동적모멘트를 이용한 Kernel Relaxation의 회귀율 향상

김은미[0], 양창호, 이배호
전남대학교 컴퓨터공학과
b1342@chollian.net[0], terabig@hotmail.com, bhlee@chonnam.ac.kr

Improvement Regression Rate of Kernel Relaxation using the Dynamic Momentum

Eun-Mi Kim[0], Chang-Ho Yang, Bae-Ho Lee
Dept. of Computer Engineering, Chonnam National University

요 약

본 논문에서는 학습 중 모멘트를 동적으로 조절하여 수렴속도와 학습 성능을 향상시키는 동적모멘트를 제안하고 회귀방법으로 동적모멘트의 성능을 재확인한다. 제안된 학습방법은 기존의 정적모멘트와는 달리 수렴 정도에 따라 현재의 학습에 과거의 학습률을 달리 반영하는 방법으로 다른 학습법에 비해 보다 유연한 초평면을 갖으며 수렴에 이르는 시간이 오래 걸리는 KR(Kernel Relaxation)에 적용하여 그 성능을 확인한다. 본 논문에서 사용한 회귀방법은 RMS 오류율을 사용하였으며 제안된 학습방법인 동적모멘트를 SVM(support vector machine)의 순차 학습방법 중 최근 발표된 KR에 적용하여 RMS 오류율을 확인하였다. 실험의 공정성을 위해 신경망 분류기 표준평가데이터인 SONAR 데이터를 사용하였으며 실험 결과 동적모멘트를 이용한 회귀율이 정적모멘트를 이용한 방법보다 향상되었음을 확인하였다.

1. 서 론

모멘트를 이용한 학습법은 과거의 학습을 현재의 학습에 반영하는 것으로 이때 모멘트는 상수이며 사용자에 의해 정의되며 이를 정적모멘트라 한다[1].

정적 모멘트는 일정기간의 학습 횟수와 학습 성능을 한계적인 비례관계로 둘 경우 임의의 한계점까지 학습 성능은 학습 횟수에 따라 증가하게 되며 수렴은 한계점과 학습 횟수가 균형을 이룬 것으로 설명할 수 있다. 순차 학습법의 발진은 한계점까지의 중간과정에서 나타나는 현상으로 볼 수 있으며 결국 이러한 한계점까지의 도달 간격을 좁혀 줌으로써 발진을 억제할 수 있게 된다. 이러한 측면에서 볼 때 정적모멘트는 상수로 선언되어 학습에 있어 수렴 정도(학습의 횟수나 성능)에 관계없이 일정하게 영향력을 가짐으로써 평균적인 효과를 기대할 수 있다. 그러나 전체학습의 수렴과 학습정도에 관계없이 동일한 모멘트 값의 영향을 받는 것으로 이는 학습 횟수의 증가와 해의 수렴에 대한 기본원리를 충분히 반영하였다고 할 수 없다.

본 논문에서는 순차 학습법에서의 수렴시 발진억제를 위해 과거의 속성값을 반영하는 모멘트 항인 상수 τ 를 추가하여 수렴정도에 따라 모멘트의 크기를 조절하여 순차학습법의 수렴속도와 학습 성능을 향상시키는 학습방법을 제안한다. 공정한 학습 성능평가를 위해 신경망 분류기 표준평가데이터인 SONAR 분류 문제를 이용하였으며[2], 회귀 문제의 평가를 위해 RMSE(root mean square error)를 사용하였다. 실험결과 수렴속도와 학습 성능 및 회귀율의 성능이 향상되었다.

2. Kernel Relaxaion

커널은 특징공간의 정확한 형태를 고려하지 않아도 최적 초평면(optimal hyperplane)을 구성하기 위해 사용할 수 있다[3]. 내적 커널의 정의는 다음과 같다.

$$\mathbf{K}(\mathbf{x}, \mathbf{x}_i) = \boldsymbol{\varphi}^T(\mathbf{x})\, \boldsymbol{\varphi}(\mathbf{x}) \quad . \qquad (2.1)$$

