decay parameter described in equation (11), and N is determined as shown in equation (30).

Table 1. k_α with respect to α and N

α \ N	100	500	1000	5000	10000
0.500	0.5013	0.5286	0.5350	0.5437	0.5457
0.550	0.5683	0.5947	0.6010	0.6097	0.6118
0.600	0.6390	0.6651	0.6715	0.6801	0.6822
0.650	0.7145	0.7411	0.7475	0.7561	0.7581
0.700	0.7972	0.8243	0.8306	0.8392	0.8413
0.750	0.8907	0.9174	0.9236	0.9322	0.9343
0.800	0.9975	1.0243	1.0309	1.0394	1.0414
0.850	1.1282	1.1535	1.1599	1.1684	1.1704
0.900	1.2949	1.3214	1.3274	1.3360	1.3380
0.910	1.3382	1.3627	1.3688	1.3773	1.3793
0.920	1.3826	1.4079	1.4140	1.4224	1.4245
0.930	1.4336	1.4579	1.4640	1.4724	1.4745
0.940	1.4887	1.5140	1.5201	1.5286	1.5307
0.950	1.5554	1.5789	1.5848	1.5933	1.5954
0.960	1.6319	1.6550	1.6616	1.6699	1.6719
0.970	1.7272	1.7504	1.7567	1.7649	1.7669
0.980	1.8566	1.8778	1.8844	1.8925	1.8945
0.990	2.0622	2.0826	2.0879	2.0962	2.0982

5. CONCLUSION

Relations of feedback error magnitude to transient and steady state behavior of variance learning neural network (VALEAN) are studied in this paper. Expectation value of feedback error is maximized to be $\pm|\Delta\sigma|$ when the difference between true variance and estimated variance is very large. In practice, this case ocurrs in the beginning of the learning process. Relation between magnitude of feedback error and transient behavior of VALEAN is studied in the part of transient state analysis. Decay parameter γ and additive parameter ε is proposed to improve the learning behavior in this study.

In steady state analysis, relations between magnitude of feedback error and probability function of estimated variance are studied. Using the result of this analysis, it is possible to expect the maximum relative error between true variance and estimated variance with desired confidence level as shown in the simple example. It is also possible to determine the decay parameter γ , since the magnitude of feedback error can be described by N. From transient and steady state analysis of VALEAN, we can design adquate value of decay parameter with respect to learning time t.

Reference

1. Hopfield, J. Neurons with graded response have collective computational properties like those of two-state neurons. Proceedings of National Academy of Sciences (U.S.A), 81, pp3088-3092., 1984.
2. Hecht-Nielsen, R. Kolmogorov mapping neural network existence theorem. IEEE International Conference on Neural Networks, 3, pp. 11-13, 1987.
3. Hornik, K., Stinchcombe, M., and White, H. Multilayer feedforward networks are universal approximators. Neural Networks, 2, pp359-366., 1989.
4. Irie, B., and Miyake, S. Capabilities of three-layered perceptrons. IEEE International Conference on Neural Networks, 1, pp641-648. , 1988.
5. Parzen, E. On the estimation of a probability density function and mode. Annuals of Mathematical Statistics, 33, pp1065-1076., 1962.
6. Rumelhart, D., Hinton, G., and Williams, R. Learning internal representations by error propagation. In D. Rumelhart, & J. McClelland(Eds.), Parallel Distributed Processing: Exploration in the Microstructure of cognition. Cambridge, MA: MIT press. 1986.
7. Weigend, A. S., and Nix, D. A. Predictions with confidence intervals (local error bars). Proc. Int. Conf. Neural Inform. Process, Seoul, Korea, pp847-852, 1994.
8. Peter M. Williams, Using Neural Networks to Model Conditional Multivariate Densities, Neural Computation, 8, pp843-854, 1996.
9. Young-Bin Cho, Variance Learning Neural Network for Confidence Estimation, KSPE, Vol 14. No. 6, in progress,
10. Donald F. Specht, Probabilistic Neural Networks and the Polynomial Adaline as Complementary Techniques for Classification, IEEE Trans. on Neural Networks Vol 1, No. 1. March, 1990.

