간소화된 오차 영확률 최대화에 기반한 학습 알고리듬의 성능분석

김 남 용[1] · 권 기 현[2*]
[1]강원대학교 전자정보통신공학과 교수   [2*]강원대학교 전자정보통신공학과 교수

Performance Analysis of Learning Algorithms Based on Simplified Maximum Zero-Error Probability

Namyong Kim[1] · Ki-Hyeon Kwon[2*]

[1]Professor, Department of Electronics, Information & Communication Engineering, Kangwon National University, Samcheok 25913, Korea

[2*]Professor, Department of Electronics, Information & Communication Engineering, Kangwon National University, Samcheok 25913, Korea

[요   약]

일반적으로 쓰이는 MSE(mean squared error) 성능 기준에 근거한 학습 알고리듬은 비 가우시안 잡음환경에서 충분한 학습 성능을 나타내지 못하는 것으로 알려져 있다. 한편, MEE (minimum error entropy)에 기반하여 파생된 학습 알고리듬들은 이러한 장애를 극복할 수 있다. 그러나 MEE의 단점 중 하나로 보다 정확한 오차 엔트로피를 계산할 수 있도록 충분한 수의 오차 샘플이 필요하며 이로 말미암아 학습 시스템이 복잡해지게 된다. 최근 오차 엔트로피 계산의 효율성을 위해 연속 오차 샘플 e_k와 e_{k-1}의 차이를 활용하는 간단한 학습 방법(SMEE, simplified MEE)에서 $e_k - e_{k-1}$의 변동 (variance)가 e_k 또는 e_{k-1} 자체가 가지는 변동보다 크다는 사실에 착안하여 e_k와 e_{k-1}을 활용하는 MZEP (maximum zero error probability)를 제안하고 그 학습 성능을 통신 환경의 등화 학습에 적용하여 실험하고 분석하였다. 이 제안된 SMZEP (simplified MZEP)는 기존의 SMEE에 비해 2배 이상 빠른 수렴 속도와 1dB 이상 낮은 정상상태 MSE 성능을 나타냈다. 이에, 제안한 방법은 효율성이 필요한 학습 시스템에서 기존의 SMEE에 비해 보다 적절한 학습 알고리듬으로 적용될 수 있다.

[Abstract]

The common Mean Squared Error (MSE)-based learning algorithms are known to yield insufficient performance in non-Gaussian noise environments. In contrast, learning algorithms developed from the Minimum Error Entropy (MEE) criterion can overcome these obstacles. One of the MEE drawbacks is known for requiring sufficient error samples to correctly calculate error entropy, which in turn makes the system complicated. A recently proposed learning method for simple and efficient calculation of error entropy utilizes the difference between two consecutive error samples e_k and e_{k-1} in an iteration times k. Inspired by the fact that the variance of $e_k - e_{k-1}$ could be larger than that of e_k or e_{k-1}, this study proposed a new simple learning algorithm based on the Maximum Zero Error Probability (MZEP) criterion and its learning performance was analyzed through adaptive equalization experiment in a communication system model. The proposed simplified MZEP (SMZEP) shows convergence faster by about two times and lower steady-state MSE by about 1 dB than the simplified MEE (SMEE), indicating that the proposed SMZEP can be more appropriate for efficiency-requiring learning systems than the existing SMEE.

Ⅰ. Introduction

Many tasks in machine learning or adaptive signal processing require robustness against noise and rapid learning speed and lower steady state error. The common mean squared error(MSE) based learning algorithms are known not to yield sufficient performance in non-Gaussian noise environments[1],[2]. On the other hand, learning algorithms developed from minimum error entropy(MEE) criterion can break through these obstacles[3],[4]. The MEE criterion is a powerful approach for robust machine learning as well as for non-Gaussian signal processing[4].

MEE criterion has been designed to create error entropy using error probability density based on Parzen density method[5]. In kernel density estimation method like Parzen density, the estimate for error density converges to the true error probability as the number of error samples increases[6]. That is, the MEE requires sufficient error samples for correct calculation of error entropy, which in turn makes the system complicated.

