

모멘트를 이용한 Support Vector Machines의 학습성능개선

조 용 현†

요 약

본 논문에서는 기율기상승법에서 해의 변화에 과거의 속성을 반영하는 모멘트 항을 추가한 수정된 kernel-adatron 알고리즘을 제안하고 이를 support vector machines의 학습기법으로 이용하였다. 이는 기율기상승법에서 일어나는 최적해로의 수렴에 따른 발진을 억제하여 그 수렴속도를 좀 더 개선시키는 모멘트의 장점과 kernel-adatron 알고리즘의 구현 용이성을 그대로 살리기 위함이다. 제안된 학습기법의 SVM을 임의의 18개 데이터 기반형 분류문제의 200명의 암환자를 2부류(초기와 악성)로 분류하는 문제에 적용하여 시뮬레이션한 결과, 기존의 QP와 Campbell 등의 kernel-adatron 알고리즘을 이용한 SVM에 의한 결과와 각각 비교할 때 학습시간과 시험 데이터의 분류률에서 더욱 우수한 성능이 있음을 확인하였다.

Improving the Training Performance of Support Vector Machines Using Momentum

Yong-Hyun Cho†

ABSTRACT

This paper proposes a modified kernel-adatron algorithm which is a learning algorithm for the support vector machines. It is added the gradient ascent to a fraction of previous solution-change value, momentum. The momentum is applied for increasing the speed of convergence by suppressing the oscillations as the optimal solution converges. It can achieve a superior property of the implementations in the kernel-adatron algorithm. The SVM using the proposed algorithm has been applied to classify the nonlinear problems with 18 data and 200 breast cancer databases by 2-class. The simulation results shows that the proposed algorithm has better performances of the learning time and the classification for test data, in comparison with those using the conventional QP and Campbell's kernel-adatron algorithm.

1. 서 론

경험적인 데이터 모델링 문제는 많은 공학적인 응용과 밀접한 관계가 있다. 여기서의 유도(induction) 과정은 시스템의 모델을 구성하는 데 이용되며, 그것으로부터 관찰되어야 될 시스템의 응답을 추론할 수 있다. 궁극적으로 관찰의 결과 많은 경험적인 모델의 성능을 지배한다 즉, 얻어진 데이터는 유한하며 표본화되나

표본화는 일정하지 않으며 문제의 고차원성 때문에 데이터는 입력공간에서 희박한 분포로 형성될 것이다. 결국 문제는 거의 늘 ill-posing하게 된다[1-2].

학습과 대규모 병렬처리에 기인한 신경망을 이용한 접근은 데이터의 overfitting에 따른 모델을 생성할 수 있어 일반화(generalization) 측면에서 어려움이 뒤따르게 된다[2-6]. 이는 학습을 위한 최적화 알고리즘과 가장 좋은 모델을 선택하는데 이용된 통계적 추도의 결과이다. 이러한 신경망이 가지는 제약들을 해결하기 위해서 support vector machine(SVM)이 제안[2-6]되

† 정 회 원 . 대구효성가톨릭대학교 컴퓨터정보통신공학부 교수
논문접수 : 1999년 8월 18일, 심사완료 : 2000년 4월 28일

었으며, 문자, 얼굴, 그리고 물체 인식 등의 실제분야에 성공적으로 적용되었다[4-7]. SVM이 overfitting을 효과적으로 막아주는 것은 VC(Vapnik-Chervonenkis) 이론으로 설명될 수 있으며[1], 그것의 학습은 볼록함수(convex function)를 최대화 함으로써 수행되는 것으로 이는 polynomial time내에 발견될 수 있는 유일해가 존재한다는 것을 의미한다. 결국 SVM은 polynomial 학습머신, radial basis 함수 망, 그리고 다층퍼셉트론(multi-layer perceptron .MLP)[2, 3, 8, 9]를 위한 또 다른 학습기법으로 생각될 수 있다 망의 가중치는 신경망의 학습기법에서 오목(non-convex)하고 조건을 포함하지 않는 최소화 문제를 푸는 것과는 달리 선형의 부등과 항등 조건을 가진 2차 프로그래밍(quadratic programming : QP) 문제를 풀므로써 찾을 수 있다. 간단한 이진 분류문제의 경우, 학습데이터에 대한 2개의 클래스 사이의 거리를 최대로 하는 초평면(hyperplane)을 찾기 위해서 고차원의 특징공간으로 사상시키는 것이다. 이는 좋은 일반화를 갖도록 시스템의 VC 차원을 최소화하는 것이다. 최대 여유(margin)를 가지는 초평면을 찾는 것은 최적화 과정을 거쳐 해결할 수 있는 QP 문제로 생각할 수 있다.

최근 수 많은 실제 분야에서 SVM이 성공적으로 적용되었으나 아직까지 machine learning에서 신경망 등과는 달리 표준도구로 인정받지 못하고 있다. 이는 QP 문제를 푸는 과정에 복잡한 계산이 요구되며 시스템을 구현하는 데도 어려움이 따르기 때문이다. 이때 요구되는 메모리량은 데이터 수의 자승에 비례하며 학습 또한 대단히 느리게 이루어진다. 특히, SVM의 학습알고리즘은 복잡하고 미묘하며 때로는 구현이 어렵다. 이러한 제약점을 해결하기 위한 많은 연구가 진행되고 있다 [3-7]. Osuna 등은 QP 문제를 여러 개의 작은 문제의 연속으로 나누어 학습시키는 chunking을 이용한 기법을 제안하였다[4]. 이는 학습데이터의 자승에 비례하는 행렬의 크기를 감소시켜 계산량을 줄이는 것이나 여전히 메모리 문제는 그대로 남아있다. 또한 이들은 학습중에 동일한 개수의 학습데이터들을 추가하거나 제거함으로써 각각 QP 문제의 행렬을 일정한 크기로 하여 학습시키는 분할(decomposition)기법을 제안하였다[5]. 여기에서도 알고리즘의 내부과정으로 수치적인 2차 프로그래밍을 수행함으로써 많은 수치적 문제에 직면하게 되고 이로 인한 시간적인 제약도 가지고 있다. 한편 Platt은 순차적인 최소최적화(sequential minimal opti-

mization) 기법을 제안하여 학습시간과 구현문제를 동시에 개선하였다[6]. 이 기법은 큰 규모의 QP 문제를 가능한한 가장 작은 규모의 문제로 나누는 것까지는 Osuna 등의 방법과 동일하나 작은 규모의 문제를 해석적으로 해결함으로써 분할기법이 가지는 제약을 해결하였다 Campbell 등은 커널(kernel) 방법과 퍼셉트론 규칙을 병합한 반복기법의 kernel-adatron 알고리즘을 제안하였다[3]. kernel-adatron 알고리즘은 adatron의 용이한 구현성과 커널에 의한 비선형 특징공간에서의 동작을 조합한 하이브리드 학습기법이다. 여기서는 특징공간에서의 여유를 최대화하기 위하여 기울기상승(gradient ascent)법을 이용하고 등식의 조건을 만족시키는 라그랑지안 계수(Lagrangian multiplier)을 얻기 위해서 할선법(secant method)을 이용하였다. 그러나 기울기상승법은 해의 변화를 라그랑지안 함수의 미분에 따라 변화시킴으로서 최적해로 수렴될 때 발견이 일어날 수도 있다. 따라서 우수한 구현성은 그대로 살리면서도 최적해로의 수렴에서 일어나는 발견을 막아 좀더 빠른 속도로 SVM을 학습시킬 수 있는 효과적인 기법의 연구가 요구된다.

