

코렌트로피 학습 알고리즘을 위한 오차분산 기반 커널 사이즈 조절

김 남 용*

Kernel Size Adjustment Based on Error Variance for Correntropy Learning Algorithms

Namyong Kim*

요 약

정보이론적 학습법의 한 성능기준인 MCC (maximum correntropy criterion)은 기울기에서 항상 커널 사이즈 제곱의 역수를 지니며 이것이 MCC 기반 학습 알고리즘을 불안정하게 만든다. 최근 커널 사이즈 제곱의 역수를 제거한 기울기에 커널 사이즈를 조절하는 방법들이 연구되었다. 그러나 충격성 잡음을 극복하고자 도입한 최소 오차 샘플 추출 과정이 커널 사이즈를 수렴 후 0이 되게 만들어 가중치 갱신은 멈추게 된다. 이 논문에서는 이 최소 오차 샘플 추출 과정을 없애고 평균 및 평탄화 과정을 거친 오차 절대값을 커널 사이즈로 계산하는 방법을 제안하였다. 제안한 방식은 적절히 조절되는 커널 사이즈로 가중치 갱신을 지속하며 기존 방식보다 더 빠른 수렴속도에 약 2 dB이상의 정상상태 MSE 향상과 0에 더 집중된 오차 분포 성능을 나타냈다.

키워드 : 코렌트로피, MCC, 충격성 잡음, 커널 사이즈, 오차 분산, 최소 샘플 추출, 평탄화 과정

Key Words : Correntropy, MCC, impulsive noise, kernel size, error variance, minimum sample extraction, smoothing process

ABSTRACT

The MCC (maximization of correntropy criterion) as one of ITL (information theoretic learning) criteria, has the inverse of kernel size squared in its slope for maximization, and this term causes the system instability. Recently, studies for adaptively adjusting the kernel size of the slope with the inverse of kernel size squared being removed. In those studies, however, the process of minimum error sample extraction employed for impulsive noise robustness leads the kernel size to zero after convergence, so that the weight adjustment cannot continue. In this paper it is proposed that without the minimum error sample extraction, an averaging and smoothing process on the absolute values of a block of error samples can create an appropriate kernel size. In the experiment, the proposed algorithm continues its weight adjustment even after convergence and yields enhanced learning performance by about 2 dB of steady state MSE with faster convergence speed compared to the conventional algorithm.

* First Author : Kangwon National Univ., Department of Electronic, Information & Communication Engineering, namyong@kangwon.ac.kr, Regular member, 종신회원

논문번호 : 202008-196-A-RN, Received August 18, 2020; Revised October 31, 2020; Accepted November 11, 2020

I. 서론

비가우시안 잡음에서 열악한 성능을 보이는 MSE (mean squared error) 기반의 학습 알고리즘에 대처될 학습법으로 커널 밀도추정법과 데이터 엔트로피에 기반한 정보이론적 학습법 (ITL, information theoretic learning)이 소개되었다^[1,2]. ITL의 일종으로, 코렌트로피 함수 (correntropy function)을 성능기준으로 하는 신호처리 또는 학습 알고리즘들이 개발되어 road noise, active headphones, transformer noise 등에 대한 ANC(active noise control)에 응용되며, 안면인식, 이미지 분류 및 인식에 쓰이는 ELM (extreme learning machine) 등, 여러 응용분야에 적용되고 있다^[3,4].

ITL 기반 알고리즘이 지닌 공통된 한 단점은 커널 사이즈 σ 의 선택이 매우 민감하고 불안정하다는 것이다^[5]. 이에 적응적으로 커널 사이즈를 조절하는 방법이 코렌트로피 최대화 (MCC, maximum correntropy criterion) 기반 알고리즘들에 대해 최근 연구되었다^[6,7]. 그러나 이들 연구에서 주의할 점은, MCC의 기울기에 근본적으로 존재하는 $1/\sigma^2$ 항이 커널 사이즈 선택을 어렵게 하고 있으므로 이 항을 제거한 비용함수를 공통으로 채택하고 있다. 그러나 $1/\sigma^2$ 을 제거하지 않은 원래의 코렌트로피를 기반으로 한 학습알고리즘들이 이미 통신 시스템 적응등화에 성공적으로 응용되고 있는 실정이다^[8].

