1998년도 대한전기학회 정기총회 및 추계학술대회 논문집 1998. 11. 28

자율 데이터 Sphering: 회귀 신경망

최 승진, 류 영기
충북대학교, 전기공학과

Unsupervised Data Sphering: Recurrent Neural Network

Seungjin CHOI, Youngki LYU
Department of Electrical Engineering, Chungbuk National University

Abstract - 데이터 sphering은 신호처리 및 신경망 분야에서 널리 이용되는 기본적인 처리과정으로, 그 목적은 주어진 데이터간에 correlation을 제거하는 것이다. 본 논문에서는 회귀 신경망을 도입하여, 데이터 sphering을 위한 새로운 on-line 알고리듬을 제시한다. 정보 이론에 입각한 risk 함수와, 최근에 제시된 natural gradient을 이용하여 새로운 데이터 sphering 알고리듬을 유도한다. 새로 제시된 알고리듬의 성능을 기존에 널리 이용되어 온 anti-Hebbian 학습 알고리듬과 비교 분석한다.

1. 서 론

데이터 sphering (whitening, decorrelation)은 매우 기본적인 데이터 처리 과정으로, 그 목적은 주어진 데이터간에 correlation을 제거하는 것이다. 신호처리 및 신경망 분야에서, 데이터 sphering은 널리 이용되어 왔는데, 대표적인 방법으로는 principal component analysis (PCA) [1,2], anti-Hebbian rule [3], Almeida-Silva 알고리듬 [4]등이 있다. 본 논문에서는 아주 간단한 local 알고리듬인 anti-Hebbian rule을 먼저 간단히 고찰을 하고, natural gradient [5]를 이용한 보다 성능이 우수한 데이터 sphering 알고리듬을 제시한다.

주어진 n 차원의 벡터를 $\boldsymbol{x}=[x_1,\ldots,x_n]$라 하자. 데이터 sphering의 목적은 벡터 $\boldsymbol{x}$를 다른 n 차원의 벡터 $\boldsymbol{y}$로 선형 변환하여 벡터 $\boldsymbol{y}$의 성분들이 모두 서로 상관관계가 없게 만드는 것이다. 이러한 목적에 맞는 선형 변환을 설계하여야 하는데, 선형 변환은 공간 필터 또는 신경망으로써 표현이 된다.

2. 본 론

본론에서는 먼저 데이터 sphering을 위한 대표적 알고리듬인 anti-Hebbian rule를 간단히 살펴보고, 본 논문의 핵심인 새로운 알고리듬을 유도한다.

2.1 Anti-Hebbian rule

Hebbian 학습 알고리듬은 대표적으로 널리 쓰이는 자율 학습 알고리듬 (unsupervised learning algorithm)으로써, 많은 인공 신경망 분야에서 널리 쓰여왔다. 대표적인 예로는 Oja's rule [6]이 있는데, Oja는 normalized Hebbian rule을 선형 근사화하여, 입력 데이터의 공분산 행렬의 가장 큰 eigenvalue에 해당하는 eigenvector를 찾아내는 알고리듬을 제시하였다. Oja의 single neuron network은 그 이후 여러 사람에 의해 확장되어 여러 개의 principal component를 추출하는 신경망이 제시되어 왔다.(참고문헌 [2] 참조).

Foldiak은 여러 principal component 추출을 위하여, 출력 노드의 값들이 서로 상관 관계가 없어야 한다는 점에 착안 anti-Hebbian rule을 도입하였다. Anti-Hebbian rule은 Hebbian rule과는 반대로 신경망의 시냅스가 데이터들간의 상관 관계가 줄어드는 방향으로 증가함으로써, 평형 상태에 도달하였을 때 신경망의 출력들은 서로 상관관계가 없게된다.