본 연구에서는 기울기상승법에서 해의 변화에 과거의 속성을 반영하는 모멘트(momentum) 항을 추가한 수정된 kernel-adatron 알고리즘을 제안하고 이를 SVM의 학습기법으로 이용하였다. 이는 기존의 기울기상승법에서 일어나는 최적해로의 수렴에 따른 발견을 억제하여 그 수렴속도를 좀더 개선시키는 모멘트의 장점과 kernel-adatron 알고리즘의 구현 용이성을 그대로 살리기 위함이다. radial basis 함수의 커널을 이용한 제안된 학습기법의 SVM을 18개 데이터의 비선형 분류문제와 실제 200명의 암환자를 2부류(초기와 악성)로 분류하는 문제에 적용하여 시뮬레이션하고 그 타당성을 확인하였으며, 표준 QP 알고리즘과 kernel-adatron 알고리즘을 각각 SVM에 적용한 결과와 비교 고찰하였다.

2. Support vector Machines

support vector machine은 기존의 신경망 등에서 이용된 경험적 위험(empirical risk)을 최소화하는 원리보다는 구조적 위험(structural risk)을 최소화하는 근사적 구현이다. 구조적 위험을 최소화하는 것은 학습데이터에 대한 오차를 최소화하는 경험적 위험을 최소화하는 것과는 달리 기대되는 위험에 대한 상한을 최소

화하는 것이다 즉, 구조적 위험을 최소화하는 방법은 학습오차 비율의 합으로 범위가 결정되는 시험오차 항과 학습미신의 VC-차원(VC-dimension)에 의존하는 항에 기반을 두고 있다 두 항의 합을 최소화함으로써 우수한 일반화 성능을 얻을 수 있다. SVM은 분류문제 뿐만 아니라 최근 회귀(regression)문제에도 적용되어 우수한 해결 능력을 보여 주었다[2, 8, 9]. 본 연구에서는 분류문제를 대상으로 소개한다.

2개의 분리된 클래스에 속하는 학습벡터의 집합을 선형적으로 분리 가능한 문제를 생각해 보자. 이는 가중치 벡터 ω 와 바이어스 θ 로 구성되는 $(\omega \cdot \mathbf{x}) + \theta = 0$ 의 초평면을 가지도록 표본들 $\{(x_i, d_i)\}_{i=1}^N$ 를 학습시키는 것이다. 여기서 x_i 는 i 번째 예에 대한 입력패턴이고, d_i 는 ± 1 의 값을 가지는 원하는 응답이다 초평면 $(\omega \cdot \mathbf{x}) + \theta = 0$ 는 다음의 조건을 만족한다 즉,

$$\begin{aligned} (\omega \cdot \mathbf{x}_i) + \theta > 0, & \text{ if } d_i = 1 \\ (\omega \cdot \mathbf{x}_i) + \theta < 0, & \text{ if } d_i = -1 \end{aligned} \quad (1)$$

이다. 식에서 ω 와 θ 를 다른 값으로 하여 표준형태(canonical form)로 나타내면 다음과 같이 표현할 수 있다. 즉,

$$\begin{aligned} (\omega \cdot \mathbf{x}_i) + \theta &\geq 1, & \text{ if } d_i = 1 \\ (\omega \cdot \mathbf{x}_i) + \theta &< -1, & \text{ if } d_i = -1 \end{aligned} \quad (2)$$

와 같다. 위의 표준형태를 단일항의 식으로 나타내면

$$d_i (\omega \cdot \mathbf{x}_i) + \theta \geq 1, \quad i = 1, 2, \dots, N \quad (3)$$

이다. 식 (3)에서 등호의 조건을 만족하는 입력패턴 (\mathbf{x}_i, d_i) 을 support vector라 하며, 개념적으로 이 벡터는 초평면에 가장 가까이 위치하여 분류하기가 어려운 벡터들이다. 따라서 분류를 위한 학습은 제약조건 식 (3)을 만족하는 최적의 초평면을 찾는 것이다 이는 제약조건을 가지는 최적화 문제로 학습패턴 $\{(x_i, d_i)\}_{i=1}^N$ 이 주어질 때 최적의 초평면을 위한 최적파라미터 ω_0 와 θ_0 를 찾는 것이다. 여기서의 최적은 최대여유를 가지는 것이며, 최대여유 초평면은 최적으로 2개의 클래스를 분리할 수 있는 초평면이다. 결국 2개의 클래스와 초평면사이의 거리에 따른 여유 $\rho(\omega, \theta)$ 는 $2/\|\omega\|$ 가 되며, 입력패턴을 최적으로 분류하는 초평면은 다음의 비용함수 $\Theta(\omega)$ 를 최소화한다 여기서 비용함수 $\Theta(\omega)$ 는

$$\Theta(\omega) = \frac{1}{2} \|\omega\|^2 \quad (4)$$

이다. 식 (4)의 비용함수는 ω 의 볼록함수이며, 제약조건 식 (3)은 ω 에 선형임을 알 수 있다.