이 논문에서는 다음 세 가지 가설을 놓고, 기존 연구^[6]의 문제점을 찾아내고 개선하고자 한다. 첫째, 커널 사이즈 선택에 있어 불안정성의 주원인은 가중치 갱신식에 나타나는 $1/\sigma^2$ 에 기인한다. 둘째, 가중치 갱신은 수렴 후에도 갱신과정을 계속해야 한다. 셋째, 커널 사이즈가 0이 될 수 있는 최소오차 선택 과정이 가중치 갱신을 멈추게 할 수 있다. 이 세 가지 가설에 대해 분석해보고 가중치 갱신이 지속될 수 있는 적절한 σ 조절이 가능한 방법을 제안하고자 한다.

II. MCC와 수정된 기울기

시스템 오차 e_k 에 대해 정의한 MCC (maximum correntropy criterion)은 다음과 같다.

$$MCC = E[\exp(\frac{-e_k^2}{2\sigma^2})] \quad (1)$$

입력 $\mathbf{X}_k = [x_k, x_{k-1}, \dots, x_{k-j}, \dots, x_{k-L+1}]^T$ 과 가

중치 $\mathbf{W}_k = [w_{0,k}, w_{1,k}, \dots, w_{j,k}, \dots, w_{L-1,k}]^T$ 로 출력을 내는 TDL (tapped delay line) 구조의 시스템의 원하는 신호 (desired signal) d_k 와 오차 e_k 는 다음과 같이 쓸 수 있다.

$$e_k = d_k - y_k = d_k - \mathbf{X}_k^T \mathbf{W}_k \quad (2)$$

(1)에서 $E[\cdot]$ 를 제거하고 구한 기울기 $\partial MCC(\mathbf{W}_k) / \partial \mathbf{W}_k = 1/\sigma^2 \exp(-e_k^2/2\sigma^2) e_k \mathbf{X}_k$ 에서 작은 커널 사이즈 σ 경우 매우 큰 값을 나타내어 갱신식을 불안정하게 만든다. 이에 논문[5]와 [6]은 $1/\sigma^2$ 을 제거한 $Slope_{MCC-H}$ 를 사용한다.

$$Slope_{MCC-H} = \exp(-e_k^2/2\sigma^2) e_k \mathbf{X}_k \quad (3)$$

논문[6]은 이 기울기(3)에 least square 방법으로 가중치를 갱신하며, 오차 밀도 추정을 최적화하는 커널 사이즈를 채택하였다.

III. 오차의 MCC 최대화를 위한 커널 사이즈

오차에 대해 MCC가 최대화되는 커널 사이즈 $\sigma_{k,H}$ 를 구하면 다음과 같다.

$$\frac{\partial MCC}{\partial e_k} = -\exp(\frac{-e_k^2}{2\sigma_{k,H}^2}) \frac{e_k (\sigma_{k,H} - e_k \partial \sigma_{k,H} / \partial e_k)}{\sigma_{k,H}^3} = 0 \quad (4)$$

$$\sigma_{k,H} = e_k \cdot \partial \sigma_{k,H} / \partial e_k \quad (5)$$

$$\sigma_{k,H} = K |e_k| \quad (6)$$

여기서 오차 절대값의 샘플 블록에서 최소 샘플만 추출하는 과정과 smoothing 과정을 수행한다[6].

$$|e_{k,\min}| = \text{Min}(|e_k|, |e_{k-1}|, \dots, |e_{k-N+1}|) \quad (7)$$

$$\bar{e}_k = \beta \bar{e}_k + (1 - \beta) \cdot |e_{k,\min}| \quad (8)$$

$$\sigma_{k,H} = K e_k \quad (9)$$

(3)과 (9)에 의해 구해진 알고리즘을 정리하면 다음과 같으며 편의상 MCC-H (MCC-Huang)라고 하겠다 [6].

$$\mathbf{W}_{k+1} = \mathbf{W}_k + \mu \cdot \exp(-e_k^2/2\sigma_{k,H}^2) e_k \mathbf{X}_k \quad (10)$$

IV. 오차확률밀도 추정법을 위한 커널 사이즈

오차에 대한 커널 밀도 추정법(kernel density estimation)은 다음과 같다⁹⁾.

$$f_{E,\sigma}^{\wedge}(e) = \frac{1}{N} \sum_{i=k-N+1}^k \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(e-e_i)^2}{2\sigma^2}\right] \quad (11)$$

추정 오차확률밀도 $f_{E}^{\wedge}(e)$ 와 실제 $f_E(e)$ 의 차이 척도 중에 분리 가능 형태를 만드는 KL divergence, $D_{KL}(\cdot)$ 에 적용하면 다음과 같다⁵⁾.

