

커널 메소드의 효과적인 학습 성능 향상

。 김 은 미, 김 수 희, * 정 태 응, 이 배 호
전남대학교 컴퓨터공학과, 전남대학교 의과대학 방사선과 *
전화 : 062-530-0719 / 핸드폰 : 016-655-1742°

Improving effective Learning Performance of Kernel method

。 Eun-Mi Kim, Su-Hee Kim, *Tae-Woong Chung, Bae-Ho Lee
Dept. of Computer Engineering, Chonnam National University
Dept. of Diagnostic Radiology, Chonnam National University Medical School *
E-mail : b1342@chollian.net°

Abstract

This paper proposes a dynamic moment algorithm to control oscillation before the convergence of the KR(Kernel Relaxation). The proposed dynamic moment algorithm can be controlled to convergence speed and performance according to the change of the dynamic moment by learning training. we used SONAR data that is a neural network classifier standard evaluation data in order to do impartial performance evaluation. The proposed algorithm has been applied to the KP(kernel perceptron), KPM(kernel perceptron with margin) and KLMS(kernel lms) as the kernel method presented recently. The simulation results of proposed algorithm have better the convergence performance than those using none and static moment.

I. 서론

커널은 저차원 공간을 다차원 공간으로 사상하여 문제를 해결하는 방법으로 본 논문에서는 이러한 커널을 적용한 모든 학습 방법을 커널메소드라 한다[3]. 대표적인 커널 메소드를 이용한 방법은 RBF 신경 회로망이 있으며, 특히 KA(Kernel Adatron)와 KR(Kernel Relaxation)은 SVM(Support Vector Machine)의 해법으로 선형, 비선형 분류 문제에 있어 효과적인 성능을 보이는 것으로 소개되었다[3]. 이들 학습방법은 각각의 학습 패턴을 통하여 가중치를 조절하는 순차학습법이다[3]. KA와 KR은 기존의 SVM이 문제의 해를 QP(Quadratic Programming)를 이용하여 구함으로써 발생하는 메모리 문제와 문제의 복잡도에 대한 새로운 접근 방법이다[3]. 특히 문제의 해결 방법은 기존의 재구성 방법들과는 달리 기존의 선형

학습법인 LMS와 Relaxation을 커널 메소드를 이용하여 비선형 분류 문제에 적용시킨 방법으로 이 또한 순차 학습법이다. 하지만 KR과 KA는 각각 최대 강하 학습법과 좌표 강하 학습법으로 이론적으로 최대 강하 학습법인 KR이 좌표 강하 학습법인 KA보다 우수한 성능을 가져야하나 지금까지는 KR과 KA가 동등한 성능이 있는 것으로 소개되었다[3]. 이는 커널 공간이 고차원의 학습 공간으로 KR은 수렴에 가까워질수록 KA보다 잦은 발진을 하며 보다 많은 학습 횟수를 요구하게 된다. 이러한 발진 현상은 KR에만 국한된 것이 아니라 순차적 학습법에서 수렴시 필연적으로 나타나게 된다. 또한 이러한 수렴 문제는 다시 커널 공간으로 그대로 확장이 되며 이때 커널 공간은 다차원의 학습공간으로 순차학습법에서의 발진 현상은 필연적이다. 선형 공간에서의 발진 문제는 이미 모멘트와 학습 계수를 조절함으로써 발진 문제를 해결 할 수 있는 것으로 이미 고전 신경망에서 소개[2] 되었으며 이러한 방법을 KA에 적용함으로써 성능 향상이 있는 것으로 소개되었다[1]. 하지만 이때 모멘트 값을 결정하기 위한 일반적인 방법은 아직 소개되지 않았다. KA는 좌표 강하 학습법으로 단순히 모멘트를 첨가함으로써 향상된 학습 성능을 보일 수도 있으나 KR의 경우는 최대 강하 학습법으로 단순한 모멘트 항의 첨가로 성능 향상을 얻기 힘들며 최적의 모멘트 값을 획득하기에도 KA에 비해 상당히 어려운 점이 있다. 이러한 문제점을 해결하기 위한 최적의 모멘트는 수렴과 학습 횟수를 효과적으로 반영 할 수 있어야 하며 모멘트는 학습에 따라 동적으로 변화되어야 한다. 본 논문에서는 커널 메소드를 이용한 순차 학습법(KP, KPM, KLMS)에서 수렴시 발생하는 발진 문제를 제안된 동적 모멘트를 적용하여 실험하였다. 실험결과 기존의 정적인 모멘트를 이용한 학습법보다 향상된 수렴속도 및 학습 성능을 확인하였다.

