

가중치 진동의 감소를 이용한 신경회로망의 학습속도 향상

Acceleration of Learning Speed in Neural Networks by Reducing Weight Oscillations

임빈철*, 박동조**

(Bin-Chul Ihm*, Dong-Jo Park**)

305-701 대전광역시 유성구 구성동 373-1

한국과학기술원, 전기 및 전자공학과

(KAIST, Department of Electrical Engineering)

Tel) +82-42-869-5438 Fax) +82-42-869-8038

E-mail) bcihm@tercel.kaist.ac.kr djpark@eekaist.kaist.ac.kr

ABSTRACT

본 논문에서는 신경회로망의 수렴속도를 높이기 위한 알고리즘을 제안한다. 전형적인 역전파 학습방식은 느린 수렴속도가 단점으로 제기되는데 이는 비용함수의 계곡부근에서 가중치의 궤적이 심한 진동현상을 보이기 때문이다. 이 문제를 해결하기 위해서 본 논문에서는 경사법에서 사용되는 갱신방향을 계곡의 진행방향을 이용하여 변경한다. 모의실험을 통하여 제안된 방법으로 가중치의 궤적에 나타나는 진동을 줄이고 수렴속도를 향상시킬 수 있음을 보인다.

I. 서론

Rumelhart와 McClelland[1]에 의해 신경회로망의 학습에 역전파 방법이 이용된 이후로 수많은 종류의 역전파 방법이 개발되어 왔다. 그러나 역전파 학습방법은 수렴속도가 낮기 때문에 이런 단점을 해결하기 위해서 오차 함수, 탐색 방향 등의 변형을 통한 기법들이 제안되어 왔다. 이러한 기법들 중에서 Jacob의 델타-바-델타 알고리즘[2]과 Rigler의 알고리즘[3]이

자주 사용되고 있다. 특히 Jacob의 하이브리드 알고리즘[2]은 빠른 수렴속도로 유명하다. 그러나 근본적으로 이런 역전파 방법은 수렴속도가 느리다는 단점이 있는데, 이는 가중치의 학습방향으로 비용함수의 경사도를 그대로 사용하기 때문이다. 느린 수렴속도의 주된 이유는 비용함수의 좁고 긴 모양의 계곡에서 가중치의 학습 궤적이 진동을 하기 때문이다. 대부분의 알고리즘에서는 진동을 줄이기 위해서 모멘텀 항을 이용한다. 그러나 모멘텀 항만으로는 진동을 효

과적으로 줄일 수 없었기 때문에 K. Ochiai 등 [4]이 수정 항을 추가하였다. 수정 항은 궤적의 성분 중에서 계곡의 진행방향과 직각인 성분을 제거함으로써 진동을 줄이는 역할을 한다. 이 방법은 학습 시간을 단축하는데 효과적이다.

본 논문에서는 역전과 방법을 이용하여 학습하는 과정 중에 발생하는 계곡에서의 진동을 줄이기 위해서 가중치의 궤적을 계곡의 진행방향에 맞춰 조정하는 방법을 제안한다. 본문에서는 제안한 방법에 대해 자세히 논한다. NAND 문제에 적용하여 제안된 방법이 기존의 방법들 보다 향상된 결과를 보이는 것을 모의실험을 통하여 예시한다.

II. 가중치 궤적의 진동 감소

일반적으로 신경회로망의 비용함수는 학습의 수렴속도를 떨어뜨리는 계곡을 가지고 있다. 계곡에서의 가중치는 그림 1에서 보는 것처럼 심한 진동을 보이게 된다.

계곡에서의 진동을 줄이기 위해서는 계곡의 하향방향에 직각인 성분을 제거하는 것이 가장 큰 관건이 된다. 이를 위해서는 우선 진동하는 동안 계곡의 하향방향 V 를 추정한다. 가중치의 갱신량 ΔW 의 방향이 V 와 다르면 ΔW 를 V 에 맞추어서 다음 스텝에서의 가중치가 계곡선에 위치하도록 한다 (그림 2 참조). ΔW 가 많이 바뀌는 경우, 즉 $\Delta W^T V < 0$ 이면 계곡의 방향이 달라졌다는 것을 의미하므로 다음의 진동이 발생할 때까지는 일반적인 모멘텀 방법을 이용하여 학습해 나간다. 이 과정을 다음과 같이 정리할 수 있다. 이후의 표기에서 VALLEY 모드는 현재의 가중치가 계곡에 위치하고 있음을 의미하며 V 는 이미 추정된 상태

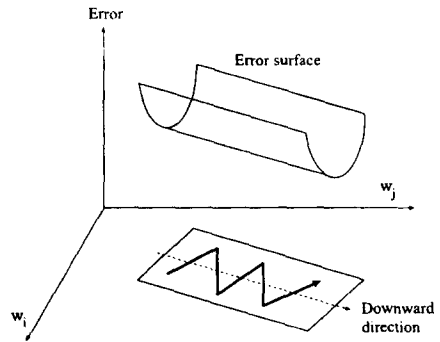


그림 1: 계곡에서 궤적의 진동

이다. NORMAL 모드는 그 외의 경우를 나타낸다.

Step 1. 초기화 작업 W_0 , $\Delta W_0 = 0$, NORMAL 모드, $k=1$, 그리고 **Step 2**로 진행한다.

Step 2. 경사도 g_k 를 계산하고 가중치의 갱신량을 계산한다.

$$d_k = -\eta g_k + \alpha \Delta W_{k-1},$$

여기에서 η 는 step size 그리고 α 는 모멘텀 계수이다.