지금까지 서술된 분류를 위한 SVM을 정리하면, 학습패턴이 주어질 때 제약조건 식 (3)을 만족하는 가중치 벡터 ω 와 바이어스 θ 를 찾는 최적화 문제인 기본문제(primal problem)로 생각할 수 있으며, 이때 ω 는 비용함수 식 (4)도 최소화한다. 이 최적화 문제를 해결하기 위하여 라그랑지안 계수법을 이용하면 다음과 같은 라그랑지안 함수 $J(\omega, \theta, \alpha)$ 을 얻을 수 있다 즉,

$$J(\omega, \theta, \alpha) = \frac{1}{2} \|\omega\|^2 - \sum_{i=1}^N \alpha_i [d_i (\omega^T x_i + \theta) - 1] \quad (5)$$

이며, 식에서 α 와 α_i 는 각각 라그랑지안 계수벡터와 그 계수들이다. 식 (5)의 최적화 문제에 대한 해는 ω 와 θ 에 대해서는 최소화되나 $\alpha_i \geq 0$ 에 대해서는 최대화되어야 한다 따라서 ω 와 θ 에 대한 $J(\omega, \theta, \alpha)$ 의 최소는 그 각각에 대한 미분으로 얻어질 수 있다. 즉,

$$\begin{aligned} \frac{\partial J}{\partial \theta} = 0 &\rightarrow \sum_{i=1}^N \alpha_i d_i = 0 \\ \frac{\partial J}{\partial \omega} = 0 &\rightarrow \omega = \sum_{i=1}^N \alpha_i x_i d_i \end{aligned} \quad (6)$$

이다. 한편 α_i 를 구하기 위해서 기본문제에 대한 라그랑지안 함수 $J(\omega, \theta, \alpha)$ 를 이원문제(dual problem)의 목적함수 $Q(\alpha)$ 로 표현할 수 있다. 여기서 목적함수 $Q(\alpha)$ 는 식 (6)을 식 (5)에 대입하여 얻어지며, 그 표현식은

$$Q(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j d_i d_j (x_i^T x_j) \quad (7)$$

과 같고, 조건은 $\sum_{i=1}^N \alpha_i d_i = 0$ 와 $\alpha_i \geq 0 (i = 1, 2, \dots, N)$ 이다. 식 (7)의 목적함수는 일반적인 QP 문제의 형태로 학습패턴의 항으로만 구성되며, 입력패턴의 내적(dot product)의 형태로 $(x_i^T x_j)$ 로 표현된다. 따라서 분류문제를 식 (7)의 이원문제로 생각하면, 이는 학습패턴 $\{(x_i, d_i)\}_{i=1}^N$ 이 주어질 때, 제약조건 $\sum_{i=1}^N \alpha_i d_i = 0$ 와 $\alpha_i \geq 0 (i = 1, 2, \dots, N)$ 을 만족하는 목적함수 식 (7)를 최대화하는 라그랑지안 계수 α_i 를 찾는 것이다

따라서 QP 알고리즘에 따라 제약조건 식 (3)하에서 목적함수 식 (7)을 최대로 하는 최적의 라그랑지안 계수 α_{10} 를 찾으려면, 최적의 가중치 벡터 ω_0 는 식 (6)에 따라 계산될 수 있고, 최적의 바이어스 θ_0 는 support vector로부터 계산될 수 있다. 그 각각의 계산식은

$$\begin{aligned} \omega_0 &= \sum_{i=1}^N \alpha_{10} x_i d_i \\ \theta_0 &= -\frac{1}{2} \omega_0 [x_r + x_s] \end{aligned} \quad (8)$$

이다. 여기서 x_r 과 x_s 는 각각 다음의 조건을 만족하는 support vector들

$$\alpha_{10}, \alpha_{20} > 0, \quad d_r = 1, \quad d_s = -1 \quad (9)$$

이다. 이때 SVM에 의한 분류식을 정리하면 다음과 같다. 즉,

$$\begin{aligned} f(\mathbf{x}) &= \text{sign}(\mathbf{w}_0 \cdot \mathbf{x} + \theta_0) \\ &= \text{sign}[\sum_{i=1}^N \alpha_{10} d_i (x \cdot x_i) + \theta_0] \end{aligned} \quad (10)$$

으로 이는 선형의 결정면을 가짐을 알 수 있다. 여기서 $\text{sign}(\cdot)$ 는 \cdot 이 양수이면 1, 그렇지 않으면 -1를 가지는 함수이다.

한편, 선형적으로 분류 가능하지 않는 문제에 대해서도 분류 가능하게 하는 일반화된 초윌면을 구성하기 위해서 음수가 아닌(non-negative) 스칼라변수 $\xi_i \geq 0$ 을 소개한다. 여기서의 ξ_i 는 잘못된 분류와 관련된 오차의 척도로 슬랙변수(slack variables)이다. 따라서 분류 불가능한 경우를 위한 슬랙변수 ξ_i 를 포함하는 제약조건은 식 (3)을 다음과 같이 수정함으로써 얻어질 수 있다. 즉,

$$d_i (\omega \cdot \mathbf{x}_i) + \theta \geq 1 - \xi_i, \quad i = 1, 2, \dots, N \quad (11)$$

이다. 또한 위의 제약조건을 만족하는 가중치 벡터 ω 와 슬랙변수 ξ_i 를 포함하는 비유함수 $\Theta(\omega, \xi)$ 는 다음과 같다. 즉,

$$\Theta(\omega, \xi) = \frac{1}{2} \|\omega\|^2 + c \sum_{i=1}^N \xi_i \quad (12)$$

이다. 여기서 c 는 양의 값으로 학습오차와 일반화 능력 사이의 상관관계를 제어하는 파라미터이다. 본 실

험에서는 c 를 여러 가지 값으로 변화시켜 실험하였다. 따라서 식 (11)의 제약조건하에서 식 (12)의 최적화 문제에 대한 라그랑지안 함수는 $J(\omega, \theta, \xi, \alpha, \beta)$ 로 표현되며, 여기서 β 는 $\xi_i \geq 0$ 의 조건을 위한 라그랑지안 계수이다. 한편 주어진 라그랑지안 함수 $J(\omega, \theta, \xi, \alpha, \beta)$ 에 대한 이원문제의 목적함수 $Q(\alpha)$ 는 식 (7)의 분리 가능한 문제에서의 목적함수와 동일하다. 이때 주어진 목적함수 $Q(\alpha)$ 를 최대화하기 위한 조건은 $\sum_{i=1}^N \alpha_i d_i = 0$ 와 $0 \leq \alpha_i \leq c$ ($i = 1, 2, \dots, N$)로 α_i 의 상한이 주어지는 것만 다르다. 여기서도 QP 과정에 따라 최적의 라그랑지안 계수 α_{10} 를 찾으려면, 최적의 가중치 벡터 ω_0 와 최적의 바이어스 θ_0 는 각각 식 (8)로 구해지며, 그 분류식은 식 (10)과 동일하다.