For this reason there has not been any research of using reduced sample points before the approach of using only the current error sample e_k and the previous one e_{k-1} for calculating the error entropy has recently been proposed[7]. The calculation reduced simplified MEE(SMEE) has shown almost equal performance unlike the belief of researchers who work on MEE[7].

The SMEE criterion contains the term $e_k - e_{k-1}$ and the variance of $e_k - e_{k-1}$ could be larger than that of e_k or e_{k-1} itself, which the maximum zero error probability(MZEP) criterion handles error samples not their differences[8]. The variance of error samples leads us to propose a new simple learning algorithm based on the MZEP criterion using only e_k and e_{k-1}. In this paper, learning performance of the proposed simplified MZEP(SMZEP) is analyzed through adaptive equalization experiment in a communication system model with severely distorted multi-path fading and impulsive noise[9].

Ⅱ. MSE Criterion and LMS Algorithm

In case of FIR linear filter, a linear combiner with L weights W_k at time k can be used for input vector $X_k = [x_k, x_{k-1}, x_{k-2}, ..., x_{k-L+1}]^T$ and output $y_k = W_k^T X_k$ [1]. Defining the error as $e_k = a_k - y_k$ where a_k is the desired value or training symbol, its MSE is derived as

$$MSE = E[e_k^2] \tag{1}$$

Using the instant error power e_k^2 instead of estimating the ensemble average of error power we obtain the weight update equation 1.

$$\begin{aligned} W_{k+1} &= W_k - \mu_{LMS} \frac{\partial e_k^2}{\partial W} \\ &= W_k + 2\mu_{LMS} e_k X_k \end{aligned} \tag{2}$$

Where μ_{LMS} is a step-size or convergence parameter that controls its stability. From the fact that LMS algorithm employs the instant error power instead of using the statistically averaged error power (1), we can refer to the LMS algorithm as a simplified version of MSE based algorithms using the ensemble average of error power.

Ⅲ. MZEP Criterion based on ITL

The ITL based performance criterion is constructed by error probability density which uses Parzen window method[5].

In this section we briefly introduce the ITL based criterion that maximizes zero-error probability $f_E(e=0)$ so as to create a concentration of error samples around zero. The estimated error probability $\hat{f}_E(e)$ can be calculated using the Parzen window method with the sample size N and the Gaussian kernel of kernel size σ[5].

$$\hat{f}_E(e) = \frac{1}{N} \sum_{i=k-N+1}^{k} G_\sigma(e - e_i) \tag{3}$$

where $G_\sigma(e_i) = \frac{1}{\sigma\sqrt{2\pi}} \exp[\frac{-e_i^2}{2\sigma^2}]$ $\tag{4}$

This criterion can be developed by minimizing quadratic probability distance $QD = \int (f_E(e) - \delta(e))^2 de$ between the PDF of error signal $f_E(e)$ and Dirac-delta function of error $\delta(e)$, so that $f_E(e)$ is forced to form the shape of $\delta(e)$.

$$QD = \int f_E^2(e)de + \int \delta^2(e)de - 2\int f_E(0)de \qquad (5)$$

The term $\int \delta^2(e)de$ is not defined mathematically but can be treated as a constant since it has no relations with the system weights.

Minimization of QD in (5) induces minimization of $\int f_E^2(e)de$ and maximization of $f_E(0)$, simultaneously. It is noticeable that minimization of $\int f_E^2(e)de$ indicates maximization of error entropy that makes error samples spread apart. It is apparent that this contradicts the MEE criterion that maximizes $\int f_E^2(e)de$[3],[4]. To avoid this contradiction, MZEP criterion that maximizes only the term $f_E(0)$ in (5) has been proposed in [8] as:

$$\max_W f_E(0) \qquad (6)$$

The estimated zero-error probability $f_E^\wedge(0)$ can be calculated by replacing e with zero in the kernel density estimation method (3) with the sample size N.