여기 $\boldsymbol{\varphi}(\mathbf{x})$는 크기가 $(m+1)$-by-1로 확장된 특징 벡터이다. 사용할 커널은 머서의 정리(Mercer's theorem)를 만족해야 한다[4]. 실험에서 머서의 정리를 만족하는 아래의 RBF 커널을 사용하였다.

$$\exp\left(-\frac{1}{2\sigma^2}\|\mathbf{x} - \mathbf{x}_i\|^2\right) \qquad (2.2)$$

이때 σ^2은 사용자에 의해 정의 될 수 있다. 커널 공간에서 커널 행렬을 패턴 행렬로 커널의 행과 열을 패턴 벡터로 정의하면 라그랑지안 승수 $\boldsymbol{\alpha}$는 패턴의 가중치가 되어 아래와 같이 $\mathbf{g}_k(\mathbf{Y})$를 정의 한다.

$$\mathbf{g}_k(\mathbf{Y}) \equiv sign(\mathbf{K}\ \boldsymbol{\alpha}) = \mathbf{d} \quad . \qquad (2.3)$$

이때 커널 판별 함수는 다음과 같다[4].

$$g_k(\mathbf{y}_j) \equiv \boldsymbol{\alpha}^t\, \mathbf{K}(:, j) = \mathbf{K}(j, :)\, \boldsymbol{\alpha} \quad . \qquad (2.4)$$

따라서 입력 공간은 방정식의 수가 미지수의 수 보다 많은 과잉결정 구조(overdetermined system)이고 커널공간은 방정식의 수와 미지수의 수가 같은 완전 결정 구조(exactly determined system)가 된다. 결국 커널 공간에서는 비 특이 행렬 $\mathbf{K}$에 대하여 해가 언제나 존재하게 된다[5].

따라서 선형판별 함수의 모든 학습 방법을 그대로 커널 공간에 적용할 수 있게 된다. 커널의 최대 강하/상승법은 아래와 같다[5].

$$\Delta \boldsymbol{\alpha} = \eta\,(\ \mathbf{d} - \boldsymbol{\alpha}'\ \mathrm{K})\quad . \qquad (2.5)$$

위 식을 요소(component)별로 표현하면 다음과 같다[5].

$$\Delta\alpha_j = \eta(d_j - \boldsymbol{\alpha}'\ \mathrm{K}(:,j))\quad . \qquad (2.6)$$

위 식은 Campbell이 제안한 KA(kernel adatron)이다[4]. 하지만 주어진 패턴 y_i에 대해 벡터 $\boldsymbol{\alpha}$의 요소 중에서 α_j만 학습하는 것으로 이 방법은 최대 강하 학습법이 아니라 좌표 강하법이다[5].

KR은 주어진 가중치 $\boldsymbol{\alpha}$ 를 학습하는 방법으로 유도 과정은 다음과 같다[5]. 식(3.1.3)에 의해 error cost function(오류 비용 함수)을 아래와 같이 정의한다[4].

$$E = \xi^2 = \frac{1}{2}\,\|(\ \mathbf{d} - \mathrm{K}\ \boldsymbol{\alpha})\|^2\quad , \qquad (2.7)$$

$$\frac{\partial E}{\partial\, \boldsymbol{\alpha}} = (\ \mathbf{d} - \mathrm{K}\ \boldsymbol{\alpha})\,(\ -\mathrm{K})\quad . \qquad (2.8)$$

따라서 위 식은 주어진 가중치 $\boldsymbol{\alpha}$ 를 학습하는 아래의 식으로 변경되며, 신경망에서 델타 규칙(Delta Rule)로 알려져 있다.

$$\Delta\, \boldsymbol{\alpha}\,|_{\ y_i} = \eta(d_j - \boldsymbol{\alpha}'\,\mathrm{K}(:,j))\,\mathrm{K}(:,j)\quad . \qquad (2.9)$$

3. 동적모멘트를 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 발진은 한계점까지의 중간 과정에서 나타나는 일련의 현상으로 볼 수 있으며 결국 이러한 한계점까지의 도달 간격을 조정하여 발진을 억제한다. 이러한 측면에서 볼 때 정적모멘트는 학습의 횟수나 성능에 관계없이 일정하게 영향력을 가짐으로써 학습효과를 충분하게 반영할 수 없게 된다.