이상에서 살펴 본 분류식은 선형의 결정면만을 설명한 것으로 모든 문제에 적용될 수는 없으며, 비선형 분류도 가능하게 하기 위해서는 좀 더 일반적인 결정면을 가지도록 하여야 한다. 이를 위한 방법으로 SVM에서는 입력벡터 \mathbf{x} 를 고차원의 특징공간 \mathbf{z} 로의 시싱을 이용한다. 이는 분리 가능한 초윌면이 입력공간에서는 비선형이나 변환을 통한 고차원의 특징공간에서는 선형적으로 표현될 수 있다는 것이다[2]. 그러나 이를 위한 목적함수 $Q(\alpha)$ 를 최대화하는 것은 고차원의 특징공간에서 내적 ($\mathbf{z}(\mathbf{x}) \cdot \mathbf{z}(\mathbf{x}_i)$)의 계산을 요구한다. 어떤 조건하에서, 내적계산은 아주 비효율적이거나 커널함수 K 를 사용함으로써 효율적인 계산이 가능하다[1, 2, 8]. 여기서 커널함수의 특징공간과의 관계는 다음과 같다. 즉,

$$K(\mathbf{x}, \mathbf{x}_i) = \mathbf{z}(\mathbf{x})^T \cdot \mathbf{z}(\mathbf{x}_i) \quad (13)$$

이며, 이때 사상을 위한 커널함수로는 polynomial, radial basis, 그리고 s-자형 함수 등이 이용된다[2]. 따라서 비선형 분류도 가능하게 하는 최적의 초윌면을 구성하기 위한 목적함수 $Q(\alpha)$ 는

$$Q(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j d_i d_j K(x_i, x_j) \quad (14)$$

로 표현되며, 이는 선형 분류의 목적함수에서 입력벡터의 내적을 커널함수로 대체한 것에 불과하다. 여기서의 조건도 $\sum_{i=1}^N \alpha_i d_i = 0$ 와 $0 \leq \alpha_i \leq c$ 로 선형적으로 분리 불가능한 경우와 동일하다. 또한 특징공간에서 최적의 분리를 위한 초윌면의 분류식도 식 (10)과 유사한 다음과 같이 나타낼 수 있다. 즉,

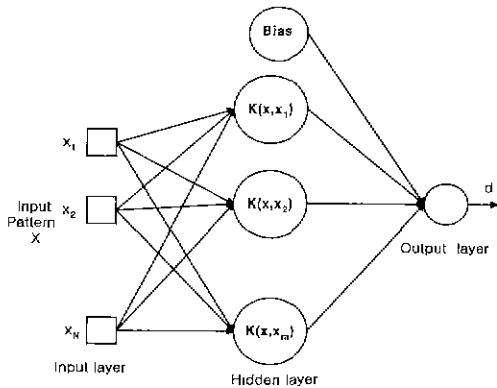
$$f(\mathbf{x}) = \text{sgn}[\sum_{i \in SV_s} \alpha_i d_i K(x, x_i) + \theta_o] \quad (15)$$

이다. 여기서 SVs는 support vector를 나타내며, 최적의 가중치 벡터 w_o 와 최적의 바이어스 θ_o 는 각각

$$w_o = \sum_{i \in SV_s} \alpha_i d_i z(x_i)$$

$$\theta_o = -\frac{1}{2} \sum_{i \in SV_s} w_o d_i [K(x_i, x_i) + K(x_i, x_i)] \quad (16)$$

로 계산될 수 있다. 다음의 (그림 1)은 SVM의 구조를 나타낸 것이다. 그림에서 $K(x, x_i)$ 는 내적 커널을 의미한다.



(그림 1) Support vector machine의 구조

지금까지 살펴본 SVM을 이용한 패턴의 분류는 가장 가까운 패턴들로부터 거리가 최대인 초평면을 찾는 것으로 요약될 수 있다. 이를 위한 과정은 분류문제를 QP 문제로 변환하여 해결하는 것이다. 그러나 QP 과정에서는 계산적인 복잡성과 구현의 어려움과 같은 제약점이 있으며, 문제의 규모가 커질수록 그 제약점은 더욱 더 심화된다. 결국 좀 더 빠르고 구현이 용이한 새로운 해결책이 요구되며, 본 연구에서는 빠르고 구현성이 뛰어난 학습알고리즘으로 알려진 kernel-adatron 알고리즘에서 최적해로의 수렴에 따른 발진을 억제하여 그 수렴속도를 좀 더 개선시키는 새로운 학습알고리즘을 제안한다.

3. 모멘트를 이용한 새로운 kernel-adatron 알고리즘

$$\sum_{i=1}^N \alpha_i d_i = 0 \text{의 조건하에서 오목한 목적함수 } Q(\alpha)$$

를 최소화하는 α 를 찾는 것은 기울기상승법의 이용으로 가능하다. 이는 변수 α_i 에 대한 2차 함수로 표현되는 등식의 선형 제약조건을 가지는 최적화 문제인 표준 QP 문제이다. 여기서 조건을 포함하는 목적함수인 라그랑지안 함수 $Q(\alpha)$ 는

$$Q(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j d_i d_j K(x_i, x_j) - \lambda \sum_{i=1}^N \alpha_i d_i \quad (17)$$

로 표현된다. 식에서 마지막 항은 $\sum_{i=1}^N \alpha_i d_i = 0$ 의 제약조건을 위한 항이며, λ 는 이를 위한 라그랑지안 계수이다. 한편, 제약조건 $0 \leq \alpha_i \leq c$ 는 α_i 의 값이 음수이면 0으로 c 보다 큰 값이면 c 로 제한하면 된다. 라그랑지안 함수 $Q(\alpha)$ 의 최대화는 미분 기호를 둔 기울기상승법으로 가능하며, 이때 최대화를 위한 α_i 의 변화량 $\delta \alpha_i$ 와 λ 를 각각 구하는 식은

$$\delta \alpha_i(t+1) = \eta \frac{\partial Q}{\partial \alpha_i(t)}$$

$$= \eta \left(1 - \sum_{i=1}^N \alpha_i(t) d_i K(x_i, x_i) - \lambda(t) d_i \right)$$

$$\lambda(t) = \lambda(t-1) + \rho \left(\sum_{i=1}^N \alpha_i(t) d_i \right) \quad (18)$$

이다. 여기서 t 는 반복수(epoch)이며, η 와 ρ 는 각각 학습률(learning rate)로 양수값이다. 식에서 $K(x_i, x_i)$ 는 대칭이며 양수로 정의되는(positive-definite) $N \times N$ 의 행렬 K 의 요소이고, $\delta \alpha_i$ 의 계산을 위해서는 미리 이들 요소값들이 계산되어야 한다. 이는 SVM을 학습시키기 위한 기존의 QP 알고리즘에서 복잡한 계산과 시스템 구현의 어려움을 일으키는 이유이다. Campbell 등이 제안한 kernel-adatron 알고리즘은 퍼셉트론과 유사한 알고리즘인 adatron과 비선형 특징공간에서의 동작을 위한 kernel을 조합한 하이브리드 학습기법이다. 여기서의 adatron은 이론적으로 최적해로의 수렴이 알고리즘의 반복에 지수함수적으로 빠르게 일어난다고 알려져 있다. [3] 결국 이 기법은 SVM의 특징공간에 adatron을 도입하는 것으로 용이한 구현성뿐만 아니라 커널에 의한 비선형 특징공간에서의 단순한 동작 특성도 동시에 얻을 수 있어, 기존 QP 알고리즘이 가지는 계산과 구현상의 제약들을 효과적으로 해결할 수 있다. 또한 여기서는 좀 더 빠른 알고리즘의 수렴을 위해 라그랑지안 계수 λ 를

식 (18)에서와는 달리 할선법을 이용하여 구하였다.

