DBpia

논문 14-39A-03-04

The Journal of Korea Information and Communications Society '14-03 Vol.39A No.03
http://dx.doi.org/10.7840/kics.2014.39A.3.149

랜덤오더 심볼열과 상호 코렌트로피를 이용한 블라인드 알고리듬의 현실적 접근

김 남 용◆

Practical Approach for Blind Algorithms Using Random-Order Symbol Sequence and Cross-Correntropy

Namyong Kim◆

요 약

상호-코렌트로피 개념은 가우시안 커널의 커널 밀도 추정법에 의해 구축된 두 가지 서로 다른 확률밀도함수의 내적으로 표현될 수 있다. N개의 랜덤 심볼열과 상호-코렌트로피의 최대화 (MCC) 에 바탕을 두고 개발된 블라인드 알고리듬은 탁월한 학습 성능을 보인다. 그러나 이 알고리듬은 MCC 기반으로 가중치를 갱신할 목적으로 설계되면서 과도한 계산 복잡도를 지니게 된다. 이 논문에서는 상호-코렌트로피의 기울기 계산을 반복적으로 수행하여 MCC 알고리듬의 계산상의 복잡도를 크게 줄이는 방법을 제안하였다. 기존의 MCC 알고리듬은 블록 처리 방식에 의해 기울기를 계산하여 $O(N^2)$의 계산량이 필요했던 반면 제안된 방법은 $O(N)$의 계산만을 수행한다. 시뮬레이션 결과로부터, 제안된 이 방법이 기존의 알고리듬과 비교하여 계산량의 부담을 크게 줄이면서도 동일한 학습 성능을 보였다.

Key Words : cross-correntropy, MCC, random symbols, blind, PDF

ABSTRACT

The cross-correntropy concept can be expressed with inner products of two different probability density functions constructed by Gaussian-kernel density estimation methods. Blind algorithms based on the maximization of the cross-correntropy (MCC) and a symbol set of randomly generated N samples yield superior learning performance, but have a huge computational complexity in the update process at the aim of weight adjustment based on the MCC. In this paper, a method of reducing the computational complexity of the MCC algorithm that calculates recursively the gradient of the cross-correntropy is proposed. The proposed method has only $O(N)$ operations per iteration while the conventional MCC algorithms that calculate its gradients by a block processing method has $O(N^2)$. In the simulation results, the proposed method shows the same learning performance while reducing its heavy calculation burden significantly.

Ⅰ. 서 론

적응 신호 처리 및 등화 응용 분야에서 성능평가 기준으로 쓰이는 평균 제곱 오차 (mean squared error, MSE) 가 블라인드 등화기에도 적용되어 널리 사용되어 왔다[1]. 그러나 최근 커널 밀도 추정법 (kernel density estimation) 과 확률 밀도 함수를 바탕으로 한 정보이론적 학습 방법 (information-theoretic

◆ First Author : 강원대학교 전자정보통신공학부, namyong@kangwon.ac.kr, 종신회원
논문번호 : KICS2014-01-006, Received January 12, 2014; Reviewed March 3, 2014; Accepted March 16, 2014

www.dbpia.co.kr

learning, ITL) 이 적응 시스템의 성능 기준으로 채택되며 더욱 향상된 성능을 발휘하고 있다[2,3]. ITL 방법 중의 한 성능기준으로 제안된 코렌트로피 (correntropy)는 랜덤변수 사이의 상관관계 (correlation)를 보다 일반화한 개념으로서 확률 밀도 함수의 고차항까지 포함하여 랜덤변수를 비선형 영역으로 변환 후 변환된 영역에서 상관관계를 다루는 함수이다[4]. 특히 상호-코렌트로피 (cross-correntropy) 함수는 서로 다른 두 랜덤 변수에 대해 다루며 커널 사이즈에 의해 조율되는 변환 영역에서 두 랜덤 변수가 서로 유사할 확률이 어느 정도인가와 관련되어있다[5].

