

# 코렌트로피 기반 학습 알고리즘의 커널 사이즈에 관한 연구

김남용

강원대학교 전자정보통신공학부

## A Study on Kernel Size Adaptation for Correntropy-based Learning Algorithms

Namyong Kim

School of Electronic, Information & Communications Eng, Kangwon Univ.

**요약** 머신 러닝 및 신호처리에 활용되고 있는 정보이론적 학습법(ITL, information theoretic learning)은 커널 사이즈( $\sigma$ ) 설정이 매우 민감한 어려움을 지닌다. ITL의 성능지표중 하나인 코렌트로피 함수를 최대화하는 성능지표에 대해, 기울기에 존재하는  $1/\sigma^2$ 를 제거한 뒤 남은 커널 사이즈에 대해 적응적으로 조절하는 방법들이 연구되었다. 이 논문에서는,  $1/\sigma^2$ 의 커널 사이즈가 실제 시스템의 민감성이나 불안정에 큰 역할을 하고 있으며 남은 부분에 존재하는 커널 사이즈에 대한 최적해는 오차의 절대값 근방에 수렴함에 따라 오히려 수렴 후 가중치 갱신을 멈추게 하는 부작용이 나타남을 밝혔다. 이에 적응적 커널 사이즈 조절 대신 적절한 상수를 선택하는 것이 보다 효과적이라는 것을 제안하였고, 실험 결과에서 동일한 수렴 속도에 약 2dB 향상된 정상상태 MSE를 보였다. 제안한 방식을 더욱 열악한 다경로 채널환경에 적용하여 실험한 결과 4dB 이상의 성능향상을 보여 제안한 방식은 열악한 상황일수록 더욱 향상된 성능을 보임을 알 수 있다.

**Abstract** The ITL (information theoretic learning) based on the kernel density estimation method that has successfully been applied to machine learning and signal processing applications has a drawback of severe sensitiveness in choosing proper kernel sizes. For the maximization of correntropy criterion (MCC) as one of the ITL-type criteria, several methods of adapting the remaining kernel size ( ) after removing the term have been studied. In this paper, it is shown that the main cause of sensitivity in choosing the kernel size derives from the term and that the adaptive adjustment of in the remaining terms leads to approach the absolute value of error, which prevents the weight adjustment from continuing. Thus, it is proposed that choosing an appropriate constant as the kernel size for the remaining terms is more effective. In addition, the experiment results when compared to the conventional algorithm show that the proposed method enhances learning performance by about 2dB of steady state MSE with the same convergence rate. In an experiment for channel models, the proposed method enhances performance by 4 dB so that the proposed method is more suitable for more complex or inferior conditions.

**Keywords** : Correntropy, MCC, Kernel size, Impulsive noise, Constant

---

\*Corresponding Author : Namyong Kim(Kangwon National Univ.)

email: [namyong@kangwon.ac.kr](mailto:namyong@kangwon.ac.kr)

Received October 7, 2020

Revised November 9, 2020

Accepted February 5, 2021

Published February 28, 2021

## 1. 서론

MSE(mean squared error) 성능기준에 기반하여 설계된 신호처리 알고리즘들은 충격성 잡음 같은 비가우시안 잡음환경하에서 열악한 성능을 보인다. 이에 커널 밀도추정법에 기반하여 비가우시안 잡음 상황에 효과적인 정보이론적 학습법(ITL, information theoretic learning)의 하나로 코렌트로피(correntropy) 성능기준이 소개되었다[1][2]. 이 코렌트로피 함수를 기반으로 개발된 필터링 알고리즘 또는 기계학습법이 여러 응용분야에 쓰이고 있다[3][4].