한편, Campbell 등의 kernel-adatron 알고리즘에서는 $\delta \alpha_i$ 의 계산에 기울기상승법을 그대로 이용하고 있어 최적해로 수렴할 때 학습률 η 에 따라서는 발진이 일어날 수도 있다 따라서 수렴에 따른 발진을 막아 줄 수만 있다면 좀 더 빠르게 최적해로 수렴시킬 수 있을 것이다. 이를 위해 본 연구에서는 $\delta \alpha_i$ 를 계산할 때 이전 변화의 속성을 가지는 모멘트를 추가하였다. 이는 현재의 계산방향이 이전의 계산방향을 따르도록 하는 것이다. 이렇게 하면 kernel-adatron 알고리즘이 가지는 우수한 구현성은 그대로 살리면서도 최적해로의 수렴에서 일어나는 발진을 억제시켜 좀 더 빠른 속도로 SVM을 학습시킬 수 있다. 제안된 모멘트를 이용한 $\delta \alpha_i$ 의 계산식은

$$\delta \alpha_{i,(t+1)} = \eta \left(1 - \sum_{j=1}^N \alpha_j(t) d_j K(x_i, x_j) - \lambda(t) d_i \right) + m(\alpha_i(t) - \alpha_i(t-1)) \quad (19)$$

와 같으며, 여기서 m 은 모멘트이다. 따라서 모멘트 항을 이용한 새로운 kernel-adatron 알고리즘의 과정을 정리하면 다음과 같다. 즉,

- 단계 1 : 라그랑지안 계수 $\alpha_i(0)$ 와 γ , 파라미터 c , 학습률 η , 모멘트 m , 그리고 알고리즘의 최대 반복수 t_{max} 를 설정한다.
- 단계 2 : $i, j = 1$ 에서 N 까지 커널요소 $K(x_i, x_j)$ 를 계산한다.
- 단계 3 : 알고리즘의 반복 수 $t = 1$ 에서 t_{max} 까지 다음 단계 4에서 8까지를 수행한다.
- 단계 4 : 만약 $t = 0$ 이면, 라그랑지안 계수 $\lambda(0) = \gamma$ 로, 만약 $t = 1$ 이면, 라그랑지안 계수 $\lambda(1) = -\gamma$ 로 그렇지 않으면 $\lambda(t) = \lambda(t-1) - \omega(t-1) \left(\frac{\lambda(t-1) - \lambda(t-2)}{\omega(t-1) - \omega(t-2)} \right)$ 을 계산한다.
- 단계 5 : $i = 1$ 에서 N 까지 다음 단계 6과 7을 수행한다.
- 단계 6 : $M_i = \sum_{j=1}^N \alpha_j(t) d_j K(x_i, x_j)$ 를 계산한다
- 단계 7 : $\delta \alpha_i(t) = \eta(1 - M_i d_i - \lambda(t) d_i) + m(\alpha_i(t) - \alpha_i(t-1))$ 를 계산하고

만약 $(\alpha_i(t) + \delta \alpha_i(t)) \leq 0$ 이면, $\alpha_i(t) = 0$ 으로

만약 $0 < (\alpha_i(t) + \delta \alpha_i(t)) < c$ 이면,

$$\alpha_i(t) = \alpha_i(t) + \delta \alpha_i(t)$$

만약 $(\alpha_i(t) + \delta \alpha_i(t)) \geq c$ 이면,

$$\alpha_i(t) = c$$
으로 한다

- 단계 8 : $\omega(t) = \sum_{i=1}^N \alpha_i(t) d_i$ 를 계산한다
- 단계 9 : 만약 반복수 t 가 t_{max} 를 넘거나 여유 ϵ 를 계산하여 근사적으로 1이면 학습을 종료하며, 그렇지 않으면 단계 3으로 간다.

$$x = \frac{1}{2} \left((i|d_i = 1, \alpha_i \leq c) (M_i) - (i|d_i = -1, \alpha_i \leq c) (M_i) \right)$$

제안된 알고리즘으로 얻어진 라그랑지안 계수 α_i 와 λ 의 값을 이용하면 시험패턴들을 분류할 수 있다. 이때 특징공간에서 최적으로 패턴들을 분리시킬 수 있는 초평면의 분류식은

$$f(x) = \text{sign} \left[\sum_{i=1}^N \alpha_i d_i K(x, x_i) + \lambda \right] \quad (20)$$

와 같다.

4. 시뮬레이션 결과 및 분석

제안된 kernel-adatron 알고리즘을 이용한 SVM의 성능을 평가하기 위해서 2가지 분류문제를 대상으로 실험하였다. 실험에서 커널함수는 가우스(Gaussian)의 radial basis 함수 $K(x, x_i) = \exp\left(-\frac{\|x-x_i\|^2}{2\sigma^2}\right)$ 를 이용하였다. 라그랑지안 계수 α_i 의 초기값은 랜덤시드(random seed)를 이용하여 0과 1 사이의 임의의 값으로 설정하였으며, 알고리즘의 종료는 반복수가 20,000번($t_{max} = 20,000$) 이상이거나 여유 ϵ 가 1 이상일 때로 하였다.

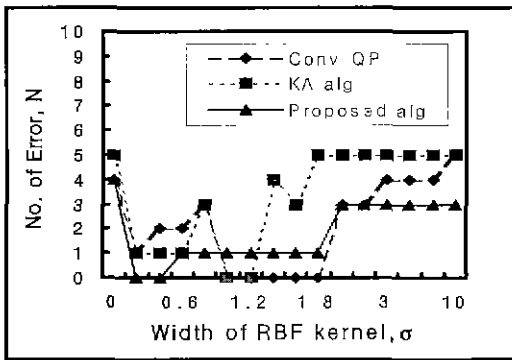
제안된 학습기법의 SVM을 임의의 비선형 분류문제와 실제 암환자를 2부류(초기와 악성)로 분류하는 문제를 대상으로 팬티엄 II-MMX200 컴퓨터를 이용하여 시뮬레이션한 후 그 타당성을 확인하였으며, 기존의 표준 QP 알고리즘과 Campbell 등의 kernel-adatron 알고리즘을 각각 이용한 SVM에 의한 결과와도 비교 고찰하였다.