$$D_{KL}(\sigma_k) = \int f_E(e) \log f_E(e) de - E[\log f_E^{\wedge}(e)] \quad (12)$$

여기서 커널 사이즈와 무관한 $f_E(e)$ 를 제거하면, (12)는 다음 (13)과 같이 정리되며 $\partial D_{KL}(\sigma_k)/\partial \sigma_k$ 를 0으로 하여 최적 커널 사이즈 $\sigma_{k,R}^2$ 가 구해진다.

$$D_{KL}^{\circ}(\sigma_k) = E\left[\log \frac{1}{N} \sum_{i=k-N+1}^k \frac{1}{\sigma_k\sqrt{2\pi}} \exp\left(-\frac{(e-e_i)^2}{2\sigma_k^2}\right)\right] \quad (13)$$

$$\sigma_{k,R}^2 = E\left[\frac{\sum_{i=k-N+1}^k (e_k - e_i)^2 \exp\left(-\frac{(e_k - e_i)^2}{2\sigma_k^2}\right)}{\sum_{i=k-N+1}^k \exp\left(-\frac{(e_k - e_i)^2}{2\sigma_k^2}\right)}\right] \quad (14)$$

한편, 오차확률밀도 추정을 위한 (14)의 최적 커널 사이즈 $\sigma_{k,R}^2$ 를 채용하면서, MCC 비용함수 (1)에 시간평균을 적용한 recursive MCC, $RMC = \sum_{i=0}^k \lambda^i \exp(-e_{k-i}^2/2\sigma_{k-i,R}^2)$ 가 논문[7]에 소개되었다. $1/\sigma_{k,R}^2$ 이 제거된 기울기 $Slope_{RMC}$ 는

다음 식 (15)와 같다[7].

$$Slope_{RMC} = \sum_{i=0}^k \lambda^i \exp\left(\frac{-e_{k-i}^2}{2\sigma_{k-i,R}^2}\right) e_{k-i} \mathbf{X}_{k-i} \quad (15)$$

여기서 주목할 것은, $Slope_{RMC}$ 에 쓰이는 커널 사이즈가 비용함수 RMC 와 무관한, 추정 오차밀도함수 최적화용 커널 사이즈라는 점이다.

V. 정상상태의 커널 사이즈 분석 및 제한

식 (6)의 $\sigma_{k,H} = K|e_k|$ 에서 오차 절대값 최소샘플 추출 과정과 평탄화 과정을 평균화 과정이라고 보면, 식(9)의 $\sigma_{k,H} = K e_k$ 의 제곱은 다음과 같이 표현될 수 있다.

$$\sigma_{k,H}^2 \approx K^2 E[e_{k,\min}^2] \quad (16)$$

충격성 잡음의 임펄스 외 영역에 대해 정상상태에서 오차가 0에 근접한다고 볼 수 있으므로

$$\sigma_{k,H}^2 \Big|_{\text{steadystate}} \approx 0 \quad (17)$$

한편, 4절에 소개된 RMC 의 식(14)에서 오차거리 제곱을 $\nabla_{k,i} = (e_k - e_i)^2$ 라고 하고, 정상상태에서 오차가 0에 근접한다고 가정하면,

$$\begin{aligned} \sigma_{k,R}^2 \Big|_{\text{steadystate}} &= E\left[\frac{\sum_{i=k-N+1}^k \nabla_{k,i} \exp\left(-\frac{\nabla_{k,i}}{2\sigma_k^2}\right)}{\sum_{i=k-N+1}^k \exp\left(-\frac{\nabla_{k,i}}{2\sigma_k^2}\right)}\right] \\ &\approx E[\nabla_{k,i}] = 2E[e_k^2] \end{aligned} \quad (18)$$

결국, 오차에 대한 MCC 최대화 기반의 $\sigma_{k,H}^2$ (16)과 추정오차확률밀도 최적화의 $\sigma_{k,R}^2$ (18)이 정상상태에서 $E[e_k^2]$ 에 비례한다는 특성을 공통으로 지닌다. 그러나 $\sigma_{k,H}^2$ 는 $|e_{k,\min}| = \text{Min}(|e_k|, |e_{k-1}|, \dots, |e_{k-N+1}|)$ 과정을 택함으로써 식(17)과 같이 0으로 수렴하는 점이

다르다.