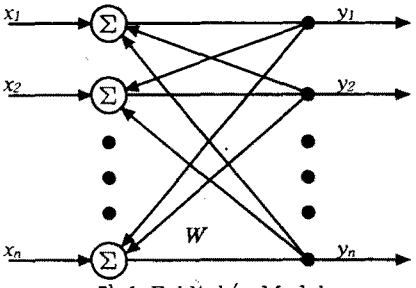

그림 1 Foldiak's Model

그림 1은 Foldiak이 제시하였던 network으로 self-feedback 연결은 고려되지 않았었다. 즉 Foldiak의 모델에서 입출력 관계는 다음과 같이 주어진다.

$$y_i = x_i + \sum_{k \neq i}^{n} w_{ik} y_k \tag{1}$$

주어진 입력 데이터의 sphering을 위하여 Foldi마이 제시하였던 anti-Hebbian rule은 다음과 같다.

$$\Delta w_{ij} = -\eta y_i y_j \tag{2}$$

2.2 Modified Anti-Hebbian rule

Anti-Hebbian rule의 stationary point는 $E\{y_i y_j\}=0$를 (여기서 $E\{\cdot\}$는 statistical expectation operator를 나타냄) 만족함으로 평형 상태에서 출력 $\{y_i\}$는 서로 상관관계가 없음을 쉽게 알 수 있다. 그러나 Anti-Hebbian rule은 출력 노드간에 decorrelation만을 하기 때문에, 각 노드의 분산이 작아짐을 제어할 수 없다. 그래서 출력 y_i의 분산을 1로 하기 위하여서는 self-feedback 연결 w_{ii}를 도입, anti-Hebbian rule에 다음을 더 추가할 수 있다.

$$\Delta w_{ii} = \eta(1 - y_i^2) \tag{3}$$

2.3 Natural Gradient-based decorrelation algorithm

확률론적 dependency를 줄이기 위해 일반적인 비선형함수로 알고리듬을 유도하지만, 뒤에서 Gaussian model인 경우 데이터 sphering 이나 decorrelation에 선형 함수로 쓰일 수 있음을 보이겠다.

정보이론의 상호정보량 의해 risk 함수 $R(\boldsymbol{W})$는 다

음과 같이 주어진다.

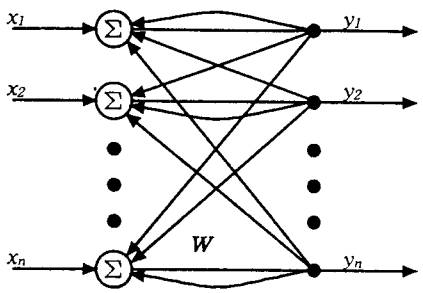

그림 2 Fully-connected Network

$$\begin{aligned} R(W) &= L(W) \\ &= \int p(y)\log\frac{p(y)}{q_1(y_1)\cdots q_n(y_n)}dy \\ &= E\left\{\log\frac{p(y)}{q_1(y_1)\cdots q_n(y_n)}\right\} \\ &= -H(y)+\sum_{i=1}^{n}H(y_i), \end{aligned} \tag{4}$$

여기서 $H(\cdot)$는 entropy를 나타내고, 다음과 같이 정의 된다.

$$\begin{aligned} H(y) &= -\int y\log p(y)dy, \\ H(y_i) &= -\int y_i\log q_i(y_i)dy_i. \end{aligned} \tag{5}$$

이것은 다음과 같이 바꾸어 표현할 수 있다.

$$H(y)=H(x)+\log|\det W|, \tag{6}$$

여기서 det 는 행렬의 행렬식이다.

그림 2의 선형 회귀 신경망의 출력 y는 다음과 같이 나타내어 진다.

$$\begin{aligned} y &= x+Wy \\ &= [I-W]^{-1}x. \end{aligned} \tag{7}$$

출력 노드 y_i들 간의 상호정보량을 최소화 하기 위한 loss 함수은 다음과 같이 주어진다.

$$\begin{aligned} L(W) = &-\sum_{i=1}^{n}\log q_i(y_i) \\ &-\log\det|(I-W)^{-1}|. \end{aligned} \tag{8}$$

비선형 함수 $f_i(y_i)=-\dfrac{d\log q_i(y_i)}{dy_i}$ 를 사용하여 다음의 식을 얻을 수 있다.

$$\begin{aligned} &d\left\{-\sum_{i=1}^{n}\log q_i(y_i)\right\} \\ &=f^T(y)[I-W]^{-1}dWy. \end{aligned} \tag{9}$$

수정된 미분계수 dV를 다음과 같이 정의한다.

$$dV=[I-W]^{-1}dW. \tag{10}$$

이 정의로 식 (9)는 다음과 같은 식으로 바뀐다.