코렌트로피를 최대화(MCC, maximum correntropy criterion)하는 많은 알고리즘들 중에 민감한 문제인 커널 사이즈에 대해 최근 적응적으로 조절하는 방법을 제안한 연구가 논문[5][6]에서 발표되었다. 그런데 이 방법들은 MCC의 기울기에 존재하는  $\frac{1}{\sigma^2}$  항을 임의로 제거한 비용함수를 공통으로 채택하고 이를 기반으로 커널 사이즈 적응 조절법을 제안하고 있다. 이에, 이 논문에서는 다음 세 가지 가설을 놓고, 기존 방식[5]와[6]의 문제점을 찾아내고 개선하고자 한다. 가설1: 코렌트로피 학습법에서 커널 사이즈의 불안정성은 가중치 갱신식의  $\frac{1}{\sigma^2}$ 에 기인한다. 가설2: 오차에 대한 가중치 갱신식의  $\exp(-e_k^2/2\sigma^2)$ 에서  $\sigma$  값은 시스템 불안정에 크게 관련이 없다. 가설3: 수렴후 가중치 갱신을 계속할 수 있게 하는 다른  $\sigma$  값이 존재한다.

이 논문의 구성은 다음과 같다. 2절에서 MCC비용함수의 정의와 최대화를 위한 기울기를 전개하고 3절과 4절에서 서로 다른 커널 사이즈 최적화 방법을, 5절에서 이런 최적화 방법이 가진 문제점을 분석하고 문제 해결 방안을 제시한다. 그리고 6절에서 실험결과를, 7절에서 결론을 다루고 있다.

## 2. MCC 비용함수와 기울기

가우시안 커널 기반의 커널 밀도 추정법 (kernel density estimation)으로 계산된 시간차  $m$ 의 두 확률밀도함수  $X_k$  과  $X_{k-m}$  에 대한 내적(inner product)으로 정의되는 코렌트로피 함수  $V_X[m]$  는 다음과 같다[1][2].

$$V_X[m] = E[G_\sigma(X_k - X_{k-m})] \quad (1)$$

여기서  $G_\sigma(\cdot)$ 는 표준편차로 커널 사이즈  $\sigma$  와 영평균을 가지는 가우시안 커널이다.

$$G_\sigma(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right) \quad (2)$$

시스템 오차  $e_k$  에 대해 가우시안 커널의 exp함수 부분만 채택하고 만든 코렌트로피 함수를 최대화하는 성능 기준인 MCC (maximum correntropy criterion)가 다음과 같다.

$$MCC = E[\exp\left(-\frac{e_k^2}{2\sigma^2}\right)] \quad (3)$$

입력 벡터  $\mathbf{X}_k = [x_k, x_{k-1}, \dots, x_{k-j}, \dots, x_{k-L+1}]^T$  와 가중치  $\mathbf{W}_k = [w_{0,k}, w_{1,k}, \dots, w_{j,k}, \dots, w_{L-1,k}]^T$  로 출력  $y_k = \mathbf{X}_k^T \mathbf{W}_k$  인 TDL (tapped delay line) 구조의 적응 시스템이라는 가정하에, 원하는 신호(desired signal)  $d_k$  와 출력의 차이인 오차  $e_k$  는 다음과 같이 표현된다.

$$e_k = d_k - y_k = d_k - \mathbf{X}_k^T \mathbf{W}_k \quad (4)$$

비용함수 (3)에서  $E[\cdot]$  를 제거하고 가중치  $\mathbf{W}_k$  에 대해 기울기를 구하면 다음과 같다.

$$\nabla MCC_1(\mathbf{W}_k) = \frac{1}{\sigma^2} \exp\left(-\frac{e_k^2}{2\sigma^2}\right) e_k \mathbf{X}_k \quad (5)$$

식(5)의 기울기  $\nabla MCC_1(\mathbf{W}_k)$  는 커널 사이즈  $\sigma$  가 작은 값일 경우 오차에 매우 불안정한 값을 산출한다. 이에 논문 [5]와 [6]에서는 커널 사이즈  $\sigma$  에 대한 기울기의 민감성을 줄이기 위해, 이 기울기에서  $\frac{1}{\sigma^2}$  을 제거하고  $\exp(-e_k^2/2\sigma^2) e_k \mathbf{X}_k$  만을 기울기로 선택하였다.