II. 커널공간에서의 학습

커널은 특징공간의 정확한 형태를 고려하지 않아도 초평면(optimal hyperplane)을 구성하기 위해 사용할 수 있다. 아래와 같이 내적 커널을 정의한다[2].

$$K(x, x_j) = \varphi^T(x)\varphi(x_j) \quad (1)$$

여기에서 $\varphi(x)$ 는 크기가 $(m+1)$ -by- 1 로 확장된 특징벡터이다. 사용할 커널은 머서의 정리를 만족해야 한다.

L_1 norm에 대해서 $b=1$ 로 하고 확장된 선형 내적 커널을 비선형 내적 커널로 치환하면 커널 공간에서의 식을 구할 수 있다. 즉, 커널공간에서의 초평면은 다음과 같다[2].

$$g_k(Y) \equiv K\alpha = d \quad (2)$$

이때 커널 판별함수는 다음과 같이 정의한다.

$$g_k(y_j) \equiv \alpha^T K(:, j) = K(j, :) \alpha \quad (3)$$

2.1 Kernel Perceptron

퍼셉트론의 훈련 알고리즘은 잘못 분류된 훈련 패턴에 대하여 가중치 벡터 α 를 다음과 같이 계산한다[2].

$$\begin{aligned} &\text{if } d_k \alpha^T K(:, k) \leq 0 \\ &\quad \text{then } \alpha(k+1) \leftarrow \alpha(k) + \eta(k) d_k K(:, k) \\ &\text{until } d_k \alpha^T K(:, k) > 0 \end{aligned} \quad (4)$$

2.2 Kernel Perceptron with Margin

커널 퍼셉트론 마진은 퍼셉트론 구조에 마진을 추가하여 마진 이하로 학습된 훈련 패턴에 대하여 가중치 α 를 학습한다[2].

$$\begin{aligned} &\text{if } d_k \alpha^T K(:, k) \leq b \\ &\quad \text{then } \alpha(k+1) \leftarrow \alpha(k) + \eta(k) d_k K(:, k) \\ &\text{until } d_k \alpha^T K(:, k) > b \end{aligned} \quad (5)$$

2.3 Kernel LMS

LMS는 전체 훈련 패턴에 대해 아래와 같이 학습하며 학습은 가중치 α 가 더 이상 갱신되지 않을 때 종료된다[2].

제 2공간에서 패턴벡터에 대한 가중치 α 의 학습은 아래와 같다.

$$\alpha(k+1) = \alpha(k) + \eta(k) \frac{d_k b - \alpha^T K(:, k)}{\|K(:, k)\|^2} K(:, k) \quad (6)$$

$$\text{until } |\eta(k)(d_k b - \alpha^T K(:, k))| < \theta$$

2.4 RBF 신경회로망

RBF 네트워크는 입력층, 은닉층, 출력층을 가진 3개의 층으로 구성되며 그 구조는 다음과 같이 나타낼 수 있다[2].

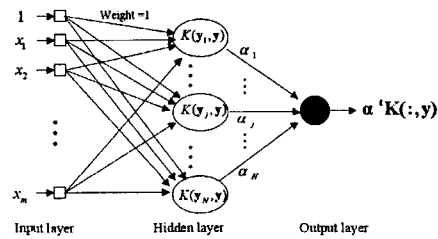


그림 3. RBF네트워크 (Radial-Basis Function Networks)

출력층은 그림과 같이 $d_i = \alpha^T K(:, y_i)$ 로 나타낼 수 있으며 본 실험에서는 머서의 정리를 만족하는 RBF 커널을 사용하였다[2].

$$\exp\left(-\frac{1}{2\sigma^2} \|x - x_i\|^2\right) \quad (7)$$

이때 σ^2 은 사용자에게 의해 정의될 수 있다.