한편, 커널 변환 함수로서 가우시안 커널을 사용한 상호-코렌트로피는 두 확률밀도함수의 내적으로 표현될 수 있으며 이것을 최대화하는 MCC (maximum cross-correntropy) 기준은 두 확률밀도함수가 서로 가장 매칭된 (matched) 형태가 되게 한다. 이 특성을 블라인드 등화기에 응용한 MCC 알고리듬은 전송 심볼점이 가지는 확률밀도함수의 모양과 같도록 수신단에서 인위적으로 무작위 샘플을 발생시켜서 MCC 기준에 적용하였다[6]. 무작위 발생 샘플을 활용하여 송신 심볼의 확률 분포와 등화기 출력 샘플의 확률분포 사이의 유클리드 거리를 최소화 하는 기존의 ITL 타입 블라인드 알고리듬보다 월등히 향상된 성능을 보였다[7].

그러나 이 MCC 알고리듬은 이중 합산 (double summation)과정에 의한 과도한 계산량이 문제점으로 남아있다. 이 과도한 계산량은 현실적 적용 및 구현에 문제가 된다. 따라서 이 논문에서는 이 MCC 알고리듬의 기울기가 가지는 이중 합산 계산에 현재 샘플 시간의 기울기와 다음 샘플 시간의 기울기 사이의 관계를 분석하여 과도한 계산량을 줄일 수 있는 방법을 제안하고자 한다.

Ⅱ. 상호 코렌트로피와 MCC 알고리듬

상호-코렌트로피는 두 개의 서로 다른 스칼라 랜덤변수 X와 Y에 대하여 다음과 같이 정의된다[5].

$$V(X,Y) = E[< \Phi(X), \Phi(Y) >_F] \qquad (1)$$

여기서 E[·]는 통계적 기대치를 나타내고 $<\cdot,\cdot>_F$ 는 무한 차원의 자가생산 커널 힐버트 공간 (an infinite dimensional reproducing kernel Hilbert space) F에서 내적 (inner product) 을 나타내며, 비선형 매핑함수 Φ는 데이터를 힐버트 공간 F로 변환시킨다[4]. 특히, 두 스칼라 랜덤변수 X와 Y가 통계적으로 독립일 때 식 (1)은 다음 식 (2)로 표현될 수 있다.

$$V(X,Y) =< E[\Phi(X)], E[\Phi(Y)] >_F \qquad (2)$$

한편, N개 샘플 $\mathrm{X}=[x_1, x_2, \cdots, x_N]^T$에 대하여 각 샘플 위치마다 가우시안 커널 $\frac{1}{\sigma\sqrt{2\pi}}\exp[\frac{-(x-x_i)^2}{2\sigma^2}]$을 대응시킬 때, N개의 샘플 $\mathrm{X}=[x_1, x_2, \cdots, x_N]^T$이 만드는 확률분포는 다음과 같이 표현될 수 있음을 알 수 있다[8].

$$f_X(x) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sigma\sqrt{2\pi}}\exp[\frac{-(x-x_i)^2}{2\sigma^2}] \qquad (3)$$

이 확률밀도 함수는 비선형 매핑함수 Φ를 가우시안 커널로 볼 때 샘플 평균이 취해진 것으로 볼 수 있다. 따라서 $E[\Phi(X)]$를 $f_x(\xi)$로, $E[\Phi(Y)]$를 $f_y(\xi)$로 대입할 수 있다. 결국, 커널 방법으로 구축된 서로 다른 두 확률밀도함수의 내적은 다음과 같이 표현되며 식 (2)와 일치하는 상호-코렌트로피의 특성을 가진다.

$$\int f_x(\xi)\cdot f_y(\xi)d\xi = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{\sigma\sqrt{2}\sqrt{2\pi}}\exp[\frac{-(x_j-y_i)^2}{4\sigma^2}] \qquad (4)$$

두 확률밀도함수의 내적으로 표현된 이 상호-코렌트로피를 최대화하게 되면, 랜덤 변수가 비선형 가우시안 커널로 매핑되어, 커널 공간에서 두 확률밀도함수의 상관정도가 최대화 된다. 이는 두 확률밀도함수가 서로 가장 매칭된 (matched) 형태가 됨을 의미한다.

