

Kohonen 학습의 입력에 잡음 주입의 효과

정혁준, 송근배, 이행세

아주대학교 전자공학과

전화 : 031-219-2376 / 핸드폰 : 019-209-9152

The Effect of Noise Injection into Inputs in the Kohonen Learning

Hyuck-jun Jung, Geunbae Song, Haingsei Lee

School of Electronics Engineering, Ajou University

E-mail : gold2@chollian.net

Abstract

This paper proposes the strategy of noise injection into inputs in the Kohonen learning algorithm (KLA) to improve the local convergence problem of the KLA. Noise strengths are high in the begin of the learning and gradually lowered as the learning proceeds. This strategy is a kind of stochastic relaxation (SR) method which is broadly used in the general optimization problems. It is convenient to implement and improves the convergence properties of the KLA with moderately increased computing time compared to the KLA. Experimental results for Gauss-Markov sources and real speech demonstrate that the proposed method can consistently provide better codebooks than the KLA.

1. 서론

Kohonen 학습법 (KLA)은 신경망 학습을 위해 Kohonen에 의해서 제안된 일종의 자율학습 클러스터링 알고리즘이다. KLA는 신호처리의 관점에서 보면, nonconvex 형태의 목적함수에 적용된 least mean squares (LMS) 알고리즘으로 볼 수 있다. 따라서 KLA는 강하법 (descent algorithm) 계열의 알고리즘들이 본질적으로 안고 있는 극소점(local minimum)으로 수렴하는 문제를 가지고 있다.

극소점 수렴 문제는 simulated annealing (SA)방법 혹은 보다 일반적인 개념으로서 stochastic relaxation (SR)방법에 의해 해결될 수 있다. 이러한 방법들의 공통점은 시스템 파라미터에 잡음을 부가하여 잡음의 강도(즉 분산)를 학습의 초기에는 큰 값에서 학습이 진행됨에 따라 서서히 줄여나간다. 이렇게 함으로써 학습은

랜덤탐색(random search)에서 점차로 직접탐색(direct search)으로 바뀌어 가며 따라서 학습의 초기에 극소점에 갇히게 되는 현상을 줄일 수 있다. 본 논문은 KLA의 극소점 문제를 개선하기 위하여 SR 개념을 도입하였다. 이 방법은 구현하기 쉬우며 KLA의 특성을 보존하고 있다. 본 논문은 가우스-마코프 소스와 음성소스에 대한 벡터 양자화 실험을 통하여 제안된 방법이 기존의 KLA 혹은 일괄처리 방식의 GLA 보다 일관되게 우수한 코드북을 생성함을 실험적으로 증명해 보인다.

2. KLA, SR의 기본 개념

2.1 KLA의 기본개념

KLA는 신경망의 자기 조직화 학습을 위해 제안되었다. 보통 self-organizing map(SOM) 알고리즘이라 하면 복수 승자 방식의 KLA를 일컫는다. 즉 SOM은 승자 출력 뉴런 뿐 아니라 그 뉴런으로부터 공간적으로 일정 거리에 드는 이웃뉴런에게까지 학습의 기회를 준다. 이렇게 함으로써 학습과정을 통해 인접한 뉴런들이 서로 닮게 되는 결과를 가져온다. 이는 고차원 입력벡터의 복잡하고 비선형적인 관계를 저차원의 단순하고 기하학적인 관계로 보존하여 표현하여 줌으로써 고차원 데이터를 시각화(visualization)하고 정보를 함축(abstract)하는 효과를 가져올 수 있다.

