

오차분포 유클리드 거리 기반 학습법의 커널 사이즈 적응

김남용

강원대학교 전자정보통신공학부

Adaptive Kernel Estimation for Learning Algorithms based on Euclidean Distance between Error Distributions

Namyong Kim

Division of Electronic, Information & Communication Eng., Kangwon University

요약 오차분포 추정을 위한 커널 사이즈는 오차확률밀도 사이의 유클리드 거리를 최소화 알고리즘의 가중치 갱신에 적합한 커널 사이즈가 될 수 없다. 이 논문에서는 MED 알고리즘의 수렴 성능 향상을 위해 적응적으로 커널 사이즈를 갱신하는 방법을 제안하였다. 제안한 방식은 MED 학습 알고리즘의 가중치 갱신을 위해 커널 사이즈에 대한 오차분포의 평균변화율을 도입하여 MED의 오차에 대한 평균전력이 감소하는 방향으로 커널 사이즈를 조절하도록 하였다. 제안된 적응 커널 추정법을 무선통신 채널의 왜곡 보상에 적용하여 학습 성능을 실험하고 그 효능을 밝혔다. 오차분포에 비례한 작은 값을 가지는 기존의 오차분포 추정 위한 최적 커널 사이즈와 달리, 제안한 방법에 의한 커널 사이즈는 MED 가중치 수렴을 위한 적절한 커널 사이즈로 수렴함을 보였다. 실험 결과로부터 제안한 방법이 MED 알고리즘의 커널 사이즈 설정에 따른 민감성을 크게 해결한 방법이라고 볼 수 있다.

Abstract The optimum kernel size for error-distribution estimation with given error samples cannot be used in the weight adjustment of minimum Euclidean distance between error distributions (MED) algorithms. In this paper, a new adaptive kernel estimation method for convergence enhancement of MED algorithms is proposed. The proposed method uses the average rate of change in error power with respect to a small interval of the kernel width for weight adjustment of the MED learning algorithm. The proposed kernel adjustment method is applied to experiments in communication channel compensation, and performance improvement is demonstrated. Unlike the conventional method yielding a very small kernel calculated through optimum estimation of error distribution, the proposed method converges to an appropriate kernel size for weight adjustment of the MED algorithm. The experimental results confirm that the proposed kernel estimation method for MED can be considered a method that can solve the sensitivity problem from choosing an appropriate kernel size for the MED algorithm.

Keywords : Adaptive kernel size, MED, Error distribution, Delta function, Euclidean distance, Averaged error power

1. 서론

최소자승오차 (mean squared error, MSE)는 가우시안 잡음의 학습 알고리즘에 비유함수로 주로 사용되고

있다. 한편 비가우시안 잡음에 강한 특성을 지닌 정보이론적 학습방법 (information theoretic learning, ITL)은 커널 밀도추정법 (kernel density estimation)과 정보 포텐셜 등을 기반으로 개발되어 통신, 의료 신호처리

*Corresponding Author : Namyong Kim(kangwon National Univ.)

email: namyong@kangwon.ac.kr

Received February 2, 2021

Accepted May 7, 2021

Revised March 18, 2021

Published May 31, 2021

및 기계학습을 위한 딥러닝에도 활용되고 있다 [1-3].

ITL 학습법의 커널 밀도추정법을 기반으로, 오차 샘플들이 형성하는 확률분포와 델타함수(delta function)의 유클리드 거리 (Euclidean distance, ED)를 최소화 (minimization of ED, MED)시켜서 오차샘플들을 0에 집결시키는 MED 알고리즘이 개발되어 우수한 성능을 보였다 [4]. 그러나 MED를 포함하여 모든 ITL 학습 알고리즘들은 커널 사이즈 (kernel size)의 선택에 매우 민감하다 [5,6]. 이에 추정오차 확률분포가 실제오차 확률분포와 가장 유사하도록 커널 사이즈를 적응적으로 찾아내는 기법이 제안되어 ITL 학습법에 적용되었다 [6]. 이 연구는 두 분포의 상이성 척도에 Kullback-Leibler (KL) divergence를 적용하였고 두 분포 사이의 KL divergence를 최소화시키면서 최적 커널 사이즈를 찾아가는 방법이다. 그러나 실제오차 확률분포에 추정오차 확률분포가 근접하도록 하는 커널 사이즈는 수렴 후 오차분산에 비례하는 아주 작은 값이 됨이 밝혀졌다 [7].