III. Dynamic Moment를 이용한 학습방법

일정 기간의 학습 횟수(Epoch)와 학습 성능을 한정적인 비례관계로 둘 경우 일정 범위 내에서 임의의 한계점까지의 학습 성능은 학습 횟수에 따라 증가하게 된다. 따라서 순차 학습법의 발전은 한계점까지의 중간 과정에서 나타나는 일련의 현상으로 볼 수 있으며 결국 이러한 한계점까지의 도달 간격을 조정하여 발전을 억제한다. 이러한 측면에서 볼 때 정적 모멘트는 학습의 횟수나 성능에 관계없이 일정하게 영향력을 가짐으로써 학습효과를 충분히 반영하기 어렵게 된다.

제한된 학습 방법은 학습 초기 모멘트의 크기를 작게 하여 이전 속성 값의 반영을 줄였으며 점차 학습 수행 횟수가 많아져 수렴에 가까워질수록 모멘트의 크기는 커지게 된다. 하지만 학습 횟수에 따라 연속적으로 모멘트를 증가시킬 경우 모멘트는 단조 증가 함수가 되어 학습 횟수가 매우 많아지게 되면 과거의 학습이 현재의 학습을 능가하게 되어 학습 성능과 수렴속도에 대해 개선 효과를 가져오지 못하게 된다.

따라서 동적 모멘트를 이용한 학습법은 수렴특성에 따라 초기 모멘트 값은 학습 횟수에 따른 증가를 보이며 수

럼에 가까워질수록 다시 모멘트항의 값을 줄여감으로써 전체적으로 수렴속도를 향상시키고 성능을 개선시키는 방법이 된다. 아래는 동적 모멘트 항(M)의 계산식이다.

$$M = \frac{m(k+1)}{\tau}$$

if $M > m$
 $M = 0; \tau = \tau^2;$
 end

(8)

τ 는 초기 모멘트의 값의 크기를 최소화하기 위해 사용한 제어 변수로 예상된 최초 한계치에서의 학습 횟수가 된다. m 은 기존의 상수 모멘트로 동적 모멘트를 이용한 학습방법에서는 적용될 모멘트의 상한값이 된다.

IV. 실험결과

본 실험에서 사용한 데이터는 신경회로망 분류기 표준 평가 데이터인 SONAR 데이터를 사용하였으며, 동적 모멘트가 모든 순차 학습법에 적용 가능한가를 알아보기 위해 모멘트 적용전, 정적 모멘트 적용, 그리고 본 논문에서 제안한 동적 모멘트를 적용한 Kernel Methods에 대한 학습 성능을 비교·평가하였다.

실험에서 RBF 사용자 정의인수 σ 는 최적의 성능을 나타내는 0.4로 하였으며 각 실험당 모멘트 적용전, 기존의 정적 모멘트 사용 및 제안된 학습법인 동적 모멘트 사용으로 구별하여 좀 더 정확하게 수렴 속도 및 학습 성능 향상에 대한 내용을 살펴보도록 하였다. 실험그래프에서 가로축은 Epoch, 세로축은 Performance로 두었으며 Epoch은 50으로 수렴 조건 또한 50으로 정하였다.