$$\nabla MCC_2(\mathbf{W}_k) = \exp\left(-\frac{e_k^2}{2\sigma^2}\right) e_k \mathbf{X}_k \quad (6)$$

논문[5]는 (6)에 gradient ascent 방법으로 가중치를 갱신하고 있으며, 오차 절대값을 사용하는 적응적 최적 커널 사이즈를 제안하였다. 논문[6]는 (6)에 least

square 방법으로 가중치를 갱신하며, 오차밀도추정법의 정확성 향상에 목적을 둔  $\sigma_k$ 의 적응 조절 방법을 제안하였다. 이 커널 조절법들의 문제점을 찾아보도록 한다.

### 3. 오차감소를 최대화를 위한 커널 사이즈와

#### Gradient ascent MCC 적용

오차감소를 최대화를 위해, 식(3)의 오차 미분값을 0으로 놓고 비례상수  $C$ 를 도입하면

$$\frac{\partial MCC}{\partial e_k} = -\exp\left(\frac{-e_k^2}{2\sigma_k^2}\right) \frac{e_k(\sigma_k - e_k \partial \sigma_k / \partial e_k)}{\sigma_k^3} = 0 \quad (7)$$

$$\sigma_k = e_k \partial \sigma_k / \partial e_k \quad (8)$$

$$\sigma_k = C |e_k| \quad (9)$$

오차 변동의 영향을 줄이기 위해  $\{|e_k|, |e_{k-1}|, \dots, |e_{k-N+1}|\}$ 에서 최소 샘플값 추출 동작  $Min(|e_k|, |e_{k-1}|, \dots, |e_{k-N+1}|)$ 을 수행하고 저역통과 필터링을 추가 실행한다. 즉,

$$\bar{e}_k = \beta \bar{e}_k + (1 - \beta) \cdot Min(|e_k|, |e_{k-1}|, \dots, |e_{k-N+1}|) \quad (10)$$

$$\sigma_k = C \bar{e}_k \quad (11)$$

$$\mathbf{W}_{k+1} = \mathbf{W}_k + \mu \cdot \exp\left(\frac{-\bar{e}_k^2}{2(C\bar{e}_k)^2}\right) e_k \mathbf{X}_k \quad (12)$$

(10), (11), (12)로 요약되는 논문[5]의 이 알고리즘을 편의상 MCC-H 라 하자. 식(10)은 최소 오차 절대값을 선택하므로 0에 더 가까운 값이 선택된다. 또한 (10)의 저역통과 필터링은 작은  $\bar{e}_k$  값을 유지시키는 역할이므로 (12)의  $\exp(-\bar{e}_k^2/2(C\bar{e}_k)^2)$ 은 0이 될 확률이 다분하다. 따라서 수렴 후 가중치 갱신은 멈춘 상태가 발생하므로 성급한 수렴에 의해 잔여오차가 있을 경우, 더 이상 잔여오차를 제거하기 불가능하다.

한편, 논문[6]은 동일한 비용함수를 기반으로 가중치 갱신을 수행하나, 오차 확률밀도 추정의 정확성을 위한 커널 사이즈는 비용함수와 무관하다고 할 수 있다. 다음 4절에서, 논문 [6]은 비용함수 없이 오차확률밀도 추정이 보다 실제와 가깝도록 하는 커널 사이즈를 MCC 기반 least square 알고리즘에 적용하였음을 설명한다.