$$f_E^\wedge(0) = \frac{1}{N}\Sigma_{i=k-N+1}^{k} G_\sigma(-e_i) \qquad (7)$$

Ⅳ. Simplified MZEP Criterion and related Algorithms

In Parzen window method as in (3), the estimate for error density $f_E^\wedge(e)$ converges to the true error probability $f_E(e)$ as N increases[6]. Also for the error entropy criterion[3], the error entropy approaches Renyi entropy as N increases. For these reasons, there has not been any research of using reduced sample points even to $N=2$ for e_k and e_{k-1}. The approach of using only e_k and e_{k-1} for calculating the error entropy has recently been proposed in [7]. The SMEE criterion $V^\circ(e_k)$ can be written as

$$V^\circ(e_k) = \frac{1}{2}[G_\sigma(e_k - e_{(k-1)}) + \frac{1}{\sigma}\sqrt{2\pi}] \qquad (8)$$

From the results of the work[7], learning performance in communication channel equalization based on the SMEE has shown to be similar to the case of sample size N being large enough. Inspired by

this idea, we propose to use only e_k and e_{k-1} for the criterion of the zero-error probability and analyze its performance compared to the SMEE in (8).

Then, eq. (7) becomes the following $f_E^\circ(0)$.

$$f_E^\circ(0) = \frac{1}{2}[G_\sigma(-e_{k)} + G_\sigma(-e_{k-1})] \qquad (9)$$

And maximization of (9) forces error samples e_k and e_{k-1} to be moved to near zero.

$$\max_W f_E^\circ(0) \qquad (10)$$

This simplified criterion (9) and (10) can be a summarized expression of SMZEP.

For the maximization of the Criterion (9) with respect to the system $y_k = W_k^T X_k$ in Section II, the gradient ascent method with the step-size μ_{SMZEP} can be employed as

$$W_{new} = W_{old} + \mu_{SMZEP}\frac{\partial f_E^\circ(0)}{\partial W} \qquad (11)$$

where the gradient can be evaluated from

$$\frac{\partial f_E^\circ(0)}{\partial W} = \frac{1}{2\sigma^2}e_k G_\sigma(-e_{k)}\frac{\partial y_k}{\partial W} + \frac{1}{2\sigma^2}e_{(k-1)}G_\sigma(-e_{(k-1)})\frac{\partial y_{k-1}}{\partial W} \qquad (12)$$

So the weight update equation (SMZEP algorithm) can be expressed as

$$W_{k+1} = W_k + \frac{\mu_{SMZEP}}{2\sigma^2}[e_k \cdot G_\sigma(-e_k) \cdot X_k + e_{k-1} \cdot G_\sigma(-e_{k-1}) \cdot X_{k-1}] \qquad (13)$$

For comparison, SMEE algorithm in [7] can be written as

$$W_{k+1} = W_k + \frac{\mu_{SMEE}}{\sigma^2}G_\sigma(e_k - e_{k-1}) \cdot (e_k - e_{k-1})(X_k - X_{k-1}) \qquad (14)$$

It can be noticed that SMZEP algorithm applies Gaussian kernel to the error samples e_k and e_{k-1} separately while SMEE deals with the difference between e_k and e_{k-1}. On the other hand, by comparing

(2) and (13), we can notice the error term of (13) has a modified form $e_k^\circ = e_k G_\sigma(-e_k)$ while (2) has only e_k. Likewise, it can be considered that the SMEE has a modified error $e_k^\Delta = (e_k - e_{k-1}) G_\sigma(e_k - e_{k-1})$ in (14). Then, SMZEP algorithm) becomes

$$W_{k+1} = W_k + \frac{\mu_{SMZEP}}{2\sigma^2}[e_k^\circ X_k + e_{k-1}^\circ X_{k-1}] \tag{15}$$

And SMEE algorithm becomes

$$W_{k+1} = W_k + \frac{\mu_{SMEE}}{\sigma^2} e_k^\Delta (X_k - X_{k-1}) \tag{16}$$

It may be worthwhile to observe how different performance would be resulted from these different features between SMZEP with e_k° and SMEE with e_k^Δ. And the role of these modified errors is needed to be analyzed.