이 MCC 개념을 훈련열을 활용하지 못하는 블라인드 등화기 알고리듬에 응용하는 한 기법으로, 전송되는 심볼점이 가지는 확률밀도함수의 정보를 활용하여 이와 같은 모양을 갖도록 수신단에서 인위적으로 무작위 샘플을 발생시켜 사용할 수 있다[6].

대부분의 선형 블라인드 등화기 출력 신호는 입력 $\mathbf{X}_K = [x_k, x_{k-1}, \cdots, x_{k-L}+1]^T$에 대해 $y_k = \mathbf{W}_k^T\mathbf{X}_k$ 형태로 표현된다. M개의 전송 심볼점 A_m이 다음과 같은 동일 확률 $1/M$ 의 독립적 이산 값들을 가진다고 할 때,

$$A_m = 2m - 1 - M\ ,\ \ m = 1,2,\ldots,M \qquad (5)$$

수신단에서는 $\mathbf{D}_N = [d_1, d_2, \ldots, d_N]^T$ 을 랜덤하게 발

www.dbpia.co.kr

생하여 MCC 기반 블라인드 알고리듬에 사용한다. 이 때 랜덤 샘플은 그 확률밀도함수가 송신 심볼점이 가지는 확률 밀도 함수와 같도록 발생된다. 이 확률밀도함수를 $f_d(\xi)$라 하고 등화기 출력샘플의 확률밀도함수를 $f_y(\xi)$라 할 때, 식 (4)에 해당하는 상호-코렌트로피는 식 (3)을 활용하여 다음과 같이 표현된다.

$$\int f_d(\xi) f_y(\xi) d\xi = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{\sigma\sqrt{2}\sqrt{2\pi}} \exp[\frac{-(d_j - y_i)^2}{4\sigma^2}] \tag{6}$$

식 (6)을 최대화하기 위하여 그 기울기를 구하면

$$\frac{\partial \int f_d(\xi) f_y(\xi) d\xi}{\partial \mathbf{W}} = \frac{1}{2N^2\sigma^3\sqrt{\pi}} \sum_{i=1,}^{N} \sum_{j=1}^{N} (d_j - y_i) \cdot \exp[\frac{-(d_j - y_i)^2}{4\sigma^2}] \cdot \mathbf{X}_i \tag{7}$$

샘플 기반 on-line 시스템에서는 사이즈 N의 버퍼에 sliding window 기법으로 매 샘플 시각마다 범위 밖 샘플을 버리고 새 샘플을 받아들인다[7]. 즉,

$$\frac{\partial \int f_d(\xi) f_y(\xi) d\xi}{\partial \mathbf{W}} = \frac{1}{2N^2\sigma^3\sqrt{\pi}} \sum_{i=k-N+1,}^{k} \sum_{j=1}^{N} (d_j - y_i) \cdot \exp[\frac{-(d_j - y_i)^2}{4\sigma^2}] \cdot \mathbf{X}_i \tag{8}$$

기울기 (8)에 안정도를 위한 파라메터 μ를 도입하고 steepest ascent 방법으로 최대를 찾아가도록 등화기 가중치를 갱신하는 식은 다음과 같이 구해진다.

$$\mathbf{W}_{k+1} = \mathbf{W}_k + \frac{\mu}{2N^2\sigma^3\sqrt{\pi}} \sum_{i=k-N+1,}^{k} \sum_{j=1}^{N} (d_j - y_i) \cdot \exp[\frac{-(d_j - y_i)^2}{4\sigma^2}] \cdot \mathbf{X}_i \tag{9}$$

자가 발생 랜덤 샘플 d_j를 활용하는 이 알고리듬을 논문 [6]에서와 같이 편의상 이 논문에서도 MCC 알고리듬이라 부르기로 한다.