### 4. 오차확률밀도 추정법을 위한 커널

#### 사이즈와 Recursive MCC 적용

$N$  개의 오차 샘플들과 (2)에 기반하여 오차확률밀도  $f_E^\wedge(e)$ 를 추정하는 커널밀도추정법 (kernel density estimation)은 다음과 같다[7].

$$f_{E,\sigma}^\wedge(e) = \frac{1}{N} \sum_{i=1}^N \frac{1}{\sigma \sqrt{2\pi}} \exp\left[-\frac{(e-e_i)^2}{2\sigma^2}\right] \quad (13)$$

이 추정분포  $f_E^\wedge(e)$ 는 실제분포  $f_E(e)$ 와 차이를 가지는데, 시스템의 비용함수와 관계없이, 추정 오차확률밀도가 실제 확률밀도에 근접하도록 하는 최적 커널 사이즈를 구하고자 한다. 차이 척도 중에 분리 가능한 형태를 만드는 KL divergence에 적용할 경우, 다음과 같이  $D_{KL}$ 로 쓸 수 있다 [8].

$$\begin{aligned} D_{KL}(\sigma_k) &= \int f_E(e) \log\left(\frac{f_E(e)}{f_E^\wedge(e)}\right) de \\ &= \int f_E(e) \log f_E(e) de - E[\log f_E^\wedge(e)] \end{aligned} \quad (14)$$

여기서 실제분포  $f_E(e)$ 는 추정을 위한 커널 사이즈와 무관하므로 (14)는 다음  $D_{KL}^\circ(\sigma_k)$ 과 같이 쓸 수 있다.

$$\begin{aligned} D_{KL}^\circ(\sigma_k) &= E[\log f_{E,\sigma}^\wedge(e)] \\ &= E\left[\log \frac{1}{N} \sum_{i=1}^N \frac{1}{\sigma_k \sqrt{2\pi}} \exp\left(-\frac{(e-e_i)^2}{2\sigma_k^2}\right)\right] \end{aligned} \quad (15)$$

분포 차이  $D_{KL}^\circ(\sigma_k)$ 의 기울기는

$$\begin{aligned} \partial D_{KL}^o(\sigma_k)/\partial \sigma_k &= E\left[\frac{\sum_{i=k-N+1}^k \exp\left(\frac{-(e-e_i)^2}{2\sigma_k^2}\right)\left(\frac{(e-e_i)^2}{\sigma_k^3} - \frac{1}{\sigma_k}\right)}{\sum_{i=k-N+1}^k \exp\left(\frac{-(e-e_i)^2}{2\sigma_k^2}\right)}\right] \quad (16) \end{aligned}$$

현실적 적용을 위해 통계적 평균  $E[\cdot]$  을 샘플 평균으로 대체하고 기울기 (16) 를 0으로 하여 샘플시간  $k$ 에서 최적 커널 사이즈  $\sigma_{k,opt}^2$  를 구하게 된다.

$$\sigma_{k,opt}^2 = \frac{1}{N} \sum_{i=k-N+1}^k \left[ \frac{\sum_{i=k-N+1}^k (e_k - e_i)^2 \exp\left(\frac{-(e_k - e_i)^2}{2\sigma_k^2}\right)}{\sum_{i=k-N+1}^k \exp\left(\frac{-(e_k - e_i)^2}{2\sigma_k^2}\right)} \right] \quad (17)$$

이렇게 시스템의 비용함수와 관계없이 설계된 (17)의  $\sigma_{k,opt}^2$  는 추정 오차확률밀도를 실제 확률밀도와 근접하도록 하는 기능을 수행한다고 볼 수 있다.

한편 오차확률밀도 추정을 위한 최적 커널 사이즈  $\sigma_{k,opt}^2$  를 채용하면서, 통계적 평균  $E[\cdot]$  대신, forgetting factor  $\lambda$  기반의 시간평균으로 대체한 MCC비용함수 (RMC, recursive MCC)  $RMC = \sum_{i=0}^k \lambda^{k-i} \exp(-e_i^2/2\sigma_{i,opt}^2)$  가 논문[6]에 소개되었다. 이 비용함수를 가중치로 미분한 기울기에 나타나는  $1/\sigma_{k,opt}^2$  을 제거하고 얻은 수정된 기울기  $\partial RMC/\partial \mathbf{w}_k$  는 다음 식(18)과 같다 [6].