SOM은 또한 벡터 양자화기 설계를 위해서도 사용되었다. <그림 1>은 벡터 양자화기로 이용될 수 있는 단층 신경망 구조를 나타낸 것이다. 그림에서 가중치벡터 $W_i = [w_{i1}, \dots, w_{iM}]$ 는 M 개의 코드벡터 중 i 번째 코드벡터에, 뉴런 출력 y_1, \dots, y_M 은 코드벡터 인덱스에

대응된다. 벡터 양자화의 관점에서 보면 복수승자 방식은 하나의 휴리스틱이 될 수 있다. 즉, 일반적으로 입력 벡터는 벡터 공간에서 특정 지역에 편중되어 분포되어 있다. 따라서 균일 분포의 랜덤 값으로 코드벡터들을 초기화시키는 것은 그리 좋은 방법이 아니다. 학습 초기에 승자뿐만 아니라 넓은 범위의 이웃 뉴런들에게까지 학습기회를 주는 것은 초기 코드 벡터들을 입력 벡터의 분포 지역으로 이동시켜줄 것이다. 그 뒤 학습이 진행됨에 따라 이웃뉴런의 범위를 줄여나가면 각 뉴런에 의해 보다 세밀한 클러스터링을 할 수 있을 것이다. 그러나 이러한 전략의 효과는 불분명하며 일관되지 않은 결과를 가져오는 것으로 알려져 있다. 이 문제와 관련한 다른 휴리스틱들은 등을 참조할 수 있다. 본 논문에서는 문제를 간단히 하기 위해 단일 승자 방식의 SOM 알고리즘에 대해서만 논의하기로 하며 이를 간단히 'KLA'라 부르기로 한다.

본 논문에서 KLA는 벡터 양자화 문제에 사용된다. Euclid 거리측도 d 의 관점에서, t 시점의 랜덤벡터 $x(t)$ 에 가장 가까운 가중치벡터(즉, 코드벡터)를 $W_c(t)$ 라 하자. 여기서 c 는 승자뉴런을 가리키며 (1)에 의해 결정된다.

$$c = \arg \min_i d(x(t), W_i(t)) \quad (1)$$

양자화기 설계의 궁극 목적은 (2)로 정의되는 평균양자화오차를 최소로 하는 코드북 $W = \{W_i\}_{i=1}^M$ 을 탐색하는 것이다.

$$D(t) \equiv E[d(x(t), W_c(t))^2] \quad (2)$$

여기서 $E[\cdot]$ 는 평균연산을 나타낸다. 만약 랜덤벡터 $x(t)$ 가 stationary하다고 하면 시간변수 t 는 생략될 수 있다. (2)의 $E[\cdot]$ 를 계산하기 위해서는 $x(t)$ 의 확률밀도함수 $p(x(t))$ 를 알아야 한다. 일반적으로 이는 사전에 알려져 있지 않다. 그리고 설명 알려져 있다하더라도 복잡한 경우가 대부분 임으로 (2)에 직접 gradient descent 방법을 적용하기는 어렵고, (2)의 최적 해를 달린 형태로 풀어내기는 더욱 어렵다. 따라서 현실적으로 반복적 근사법으로 준최적의 해를 탐색하게된다. KLA는 stochastic gradient descent (SGD)라 불리는 반복적 근사법이다. 이 경우에 기준이 되는 오차함수는 (2)의 형태가 아닌 (3)의 형태의 순시 오차함수이다.

$$\hat{D}(t) \equiv d(x(t), W_c(t))^2 = \|x(t) - W_c(t)\|^2 \quad (3)$$

여기서 $\hat{D}(t)$ 는 t 시점에서의 순시 오차를 나타낸다. (3)을 최소화하기 위한 근사적 최적화 알고리즘은 다음

과 같다.

$$W_c(t+1) = W_c(t) - (1/2)\alpha(t) \frac{\partial \hat{D}(t)}{\partial W_c(t)} \quad (4)$$

여기서 $\alpha(t)$ 는 학습률을 나타내는 0과 1 사이의 작은 스칼라 함수이다. (4)의 오른쪽의 편미분을 풀면 (5), (6)과 같은 잘 알려진 KLA를 얻게된다.