Ⅲ. 제안된 기울기 계산방식의 MCC 알고리듬

MCC 알고리듬의 문제점은 많은 계산량을 들 수 있다. 현실적 구현에 장애가 되는 이 계산량을 줄일 수 있는 방법 중 하나로 이 장에서는 MCC 알고리듬의 기울기 (8) 이 가지는 이중 합산 (double summation) 계산에 현재 샘플 시간의 기울기와 다음 샘플 시간의 기울기 사이의 관계를 분석하고 다음 기울기 계산에 현재 계산된 기울기를 활용할 수 있는 방안을 제안한다.

블록 데이터 처리를 위한 버퍼에 데이터가 다 차지 않은 상태를 초기상태(initial) 라 하고 다 찬 상태를 안정상태 (steady) 라고 정의할 때, 등화기 의 기울기 계산 상태를 초기상태와 안정상태를 나누어 반복적 기울기 계산방식을 제안하고자 한다. $1 \le k \le N$에서는 블록 데이터 처리를 위한 버퍼 속에 k 개의 샘플들이 자리하고 있고 현재 출력 샘플 y_k와 함께 사용되어 현재 시각 k에서 기울기 $\nabla_k|_{initail} = \left.\frac{\partial \int f_d(\xi) f_y(\xi) d\xi}{\partial \mathbf{W}}\right|_k$는 다음과 같이 표현될 수 있다. 이 때, 최초 기울기 $\nabla_0|_{initail} = \nabla_1|_{initail} = 0$ 으로 한다.

$$\nabla_k|_{initail} = \frac{1}{2Nk\sigma^3\sqrt{\pi}} \sum_{i=1,}^{k} \sum_{j=1}^{N} (d_j - y_i) \cdot \exp[\frac{-(d_j - y_i)^2}{4\sigma^2}] \cdot \mathbf{X}_i \tag{10}$$

그리고 등화기 블록 데이터를 위한 버퍼가 다 채워진 정상상태, 즉 $k \ge N+1$에서는 버퍼 속에 항상 N 개의 샘플들이 있고 기울기는 다음과 같다.

$$\nabla_k|_{steady} = \frac{1}{2N^2\sigma^3\sqrt{\pi}} \sum_{i=k-N+1,}^{k} \sum_{j=1}^{N} (d_j - y_i) \cdot \exp[\frac{-(d_j - y_i)^2}{4\sigma^2}] \cdot \mathbf{X}_i \tag{11}$$

먼저, 초기상태에 대하여, 다음 샘플 시각 k+1 에서 기울기 $\nabla_{k+1}|_{initail}$ 은 다음과 같게 된다.

$$\nabla_{k+1}|_{initail} = \frac{1}{2N(k+1)\sigma^3\sqrt{\pi}} \sum_{i=1,}^{k+1} \sum_{j=1}^{N} (d_j - y_i) \cdot \exp[\frac{-(d_j - y_i)^2}{4\sigma^2}] \cdot \mathbf{X}_i = \frac{k}{2Nk(k+1)\sigma^3\sqrt{\pi}} \left\{ \sum_{i=1,}^{k} \sum_{j=1}^{N} (d_j - y_i) \right.$$

www.dbpia.co.kr

$$\cdot \exp[\frac{-(d_j - y_i)^2}{4\sigma^2}] \cdot \mathbf{X}_i + \sum_{j=1}^{N} (d_j - y_{k+1}) \cdot \exp[\frac{-(d_j - y_{k+1})^2}{4\sigma^2}] \cdot \mathbf{X}_{k+1} \Bigg\} \quad (12)$$

여기서 $\sum_{i=1,}^{k} \sum_{j=1}^{N} (d_j - y_i) \cdot \exp[\frac{-(d_j - y_i)^2}{4\sigma^2}] \cdot \mathbf{X}_i$ 는 $2Nk\sigma^3\sqrt{\pi} \cdot \nabla_k|_{initail}$ 과 같으므로

$$\nabla_{k+1}|_{initail} = \frac{k}{k+1} \nabla_k|_{initail} + \frac{1}{2N(k+1)\sigma^3\sqrt{\pi}} \sum_{j=1}^{N} (d_j - y_{k+1}) \cdot \exp[\frac{-(d_j - y_{k+1})^2}{4\sigma^2}] \cdot \mathbf{X}_{k+1} \quad (13)$$