Appendix

Let $1+r(n,N)$ be defined by the following equation

$$1+r(n,N)=\frac{(3N^2+n^2+n)(\sqrt{3N^2+n^2}-n)}{n(3N^2+n^2+n)+3N^2} \quad (A.1)$$

where n and N are positive integers.

Theorem 1. $1+r(n,N)$ is monotonic decreasing function

that is

$$\frac{1+r(n+1,N)}{1+r(n,N)}<1 \quad (A.2)$$

Proof

Let $A_n=3N^2+n^2$ and $B_n=A_n+n$. . Then left hand side of equation (A.2) can be written as follows.

$$\frac{1+r(n+1,N)}{1+r(n,N)}=\frac{\sqrt{A_{n+1}}-(n+1)}{\sqrt{A_n}-n}\cdot \frac{nB_nB_{n+1}+3N^2B_{n+1}}{(n+1)B_nB_{n+1}+3N^2B_n}. \quad (A.3)$$

Suppose that

$$\frac{\sqrt{A_{n+1}}-(n+1)}{\sqrt{A_n}-n}<1 \quad (A.4)$$

and

$$\frac{nB_nB_{n+1}+3N^2B_{n+1}}{(n+1)B_nB_{n+1}+3N^2B_n}<1 \quad (A.5)$$

Then $1+r(n,N)$ can be said to be monotonic decreasing function.

equation (A.4) can be written

$$\sqrt{A_{n+1}} < \sqrt{A_n} + 1 \qquad (A.6)$$

or

$$2n < 2\sqrt{3N^2 + n^2} \qquad (A.7)$$

From the equation (A.7), it is clear that (A.4) is true.

Equation (A.5) can be written

$$3N^2(B_{n+1} - B_n) < B_n B_{n+1} \qquad (A.8)$$

or

$$3N^2(2n+2) < (3N^2 + n^2 + n)(3N^2 + n^2 + 2n + 3) \qquad (A.9)$$

From the equation (A.9), it is clear that (A.5) is true.

$1 + r(n,N)$ is monotonic decreasing function, since (A.4) and (A.5) are true.

Theorem 2. $1 + r(n,N)$ **has the following value.**

$$1 + r(0,N) = \sqrt{3}N \qquad (A.10)$$

$$1 + r(N,N) = \frac{4N+1}{4N+4} \qquad (A.11)$$

$$1 + r(\infty,N) = 0 \qquad (A.12)$$

Proof

It is axiomatic that (A.10), (A.11), and (A.12) is true.

Theorem 3. The following equation is satisfied for any positive integer N

$$0 < 1 + r(n,N) < 1 \quad n > N \qquad (A.13)$$

Proof

Using the theorem 1 and the Equation (A.11), and (A.12), the proof of (A.13) is completed.

Theorem 4. The following equation is satisfied for positive integer N

$$0 < \frac{1}{1 + r(n,N) < 1} \quad n < N \qquad (A.14)$$

Proof

$1 + r(n,N)$ at $n = N - 1$ can be written

$$1 + r(N-1,N) = \frac{(4N^2 - N)(\sqrt{4N^2 - 2N + 1} - N + 1)}{4N^3 - 2N^2 + N} \qquad (A.15)$$

$$> \frac{(4N-1)(2N-1-N+1)}{4N^2 - 2N + 1} = \frac{N(4N-1)}{4N^2 - 2N + 1} \geq 1$$

Using the theorem 1 and the Equation (A.10), and (A.15), the proof of (A.14) is completed.