이 논문에서는 기존 적응 추정 방법으로 얻게 되는 극히 작은 값의 커널 사이즈로는 MED 학습법의 오차를 줄이지 못함을 밝히고 자승평균오차의 변화율을 기반으로 MED 가중치 갱신을 위한 커널 사이즈 적응적 추정 방법을 제안하고자 한다.

2. 확률밀도 추정법을 위한 커널 사이즈

N 개의 오차 샘플들 $\{e_{k-N+1}, e_{k-N}, \dots, e_k\}$ 과 커널 사이즈 σ 의 가우시안 커널을 활용하여 추정된 오차 확률밀도 $f_{E,\sigma}^{\wedge}(e)$ 는 다음과 같다 [8].

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

실제오차 확률분포 $f_E(e)$ 와 추정오차 확률밀도 $f_E^{\wedge}(e)$ 의 거리를 KL divergence (D_{KL}) 기준에 따라 그 차이를 구하면 다음과 같다 [6].

$$D_{KL} = \int f_E(e) \log\left(\frac{f_E(e)}{f_E^{\wedge}(e)}\right) de$$

$$= \int f_E(e) \log f_E(e) de - E[\log f_E^{\wedge}(e)] \quad (2)$$

변수와 무관한 $f_E(e)$ 는 제거될 수 있으므로

$$D_{KL} = -E[\log f_E^{\wedge}(e)] \quad (3)$$

식(1)의 $f_{E,\sigma}^{\wedge}(e)$ 를 (3)에 대입하면

$$D_{KL} = -E\left[\log \frac{1}{N} \sum_{i=k-N+1}^k \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(e-e_i)^2}{2\sigma^2}\right]\right] \quad (4)$$

D_{KL} 를 σ 로 미분하면,

$$\frac{\partial D_{KL}}{\partial \sigma}$$

$$= E\left[\frac{\sum_{i=k-N+1}^k \exp\left(-\frac{(e-e_i)^2}{2\sigma_k^2}\right) \left(\frac{(e-e_i)^2}{\sigma_k^3} - \frac{1}{\sigma_k}\right)}{\sum_{i=k-N+1}^k \exp\left(-\frac{(e-e_i)^2}{2\sigma_k^2}\right)}\right] \quad (5)$$

식 (5)를 $\frac{\partial D_{KL}}{\partial \sigma} = 0$ 으로 놓고 정리하면 다음과 같은 최적 $\sigma_{k,opt}^2$ 을 얻게 된다.

$$\sigma_{k,opt}^2 = E\left[\frac{\sum_{i=k-N+1}^k (e_k - e_i)^2 \exp\left(-\frac{(e_k - e_i)^2}{2\sigma_k^2}\right)}{\sum_{i=k-N+1}^k \exp\left(-\frac{(e_k - e_i)^2}{2\sigma_k^2}\right)}\right] \quad (6)$$

식 (6)의 $\sigma_{k,opt}^2$ 는 추정 확률밀도가 실제 확률밀도와 가장 유사하도록 하는 커널 사이즈로 볼 수 있다. 그러나 정상상태에서 $\sigma_{k,opt}^2$ 의 영향을 관찰하면 다음과 같은 문제점을 발견하게 된다 [7]. 시스템의 정상상태 (steady state)의 오차는 거의 0에 이르게 된다고 가정하면 오차 거리제곱 $(e_k - e_i)^2$ 은 아주 작은 값을 갖게 된다.

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

$$\approx E[(e_k - e_i)^2] \quad (7)$$

정상상태에서 오차 샘플 상관도가 0이라 가정하면,

$\sigma_{k,opt}^2$ 는 오차분산 σ_E^2 의 두 배를 가지게 된다.

$$\sigma_{k,opt}^2 \approx 2\sigma_E^2 \quad (8)$$

그러나 이 작은 값의 커널 사이즈는 오히려 정상상태 오차를 증가시킬 수 있다. 다음 절에서 제안한 적응 커널 사이즈는 정상상태에서 (8)의 값으로 수렴하지 않고 다른 적정 범위로 수렴함을 실험에서 관찰하였다.