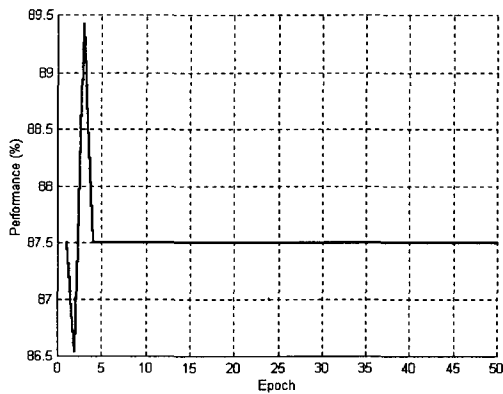


그림 2. 모멘트 사용전 퍼셉트론

(그림 2)(그림 3)(그림 4)은 커널 퍼셉트론에 대한 각각의 모멘트 사용하지 않은 경우, 정적모멘트 사용한 경우, 동적모멘트 사용한 경우를 나타낸 것이다.

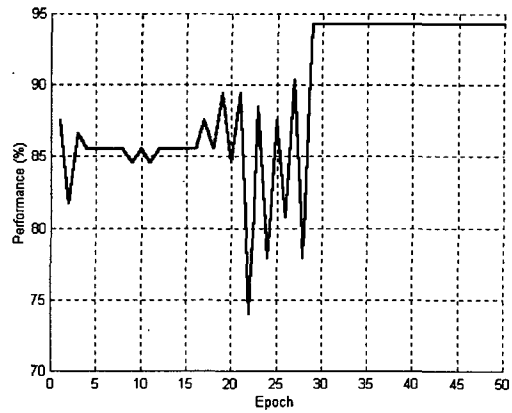


그림 3. 정적모멘트 사용 퍼셉트론

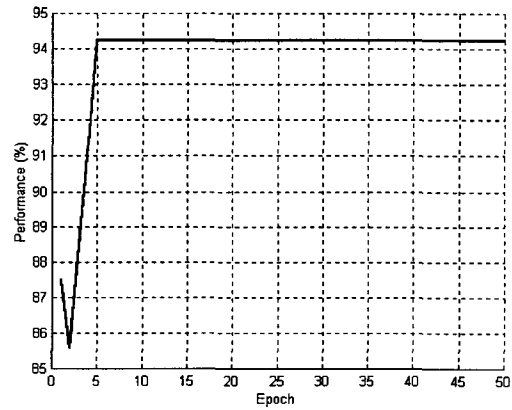


그림 4. 동적모멘트 사용 퍼셉트론

커널메소드의 퍼셉트론에 정적 모멘트와 동적 모멘트를 사용해 본 결과 (그림 2)에서처럼 모멘트 사용하지 않은 경우 Epoch 4에서 수렴이 일어났으며 이때 성능은 87.5%이다. (그림 3)에 정적 모멘트의 사용으로 Epoch 29에서 94.23%의 성능향상을 볼 수 있었다. 이때 정적모멘트에서의 최적의 모멘트 값은 1.202로 나타났다. 동적 모멘트를 적용한 (그림 4)는 최적의 모멘트 2.9를 적용하였으며 Epoch 10에서 95.19%의 성능을 나타내었다. 결과적으로 정적 모멘트보다 동적모멘트가 수렴 속도 및 학습 성능 모두 향상되었을 뿐만아니라 정적 모멘트에서 보이는 수렴전 oscillation 현상도 동적 모멘트 적용후 줄어들었다.

아래의 (그림 5)(그림 6)(그림 7)은 커널 퍼셉트론 마진에 대한 각각의 경우를 나타낸 그래프로 이때 정적 모멘트와 동적 모멘트는 각각 $7.0e-2$ 와 5.002로 나타났으며

모멘트 적용전 경우 수렴이 Epoch 8에서 일어났으며 그때 성능은 95.19%로 나타났다.