$$\partial RMC/\partial \mathbf{w}_k = \sum_{i=0}^k \lambda^{k-i} \exp\left(\frac{-e_i^2}{2\sigma_{i,opt}^2}\right) e_i \mathbf{X}_i \quad (18)$$

여기서 주목할 점은 MCC-H알고리즘이 사용한 (6)의 기울기  $\nabla MCC_2(\mathbf{w}_k)$ 와 RMC 알고리즘이 사용한 (18)의 기울기  $\partial RMC/\partial \mathbf{w}_k$  가 비용함수를 가중치로 미분하여 얻는 원래의 기울기로부터  $1/\sigma_{k,opt}^2$  을 동일하게 제거하였다는 점이다.

### 5. 정상상태에서 커널 사이즈 분석

식(17)에서  $(e_k - e_i)^2$ 을  $\nabla_{k,i} = (e_k - e_i)^2$ 라고 정의하고 시스템의 정상상태 (steady state)에서 오차가 0에 근접한다고 가정하면

$$\sigma_{k,opt}^2 = \frac{1}{N} \sum_{i=k-N+1}^k \left[ \frac{\sum_{i=k-N+1}^k \nabla_{k,i} \exp\left(\frac{-\nabla_{k,i}}{2\sigma_k^2}\right)}{\sum_{i=k-N+1}^k \exp\left(\frac{-\nabla_{k,i}}{2\sigma_k^2}\right)} \right] \quad (19)$$

$$\begin{aligned} \sigma_{k,opt}^2 &\approx \frac{1}{N} \sum_{i=k-N+1}^k \left[ \frac{\sum_{i=k-N+1}^k \nabla_{k,i} \left(1 - \frac{\nabla_{k,i}}{2\sigma_k^2}\right)}{\sum_{i=k-N+1}^k \left(1 - \frac{\nabla_{k,i}}{2\sigma_k^2}\right)} \right] \\ &= \frac{1}{N} \sum_{i=k-N+1}^k \nabla_{k,i} = \frac{1}{N} \sum_{i=k-N+1}^k (e_k - e_i)^2 \quad (20) \end{aligned}$$

샘플 평균을 통계적(statistic) 평균  $E[\cdot]$ 으로 대체하고 정상상태에서 오차들이 서로 상관관계가 없다고 가정하면

$$\sigma_{k,opt}^2 \approx 2E[e_k^2] \quad (21)$$

한편, 식(9)의  $|e_k|_{\min} = \text{Min}(|e_k|, |e_{k-1}|, \dots, |e_{k-N+1}|)$  과 one-pole 저역통과 필터링 과정을 일종의 평균화 과정이라고 보면, 식(11)로부터

$$\sigma_k^2 \approx C^2 E[|e_k|_{\min}^2] \quad (22)$$

결국, (21)과 (22)에서 두 커널 사이즈는 수렴 후 아주 작은 값의 오차전력에 비례하는 공통특성을 지니므로 식(12)의  $\exp(-e_k^2/2(Ce_k^-)^2)$ 을  $\exp(-\infty) \approx 0$ 로 만들게 된다. 따라서, 식(12)의 가중치는 더 이상 갱신되지 못하고  $\mathbf{w}_{k+1} = \mathbf{w}_k$ 를 유지하게 되어 잔여 오차를 줄일 수 없다. 따라서 성급한 수렴이 이루어질수록 정상상태의 오차전력은 더 크게 남게 된다. 이 과정을 다음 절에서 실험으로 다루어 본다.

이러한 잔여오차를 더 줄일 수 있도록 하기위해, 무의미한  $|e_k|_{\min}$  과정과 one-pole 저역통과 필터링을 빼고,

커널 사이즈를 적절한 상수  $Const$  로 놓아 수렴 후에도 가중치 갱신과정을 계속할 수 있도록 다음의 M-MCC-H (Modified MCC-H) 알고리즘을 제안한다. 즉,

$$\mathbf{W}_{k+1} = \mathbf{W}_k + \mu \cdot \exp\left(\frac{-e_k^2}{2 \cdot Const^2}\right) e_k \mathbf{X}_k \quad (23)$$

다음 절에서는, 충격성 잡음이 있는 다중경로 채널의 equalization 실험환경에서, MCC-H가 가지는 수렴특성, 커널사이즈,  $\exp(-e_k^2/2\sigma^2)$  값의 추이 등을 제안한 M-MCC-H와 함께 비교 분석하였다.