제안된 학습방법은 학습 초기모멘트의 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 증가하게 되며 과거의 학습이 현재의 학습을 능가하게 되어 학습성능과 수렴속도에 대해 개선효과를 가져오지 못하게 된다.

따라서 동적모멘트를 이용한 학습법은 수렴특성에 따라 초기 모멘트 값은 학습 횟수에 따른 증가를 보이며 수렴에 가까워질수록 다시 모멘트항의 값을 줄여감으로써 전체적으로 수렴속도를 향상시키고 성능을 개선시키는 방법이 된다. 아래는 동적모멘트 항 **M**의 계산식이다.

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

$$\text{if } \mathbf{M} > m$$

$$\mathbf{M} = 0;\ \tau = \tau^2;$$

$$end \qquad (3.1)$$

τ는 초기 모멘트의 값의 크기를 최소화하기 위해 사용한 제어 변수로 예상된 최초 한계치에서의 학습 횟수가 된다. m은 기존의 상수 모멘트로 동적모멘트를 이용한 학습방법에서는 적용될 모멘트의 상한값이 된다.

4. Root Mean Square Error

RMSE는 MSE의 절대값인 표준 에러로 성능을 평가하기 위한 회귀 문제의 한 방법이다.

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\ Y_t^s - Y_t^a)^2} \qquad (4.1)$$

$\langle\ Y_t^s = Y_t$ 의 결과값, $Y_t^a = Y_a$ 의 실제값$\rangle$

$Y_t^s = Y_t^a$ 일 경우 100%의 성능을 갖게 되며 이때 RMSE는 0 이 된다. 즉 RMSE 가 작을수록 성능은 향상되어진다.

5. 실험 및 결과

본 실험은 Windows 2000 서버를 사용하였으며, 시스템의 구현을 위해 Matlab Ver. 6.1을 사용하였다. 본 실험에서 사용한 데이터는 신경망 분류기 표준 평가 데이터인 SONAR 데이터[2]로 차원의 수는 60차원, 데이터의 총 개수는 208개로 처음 104개를 학습 데이터로 나머지 104개를 테스트 데이터로 사용하여 실험하였다.

실험은 제안된 동적모멘트의 RMSE를 평가하기 위한 것으로 KR에 각각 정적모멘트와 동적모멘트를 적용하여 RMSE를 비교하였다. KR에 사용된 정적모멘트의 값과 동적모멘트의 값은 0~2까지 0.1씩 증가시키며 각각의 모멘트 값에 대해 매회 50번씩 학습한 결과 중 가장 우수한 학습 성능을 갖는 모멘트 값을 확인하였다.

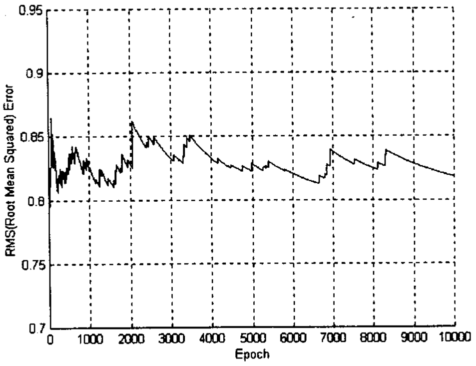
이하 실험에서 정적모멘트 값과 동적모멘트 값은 실험결과 최적의 분류성능을 보인 KR에서는 4.0e-3와 1.7e-2를 적용하였다.

본 실험은 RBF함수에서 사용자 정의인수 σ값의 변화에 따른 학습 성능을 알아보기 위해 σ를 0.1에서 0.9까지 0.1씩 변화시켰으며 각 Epoch당 학습 성능을 알아보기 위해 Epoch를 50번에서 200번까지 변화시켜 실험하였다.

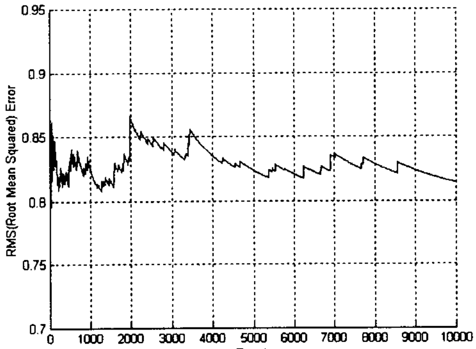
이하 모멘트 사용하지 않은 것을 NM(not moment), 정적모멘트를 적용한 것을 SM(static moment), 동적모멘트를 적용한 것을 DM(dynamic moment)라 칭한다.