$$\begin{aligned} W_c(t+1) &= W_c(t) + \Delta W_c(t) \\ &= W_c(t) + \alpha(t)[x(t) - W_c(t)], \quad (5) \\ (0 < \alpha(t) < 1) \end{aligned}$$

$$W_i(t+1) = W_i(t), \quad \text{그 밖의 } i \neq c. \quad (6)$$

(5)와 (6)의 학습법을 관찰하여 보면, 승자뉴런의 가중치벡터는 입력벡터 쪽으로 가까이 다가가게 하여 입력 벡터와 더욱 가까워지게 하고 패자뉴런의 가중치벡터는 제자리에 두어 변동 없게 하는 식의 학습임을 알 수 있다. 최적의 학습률 함수 $\alpha(t)$ 에 관해서는 실험적으로 결정하게 된다. 보통 선형이나 지수 함수적으로 감소하는 형태면 무난하다. 이 문제와 관련하여 (7) 형태의 hyperbolic 함수가 가장 좋은 학습률 함수가 될 수 있음을 (부분적으로) 증명되어있다.

$$\alpha(t) = \frac{1}{t} \quad (7)$$

따라서 본 논문에서도 이 함수를 사용하였다.

2.2 Stochastic Relaxation

SR 방법은 그 내용이 상당히 일반적이다. 어떤 시스템의 t 시점에서의 상태 $W(t)$ 와 그 시점 그 상태의 에너지 함수 $E[W(t)]$ 가 정의 되어있다고 하자. 우리의 목적은 $E[\cdot]$ 를 최소로 하는 시스템 상태 W 를 구하는 것이다. 이 문제에 SR 방법을 적용함에 있어서의 핵심은 다음의 두 가지에 있다.

- 1) Perturbation: 어떤 상태 $W(t)$ 에서 다음 상태 $W(t+1)$ 로 랜덤하게 천이 한다.
- 2) Annealing: 학습의 초기에는 perturbation의 강도가 크게 하고 학습이 진행됨에 따라 서서히 줄여 0으로 수렴하게 한다.

위 1)항에서 perturbation하는 방법은 랜덤한 성분을 포함하기만 한다면 설계자의 자유에 따라 다양하게 정의할 수 있다. 2)항에서 perturbation의 강도를 줄인다는 것은 천이과정에서 랜덤한 성분의 비중을 점점 줄여들게 하고 결정적인 성분의 비중을 크게 한다는 것을 뜻한다. 줄여나가는 속도를 미리 정한 것을 annealing

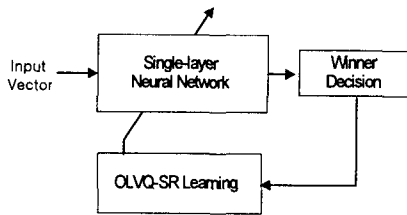
스케줄이라고 한다.

3. 잡음 주입된 KLA를 이용한 벡터 양자화

본 절에서는 '잡음 주입된 KLA를 이용한 벡터 양자화 기 설계 방법'을 논하고자한다. 우리는 이 방법을 OLVQ-SR이라 부르기로 한다. 이 방법은 입력벡터에 잡음을 섞음으로써 간단히 구현될 수 있다. 즉,

$$x(t) = x(t) + \xi(T(n)) \quad (8)$$

여기서 ξ 는 평균 0이고 분산은 온도라 불리는 변수 T 에 의해 결정되는 i.i.d. 가우시안 벡터잡음을 사용한다.



<그림 2> OLVQ-SR 벡터 양자화 기 설계 블록도

이와 같이 입력벡터에 잡음을 섞는 것은 (1)의 연산에 의한 가장 가까운 코드벡터(nearest neighbor, NN)의 선정을 perturbation 하는 것이다. 제안된 방법의 구체적인 순서는 <그림 3>과 같다.

1) 초기화:

$$t=1, \quad T=T_0, \\ W_1^{(1)}, \dots, W_M^{(1)}.$$

2) Nearest Neighbor Perturbation:

$$x(t) = x(t) + \xi(T(n)).$$

3) Centroid 계산:

$$W_c(t+1) = W_c(t) + \alpha Koh$$

4) 열 평형:

If Thermal Equilibrium, Goto 6).
Else
 $t = t + 1,$
Goto 2).

5) 온도 냉각:

If $T(n) < End\ Temperature,$ Stop
Else
 $T = T(n+1),$
Goto 2).