## 6. 실험결과 및 고찰

MCC학습법의 학습곡선 수렴 상태 및 성능을 비교하기 위해 충격성 잡음이 있는 다중경로 채널의 equalization 학습성능으로 평가하였다. 송신 심볼은 심볼 집합  $\{d_1 = -3, d_2 = -1, d_3 = 1, d_4 = 3\}$  에서 무작위로 보내지며 채널  $CH1(z) = 0.26 + 0.93z^{-1} + 0.26z^{-2}$  과  $CH2(z) = 0.304 + 0.903z^{-1} + 0.304z^{-2}$  의 두 다경로 채널 모델에 대해 실험하며 충격성 잡음  $n_k$  는 다음 확률분포에 따라 랜덤하게 발생된다 [1] 채널은  $H(z) = 0.26 + 0.93z^{-1} + 0.26z^{-2}$ , 충격성 잡음  $n_k$  는 다음 확률분포에 따라 랜덤하게 발생된다 [1].

$$f_{NOISE}(n_k) = \frac{1-\varepsilon}{\sigma_1\sqrt{2\pi}} \exp\left[-\frac{n_k^2}{2\sigma_1^2}\right] + \frac{\varepsilon}{\sigma_2\sqrt{2\pi}} \exp\left[-\frac{n_k^2}{2\sigma_2^2}\right] \quad (24)$$

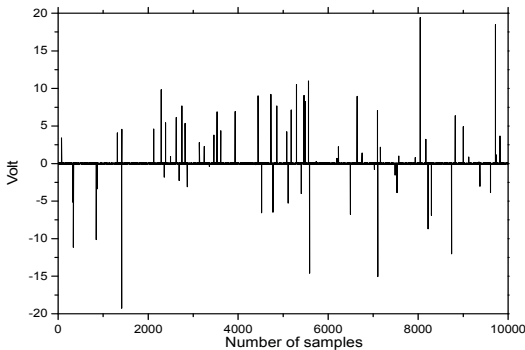


Fig. 1. A sample of impulsive noise for the simulation.

여기서  $\varepsilon < 1$ ,  $\sigma_1 = \sigma_{GN}$ ,  $\sigma_2 = \sqrt{\sigma_{GN}^2 + \sigma_{IN}^2}$ , 배경 백색잡음 전력  $\sigma_{GN}^2 = 0.001$ , 충격성 잡음의 전력  $\sigma_{IN}^2 = 50$ , 임펄스 발생률  $\varepsilon = 0.02$ 로 하였다. 랜덤 발생 잡음의 한 샘플을 그림 1에 나타냈다. 가중치 개수는  $L=11$ , 오차 블록의  $N=20$ , MCC-H를 위한 초기 커널 사이즈= 50,  $\beta = 0.98$ ,  $\mu = 0.005$ , 비교 대상인 LMS의  $\mu = 0.001$ 이다. 제안한 M-MCC-H의  $Const = 1.0$ 으로 하였다.

먼저 그림1의 동일한 잡음하에 MCC-H와 M-MCC-H의 오차 곡선을 그림 2에 비교하였다. 두 알고리즘 모두 강한 임펄스에도 안정된 오차 신호를 만들고 있다. 그러나 배경 백색 잡음만 존재하는 구간에서 오차폭을 비교하면 MCC-H는 더 줄일 수 있는 오차가 있음에도 더 줄이지 못하고 있음을 알 수 있다. 두 알고리즘의 차이가 오직  $\sigma$  를  $\sigma_k = C e_k$  로 할 것인지 상수로 둘 것인지를므로 커널 사이즈 추이를 아래 그림3에서 비교하였다. 초기치 50부터 시작한 MCC-H의 커널 사이즈는 급격히 줄어드는 모습을 보이며 3000샘플부터 거의 0을 유지하고 있다. 식(10)에서 MCC-H의 커널 사이즈가 최소 오차 값을 선택하므로 (0에 가까운 오차를 선택하므로) 타당한 결과라고 볼 수 있다.