안정상태의 기울기 반복 계산은 다음과 같다. 다음 샘플 시각 k+1 에서 기울기 $\nabla_{k+1}|_{steady}$ 는 다음과 같게 된다.

$$\nabla_{k+1}|_{steady} = \frac{1}{2N^2\sigma^3\sqrt{\pi}} \Bigg\{ \sum_{i=k-N+1,}^{k} \sum_{j=1}^{N} (d_j - y_i) \cdot \exp[\frac{-(d_j - y_i)^2}{4\sigma^2}] \cdot \mathbf{X}_i + \sum_{j=1}^{N} (d_j - y_{k+1}) \cdot \exp[\frac{-(d_j - y_{k+1})^2}{4\sigma^2}] \cdot \mathbf{X}_{k+1} - \sum_{j=1}^{N} (d_j - y_{k-N+2}) \cdot \exp[\frac{-(d_j - y_{k-N+2})^2}{4\sigma^2}] \cdot \mathbf{X}_{k-N+2} \Bigg\} \quad (14)$$

여기서 $\sum_{i=k-N+1,}^{k} \sum_{j=1}^{N} (d_j - y_i) \exp[\frac{-(d_j - y_i)^2}{4\sigma^2}] \cdot \mathbf{X}_i$ 는 $2N^2\sigma^3\sqrt{\pi} \cdot \nabla_k|_{steady}$ 와 같으므로, 식 (14)의 $\nabla_{k+1}|_{steady}$ 는 다음과 같이 정리된다.

$$\nabla_{k+1}|_{steady} = \nabla_k|_{steady} + \frac{1}{2N^2\sigma^3\sqrt{\pi}} \sum_{j=1}^{N} (d_j - y_{k+1}) \cdot \exp[\frac{-(d_j - y_{k+1})^2}{4\sigma^2}] \cdot \mathbf{X}_{k+1} - \frac{1}{2N^2\sigma^3\sqrt{\pi}} \sum_{j=1}^{N} (d_j - y_{k-N+2}) \cdot \exp[\frac{-(d_j - y_{k-N+2})^2}{4\sigma^2}] \cdot \mathbf{X}_{k-N+2} \quad (15)$$

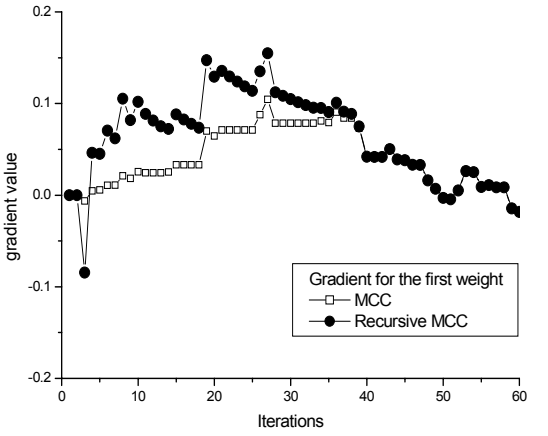

그림 1. 데이터 블록 $N=40$ 에서 첫 가중치의 기울기 학습 곡선.
Fig. 1. Gradient learning curves for the first weight with the data block size $N=40$.

식 (8)의 MCC 방식은 N 개의 블록 샘플들을 저장할 수 있는 버퍼에 매 샘플시간 마다 들어오는 샘플과 나가는 샘플의 슬라이딩 윈도우 방식으로 처리된다. 여기서 이중 합산 계산에 의해 계산량은 $O(N^2)$이 되는데 비하여, 식 (13) 과 (15) 로 요약되는 반복적 기울기 계산 방식 (이 논문에서는 recursive MCC로 편의상 정의한다) 는 원래의 기울기 (8)과 일치하면서도 계산량이 $O(N)$임을 알 수 있다.