만약 (3)의 $Slope_{MCC-H}$ 와 (15)의 기울기 $Slope_{RMC}$ 에서 $1/\sigma^2$ 과 $1/\sigma_{k,R}^2$ 을 각각 제거하지 않은 원래 기울기를 그대로 사용한다면, 이렇게 작은 값의 $\sigma_{k,H}$ 이나 $\sigma_{k,R}$ 은 기울기를 극히 큰 값으로 만들므로 매우 불안정한 시스템이 된다. 또한, $Slope_{MCC-H}$ 의 $\exp(-e_k^2/2\sigma^2)$ 와 $Slope_{RMC}$ 의 $\exp(-e_{k-1}^2/2\sigma_{k-1,R}^2)$ 는 어떤 커널 사이즈에 대해서도 (0~1) 범위 안의 값이 되므로 시스템의 불안정 또는 시스템 성능을 크게 좌우한다고 보기 어렵다.

한편, MCC-H의 기울기에서 나타나는 $1/\sigma^2$ 이 제거된 수정된 $Slope_{MCC-H}$ 을 사용함으로써 작은 커널 사이즈 선택의 불안정성이 해소된 상태이므로, 이 기울기 $Slope_{MCC-H}$ 를 기반으로 하고 정상상태에서도 오차의 변동을 고려하여 가중치 갱신을 계속할 수 있도록, (17)과 (18)대신, 정상상태에서 커널 사이즈의 제공이 오차 분산값 $E[e_k^2]$ 되는 $\sigma_{k,proposed}$ 을 제안한다. 이를 위해 다음과 같이 오차 절대값의 샘플평균을 구하고 충격성 잡음에 강인성을 위해 (19)와 같이 샘플 평균된 $|e_{k,ave}|$ 를 평탄화시킨 값을 $\sigma_{k,proposed}$ 로 한다. 즉, 제안된 새로운 알고리즘은 다음과 같이 (19), (20), (21)로 정리된다.

$$|e_{k,ave}| = \frac{1}{N} \sum_{i=k-N+1}^k |e_i| \quad (19)$$

$$\sigma_{k+1,proposed} = \beta \sigma_{k,proposed} + (1-\beta) \cdot |e_{k,ave}| \quad (20)$$

$$\mathbf{W}_{k+1} = \mathbf{W}_k + \mu \cdot \exp(-e_k^2/2\sigma_{k,proposed}^2) e_k \mathbf{X}_k \quad (21)$$

다음 절에서는, 충격성 잡음 환경하에 다중경로 채널의 equalization 실험으로 MCC-H와 제안한 (21) (MCC-proposed)에 대해 특성 및 학습 성능을 비교 분석하였다.

VI. 실험 결과 및 고찰

시간 k 에서 전송되는 송신 데이터 d_k 는

$\{d_1 = -3, d_2 = -1, d_3 = 1, d_4 = 3\}$ 심볼 집합에서 랜덤하게 선택되어 송신되어 다중경로 채널 $H(z) = 0.26 + 0.93z^{-1} + 0.26z^{-2}$ 을 통과하고 [2]의 확률밀도 $f_{NOISE}(n_k)$ 에 따라 생성된 충격성 잡음 n_k 에 더해져서 수신된다. 수신데이터는 가중치 \mathbf{W} 의 개수는 $L=11$ 인 TDL 적응필터, $y_k = \mathbf{X}_k^T \mathbf{W}_k$ 에 입력 x_k 로 작용한다.

$$f_{NOISE}(n_k) = \frac{1-\varepsilon}{\sigma_1 \sqrt{2\pi}} \exp\left[-\frac{n_k^2}{2\sigma_1^2}\right] + \frac{\varepsilon}{\sigma_2 \sqrt{2\pi}} \exp\left[-\frac{n_k^2}{2\sigma_2^2}\right] \quad (22)$$

여기서 임펄스 발생률 $\varepsilon = 0.03$, 배경 백색잡음 전력 $\sigma_1^2 = 0.001$, 임펄스 잡음 전력 $\sigma_{IN}^2 = 50$ 이 더해진 $\sigma_2 = \sqrt{\sigma_1^2 + \sigma_{IN}^2}$ 이며, 발생된 잡음의 한 샘플을 그림 1에 나타냈다.

비교 대상으로 LMS (least mean square)의 스텝사이즈는 0.001이며, 초기 커널 사이즈 $\sigma_{0,H} = 50$, $\sigma_{0,proposed} = 1$ 로 하였고, MCC-H와 MCC-proposed에 공히 $N=20$, $\beta = 0.98$, $\mu = 0.005$ 로 하였다.