<그림 3> 온라인 코드북 설계를 위한 OLVQ-SR 방법
<그림 3>에서 단계 3)의 Koh는 (5)에서 주어진 Kohonen 학습량을 뜻하는데 여기서 주의해야할 점은 Kohonen 학습량을 계산하는데 사용하는 입력벡터는 잡음에 오염된 벡터가 아닌 오염되기 전의 순수한 입력벡터 $x(t)$ 을 사용한다는 것이다. α 는 학습률을 뜻한다. 학습이 진행됨에 따라 온도가 서서히 줄어들게 되고 잡음 벡터 ξ 의 영향도 줄어들게 된다. 따라서 입력벡터는 NN 규칙에 따라 본래의 승자뉴런에 할당되게 되고 학습은 점차로 순수한 Kohonen 학습법으로 근사하게 된다.

4. 실험 및 결과 분석

제안된 방법 즉, OLVQ-SR과 KLA, GLA의 성능을 비교하기 위해 두 가지 실험을 하였다. 첫째는 가우스-마코프 소스에 대한 양자화 실험이다. 이는 벡터 양자화 기 성능실험에 흔히 사용되는 소스로 실제 데이터(음성이나 화상)에 대한 간단한 확률적 모델로서 사용될 수 있다. 둘째로는 음성소스에 대한 양자화실험이다.

4.1 가우스-마코프 소스에 대한 양자화 실험

본 실험은 상관계수 ρ 가 각각 0, 0.5, 0.9인 세 가지 경우의 1차 가우스-마코프 소스에 대해 조사하였다. 이들 소스는 (13)에 의해 생성된다.

$$x_{i+1} = \rho x_i + w_i \quad (13)$$

여기서 w_i 는 평균 0, 분산 1인 i.i.d. 가우시안 잡음이다. 이렇게 생성된 x_i 는 평균과 분산이 각각 0과 $1/(1-\rho^2)$ 인 가우시안 소스가 된다. 각 상관계수에 대해 총 16384(2^{14}) 샘플을 채집하여 벡터를 구성한 뒤(2차원일 경우 9182개 4차원일 경우 4096개) 순차적으로 인가하며 학습에 사용하였다. KLA의 최대 epoch 수는 실험 결과를 토대로 200회로 설정하였다. 그리고 제안된 방법과 형평을 기하기 위해 KLA의 점수는 초기조건을 바꿔가며 30 재시행을 실시하여 최고점으로 기록하였다.

각 소스에 대해 벡터의 차원과 코드북의 크기를 변화시켜가며 제안된 방법들과 KLA 성능을 비교한 것이 <표 1>에 주어져 있다. OLVQ-SR은 KLA에 대해 최대 2.83(dB)의 성능우위를 나타내고 있다. 이는 코드북 성능이 SNR의 관점에서 약 2배정도 개선되었음을 뜻한다. 그러나 코드북 크기가 64인 경우제안된 방법들과 KLA의 성능에 별 차이가 없었다. 이는 간단한 양자화 문제일 경우 KLA도 좋은 성능을 보이며 적절한 해법이 될 수 있음을 시사하는 것이다.

〈표 1〉 가우스-마코프 소스에 대한 KLA와 OLVQ-SR의 SNR 성능비교. 단위는 dB 이다.

ρ	벡터 차원	코드북 크기	GLA	KLA	OLVQ -SR
0	2	64	15.63	16.46	15.66
		128	17.94	18.96	19.02
		256	19.59	21.78	22.14
	4	64	7.91	7.95	8.00
		128	9.62	9.77	9.92
		256	11.39	12.03	12.22
0.5	2	64	15.60	16.96	16.33
		128	18.29	19.47	19.60
		256	19.75	21.93	23.20
	4	64	8.78	8.88	8.94
		128	10.30	10.63	10.81
		256	11.88	12.88	13.10
0.9	2	64	17.56	19.68	19.43
		128	20.25	22.11	22.58
		256	20.97	23.11	25.94
	4	64	12.85	13.32	13.36
		128	13.78	14.98	15.10
		256	14.37	16.02	17.32

〈표 2〉 음성 소스에 대한 KLA와 OLVQ-SR의 SNR 성능비교. 단위는 dB 이다.