Ⅴ. Performance Results and Discussion

In this Section, the learning performance of the SMEE and the SMZEP algorithm is discussed and resulting properties are analyzed in the experiment of communication channel equalization. The experimental communication system is shown in Fig. 1.
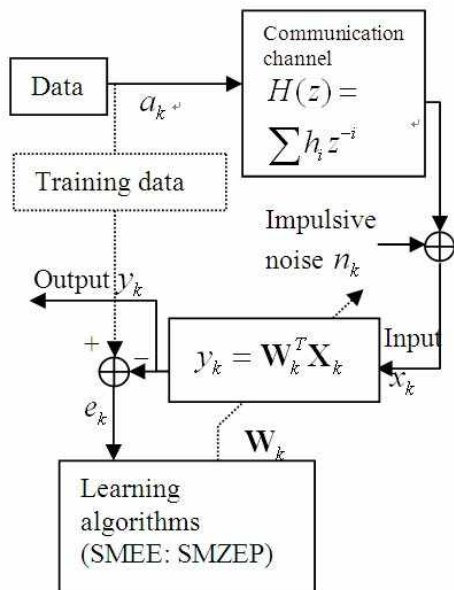

Fig. 1. A Base-band communication system for the experiment

The experimental setup is composed of a data generation (random symbol a_k is generated), communication channel $H(z)$, and a receiver equipped with a linear combiner $y_k = W_k^T X_k$. The random symbol a_k at time k is equiprobably chosen among the symbol set $\{\pm 1, \pm 3\}$ and distorted through the multipath channel $H(z)$ and corrupted by impulsive noise n_k.

The transfer function $H(z)$ has two different channel model $H1$ and $H2$ [9].

$$H1 : H_1(z) = 0.26 + 0.93z^{-1} + 0.26z^{-2} \tag{17}$$
$$H2 : H_2(z) = 0.304 + 0.903z^{-1} + 0.304z^{-2} \tag{18}$$

Both of them have spectral nulls and induce severe intersymbol interference (ISI). The impulsive noise consists of impulses and white Gaussian noise. The impulses are generated by Poisson process with variance 50 and occurrence rate 0.03[10],[11]. The variance of the background white Gaussian noise is 0.001. A sample of the impulsive noise is depicted in Fig. 2.
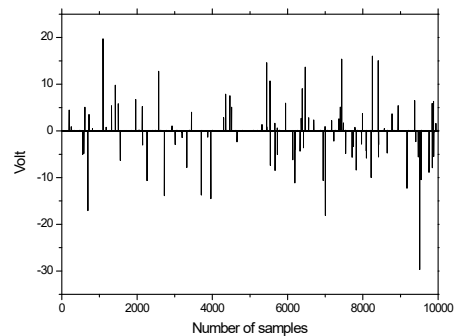

Fig. 2. A sample of impulsive noise used in the experiment

The linear combiner in the receiver has 11 weights which are updated by SMEE or SMZEP with the common step size $\mu_{SMZEP} = \mu_{SMEE} = 0.02$ and the kernel size $\sigma = 0.8$. Fig. 3 shows the MSE learning curves of SMEE and SMZEP for two different channel models. It is observed that the learning curve of LMS in (2) cannot converge below −16 dB. On the other hand, the SMEE and SMZEP algorithms show fast and stable convergence to below −23dB. In H1, the SMZEP in blue converges in about 1200 samples and the SMEE in red does in about 2000 samples. The steady state MSE is −26dB for SMZEP but −25dB for SMEE. In both convergence speed and lowest MSE, the proposed SMZEP shows superior performance. Similar results

are observed in channel model H2 showing SMZEP has a lower minimum MSE.

Fig. 4 for the comparison of system error distribution shows their performance differences more clearly. The error value on the horizontal axis is defined as the difference between the transmitted symbol and its corresponding receiver system output. The error probability is on the vertical axis. As shown in Fig. 4, the system error samples of SMEE and SMZEP produces error distributions highly concentrated around zero. Especially the proposed SMZEP has significantly narrower bell shape of error distribution than the SMEE in both channel models revealing that more of the error samples of SMZEP are around zero than those of SMEE.

To investigate the cause of the performance difference in more detail, we present the fluctuation (variance) of the center weight of each algorithm in Fig. 5. After sample number 1000 the center weight of both algorithms reaches a steady state but shows different behaviors. The variance of the center weight of SMEE stays at about 0.015 while that of SMZEP indicates 0.0025 which means the center weight of SMEE is 6 times more likely to suffer severe fluctuations.

