Journal of Korean
Data & Information Science Society
2002, Vol. 13, No.2 pp. 55~64

The Efficiency of Boosting on SVM

Kyungha Seok[1], Tae Wook Ryu[2]

Abstract

In this paper, we introduce SVM(suppprt vector machine) developed to solve the problem of generalization of neural networks. We also introduce boosting algorithm which is a general method to improve accuracy of some given learning algorithm. We propose a new algorithm combining SVM and boosting to solve classification problem. Through the experiment with real and simulated data sets, we can obtain better performance of the proposed algorithm.

Keywords : SVM, Boosting, Classification, Learning algorithm

1. 서론

신경망을 이용한 문제 해결방법은 아주 폭넓게 이용되고 있는 현실이다. 그러나 신경망을 학습시키는 과정에서 과대적합(overfitting)을 따르는 모형을 생성할 수 있어 일반화에 어려움이 많은 실정이고[5], 또한 학습에 중요한 역할을 하는 모수(parameter)들을 설정하는 과정이 객관적이고 분석적인 방법을 거치는 것이 아니라 사용자의 경험에 의존하는 문제점을 가진다. 그러므로 문제에 따라 사용자에 따라 해석이 달라지는 문제점을 가지게 된다. 그 뿐만 아니라 국소최적점에 빠지는 학습으로 인해 최적의 해를 구하지 못하는 예가 종종 있다.

이러한 신경망이 가진 단점을 보완하기 위하여 개발된 SVM은 VC이론을 근거로 하여 개발되었다[5,6,10,11]. 그로 인해 예측을 잘하고 또한 벌칙항(penalty)을 이용하여 과대적합을 피하는데 성공적이라는 것이 여러 연구의 결과에서 보여주고 있다[1,5,9]. SVM은 볼록함수(convex function)를 최소화하는 학습을 진행하기 때문에 신경망과는 달리 유일한 최적의 해를 구할 수 있다는 장점을 가지고 있다. 또한 함수근사의 문제에 있어서 이상치에 둔감하다는 장점도 가지고 있다.

부스팅기법은 분류문제를 위해 Yoav Freun와 Robert Shapire[3]가 1995년 개발했

---

1. Associate Professor, Data Science Dept., Inje University, Obangdong, Kimhae, 621-749, Korea.
E-mail : skh@stat.inje.ac.kr
2. Equity Asset Dept., Dime Investment ltd., CCMM Building, 12, Youido-dong, Youngdungpo-Gu, Seoul, 150-010 Korea

는데, 훈련용 자료를 복원 재 추출하여 여러 개의 모형을 만든다. 그리고 생성된 모형의 가중평균을 최종모형으로 결정하는 방법이다. 이 방법이 오차를 줄일 수 있다는 것이 여러 연구 결과에서 밝혀졌다.

본 논문에서는 부스팅 기법을 SVM에 적용하여 분류문제에서 오분류율을 줄일 수 있음을 제안한다. 제2절에서는 서포트벡터 분류기에 대해 간단히 소개를 하고, 제3절에서는 부스팅에 대한 소개를 한다. 그리고 4절에서는 실험결과를 소개한다. 신경망과 의사결정나무모형을 비교해 본 결과 실자료[4](Wisconsin breast cancer, German credit data, Iris data)와 모의시험자료에서 오분류율이 확연히 줄어듦을 알 수 있었다.

2. 서포트벡터 분류기

SVM은 원래 분류를 위해 Vapnik과 공동연구자들에 의해 개발되었으며 많은 응용분야에서 좋은 결과를 보여주고 있어 그 이용이 점점 더 확대되고 있는 실정이다 [1,5,6,9,10]. 훈련자료 $\{(\mathbf{x}_i, y_i), i=1,\ldots,n)\} \subset X\times\{-1,1\}$가 주어졌다고 가정한다. 여기서 X는 d차원의 입력벡터공간 R^d를 나타낸다. 모든 훈련자료가 선형분리가 가능한 경우에, 서포트벡트 분류기(Support Vector Classifier : SVC)는 다음의 식

$$min_{\mathbf{w},b} \langle \mathbf{w}\cdot\mathbf{w}\rangle \tag{1}$$
$$s.t.\quad y_i\{\langle \mathbf{w}\cdot\mathbf{x}_i\rangle + b\}\ge 1$$

을 만족하는 maximal margin classifier, 즉 $sign(f(\mathbf{x}))$, $f(\mathbf{x}) = \langle \mathbf{w}\cdot\mathbf{x}\rangle + b$ 로 이해를 하면 된다. 여기에서 $\langle \mathbf{x}\cdot\mathbf{y}\rangle$는 $\mathbf{x}$ 와 $\mathbf{y}$의 내적을 나타낸다. 식(1)의 최소화 문제를 라그랑제배수 α_i를 이용하여 표현하면

$$max_{\alpha} L(\mathbf{w}, b, \alpha) = \frac{1}{2}\langle \mathbf{w}\cdot\mathbf{w}\rangle - \sum_{i=1}^{n}\alpha_i[y_i(\langle \mathbf{w}_i\cdot\mathbf{x}_i\rangle + b) - 1] \tag{2}$$