Ⅳ. 시뮬레이션 결과 및 토론

MCC 알고리듬과 제안된 반복적 기울기 연산 방법 recursive MCC가 동일한 기울기 결과를 가지는지 살펴보기 위해, 논문 [6]의 시뮬레이션에 쓰였던 동일한 상황하에서 여러 가중치들의 기울기 학습 곡선을 비교 분석하였다. 그림 1은 등화기 의 가중치 w_0 갱신에 쓰이는 기울기에 대해 MCC 방식과 recursive MCC 방식을 비교하였다. 초기 상태에 해당하는 샘플 시간 구간 $1 \le k \le N$, 즉, N=40 샘플까지 MCC 방식과 recursive MCC 방식은 서로 다른 기울기 계산 값을 보이고 있다. 그러나 버퍼가 다 채워진 정상상태인 $k \ge N+1$ 즉, 샘플 시간 41 부터 MCC 방식과 recursive MCC 방식은 완전히 동일한 기울기 계산 값을 보이고 있다. 이 결과를 확증하기 위해 모든 가중치에 대하여 실험 관찰하는 대신 두 번째와 세 번째 가중치, 그리고 중앙 가중치의 기울기 계산 값의 학습

www.dbpia.co.kr

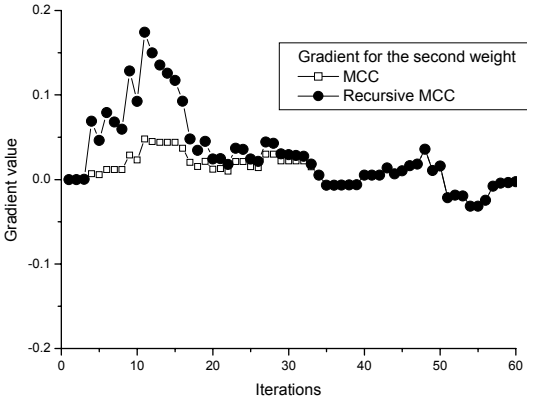

그림 2. 데이터 블록 $N=40$에서 두 번째 가중치의 기울기 학습 곡선.
Fig. 2. Gradient learning curves for the second weight with the data block size $N=40$

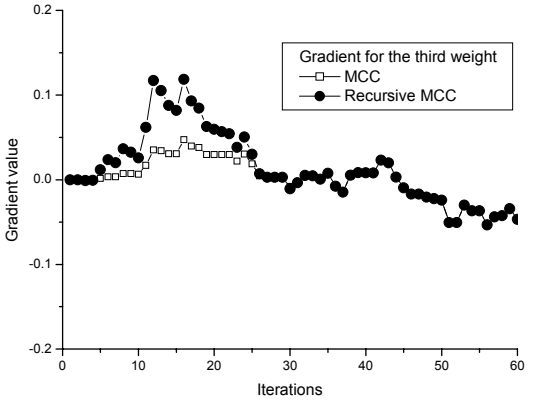

그림 3. 데이터 블록 $N=40$에서 세 번째 가중치의 기울기 학습 곡선.
Fig. 3. Gradient learning curves for the third weight with the data block size $N=40$

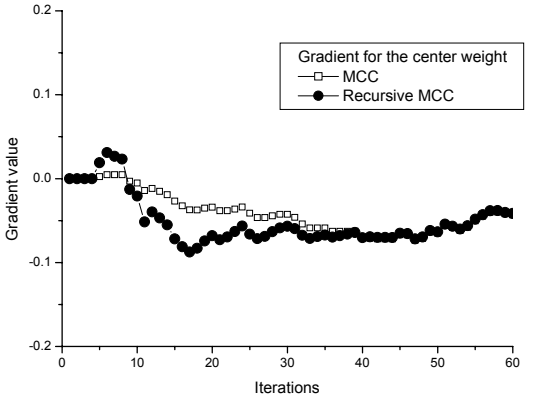

그림 4. 데이터 블록 $N=40$에서 가운데 가중치의 기울기 학습 곡선.
Fig. 4. Gradient learning curves for the center weight with the data block size $N=40$

곡선을 그림 2 부터 그림 4 까지 나타냈다. 예상대로 정상상태인 샘플시간 40 부터 ($k \geq N+1$) MCC 방식과 recursive MCC 방식은 기울기 계산 값이 모든 가중치에 대해 일치함을 알 수 있다.