$$d\left\{-\sum_{i=1}^{n}\log q_i(y_i)\right\}=f^T(y)dVy. \tag{11}$$

동시에, 다음 식의 성립을 보일 수 있다.

$$d\{\log\det|(I-W)^{-1}|\} = \mathrm{Tr}\{dV\}. \tag{12}$$

그러므로 식 (11)과 식 (12)를 조합하여 다음의 식을 얻는다.

$$dL(W)=f^T(y)dVy-\mathrm{Tr}\{dV\}. \tag{13}$$

식 (13)의 미분은 수정된 계수 dV에 관하여 표현된다. dV_{ij}는 dW_{ij}의 선형 결합에 의해 형성되어 있기 때문에, dV는 식 (8)을 최소화 하는 올바른 탐색방향을 나타낸다. 이것은 확률론적인 새로운 알고리듬을 이끌어 낸다.

$$\begin{aligned} \triangle V(t) &= V(t+1)-V(t) \\ &= -\eta(t)\frac{dL(W(t))}{dV(t)} \\ &= \eta t\{I-f(y(t))\,y^T(t)\}. \end{aligned} \tag{14}$$

그러므로, $W(t)$에 관한 학습 알고리즘은 다음과 같이 주어진다.

$$\begin{aligned} \triangle W(t) &= W(t+1)-W(t) \\ &= [I-W(t)]\triangle V(t) \\ &= \eta_t[I-W(t)]\{I-f(y(t))\,y^T(t)\}. \end{aligned} \tag{15}$$

Marginal 확률 밀도 함수 $q_i(y_i)$가 Gaussian 밀도 함수일 경우 즉, $q_i(y_i)=\dfrac{1}{\sqrt{2\pi}}e^{\frac{1}{2}y_i^2}$일 때 $f(y(t))$는 선형 함수 $y(t)$로 데이터 sphering 선형 학습 알고리듬에 이용될 수 있다. 비선형 함수 $f(y(t))$는 Independent Component Analysis(ICA) [6]에 적용될 수 있다.

3. 모의 실험

랜덤하게 발생된 세 개의 원신호가 선형 변환되어 얻어진 벡터 x의 공분산 행렬 Rx 는 다음과 같다.

$$Rx=\begin{bmatrix} 0.5025 & 0.4982 & 0.4987 \\ 0.4982 & 0.5018 & 0.4979 \\ 0.4987 & 0.4979 & 0.4974 \end{bmatrix}$$

우리는 세 개의 알고리듬을 비교한다. 그림 3는 Anti-Hebbian rule(AH) 의 weight matrix 의 궤도이고, 그림 4은 Modified Anti-Hebbian rule(MAH) 의 weight matrix의 궤도이고, 그림 5는 이 논문에서 제안한 새로운 알고리듬(NA)의 궤도이다.

원신호를 찾기위해, 그림 2의 회귀 network으로 실험하였다. 상수 학습률은 여러번의 모의 실험을 통해 적합한 값으로 $\eta_{AH}=0.004$, $\eta_{MAH}=0.001$, $\eta_{NA}=0.007$로 설정하였다.

Performance Index 로 y_1과 y_2의 값을 windowing 하면서 cross-correlation 값을 표시한 그림이 그림 6에 나타나 있다.

$$Ry_{AH}=\begin{bmatrix} 0.2437 & 0.0014 & -0.0014 \\ 0.0014 & 0.2604 & -0.0002 \\ -0.0014 & -0.0002 & 0.1008 \end{bmatrix}$$

$$Ry_{MAH}=\begin{bmatrix} 1.0053 & 0.0061 & -0.0025 \\ 0.0061 & 1.0079 & 0.0053 \\ -0.0025 & 0.0053 & 1.0021 \end{bmatrix}$$

$$Ry_{NA}=\begin{bmatrix} 1.0160 & 0.0085 & -0.0087 \\ 0.0085 & 1.0219 & -0.0015 \\ -0.0087 & -0.0015 & 1.0153 \end{bmatrix}$$

위의 행렬은 세 개의 알고리듬이 수렴한 이후의 공분산 행렬을 나타내고 있다. 이 행렬에서 알 수 있듯이 Anti-Hebbian rule의 공분산 행렬의 대각성분이 작게 나옴을 알 수 있다.