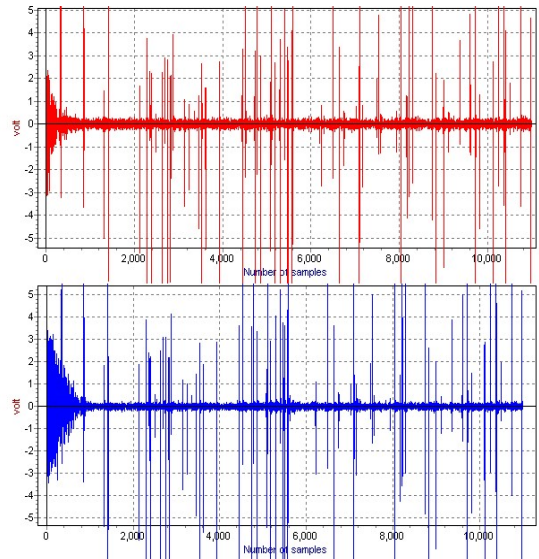


Fig. 2. Error signal comparison (위: MCC-H, 아래: M-MCC-H).

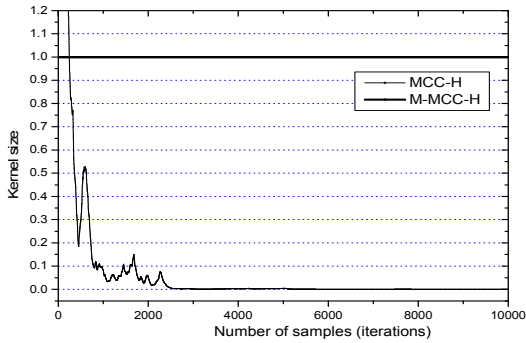


Fig. 3. kernel size learning curves.

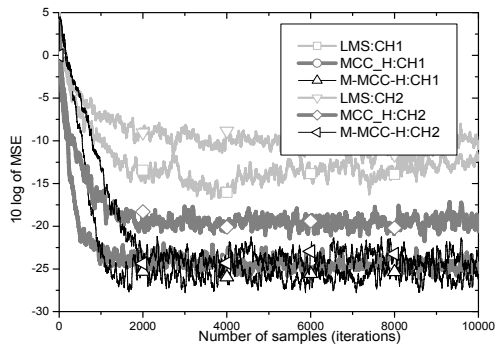


Fig. 4. MSE learning curves.

또한, 수렴 후 MCC-H는 임펄스 잡음 유무에 관계없이 가중치 갱신을 중단함을 알 수 있다. 그러나 제한한 M-MCC-H는 가중치 갱신을 계속하여 잔여 오차를 줄일 수 있음을 그림 4에 나타냈다. MSE 성능기준 기반의 LMS알고리즘은 충격성 잡음하에서 만족할 만한 수렴을 이루지 못하고 있으나 MCC-H와 M-MCC-H 모두 빠른 수렴과 낮은 정상상태 MSE성능을 나타내고 있다. 1000 샘플 정도를 수렴한 상태로 볼 수 있으므로 두 알고리즘은 같은 수렴속도를 가진다고 판단된다. 그럼에도 제한한 M-MCC-H가 CH1에서 약 2dB, 더 열악한 CH2 에서 약 4dB 이상 더 낮은 정상상태 MSE를 보이고 있다.