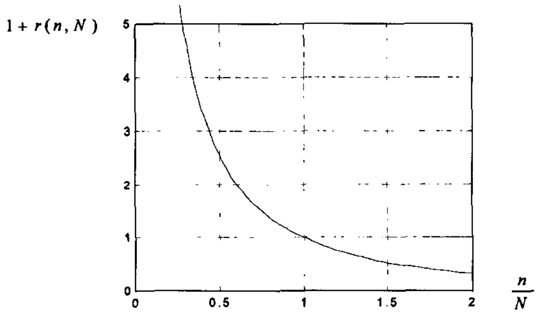

Fig. A.1 plot of *1+r(n,N)* with respect to *n/N* for N=100, 1000, 10000 respectively

Theorem 5. Probability $p(n)$ in steady state satisfies the following equation.

$$p(v^+, n) = p(v^-, n+1) \qquad (A.16)$$

Proof

Probability $p(n)$ in steady state satisfies the following equation.

$$p^t(n) = p^t(v^+, n) + p^t(v^-, n) \qquad (a.17)$$

$$p^{t+1}(n) = p^t(v^+, n-1) + p^t(v^-, n+1) \qquad (A.18)$$

where $p^t(n)$ is probability function at time t.

The following equation should be satisfied since probability $p(n)$ is in steady state.

$$p(n) = p^t(n) = p^{t+1}(n) \quad (A.19)$$

or

$$p(v^+, n-1) - p(v^-, n) = p(v^+, n) - p(v^-, n+1) \quad (A.20)$$

for any positive integer n.
(A.20) can be written

$$p(v^+, n) - p(v^-, n+1) = C \quad (A.21)$$

where C is independent of n.
(A.21) can be written as follows when $n=0$

$$p(v^+, 0) - p(v^-, 1) = C \quad (A.22)$$

or

$$p(0) = p(v^+, 0) = C + p(v^-, 1) \quad (A.23)$$

(A.18) can be written as follows when $n=0$.

$$\begin{aligned} p(0) &= p(v^+, -1) + p(v^{-1}, 1) \\ &= p(v^{-1}, 1) \quad (A.24) \end{aligned}$$

Using (A.23) and (A.24), C should be 0. This completes the proof.

Theorem 6. Let the probability distribution function $p(n_1, n_2)$ be defined by the following equation.

$$p(n_1, n_2) = \sum_{k=n_1}^{n_2} p(k) \quad (A.25)$$

$P(n_t + 1, \infty)$ satisfies the following inequality equation.

$$P(n_t + 1, \infty) < -(1 + \frac{1}{r(n_t, N)}) p(n_t) \quad (A.26)$$

where $n_t \geq N$.

Proof

$$p(n+k) = p(n) \cdot \prod_{i=0}^{k-1} (1 + r(n+i, N)) \quad k = 1, 2, \ldots$$

or (A.27)

$$\begin{aligned} p(n_t + 1, \infty) &= \sum_{k=1}^{\infty} p(n_t + k) \quad (A.28) \\ &= p(n_t) \sum_{k=1}^{\infty} \prod_{i=0}^{k-1} (1 + r(n+i, N)) \end{aligned}$$

Using the theorem 1 and 3, equation (A.28) may be described by the following equation.

$$P(n_t + 1, \infty) < p(n_t) \sum_{k=1}^{\infty} (1 + r(n_t, N))^k \quad (A.29)$$

Using the equation (A.29), the proof of (A.26) is completed.

Theorem 7. ***$P(0, n_h - 1)$* satisfies the following inequality equation.**

$$P(0, n_h - 1) < \frac{1 - r_h^{n_h}(n_n, N)}{1 - r_h(n_h, N)} r_h(n_h, N) p(n_h) \quad (A.30)$$

where $n_h \leq N$.

Proof

From the equation (25), the following equation is satisfied.

$$p(n-k) = p(n) \prod_{i=1}^{k} r_h(n-i, N) k = 1, 2, \ldots \quad (A.31)$$

where $r_h(n-1, N) = \dfrac{1}{1 + r(n-1, N)}$.

$$\begin{aligned} P(0, n_h - 1) &= \sum_{k=0}^{n_h - 1} p(k) \\ &= \sum_{k=1}^{n_h} p(n_h - k) \quad (A.32) \\ &= p(n_h) \sum_{k=1}^{n_h} (\prod_{k=1}^{k} r_h(n_h - i, N)) \end{aligned}$$

Using the theorem 1 and 4,

$$P(0, n_h - 1) < p(n_h) \sum_{k=1}^{n_h} r_h^k(n_h, N) \quad (A.33)$$

Using (A.33), the proof of (A.30) is completed.