Step 3. W_k 에서 진동이 발생했는지 검사한다

$$\Theta = \Delta W_{k-1}^T d_k$$

NORMAL 모드였고 $\Theta < 0$ 이면, W_k 를 m 번째 진동위치 W_m^{osc} 로 간주. 이러한 진동위치 $\{W_m^{osc}\}$ 의 개수가 N 이면 VALLEY 모드로 전환하고 **Step 4**로 진행한다. VALLEY 모드였고 $\Theta > 0$ 이면 **Step 4**로 진행한다. 그 외의 경우에는 NORMAL 모드로 전환하고 $\Delta W_k = d_k$ 로 한 다음 **Step 5**로 진행한다.

Step 4. 계곡의 새로운 하향방향 V 추정한다.

$$V = f(W_1^{osc}, W_2^{osc}, \dots, W_N^{osc})$$

$f(\cdot)$ 는 임의의 커브피팅 함수이다. d_k 를 V 에 맞추어 변형한다.

$$\Delta W_k = \frac{1}{2} d_k + \frac{d_k^T V}{2 \|V\|^2} V$$

이 과정은 그림 2에서 묘사되고 있다.

Step 5. 최종적인 가중치의 갱신:

$$W_{k+1} = W_k + \Delta W_k$$

$k = k+1$ 로 하고 나서 Step 2부터 반복한다.

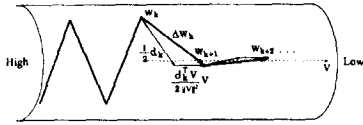


그림 2: 계곡의 하향방향에 맞추어 계곡의 진행 방향을 수정

III. 모의실험 결과

제안된 방법의 타당성을 보이기 위해서 두 개의 입력과 한 개의 출력을 갖는 간단한 구조의 신경망을 이용하여 NAND 문제를 해결하는 경우에 기존의 방법들과 제안된 방법의 학습속도를 비교해 보았다. 이 문제는 2차원의 그래프를 통하여 시각적으로 결과를 보일 수 있는 적당한 예다. 비용함수가 좁고 긴 계곡을 갖기 때문에 진동이 쉽게 발생한다. 가중치의 초기치는 (3.9, 3.2)로 하였고 V 는 간단하게

$$V = \frac{W_3^{osc} + W_2^{osc}}{2} - \frac{W_2^{osc} + W_1^{osc}}{2} = \frac{W_3^{osc} - W_1^{osc}}{2}$$

와 같이 추정하였다. 제안한 방법을 일반적인 모멘텀 방법과 kick-out 알고리즘과 비교하였다.

그림 3은 제안한 방법과 기존의 방법들과의 수렴속도를 보여주고 있다. 그림에서 보듯이 세

개의 방법 중에서 제안한 방법이 다른 두개의 방법들보다 수렴이 빨리 이루어진다는 것을 알 수 있다. 그림 4에서는 kick-out 알고리즘과 제안한 알고리즘을 이용하여 궤적이 진행되는 것을 볼 수 있다. 제안한 알고리즘이 kick-out 알고리즘보다 진동이 적다는 것을 알 수 있다.

IV. 결론

본 논문에서는 신경회로망의 수렴속도를 향상시키기 위한 방법을 제안하였다. 가중치의 학습 궤적이 비용함수의 계곡에서 진동으로 인하여 느리게 진행되는 것을 피하기 위하여 계곡의 하향방향으로 가중치의 궤적을 수정하도록 하였다. 계곡의 하향방향을 추정하기 위하여 진동이 발생하는 위치들로부터 커브피팅의 방법을 이용하였고, 궤적의 방향성분 중에서 계곡의 하향방향과 수직인 성분을 제거하도록 수정항을 삽입하였다. 이를 통하여 궤적이 계곡의 중앙선에 위치하도록 조정하게 된다. 모의 실험을 통하여 길고 좁은 계곡이 있는 비용함수에서 수렴속도가 증가함을 알 수 있다.

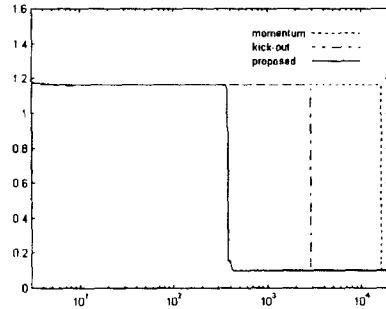
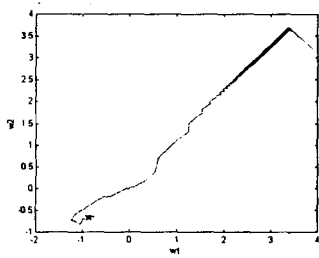
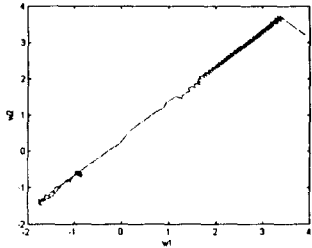


그림 3: NAND 문제를 통한 각 알고리즘의 수렴속도 비교



(a) kick-out 알고리즘



(b) 제안한 알고리즘

그림 4: 제안한 알고리즘과 kick-out 알고리즘의 가중치 궤적의 진동비교

참고문헌

- [1] D. E. Rumelhart, and J. L. McClelland, *Parallel distributed processing, I*, Cambridge, MA: MIT Press, 1986
- [2] R. A. Jacobs, "Increased rates of convergence through learning rate adaptation," *Neural Networks*, 1988, 1, pp. 295-307
- [3] T. P. Vogl, J. K. Mangis, A. K. Rigler, W. T. Zink, and D. L. Alkon, "Accelerating the convergence of the back propagation method," *Biological Cybernetics*, 1988, 59, pp. 257-263
- [4] K. Ochiai, N. Toda, and S. Usui, "Kick-out learning algorithm to reduce the oscillation of weights," *Neural Networks*, 1994, 7, pp. 797-807