4.1 비선형 분류문제

제안된 알고리즘의 타당성을 확인하기 위해서 18가지의 입력벡터와 원하는 출력값으로 구성된 선형적으로는 분류될 수 없는 임의의 문제를 구성하였다. <표 1>은 실험에 이용된 9개의 학습데이터와 시험데이터를 각각 나타낸 것이다. 그림에서 학습데이터의 경우 처음 7가지 상태까지는 선형적으로 분리 가능하나 마지막 2가지 상태를 추가함으로써 이 문제는 비선형 분류문제로 바뀐다

<표 1> 18가지 상태의 비선형 분류문제

Training Data			Test Data		
x_1	x_2	Class	x_1	x_2	Class
1	1	-1	2	3	1
3	3	1	2	1	-1
1	3	1	1	2.5	1
3	1	-1	1.5	1	-1
2	2.5	1	3	2	-1
3	2.5	-1	2.5	3	1
4	3	-1	4	2.5	-1
1.5	1.5	1	2.5	2.5	-1
1	2	-1	4	2	1



(그림 2) radial basis 함수의 폭 σ 의 변화에 따른 SVM의 일반화 오차 E(18개 데이터)

(그림 2)는 라그랑지 계수 $\gamma=1.0$, 파라미터 $c=0.1$, 학습률 $\eta=1.0$, 그리고 모멘트 $m=0.1$ 로 하여 radial basis 함수의 폭 σ 의 변화에 따른 시험데이터에 의한 SVM의 일반화 성능을 나타낸 것이다. 그림에서 보면 σ 가 1.0에서 1.8 사이에서는 QP 알고리즘이 상대적으로 약간 우수한 일반화 성능을 보이고 있다. 그러나 제안된 알고리즘의 경우는 다른 알고리즘들에 비하여 σ 의 변화에 따른 일반화 성능의 변화가 상대적으로

적게 일어나는 것으로 보아 커널함수의 내부 파라미터에 덜 의존됨을 알 수 있다. 특히, kernel-adatron 알고리즘 경우는 σ 의 변화에 따른 일반화 성능의 변화가 가장 심함을 알 수 있다. 그림의 결과로부터 σ 의 값은 0.2에서 1.8 사이의 값이 적당하며 3가지 알고리즘 모두에서는 1.0 정도가 가장 적합함을 알 수 있다. 하지만 그림에서의 단순한 일반화 성능비교 만으로는 알고리즘의 우수성을 판단할 수는 없다.

<표 2> radial basis 함수의 폭 σ 의 변화에 따른 학습알고리즘의 수행회수와 일반화 성능

σ	Conv QP			KA alg			Proposed alg.		
	N_{QP}	E_{QP}	(n_p, n_n)	N_{KA}	E_{KA}	(n_p, n_n)	N_p	E_p	(n_p, n_n)
0	20000	0.44	8(3, 5)	20000	0.56	8(4, 4)	2	0.44	8(4, 4)
0.2	20000	0.11	5(1, 4)	2	0.11	3(3, 0)	2	0	9(4, 5)
0.4	20000	0.22	5(1, 4)	2	0.11	3(3, 0)	2	0	9(4, 5)
0.6	20000	0.22	5(1, 4)	2	0.11	5(2, 3)	2	0.11	7(3, 4)
0.8	20000	0.33	7(3, 4)	7	0.33	8(3, 5)	2	0.11	7(3, 4)
1.0	233	0	8(3, 5)	9	0	7(3, 4)	2	0.11	9(4, 5)
1.2	119	0	8(3, 5)	17	0	7(3, 4)	2	0.11	9(4, 5)
1.4	20000	0	7(3, 4)	2	0.44	5(2, 3)	2	0.11	8(4, 4)
1.8	20000	0	6(3, 3)	3	0.33	5(2, 3)	2	0.11	9(4, 5)
2.0	20000	0.33	6(3, 3)	3	0.56	8(4, 4)	2	0.33	9(4, 5)
2.5	20000	0.33	6(3, 3)	2	0.56	8(4, 4)	2	0.33	8(4, 4)
3.0	20000	0.44	6(3, 3)	3	0.56	8(4, 4)	2	0.33	8(4, 4)
5.0	20000	0.44	6(3, 3)	3	0.56	6(2, 4)	2	0.33	8(4, 4)
10.0	20000	0.56	9(4, 5)	3	0.56	9(4, 5)	2	0.33	8(4, 4)

<표 2>는 radial basis 함수의 폭 σ 의 변화에 따른 학습알고리즘의 수행회수 N과 일반화 성능인 분류오차 E, 그리고 SV 수 n를 각각 나타낸 것이다. 여기서 n_p 와 n_n 는 각각 +1과 -1의 부류에 해당하는 SV 수이다. 표에서 기존의 QP 알고리즘은 함수의 폭 σ 가 1.0에서 1.8까지는 시험데이터에 대한 분류오차가 0이나 학습을 위한 반복수가 다른 2가지 알고리즘에 비해 상대적으로 훨씬 많음을 알 수 있다. 또한 kernel-adatron과 제안된 알고리즘의 경우는 기존의 QP 알고리즘에 비해 매우 빠른 학습속도를 보이며, 이는 kernel-adatron과 제안된 알고리즘 각각에 이용되는 adatron의 속성 때문이나 특히 제안된 알고리즘의 경우 학습을 위한 반복수가 일정하게 나타나 σ 값의 변화에 거의 영향을 받지 않으며, kernel-adatron보다 최대 약 8배 정도 더 빠른 학습속도를 보이고 있다. 이 결과는 제안된 알고리즘에서 모멘트 항이 기존의 kernel-adatron 알고리즘에서의 단순한 기울기상승에 따른 발전을 제어하는 속성이 있음

도 확인할 수 있다. 한편 제안된 알고리즘에서도 σ 값의 변화에 따라 분류오차와 SV의 수는 변화됨을 확인할 수 있다. 그러나 SVM의 일반화 성능인 분류오차의 경우 제안된 방법은 기존의 QP나 kernel-adatron 알고리즘에 비해서 우수한 성능이 있음을 확인할 수 있다. SV의 개수에서는 kernel-adatron 알고리즘이 다른 2가지 알고리즘에 비해 다소 작은 수를 보이고 있다.

한편 라그랑지안 함수 $J(\omega, \theta, \alpha)$ 의 계수 α 의 상한을 결정하는 파라미터 c 와 일반화 성능과의 관계를 살펴보기 위한 실험에서는 c 가 일반화의 성능에 거의 영향을 미치지 못함을 3가지 알고리즘 모두의 결과로부터 확인하였다. 이 규모의 비선형 분류문제에서 α 의 변화범위는 0에서 2.0 사이의 값을 가짐을 알 수 있다. 실험에서 c 값의 변화범위는 0.2에서 2.0 사이의 값이 적절한 값으로 나타났으며, 그 사이의 값에서는 거의 동일한 학습반복수와 분류오차율을 나타내었다. 또한 학습률 η 의 변화에 따른 SVM의 성능도 c 에서와 마찬가지로 3가지 알고리즘 모두에서 SVM의 전체 성능에는 크게 영향을 미치지 못하였다.