MCC-H의 $\sigma_{k,H}$ 는 급격히 줄어들며 약 2000 샘플부터 거의 0을 유지하고 있다. 커널 사이즈가 최소 오차값을 선택하므로 0에 가까워진다는 식(17)의 분석과 일치하는 결과이다. 따라서 가중치 갱신식(10)의 $\exp(-e_k^2/2\sigma^2)$ 가 0이 되어 더이상 가중치 갱신을 멈추게 된다. 그러나 MCC-proposed의 경우, $\sigma_{k,proposed}$ 는 초기값 1에서 시작하여 1.4로 증가했다가 급히 줄

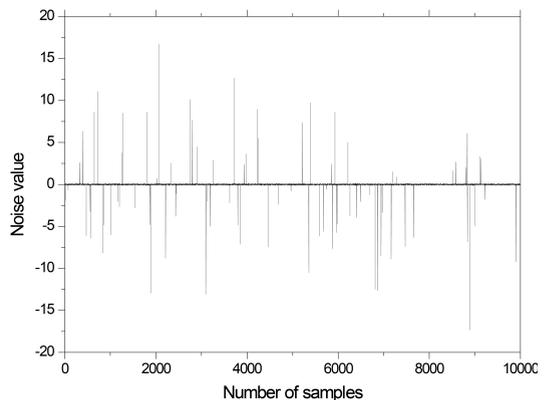


그림 1. 충격성 잡음의 한 표본
Fig. 1. A sample of impulsive noise for the simulation

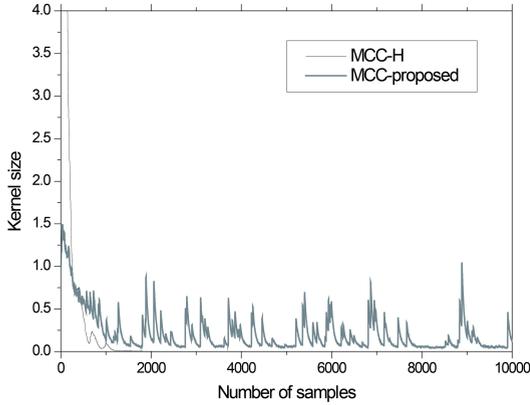


그림 2. 커널 사이즈의 학습 곡선
Fig. 2. Kernel size learning curves

어들과 있고 매 임펄스를 만날 때마다 변동을 보이지만 빠르게 다시 수렴하며 임펄스 없는 백색잡음 구간에는 약 0.07을 유지하고 있다. 오차 데이터로부터 계산된 오차 분산은 0.00597이므로 평균 오차절대값은 0.07726이 된다. 따라서 이 값은 실험 결과인

$\sigma_{k,proposed} |_{steadystate} = 0.07$ 과 매우 근사한 값을 가지므로 식(19)와 (20)의 제안된 커널 사이즈의 제곱, $\sigma_{k,proposed}^2$ 는 평균 오차전력 $E[e_k^2]$ 을 유지한다고 볼 수 있다. 또한, 제안된 $\sigma_{k,proposed}$ 가 평균 오차절대값을 적절히 유지함에 따라 가중치 갱신은 계속될 수 있게 된다. 그러므로 잔여 오차를 줄일 수 있게 된 MCC-proposed가 더 향상된 정상상태 MSE 성능을 가지게 된다. 이 결과를 그림 3에 나타냈으며 제안된 알고리즘은 더 빠른 수렴속도에, 약 2 dB이상 정상상태 MSE 격차를 보이고 있다.

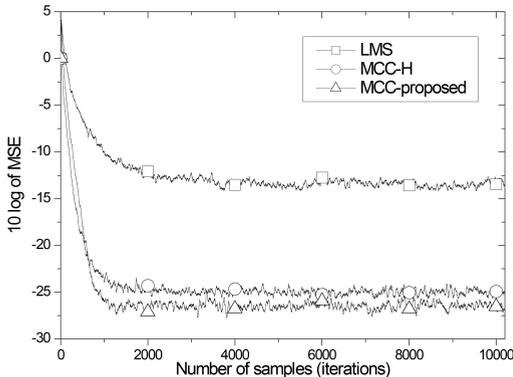


그림 3. MSE 학습 곡선
Fig. 3. MSE learning curves

그림 4는 0을 중심으로 -0.4부터 0.4까지 오차 샘플 값들의 빈도 확률을 나타내고 있는데 제안된 MCC-proposed의 잔여오차샘플들이 0에 집중-밀집된 정도에서 MCC-H 보다 현격한 차이를 보임을 알 수 있다.