벡터 차원	코드북 크기	GLA	KLA	OLVQ -SR
2	64	20.93	22.46	22.51
	128	21.95	24.10	25.96
	256	22.83	24.89	27.65
4	64	15.95	17.21	18.12
	128	16.94	18.39	20.65
	256	17.65	18.96	23.97

4.2 음성 소스에 대한 양자화 실험

음성은 SB-64 사운드카드로 모노채널, 11kHz 샘플링 주파수, 16bit 양자화 레벨로 총 16384 샘플을 채집하였다. 이렇게 채집된 샘플들을 다시 평균 0, 분산 1로 표준화한 뒤에 양자화기 설계를 위한 학습데이터로 사용하였다.

〈표 2〉에서 보면, 음성 데이터의 경우에 KLA에 대한 제안된 방법의 우수성이 보다 분명히 드러남을 알 수 있다. 이는 패턴의 분포가 복잡해짐에 따라 통계적 학습이 보다 효과를 발휘하는 것으로 볼 수 있다. 대체적으로 여기서도 벡터 차원, 코드북의 크기가 증가함에 따라 제안된 방법들과 KLA 간의 차이가 커지고 있다. OLVQ-SA는 KLA에 비해 최대 5.01에서 최저 0.05까지의 성능 우위를 나타내고 있다. 이를 코드북의 크기와 관련지어 생각해 볼 때, KLA에 비해 코드북의 크기를 절반으로 줄여도 제안된 방법이 보다 우수한 코드북을 생성할 수 있음을 보여주는 것이다. 또한 이는 샘플 당 1 bit 이상의 전송률 이득이 가능하다는 것을 뜻한다. KLA는 코드북의 크기가 128에서 256으로 증가할 때 SNR의 증가율이 급격히 둔화되고 있다. 이는 문제가 복잡

해짐에 따라 KLA에 의한 극소점 근방으로의 접근이 더욱 어려워지고 있다는 것을 시사한다. 반면 제안된 방법들은 코드북 크기에 대해 거의 등 간격의 SNR 개선을 보이고 있다. 이는 복잡해지는 문제에 대해서 잘 대응하고 있음을 시사한다.

5. 결론

본 논문은 KLA의 입력에 잡음을 추가하는 전략이 KLA의 수렴특성을 개선시킨다는 것을 실험적으로 보였다. 주입되는 잡음은 학습의 초기에는 큰 값에서 학습이 진행됨에 따라 서서히 줄어나간다. 이렇게 함으로써 학습은 랜덤탐색에서 점차로 본래의 KLA로 수렴해 간다. 제안된 방법은 구현하기 편리하며 실제데이터에 대해 계산량의 큰 증가 없이 좋은 성능을 나타내었다.

참고 문헌

- [1] T. Kohonen, *Self-Organizing Maps*, Springer-Verlag, 1995.
- [2] B. Widrow and S. D. Sterns, *Adaptive Signal Processing*, Englewood Cliffs, NJ: Prentice-Hall, 1985.
- [3] E. Erwin, K. Obermayer and K. Schulten, "Self-organizing maps: ordering, convergence properties and energy functions," *Biol. Cybern.*, vol. 67, pp. 47-55, 1992.
- [4] N. M. Nasrabadi and Y. Feng, "Vector quantization of images based upon the Kohonen self-organization feature maps," in *Proc. 2nd ICNN Conf.*, vol. I, pp. 101-105, 1988.
- [5] Jun Wang, Ce Zhu, Chenwu Wu and Zhenya He, "Neural network approaches to fast and low rate vector quantization," in *Proc. IEEE ISCAS'95* vol. 1, USA, 1995.
- [6] SriGouri Kamarsu and H. C. Card, "Vector quantization of speech using artificial neural learning," in *Proc. IEEE Pacific RIM Conf. Communications, Computers & Signal Processing*, Canada, 1995.
- [7] D. DeSieno, "Adding a conscience to competitive learning," in *Proc 2nd ICNN Conf.*, vol. I, pp. 117-124, 1988.
- [8] H. Ritter and K. Shulten, "Kohonen's self-organizing maps: Exploring their computational capabilities," in *Proc 2nd ICNN Conf.*, vol. I, pp. 109-116, 1988.