As discussed in Section IV, the modified error $e_k^{\circ} = e_k G_\sigma(-e_k)$ in SMZEP and $e_k^{\Delta} = (e_k - e_{k-1}) G_\sigma(e_k - e_{k-1})$ in SMEE are compared in Fig. 6 in the aspect of fluctuation, that is, variance of modified error. The variance of e_k^{Δ} converges to 0.042 while that of e_k° does to 0.017 which is lower by more than 2 times. It can be understood that the bigger variance of e_k^{Δ} from SMEE causes the performance degradation in MSE convergence and error distribution compared to SMZEP.
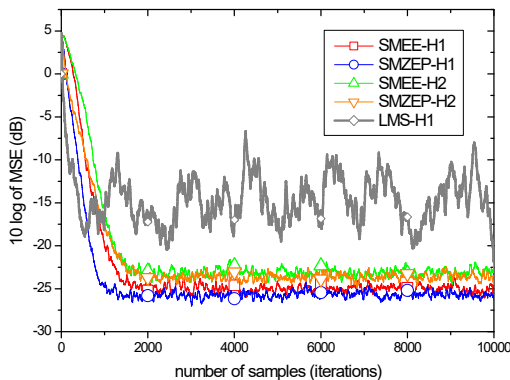

Fig. 3. MSE learning curves (red: SMEE-H1, blue: SMZEP-H1, green: SMEE-H2, orange: SMZEP-H2, gray: LMS-H1)
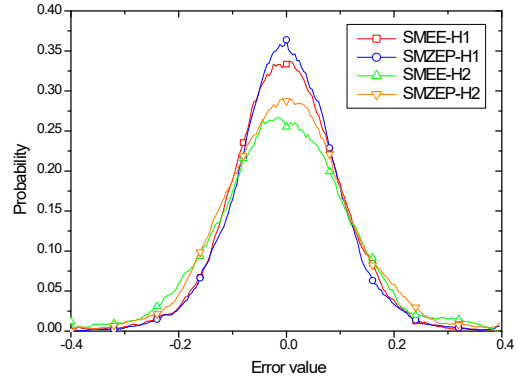

Fig. 4. Probability distribution for error samples ranging from -0.4 to 0.4 (red: SMEE-H1, blue: SMZEP-H1, green: SMEE-H2, Orange: SMZEP-H2)
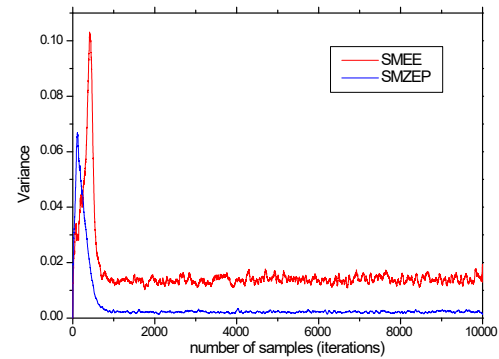

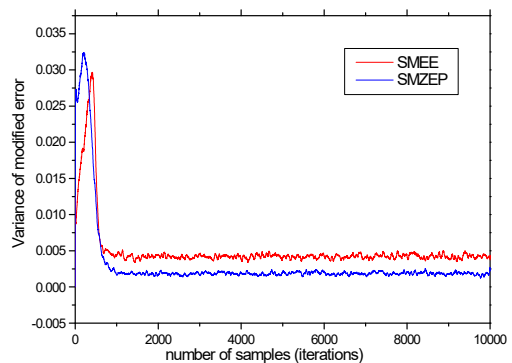**Fig. 5.** Variance of the center weight (red: SMEE, blue: SMZEP)

Fig. 6. Variance of modified error (red: SMEE, blue: SMZEP)