V. 결 론

확률밀도함수를 바탕으로 한 ITL 타입 성능기준은 등화기 나 적응 신호처리 분야에서 기존의 MSE 기준에 의해 개발된 많은 알고리듬 보다 많은 우수한 특성과 성능을 가진다. 한 블록의 샘플 데이터에 대해 커널 밀도 추정법으로 구해진 확률밀도 함수는 비선형 매핑 함수를 가우시안 커널로 하여 자가생산 커널 힐버트 공간에 변환된 상태로 볼 수 있으며 또한 서로 다른 두 확률밀도의 내적은 상호-코렌트로피의 특성을 가짐을 알 수 있었다. 두 확률밀도함수의 내적으로 표현된 이 상호-코렌트로피를 최대화하도록 설계된 MCC 알고리듬은 블라인드 등화 에서 탁월한 성능을 보이나 과중한 계산량의 부담이 알고리듬의 현실적 적용 및 응용에 하나의 문제점이었다. 이 논문에서는 확률분포 상관함수 알고리듬의 계산량을 줄이는 방법을 제안하였다. 기존의 MCC 방식은 매 샘플시간마다 N 개의 블록 샘플들을 처리하면서 계산량은 $O(N^2)$이 되는데 비하여, 이 논문에서 제안한 recursive MCC 방식은 $O(N)$의 계산량을 가진다. 실험에서 recursive MCC 의 가중치 기울기 값은 처음 블록 데이터 버퍼가 다 채워진 이후부터 원래의 MCC 방식과 완전히 일치하는 기울기 학습곡선을 나타냈다. 따라서 이 논문에서 제안한 recursive MCC 알고리듬이 통신 등화 및 블라인드 신호처리 응용 분야의 현실적 구현에 큰 장점을 가진다고 할 수 있다.

References

[1] J. Lim and Y. Pyeon, “A variable step sized constant modulus algorithm based on an averaged gradient,” *J. KICS*, vol. 35, no. 4, pp. 79-82, Apr. 2010.

[2] N. Kim, “A study on the complex-channel blind equalization using ITL algorithms,” *J. KICS,* vol. 35, no. 8, pp. 760-767, Aug. 2010.

[3] J. C. Principe, D. Xu and J. Fisher, *Information Theoretic Learning : Unsupervised adaptive filtering*, Wiley, NewYork, USA, pp. 265-319, 2000.

www.dbpia.co.kr

[4] I. Santamaria, P. P. Pokharel, and J. C. Principe, "Generalized correlation function: Definition, properties, and application to blind equalization," *IEEE Trans. Signal Processing*, vol. 54, no. 6, pp. 2187-2198, Jan. 2006.

[5] Weifeng Liu, P. P. Pokharel, and J. C. Principe, "Correntropy: properties and applications in non-Gaussian signal processing," *IEEE Trans. Signal Processing*, vol. 55, no. 11, pp. 5286-5298, Nov. 2007.

[6] N. Kim, S. Kang, and D. Hong, "Blind equalization based on maximum cross-correntropy criterion using a set of randomly generated symbols," *J. KICS*, vol. 35, no. 1, pp. 33-39, Jan. 2010.

[7] N. Kim and L. Yang, "A new criterion of information theoretic optimization and application to blind channel equalization," *J. Korean Soc. Internet Inf.*, vol. 10, no. 1, pp. 11-17, Feb. 2009.

[8] 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 8월 : 연세대학교 전자공학과 박사

1992년 8월~1998년 2월 : 관동대학교 전자통신공학과 부교수

1998년 3월~현재 : 강원대학교 공학대학 전자정보통신공학부 교수

<관심분야> Adaptive equalization algorithms, ITL algorithms, Odor-sensing signal processing.

www.dbpia.co.kr