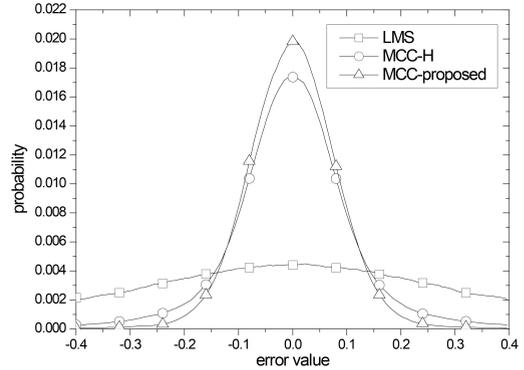


그림 4. 오차분포 성능비교
Fig. 4. Error distribution comparison

VII. 결 론

ITL 학습법의 하나인 Maximum correntropy criterion기반의 기존 연구에서 충격성 잡음을 극복하고자 도입한 최소 오차샘플 추출과정이 커널 사이즈를 정상상태에서 0 되게 만들므로 가중치 갱신을 멈추게 함을 밝혀내고 이를 보완하고자 이 논문에서는 평균 및 평탄화 과정을 거친 오차 절대값을 커널 사이즈로 채택하는 방법을 제안하였다. 제안한 방식을 통신채널 등화에 적용하여 적절한 시변 커널 사이즈로 가중치 갱신을 지속함에 의해 기존 방식보다 더 빠른 수렴속도에 약 2 dB이상의 정상상태 MSE 향상과 0에 보다 집중된 오차분포성능을 나타냈다. 이에, 커널 사이즈 계산에서 최소 오차샘플 추출과정을 제외하는 것이 보다 효과적이라는 것을 통신채널 등화실험에서 결과로 입증하였다. 따라서 이 방식은 충격성 잡음 같은 비가우시안 잡음의 통신 채널에서 적응 등화 신호 처리에 성공적으로 응용될 수 있다.

앞으로 뒤따라야 할 연구에서, MCC의 기울기에서 제거되었던 커널 사이즈 제곱의 역수를 원래대로 유지하며 이에 맞는 커널 사이즈 조절 방법이 연구되어야 한다.

References

- [1] W. Liu, P. Pokharel, and J. Principe, "Correntropy: Properties and applications in non-gaussian signal processing," *IEEE Trans. Sign. Process.*, vol. 55, pp. 5286-5298, Nov. 2007.
- [2] I. Santamaria, P. Pokharel, and J. Principe, "Generalized correlation function: Definition, properties, and application to blind equalization," *IEEE Trans. Sign. Process.*, vol. 54, pp. 2187-2197, Jun. 2006.
- [3] L. Chen and P. Honeine, "Correntropy-based robust multilayer extreme learning machines," *Pattern Recognition*, Elsevier, vol. 84, pp. 357-370, Dec. 2018.
- [4] S. Sen, "Correntropy based IPKF filter for parameter estimation in presence of non-stationary noise process," *IFAC-Papers On Line*, Elsevier, vol. 51, pp. 420-427, 2018.
- [5] A. Singh and J. Principe, "Information theoretic learning with adaptive kernels," *Signal Process.*, vol. 91, pp. 203-213, 2011.
- [6] F. Huang, J. Zhang, and S. Zhang, "Adaptive filtering under a variable kernel width maximum correntropy criterion," *IEEE Trans. Cir. and Syst.*, vol. 64, pp. 1247-1251, Oct. 2017.
- [7] H. Radmanesh and M. Hajiabadi, "Recursive maximum correntropy learning algorithm with adaptive kernel size," *IEEE Trans. Cir. and Syst.*, vol. 65, pp. 958-962, Jul. 2018.
- [8] N. Kim, "Performance analysis of correntropy-based blind algorithms robust to impulsive noise," *J. KICS*, vol. 40, pp. 2324-2330, Dec. 2015.
- [9] E. Parzen, "On the estimation of a probability density function and the mode," *Ann. Math. Stat.*, vol. 33, p. 1065, 1962.

김 남 용 (Namyong Kim)



1986년 2월 : 연세대학교 전자공학과 졸업
 1988년 2월 : 연세대학교 전자공학과 석사
 1991년 2월 : 연세대학교 전자공학과 박사
 1992년 2월~1998년 2월 : 카톨릭관동대학교 부교수

1998년 2월~현재 : 강원대학교 공학대학 전자정보통신공학부 교수

<관심분야> 정보이론적 학습 (ITL), 통신신호처리
 [ORCID:0000-0003-4397-0456]