### 7. 결론

MCC 성능지표로 학습 알고리즘을 설계할 때 가중치에 대한 MCC의 기울기는 커널 사이즈 제공의 역수를 항상 지닌다. 이 점이 MCC 기반 학습 알고리즘의 커널 사

이즈 선택의 어려움을 낳는다. 따라서 이 부분을 제거하고 나머지 부분에 나타나는 커널 사이즈에 대해 최적값을 찾는 방법들이 연구되었다. 이 논문에서는, 커널 사이즈 제공의 역수를 제외하고 남은 부분에 존재하는 커널 사이즈에 대해 최적값을 적응적으로 찾는 방법이 시스템의 민감성이나 불안정에 큰 역할을 하지 않을 뿐 만 아니라, 오차의 RMS값에 비례하는 값으로 수렴함에 따라 오히려 수렴 후 가중치 갱신을 멈추게 하는 부작용이 나타남을 보였다. 이에 최적 커널 사이즈로 적절한 상수를 선택하는 것이 보다 효과적이라는 것을 실험 결과로 입증하였다. 앞으로 뒤따라야 할 연구에서, MCC의 기울기에서 제거되었던  $1/\sigma_k^2$  을 원래 대로 유지하며 이에 맞는 커널 추정 알고리즘이 개발되어야 한다.

### References

- [1] I. Santamaria, P. Pokharel, and J. Principe, "Generalized correlation function: Definition, properties, and application to blind equalization," *IEEE Trans. Signal Processing*, vol.54, pp. 2187-2197, June 2006. DOI : <https://doi.org/10.1109/tsp.2006.872524>
- [2] W. Liu, P. Pokharel, and J. Principe, "Correntropy: Properties and Applications in Non-Gaussian Signal Processing," *IEEE Trans. Signal Processing*, vol. 55, pp. 5286-5298, Nov. 2007. DOI : <https://doi.org/10.1109/tsp.2007.896065>
- [3] S. Sen, "Correntropy based IPKF filter for parameter estimation in presence of non-stationary noise process," *IFAC-PapersOnLine, Elsevier*, vol. 51, pp. 420-427, 2018. DOI : <https://doi.org/10.1016/j.ifacol.2018.09.611>
- [4] L. Chen, P. Honeine, "Correntropy-based robust multilayer extreme learning machines," *Pattern Recognition, Elsevier*, vol. 84, pp. 357-370, Dec. 2018. DOI : <https://doi.org/10.1016/j.patcog.2018.07.011>
- [5] F. Huang, J. Zhang, and S. Zhang, "Adaptive filtering under a variable kernel width maximum correntropy criterion," *IEEE Trans. Circuit and Systems*, vol. 64, pp. 1247-1251, Oct. 2017. DOI : <https://doi.org/10.1109/tcsii.2017.2671339>
- [6] H. Radmanesh, and M. Hajiabadi, "Recursive maximum correntropy learning algorithm with adaptive kernel size," *IEEE Trans. Circuit and Systems*, vol. 65, pp. 958-962, July 2018. DOI : <https://doi.org/10.1109/tcsii.2017.2778038>
- [7] E. Parzen, "On estimation of a probability density function and the mode," *Ann. Math. Stat.* vol. 33, p. 1065, 1962.

DOI : <https://doi.org/10.1214/aoms/1177704472>

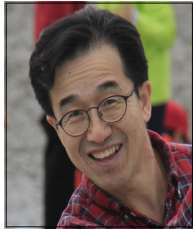
- [8] A. Singh and J. Principe, "Information theoretic learning with adaptive kernels," Signal Process., vol. 91, 203-213, 2011.

DOI : <https://doi.org/10.1016/j.sigpro.2010.06.023>

---

김 남 용(Namyong Kim)

[정회원]



- 1986년 2월 : 연세대학교 전자공학과 (공학학사)
- 1988년 2월 : 연세대학교 전자공학과 (공학박사)
- 1992년 2월 ~ 1998년 2월 : 카톨릭관동대학교 부교수

- 1998년 2월 ~ 현재 : 강원대학교 공학대학 전자정보통신공학부 교수

<관심분야>

정보이론적 학습, 정보통신