3. MED 학습법을 위한 적응 커널 사이즈

수신정보를 다룰 정보처리기로 가중치 \mathbf{W}_k 의 선형합

성기 (linear combiner)를 채용한다고 할 때, 입력 $\mathbf{X}_k = [x_k, x_{k-1}, x_{k-2}, \dots, x_{k-L+1}]^T$ 에 대해 출력은 $y_k = \mathbf{W}_k^T \mathbf{X}_k$ 로 표현될 수 있다. 훈련열과 출력의 차이인 오차에 대해, 추정 오차확률밀도 $f_E^\wedge(e)$ 가 매우 좁은 종모양 분포, 즉 $\delta(e)$ 로 만들어질수록 오차값은 거의 0에 집결한다고 볼 수 있다. 이에 추정오차 확률밀도 $f_E^\wedge(e)$ 와 델타함수 $\delta(e)$ 의 유클리드 거리 $ED[f_E^\wedge(e), \delta(e)]$ 를 다음과 같이 정의하고 이 거리를 최소화시키는 학습 알고리즘을 설계할 수 있다.

$$ED[f_E^\wedge(e), \delta(e)] = \int [f_E^\wedge(e)de - \int \delta(e)]^2 de$$

$$= \int f_E^{\wedge 2}(e)de + \int \delta^2(e)de - 2f_E^\wedge(0) \quad (9)$$

여기서 $\int \delta^2(e)de$ 은 가중치 갱신과 무관하므로 상수로 놓고 $ED[f_E^\wedge(e), \delta(e)]$ 를 미분하면

$$\left. \frac{\partial ED[f_E^\wedge(e), \delta(e)]}{\partial \mathbf{W}} \right|_k = A_k - B_k \quad (10)$$

여기서

$$A_k = \frac{1}{4\sqrt{\pi}\sigma^3 N^2} \sum_{i=k-N+1}^k \sum_{j=k-N+1}^k (e_j - e_i) \cdot \exp\left[\frac{-(e_j - e_i)^2}{4\sigma^2}\right] \cdot (\mathbf{X}_j - \mathbf{X}_i) \quad (11)$$

$$B_k = \frac{\sqrt{2}}{\sigma^3 N \sqrt{\pi}} \sum_{i=k-N+1}^k e_i \cdot \mathbf{X}_i \cdot \exp\left[\frac{-e_i^2}{2\sigma^2}\right] \quad (12)$$

스텝 사이즈 μ_{MED} 와 고정 커널 사이즈 σ 를 사용하는 MED 알고리즘을 요약하면 다음과 같다[4].

$$\mathbf{W}_{k+1}(\sigma) = \mathbf{W}_k(\sigma) - \mu_{MED} A_k(\sigma) + \mu_{MED} B_k(\sigma) \quad (13)$$

식 (11)과 (12)의 분모에 σ^3 의 영향으로 작은 값의 커널 사이즈는 매우 큰 값의 $A_k(\sigma)$ 와 $B_k(\sigma)$ 를 유발할 수 있다. 따라서 식(8)의 $\sigma_{k,opt}$ 를 MED 알고리즘의 가중치 갱신에 커널 사이즈로 쓸 경우, 오차가 오히려 증가하는 역효과를 낳게 된다. 이에 새로운 커널 사이즈 σ_k 를 사용하는 $MED(\sigma_k)$ 를 제안하고자 한다.

$$\mathbf{W}_{k+1}(\sigma_{k+1}) = \mathbf{W}_k(\sigma_k)$$

$$- \mu_{MED} A_k(\sigma_{k+1}) + \mu_{MED} B_k(\sigma_{k+1}) \quad (14)$$

그림 1은 제안한 커널 추정 방식이 표현된 수신 시스템이다. 샘플시간 k 에서 두 가지 필터 $y_k = \mathbf{W}_k^T \mathbf{X}_k$ 와 $y_{\Delta,k} = \mathbf{W}_{\Delta,k}^T \mathbf{X}_k$ 가 병렬로 작동한다. 커널 사이즈 σ_k 와, 간격 Δ 을 가지는 $\sigma_{\Delta,k} = \sigma_k + \Delta$ 로 동작하는 두 알고리즘 (14)와 (15)는 각 가중치를 따로 갱신한다.

$$\mathbf{W}_{\Delta,k+1}(\sigma_{\Delta,k+1}) = \mathbf{W}_{\Delta,k}(\sigma_{\Delta,k})$$

$$- \mu_{MED} A_k(\sigma_{\Delta,k+1}) + \mu_{MED} B_k(\sigma_{\Delta,k+1}) \quad (15)$$