4.2 암환자 분류문제

실험에 이용한 암환자 분류문제의 데이터는 미국 위스콘신 대학병원에서 제공되는 "Wisconsin breast cancer databases"를 대상으로 하였다[10]. 원 데이터베이스는 699명의 환자를 대상으로 9개의 정수값(1-10)으로 표현된 조건속성과 초기와 말기(초기: 2, 악성: 4)를 나타내는 1개의 분류속성으로 구성된다. 그리고 699명 중에서 16명의 환자에 대해서는 그 속성값들을 빠뜨렸다. 본 실험에서는 속성값이 나타나 있지 않은 16명의 환자를 제외한 나머지 683명의 환자 중 임의의 200명을 대상으로 하고, 100명은 학습을 위한 데이터로 나머지 100명은 시험을 위한 데이터로 이용하였다. 또한 실험에 이용된 데이터들에서 환자의 식별번호인 첫 번째 속성은 제거하였고, 모든 조건속성들은 정수 10으로 나누었으며, 분류속성은 2(초기)이면 -1로 4(악성)이면 1로 표현하였다. 다음의 (그림 3)은 5명의 환자를 대상으로 실제의 데이터와 본 실험에서 사용되는 데이터를 나타내는 예이다. 그림에서 2는 그 속성 데이터가 알려지 있지 않아 실험에서는 학습 데이터로 이용되지 않은 경우이다.

(그림 4)는 라그랑지 계수 $\gamma = 1.0$, 파라미터 $c = 0.1$, 학습률 $\eta = 3.0$, 그리고 모멘트 $m = 0.1$ 로 하여 radial

basis 함수의 폭 σ 의 변화에 따른 시험데이터에 대한 일반화 성능을 보인 것이다. (그림 4)를 (그림 2)의 18개 데이터 분류 결과와 비교할 때 규모가 커질수록 분류오차 0인 빈도가 작게 일어나 SVM의 일반화 성능은 더욱 더 떨어지는 결과를 보이고 있다. 그림에서 보면 기존의 QP 알고리즘에 비해 다른 2개 알고리즘의 일반화 성능이 상대적으로 우수함을 확인할 수 있다. 이는 문제의 규모가 키짐에 따라 QP 알고리즘의 경우 떨어지는 성능을 보이나 나머지 2가지 알고리즘은 개선된 성능을 가짐을 알 수 있다. 특히, 제안된 알고리즘의 경우는 Campbell의 kernel-adatron 알고리즘보다도 더욱 우수한 일반화 성능을 가지며, 그 성능변화가 상대적으로 적어 커널함수의 내부 파라미터인 σ 의 변화에 덜 의존함을 알 수 있다. 기존 QP 알고리즘의 경우는 σ 의 변화에 따른 일반화 성능의 변화가 가장 심하게 변화함을 알 수 있다. 이는 QP 알고리즘이 커널함수의 내부 파라미터에 가장 민감하게 반응함도

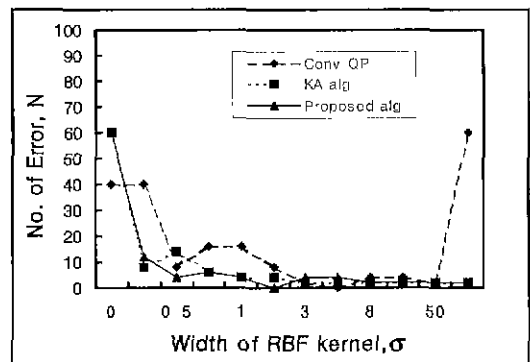
1081791,	6,	2,	1,	1,	1,	1,	7,	1,	1,	2
1084584,	5,	4,	4,	9,	2,	10,	5,	6,	1,	4
1091262,	2,	5,	3,	3,	6,	7,	7,	5,	1,	4
1096800,	6,	6,	6,	9,	6,	2,	7,	8,	1,	2
1099510,	10,	4,	3,	1,	3,	3,	6,	5,	2,	4

(a) 실제데이터

0.6,	0.2,	0.1,	0.1,	0.1,	0.1,	0.7,	0.1,	0.1,	-1
0.5,	0.4,	0.4,	0.9,	0.2,	1,	0.5,	0.6,	0.1,	1
0.2,	0.5,	0.3,	0.3,	0.6,	0.7,	0.7,	0.5,	0.1,	1
0.6,	0.6,	0.6,	0.9,	0.6,	2,	0.7,	0.8,	0.1,	-1
1,	0.4,	0.3,	0.1,	0.3,	0.3,	0.6,	0.5,	0.2,	1

(b) 실험용 데이터

(그림 3) 5명의 환자에 대한 속성데이터 예



(그림 4) radial basis 함수의 폭 σ 의 변화에 따른 SVM의 일반화 오차 E(200개 데이터)

확인할 수 있다. 그림의 결과로부터 σ 의 값은 1.0에서 50 사이의 값이 적당한 값이다 (그림 4)에서도 (그림 2)에서도 동일한 결과를 보이고 있다.

<표 3>도 radial basis 함수의 폭 σ 의 변화에 따른 SVM의 학습알고리즘의 수행회수 N과 일반화 성능인 분류오차 E, 그리고 SV 수 n를 각각 나타낸 것이다. 표에서도 제안된 알고리즘이 다른 2가지 알고리즘에 비해 빠르면서도 우수한 일반화 성능이 있음을 확인할 수 있다. 특히, kernel-adatron에서 시험데이터를 완전하게 분류할 수 있는 경우가 하나도 나타나지 않는 것은 학습시에 미리 결정해야 하는 파라미터의 설정이 상대적으로 어려운 것으로 추측된다. 한편 함수의 폭 σ 의 변화에 따른 분류오차 E가 0.1 미만인 경우에 대해서만 살펴보면, 제안된 알고리즘이 가장 넓은 변화 폭을 가짐을 알 수 있으며, 학습속도 측면에서도 빠른 결과를 보이고 있다. 이는 <표 3>에서처럼 제안된 알고리즘은 σ 값의 변화에 가장 영향을 적게 받기 때문이다. 동일한 일반화 성능에서 제안된 방법은 기존의 QP 알고리즘에 비해 약 26배 정도, kernel-adatron 알고리즘에 비해서는 약 1.5배 이상 빠르다고 할 수 있다. SV의 개수도 제안된 알고리즘이 다른 2가지 알고리즘에 비해 상대적으로 적음을 알 수 있다. 따라서 제안된 알고리즘은 기존의 QP 알고리즘과 kernel-adatron 알고리즘에서 발생하는 학습속도와 일반화 성능 두 가지 측면을 함께 개선시킬 수 있다.