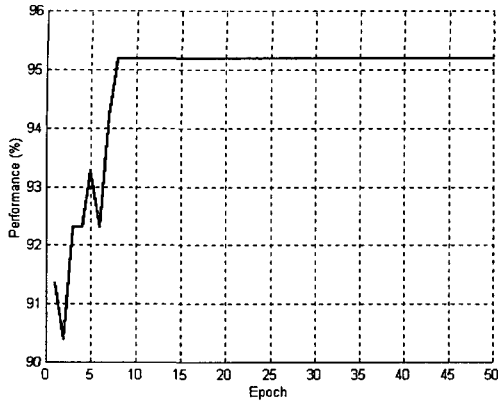


그림 5. 모멘트 적용전 퍼셉트론 마진

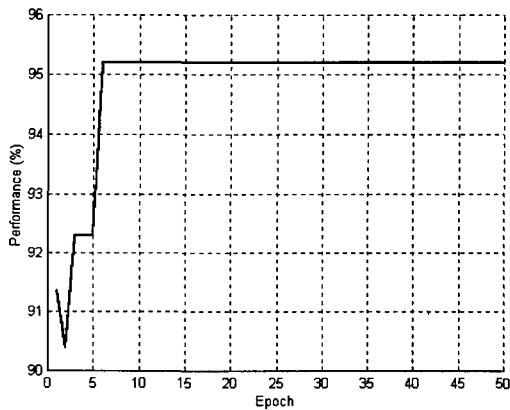


그림 6. 정적모멘트 적용 퍼셉트론 마진

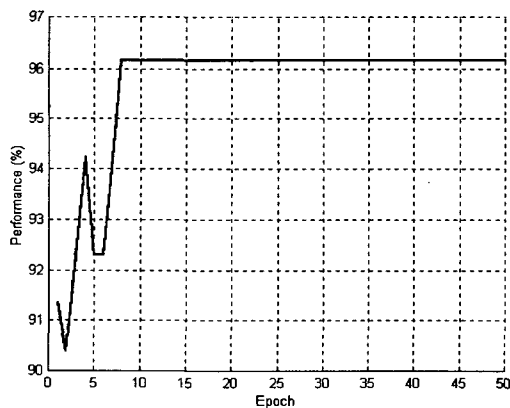


그림 7. 동적모멘트 적용 퍼셉트론 마진

정적 모멘트 적용후 성능 향상을 가져오지는 못했지만 진동 현상의 감소 및 수렴 속도의 다소 빨라진 현상을 볼 수 있었던 반면 동적 모멘트의 적용으로 Epoch 8에서 96.19%의 성능을 확인함으로써 첫번째 실험의 커널 퍼셉트론과 거의 유사하게 수렴전 발산현상이 줄어들고 동시에 수렴 속도 및 학습 성능에서 모두 향상되었음을 볼 수 있었다.

V. 결론

본 논문에서는 SVM의 순차적 학습법 중 Kernel Method에서의 학습 성능 향상 및 수렴 속도 개선을 위해 동적 모멘트를 제안하였다. 제안된 학습방법은 수렴에 따른 모멘트의 값을 동적으로 변경하는 방법으로 Kernel Perceptron, Kernel Perceptron with Margin, Kernel LMS에 적용하여 실험하였다. 실험결과 KR[4]과 마찬가지로 정적 모멘트에 비해 동적 모멘트가 모든 Kernel method에서도 수렴전 발산을 줄여 수렴 시간을 단축시키고 학습 성능을 향상시켰다.

참고문헌

- [1] Yong Hyun Cho, "Improving the Training Performance of Support Vector Machines Using Momentum", 한국정보처리학회 논문집 제 7권 제 5호, pp. 1446-1455, 5, 2000.
- [2] R.O.Duda, P.E. Hart, D.G. Strok, *Pattern Classification*, Second Edition by John Wiley & Sons, Inc, 2001.
- [3] Jae Hung Yoo, Jong Cheol Jeong, "Sparse Representation Learning of Kernel Space Using the Kernel Relaxation Procedure", Korea Fuzzy Logic and Intelligent Systems Society, 2001년도 추계 학술대회 학술 발표 논문집, Vol . 11, No . 2, pp. 60-64, 12, 2001.
- [4] Eun-Mi Kim, Bae-Ho Lee, "Improving Learning Performance of Support Vector Machine using the Dynamic Moment", THE 12th JOINT CONFERENCE ON COMMUNICATIONS AND INFORMATION, II-A.5.1~4, 4, 2002