동일한 훈련열에 대한 출력 y_k 의 오차 e_k , 출력 $y_{\Delta,k}$ 의 $e_{\Delta,k}$ 는 제공되고 샘플 사이즈 N 의 sliding window 과정을 거쳐 평균 오차 전력 $\varepsilon_{MED}(e_k)$ 와 $\varepsilon_{MED}(e_{\Delta,k})$ 이 된다. 이 커널 사이즈 추정법의 주된 요점은, 병행한 두 필터의 평균오차전력 $\varepsilon_{MED}(e_k)$ 와 $\varepsilon_{MED}(e_{\Delta,k})$ 을 비교하여 더 작은 오차전력 방향으로 커널 사이즈 σ_k 를 조금씩 옮기자는 것이다. 이에 평균변화율

$\nabla \varepsilon_{MED}(e_k) = \frac{\varepsilon_{MED}(e_{\Delta,k}) - \varepsilon_{MED}(e_k)}{\Delta}$ 와 스텝 사이즈 μ_σ 를 도입하여 아래와 같은 적응 방정식을 얻을 수 있다.

$$\sigma_{k+1} = \sigma_k - \mu_\sigma \cdot \nabla \varepsilon_{MED}(e_k) \quad (16)$$

여기서,

$$\varepsilon_{MED}(e_k) = \frac{1}{N} \sum_{i=k-N+1}^k e_i^2 \quad (17)$$

$$\varepsilon_{MED}(e_{\Delta,k}) = \frac{1}{N} \sum_{i=k-N+1}^k e_{\Delta,i}^2 \quad (18)$$

이렇게 갱신된 σ_{k+1} 와 $\sigma_{\Delta,k+1} = \sigma_{k+1} + \Delta$ 는 그림 1의 병행 필터에 각각 입력되어 가중치 갱신에 사용된다. 이 제안한 학습법을 MED-AKS (MED with adaptive kernel size)라고 하겠다.

4. 실험 결과 및 고찰

다중경로 채널을 통과하며 왜곡된 신호로부터 원 데이터를 복원하는 무선통신 수신시스템에 제안한 적응 커널 추정 MED 학습법을 적용하여 실험하였다. 송신단은 심

볼 집합 $\{-3, -2, 1, 3\}$ 에서 무작위 심볼이 전송되며 채널은 $H(z) = 0.26 + 0.93z^{-1} + 0.26z^{-2}$ 으로 사용하였다 [9]. 채널 통해 왜곡된 신호는 논문 [4,10]에서 제시한 방법의 충격성 잡음과 결합되어 수신단에 입력된다.

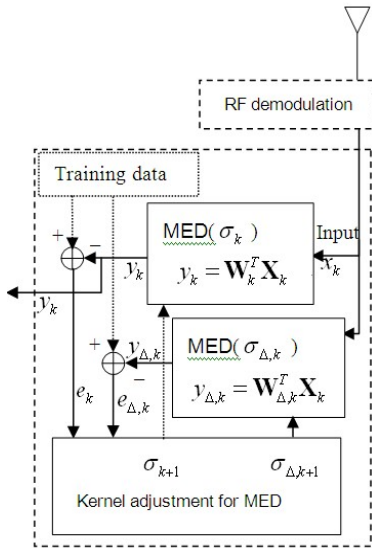


Fig. 1. Base-band receiver system employing MED algorithm with the proposed adaptive kernel estimation.

충격성 잡음은 분산 0.001의 백색잡음에 발생을 0.03, 분산 50의 충격파가 더해진다. 그림 2에 잡음 샘플을 나타냈다.

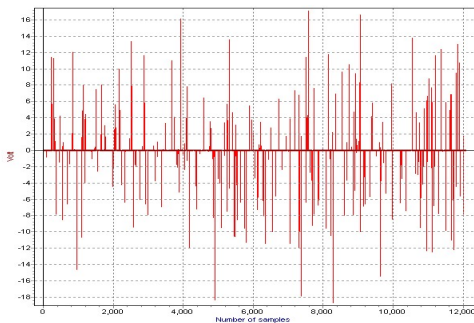


Fig. 2. A sample of impulsive noise for the experiment.