<표 3> 커널함수의 폭 σ 의 변화에 따른 학습알고리즘의 수행회수와 일반화 성능

σ	Conv. QP			KA alg			Proposed alg.		
	N_{QP}	E_{QP}	n_{QP} (n_p, n_n)	N_{KA}	E_{KA}	n_{KA} (n_p, n_n)	N_P	E_P	n_P (n_p, n_n)
0	20000	0.4	20(10,10)	20000	0.6	40(20,20)	2	0.6	40(20,20)
0.2	30	0.4	50(34,16)	4	0.08	20(2,18)	4000	0.12	44(28,16)
0.5	30	0.08	24(16, 8)	3	0.14	26(18, 8)	18	0.04	18(8,10)
0.8	30	0.16	18(10, 8)	3	0.06	14(8, 6)	2	0.06	14(8, 6)
1.0	30	0.16	14(6, 8)	3	0.04	16(8, 8)	2	0.04	16(8, 8)
2.0	30	0.08	12(8, 4)	2	0.02	14(6, 8)	2	0	22(10,12)
3.0	30	0.02	12(8, 4)	2	0.02	26(12,14)	2	0.04	14(6, 8)
5.0	53	0	18(10, 8)	2	0.02	40(20,20)	2	0.04	18(8, 10)
8.0	20000	0.04	4(2, 2)	2	0.02	40(20,20)	2	0.02	26(12,14)
10.0	109	0.04	40(20,20)	2	0.02	40(20,20)	2	0.02	40(20,20)
50.0	20000	0.02	24(4,20)	2	0.02	40(20,20)	2	0.02	40(20,20)
100.0	20000	0.6	24(4,20)	2	0.02	40(20,20)	2	0.02	40(20,20)

여기서도 학습오차와 일반화 능력 사이의 상관관계

를 제어하는 파라미터인 α 의 상한을 결정하는 파라미터 c 와 학습률 η 의 영향은 18개 데이터의 비선형 분류문제에서와 유사한 결과를 얻을 수 있었다. 결국 제안된 방법은 기존의 QP 알고리즘이 가지는 학습속도와 구현의 어려움 및 Campbell의 kernel-adatron 알고리즘에서 발생하는 발전현상을 동시에 해결할 수 있음을 확인할 수 있다. 특히, 제안된 알고리즘의 경우 18개 데이터의 비선형 분류와 200명의 암환자 분류문제의 실험결과로부터 학습속도의 변화가 거의 없어 상대적으로 큰 규모의 문제에서 더욱 더 우수한 성능이 있음을 확인할 수 있다.

5. 결 론

본 연구에서는 기술기상승법에서 해의 변화에 과거의 속성을 반영하는 모멘트 항을 추가한 수정된 kernel-adatron 알고리즘을 제안하고 이를 SVM의 학습방법으로 이용하였다. 이렇게 하면 기술기상승법에서 일어나는 최적해로의 수렴에 따른 발진을 억제하여 그 수렴속도를 좀 더 개선시키는 모멘트의 장점과 kernel-adatron 알고리즘의 구현 용이성을 그대로 살릴 수 있다.

제안된 학습알고리즘의 SVM을 임의의 18개의 비선형 데이터를 분류하는 문제와 실제 200명의 암환자를 2부류(초기와 악성)로 분류하는 문제에 각각 적용하여 시뮬레이션한 결과, 기존의 QP 알고리즘과 Campbell 등의 kernel-adatron 알고리즘을 이용한 SVM과 비교할 때 학습시간과 시험데이터의 분류오차로 표현된 일반화 성능에서 더욱 우수한 결과를 보였다. 특히, 문제의 규모가 클수록 학습속도 면에서 우수한 성능을 보였다.

향후 제안된 알고리즘의 SVM을 좀 큰 규모의 문제와 희귀문제 등과 같은 좀 더 다양한 분야에의 연구가 계속 진행되어야 할 것이며, SVM의 구현 및 적용적인 최적의 모멘트 값 설정에 대한 연구도 남아 있다.

참 고 문 헌

[1] V Vapnik, The Nature of Statistical Learning Theory, Springer Verlag, 1995.
 [2] M.O Stitson, J.A.E Weston, A. Gammerman, V. Vovk, and V. Vapnik, "Theory of Support Vector Machines." Technical report CSD-TR-96-17. Royal

Holloway. Univ. of London, May 1998.

[3] C. Campbell and N. Cristianini, "Simple Learning Algorithms for Training Support Vector Machines." <http://lara.enm.bris.ac.uk/cig/gzipped/KA-ieee.pi.gz>

[4] E.E. Osuna, R. Freund, and F. Girosi, "Training Support Vector Machines : An Application to Face Detection," Proc. Computer Vision and Pattern Recognition '97, Puerto Rico, June 1997.

[5] E.E. Osuna, R. Freund, and F. Girosi, "An Improved Training Algorithm for Support Vector Machines," Proc. of IEEE NNSP'97, New York, Sep 1997.

[6] J.C. Platt. "Fast Training of Support Vector Machines Using Sequential Minimal Optimization," In Advances in Kernel Methods : Support Vector Learning, ed., MIT Press. Cambridge, 1998.

[7] C. Cortes, and V. Vapnik, "Support Vector Networks," Machine Learning, pp 273-297, 1995.

[8] S. Haykin. Neural Networks : A Comprehensive Foundation. Prentice-Hall Inc., London, pp.318-349, 1999.

[9] A. Cichocki, and R. Unbehauen, "Neural Networks

for Signal Optimization and Signal Processing," John Wiley & Sons, New York, 1992.

[10] W. H. Wolberg, "Wisconsin Breast Cancer Database," Univ. of Wisconsin Hospital, July 1992(<ftp://mlftp.dice.unica.it/pub/mac...sin/breast-cancer-wisconsin.data>)



조 용 현

e-mail : yhcho@cuth.cataegu.ac.kr

1979년 경북대학교 공과대학

전자공학과(공학사)

1981년 경북대학교 대학원 전자

공학과(공학석사)

1993년 경북대학교 대학원 전자

공학과(공학박사)

1983년~1984년 삼성전자(주) 사원

1984년~1987년 한국전자통신연구소 연구원

1987년~1997년 영남전문대학 전자과 부교수

1997년~현재 대구효성가톨릭대학교 컴퓨터정보통신공학부 조교수

관심분야 : 신경망, 병렬분산처리, 신호처리, 영상인식, 교환기 등