Ⅵ. Conclusion

In Non-Gaussian noise environments, the common criterion MSE does not provide its' related learning algorithms with acceptable learning performance. The approach based on ITL breaks through these problems

but it has been known to require sufficient data sample size that makes the system complicated. A recent study proposed a simplified MEE that utilizes only two error samples e_k and e_{k-1} in an iteration. It yields sufficiently good performance even with two error samples. In this paper, the zero error probability criterion using only two error samples e_k ande_(k-1)like the SMEE is proposed. The proposed SMZEP shows faster convergence by about 2 times and lower steady state MSE by about 1 dB than the SMEE. Also more of the system error samples of SMZEP are concentrated on zero than those of SMEE. From the analysis of the cause of the performance difference, it is found that the modified error of SMZEP produces less fluctuation, which leads to less fluctuation of weight after convergence than that of the SMEE. So we conclude that the proposed SMZEP can be an effective candidate for adaptive learning systems that may employ the existing SMEE algorithm.

감사의 글

참고문헌

[1] S. Haykin, *Adaptive Filter Theory*, 4th ed. Upper Saddle River, NJ: Prentice Hall, 2001.

[2] J. Principe, D. Xu, and J. Fisher, Information Theoretic Learning, in *Unsupervised Adaptive Filtering*, New York, NY: Wiley, pp. 265-319, 2000.

[3] D. Erdogmus and J. C. Principe, "An Error-Entropy Minimization Algorithm for Supervised Training of Nonlinear Adaptive Systems," *IEEE Transactions on Signal Processing*, Vol. 50, No. 7, pp. 1780-1786, July 2002. https://doi.org/10.1109/TSP.2002.1011217

[4] Y. Li, B. Chen, N. Yoshimura, and Y. Koike, "Restricted Minimum Error Entropy Criterion for Robust Classification," *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 33, No. 11, pp. 6599-6612, November 2022. https://doi.org/10.1109/TNNLS.2021.3082571

[5] E. Parzen, "On Estimation of a Probability Density Function and Mode," *The Annals of Mathematical Statistics*, Vol. 33, No. 3, pp. 1065-1076, September 1962. https://doi.org/10.1214/aoms/1177704472

[6] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York, NY: John Wiley & Sons, 2012.

[7] N. Kim and K. Kwon, "A Simplified Minimum Error Entropy Criterion and Related Adaptive Equalizer Algorithms," *The Journal of Korean Institute of Communications and Information Sciences*, Vol. 48. No. 3, pp. 312-318, March 2023. https://doi.org/10.7840/kics.2023.48.3.312

[8] N. Kim and K. Kwon, "Performance Enhancement of Blind Algorithms Based on Maximum Zero Error Probability," *Journal of Theoretical and Applied Information Technology*, Vol. 95, No. 22, pp. 6056-6067, November 2017.

[9] J. Proakis, *Digital Communications*, 2nd ed. New York, NY: McGraw-Hill, 1989.

[10] I. Santamaria, P. P. Pokharel, and J. C. Principe, "Generalized Correlation Function: Definition, Properties, and Application to Blind Equalization," *IEEE Transactions on Signal Processing*, Vol. 54, No. 6, pp. 2187-2197, June 2006. https://doi.org/10.1109/TSP.2006.872524

[11] N. Kim, "Kernel Size Adjustment Based on Error Variance for Correntropy Learning Algorithms," *The Journal of Korean Institute of Communications and Information Sciences*, Vol. 46, No. 2, pp. 225-230, February 2021. https://doi.org/10.7840/kics.2021.46.2.225

김남용(Namyong Kim)

1986년 : 연세대학교 전자공학과(학사)
1988년 : 연세대학교 대학원 전자공학과 (석사)
1991년 : 연세대학교 대학원 전자공학과 (박사)

1998년~현　재: 강원대학교 교수
※관심분야 : 통신 신호처리, 정보이론적 학습

권기현(Ki-Heon Kwon)

1993년 : 강원대학교 컴퓨터과학과(학사)
1995년 : 강원대학교 대학원 컴퓨터과학과(석사)
2000년 : 강원대학교 대학원 컴퓨터과학과(박사)

1998년~2002년: 동원대학 인터넷정보과 교수
2002년~현　재: 강원대학교 전자정보통신공학과 교수
※관심분야 : 패턴 인식 (Pattern Recognition), 사물 인터넷 (IoT) 응용, 기계학습