가중치 개수 $L=11$, 샘플 사이즈 $N = 24$, MED 알고리즘의 가중치 갱신 스텝 사이즈는 0.005, 커널 사이즈 조절을 위한 스텝 사이즈는 0.0008, 커널 사이즈 간격은 0.005이다. 그림 3은 제안한 방식의 초기설정 값에 대한

커널 사이즈 학습 곡선을 나타냈다. 0.7부터 1.8까지 초기설정 커널 사이즈에 대한 학습 곡선에서 약 0.87이 MED 알고리즘의 최적 커널 사이즈임을 알 수 있다. 초기설정 커널 사이즈가 큰 값인 경우는 느리지만 안정된 수렴을, 작은 값의 경우 빠르나 초기 요동에 불안정할 수 있다. 그러나 어느 경우에서나 수렴한 커널 사이즈는 식 (8)을 따르지 않고 큰 값으로서 정상상태 커널 사이즈 0.87로 수렴함을 알 수 있다. 이는 2절에서 언급한 바와 같이 알고리즘의 수렴을 위한 적정 커널 사이즈는 (8)의 결과와 일치하지 않음을 보여준다.

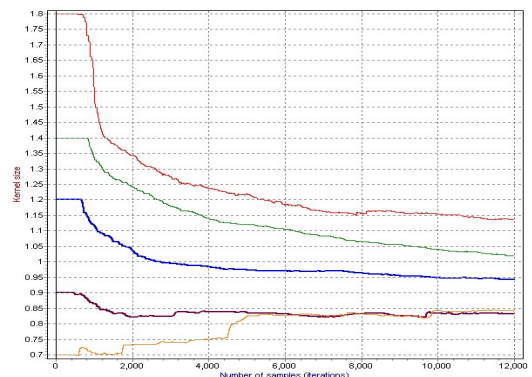


Fig. 3. Learning curves of kernel size adaptation by the proposed method

초기 커널 사이즈 $\sigma_o = 0.7$ 경우는 그림 4에, 0.9경우는 그림 5에 기존 알고리즘 ($\sigma_o = \sigma$, 빨간색)과 제안한 알고리즘(파란색)의 학습성능을 비교하였다.

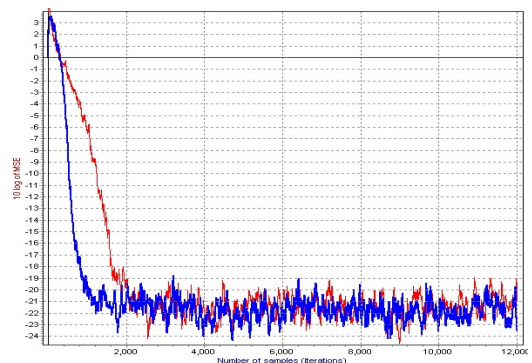


Fig. 4. Learning curves of the conventional MED (red) and proposed MED (blue) for $\sigma_o = 0.7$.

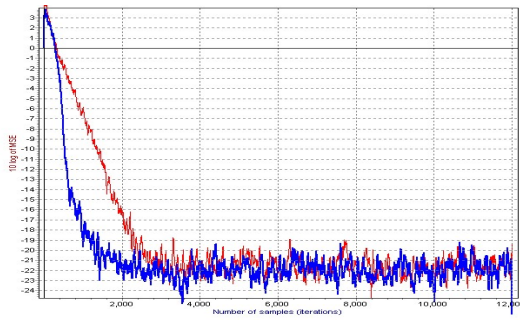


Fig. 5. Learning curves of the conventional MED (red) and proposed MED (blue) for $\sigma_o = 0.9$.

두 알고리즘의 수렴 비교에서 제안한 방식은 두 배 이상 더 빠른 수렴 속도를 나타내고 있다.

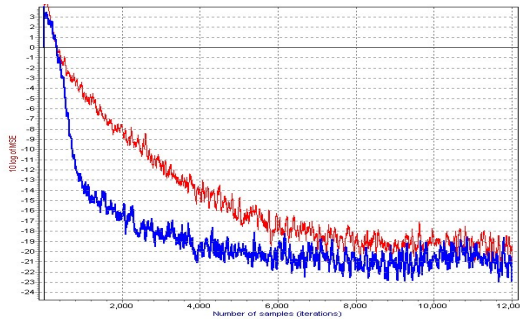


Fig. 6. Learning curves of the conventional MED (red) and proposed MED (blue) for $\sigma_o = 1.4$.