(그림 2)~(그림 4)는 KR에 모멘트를 사용하지 않은 경우와 최적의 모멘트를 사용한 기존의 정적모멘트 및 본 논문에서 제안한 동적모멘트 각각에 대한 RMS 오류율을 비교한 것이다.

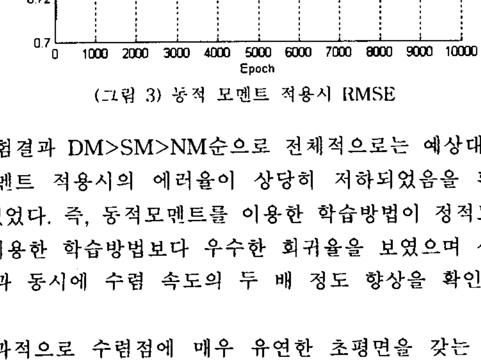
전체적인 실험은 σ값이 0.3과 0.4일 경우만을 나타내었다.



(그림 1) 모멘트 적용 전 RMSE



(그림 2) 정적 모멘트 적용시 RMSE



(그림 3) 동적 모멘트 적용시 RMSE

실험결과 DM>SM>NM순으로 전체적으로는 예상대로 동적모멘트 적용시의 에러율이 상당히 저하되었음을 확인할 수 있었다. 즉, 동적모멘트를 이용한 학습방법이 정적모멘트를 이용한 학습방법보다 우수한 회귀율을 보였으며 성능의 향상과 동시에 수렴 속도의 두 배 정도 향상을 확인할 수 있다.

결과적으로 수렴점에 매우 유연한 초평면을 갖는 KR의 학습에 있어서 정적모멘트를 이용한 방법보다 동적모멘트를 이용한 학습방법이 학습성능 향상에 효과적인 방법이 됨을 RMSE를 통해 확인할 수 있다.

6. 결론

본 논문에서는 SVM의 순차 학습방법에 있어서 수렴율 개선을 위해 동적모멘트를 이용한 학습방법을 제안하였다. 제안된 동적모멘트를 이용한 학습법을 SVM의 순차 학습 방법인 KR을 이용하여 실험하였다. 실험결과 수렴점에 매우 유연한 초평면을 갖는 KR의 학습에 있어서 기존의 모멘트를 적용하지 않은 KR의 RMS 에러율이 8.1834e-001에 비해 모멘트를 적용한 후 8.1441e-001가 되는 것을 알 수 있었으며 본 논문에서 제안한 동적모멘트를 이용한 KR의 RMS 에러율은 7.3943e-001로 기존의 정적모멘트에 비해 RMS 에러율의 감소로 동적모멘트의 우수함을 확인할 수 있었다.

그러나 제안된 동적모멘트를 이용한 학습법에서의 τ 값이 성능 향상에 직접적인 영향을 미치는가에 대해서 유동적인 τ 값의 변화에 따른 연구가 필요하다.

참 고 문 헌

[1] 조용현, "모멘트를 이용한 Support Vector Machines의 학습성능개선", 한국정보처리학회 논문지 제 7권 제 5호, pp. 1446-1455, 5, 2000

[2] Y. Li et al. , "The relaxed online maximum margin algorithm.", In Advances in NIPS 13, 1999.

[3] J. C. Platt, "Fast training of support vector machines using sequential minimal optimization," Advances in Kernel Methods - Support Vector Learning, B. Scholkopf, C. Burges, A. Smola, editors, pp. 185-208, MIT-Press, 1998.

[4] R. O. Duda, P. E. Hart, D. G. Stork, *Pattern Classification*, Second Edition by John Wiley & Sons, Inc, 2001.

[5] 류재홍, 정종철 "커널 이완절차에 의한 커널 공간 의 저밀도 표현 학습," 한국퍼지 및 지능시스템학회, 2001년도 추계 학술대회 학술 발표 논문집, Vol. 11, No. 2, pp. 60-64, 2001.12