4. 결 론

본 논문에서는 정보이론에 입각하여 데이터 sphering을 위한 risk 함수를 제시하였고, 최근에 Amari [5]에 의해 제시된 natural gradient를 도입, 새로운 데이터 sphering 알고리듬을 유도하였다. 제시된 데이터 sphering 알고리듬은 fully-connected 회귀 신경망을 학습하여, 기존의 anti-Hebbian rule보다 우수성을 모의 실험을 통하여 입증하였다.

[참 고 문 헌]

[1] E. Oja, "Neural networks, principal component analysis and subspaces," International Journal of Neural Systems, vol. 1, pp. 61-68, 1989.
[2] K. I. Diamantaras and S. Y. Kung, Principal Component Neural Networks: Theory and Applications, 1996.
[3] P. Foldiak, "Adaptive network for optimal linear feature extraction," In International Joint Conference on Neural Networks, pp. 401-405, 1989.
[4] L. B. Almeida and F. M. Silva, "Adaptive decorrelation," In Artificial Neural Networks, pp. 149-156, 1992.
[5] S. Amari, "Natural gradient works efficiently in learning," Neural Computation, vol. 10, pp. 251-276, 1998.
[6] S. Choi, "Neural learning algorithms for independent component analysis," Journal of IEEE Korea Council, vol. 2, no. 1, pp. 24-33, 1998.

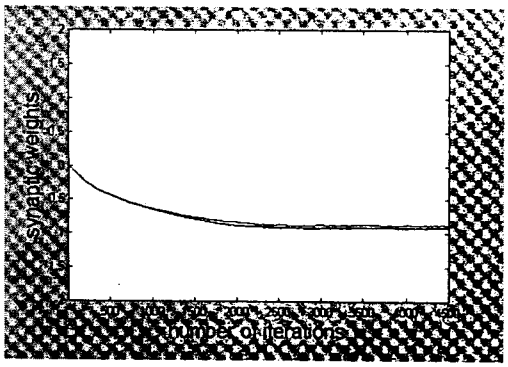

그림 3 Anti-Hebbian rule

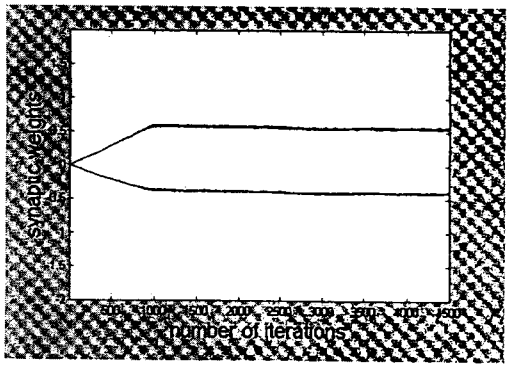

그림 4 Modified Anti-Hebbian rule

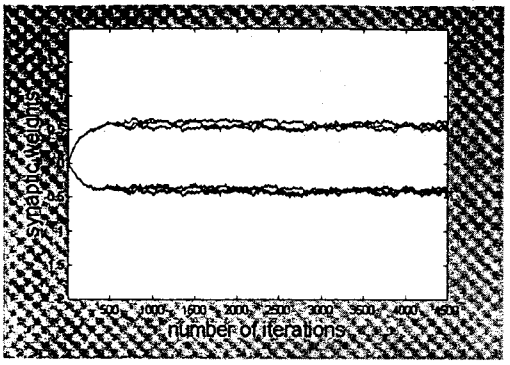

그림 5 New algorithm using Natural Gradient

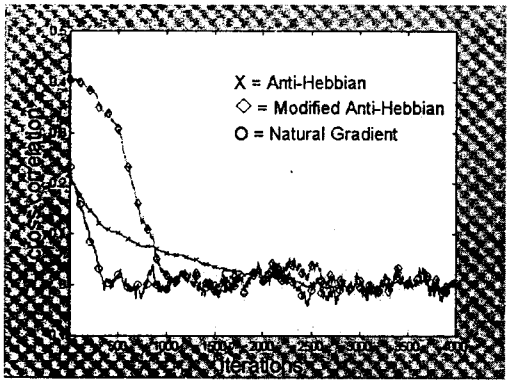

그림 6 Performance Index