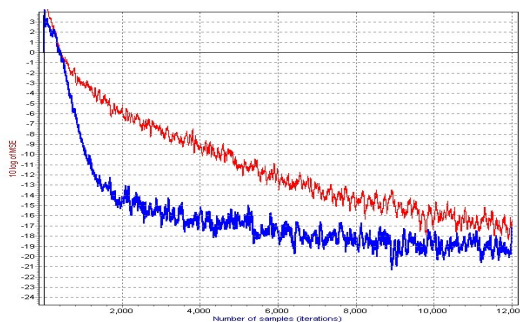


Fig. 7. Learning curves of the conventional MED (red) and proposed MED (blue) for $\sigma_o = 1.8$.

초기 커널 사이즈 1.4인 그림 6과 1.8의 경우인 그림 7에서 수렴은 함께 느려지는 경향이 있으나 수렴 속도의 차이와 정상상태 MSE의 차이가 모두 더욱 커지는 것을 알 수 있다.

5. 결론

ITL 기반의 오차 확률밀도 유클리드 거리 최소화 학습법 (MED)는 가중치 갱신을 위한 커널 사이즈가 오차 분포 추정을 위한 커널 사이즈와 다른 값이 되므로 가중치 학습을 위해 달리 설계되어야 함이 이 논문에서 밝혀졌다. 이에 MED 알고리즘의 가중치 갱신을 위한 커널 사이즈 추정에 오차분산의 평균변화율을 도입하여 적응적으로 갱신하는 방법을 제안하였다. 제안된 적응 커널 추정법을 무선통신 채널의 왜곡 보상에 적용하여 학습 성능을 실험하고 그 효능을 입증하였다. 실험 결과에서 불안정이나 느린 수렴을 유발하는 넓은 영역의 초기 커널 사이즈에 대해서 안정 영역 커널값으로 빠르게 수렴시키는 특성을 보였다. 따라서 MED 알고리즘의 민감했던 기존 커널 사이즈 설정 문제를 크게 해결한 방법으로서 여러 학습법 응용에 효과적으로 사용될 수 있다고 판단된다.

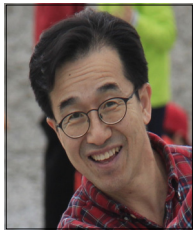
References

- [1] N. Kim, "Robustness to Impulsive Noise of Algorithms based on Cross-Information Potential and Delta Functions," *Journal of Internet Computing and Services*, vol. 17, pp. 1-6, Dec. 2017. DOI: <https://doi.org/10.7472/jksii.2016.17.2.11>
- [2] T. Ogunfunmi and M. Deb, "On the PDF estimation for information theoretic learning for neural networks," *Proceedings of APSIPA-ASC2018*, Honolulu, USA, pp. 1215-1221, Nov. 2018. DOI: <https://doi.org/10.23919/APSIPA.2018.8659642>
- [3] 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>
- [4] N. Kim and G. Lee, "Performance enhancement of algorithms based on error distributions under impulsive noise," *JICS*, vol. 19, pp. 49-56, June, 2018. DOI: <https://doi.org/10.7472/jksii.2018.19.3.49>
- [5] J. Principe, *Information Theoretic Learning*, Springer, New York, 2010, pp. 103-140. DOI: <https://doi.org/10.1007/978-1-4419-1570-2>
- [6] H. Radmanesh, M. Hajiabadi, "Recursive maximum correntropy learning algorithm with adaptive kernel size," *IEEE Trans. Circuits and Systems*, vol. 65, pp. 958-963, July 2018. DOI: <https://doi.org/10.1109/tcsii.2017.2778038>
- [7] N. Kim, "A study on kernel size adaptation for

- correntropy-based learning methods," *Journal of the Korea Academia-Industrial Cooperation Society*, in press, 2021.
- [8] E. Parzen, "On the 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>
- [9] J. Proakis, *Digital Communications*, McGraw-Hill, NY, 1989, pp 438-439. ISBN10:0070517266
- [10] 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>

김 남 용(Namyong Kim)

[정회원]



- 1986년 2월 : 연세대학교 전자공학과 (공학학사)
- 1988년 2월 : 연세대학교 전자공학과 (공학박사)
- 1992년 2월 ~ 1998년 2월 : 카톨릭 관동대학교 부교수
- 1998년 2월 ~ 현재 : 강원대학교 공학대학 전자정보통신공학부 교수

〈관심분야〉

정보이론적 학습, 정보통신 신호처리