로 된다. (2)식을 $\mathbf{w}$와 b에 대해서 편 미분을 한 후 식을 풀면

$$\mathbf{w} = \sum_{i=1}^{n} y_i\alpha_i\mathbf{x}_i, \tag{3}$$
$$\sum_{i=1}^{n} y_i\alpha_i = 0$$

를 얻을 수 있다. (3)식을 (2)식에 대입하면

$$max_{\alpha} L(\mathbf{w}, b, \alpha) = \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{n} y_i y_j\alpha_i\alpha_j\langle \mathbf{x}_i\cdot\mathbf{x}_j\rangle \tag{4}$$
$$s.t.\quad \sum_{i=1}^{n} y_i\alpha_i = 0,\quad \alpha_i > 0, i=1,\ldots,n.$$

과 같은 문제로 귀결된다. 즉, (4)식을 최대화하는 $\boldsymbol{\alpha}$를 구하여 (3)식으로부터 $\mathbf{w}$를 구한다. 그리고 (1)식으로부터 b를 구하면 우리가 원하는 f가 구해진다. 이렇게 구해진 f를 $\boldsymbol{\alpha}$를 이용하여 표현하면

$$f(x) = \sum_{i=1}^{n} y_i \alpha_i \langle \mathbf{x}_i \cdot \mathbf{x} \rangle + b \tag{5}$$

로 나타난다. 그런데 $\alpha^*, \mathbf{w}^*$, 그리고 b^*가 위에서 구해진 최적의 해라면 Karush-Kuhn -Tucker 의 조건에 의해

$$\alpha_i^*[y_i(\langle \mathbf{w}^*_i \cdot \mathbf{x}_i \rangle + b^*) - 1] = 0,\ i = 1, \ldots, n.$$

를 만족해야한다. 즉, $[y_i(\langle \mathbf{w}^*_i \cdot \mathbf{x}_i \rangle + b^*) - 1] \neq 0$인 자료에 대응하는 α_i^*는 0이 되어야 하고 $\alpha_i^* \neq 0$에 대응하는 자료는 $[y_i(\langle \mathbf{w}^*_i \cdot \mathbf{x}_i \rangle + b^*) - 1] = 0$를 만족해야 한다. 여기에서 $\alpha_i \neq 0$에 대응하는 자료들을 서포트벡터(Support Vector, SV)라고 한다. 그러므로 (5)식의 f는

$$f(x) = \sum_{i \in SV} y_i \alpha_i \langle \mathbf{x}_i \cdot \mathbf{x} \rangle + b$$

로 표현된다.

훈련자료에 잡음이 있어 선형분리가 불가능 한 경우 커널 K와 일반화모수 C를 사용하여

$$f(x) = \sum_{i \in SV} y_i \alpha_i K(\mathbf{x}_i, \mathbf{x}) + b \quad , 0 < \alpha_i < C,\ i = 1, \ldots, n.$$

를 구할 수 있다[2,10]. 많이 사용되고 있는 커널은 RBF(Radial Basis Function)

$$K(\mathbf{x}, \mathbf{y}) = \exp\left\{ -\frac{\|\mathbf{x} - \mathbf{y}\|^2}{\sigma^2} \right\}$$

이다. 여기에서 σ는 커널모수로써 SVC의 수행능력에 많은 영향을 끼친다.

3. 부스팅(Boosting)

부스팅 기법은 1984년 Valiant[8]에 의해 처음으로 개념이 소개되고 난 이후 여러 분야에서 적용을 하여 좋은 결과를 얻었다. 그 이후로 부스팅 기법을 발전시키기 위한 연구가 계속 되었는데 Freun와 Shapire가 1995년 개발한 적응식 부스팅(AdaBoosting)기법[3]이 현재 많이 사용되고 좋은 결과를 보이는 것으로 알려져 있다. 본 연구에서는 Freun와 Shapire의 1999년 논문[4]을 참고하여 설명한다.

부스팅은 자료가 $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)$로 주어 졌을 때, 부스트랩기법을 이용하여 여러 개의 분류기를 만들고 그들의 가중평균으로 최종 분류기를 만드는 방법이다. 부스트랩 개념과 조금 다른 것은 각각의 자료가 재 추출 될 확률이 추정결과에 따라서 달라진다는 것이다. 즉, 제대로 분류된 데이터에 대해서는 재 추출될 확률을 줄여주

고 그렇지 않은 데이터에 대해서는 재 추출될 확률을 크게 하는 것이다. 이해를 돕기 위해 알고리즘을 소개한다[4].

여기서 a_t 는 분류기 h_t의 가중치를 나타내는 중요한 값이다. 만약 ε_t 가 1/2작다면, $a_t \geq 0$이고, ε_t 이 작아진다면 a_t 는 커질 것이다. h_t는 초기분류기(weak learner)로 표현을 하는데 신경망, 의사결정나무 등 분류를 위한 알고리즘이 사용된다. 이 알고리즘에서는 재 추출하다 횟수 T가 중요한 역할을 한다. T의 크기에 따라 최종 분류기의 수행능력에 차이가 난다. 또한 a_t를 어떻게 설정하느냐에 따라서도 최종 분류기의 수행능력이 많은 차이를 보인다([3],[4]).

1. 훈련용 데이터 $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)$두고 $(\mathbf{x}_i, y_i)$가 재 추출하다 될 확률을

   $$D_1(i) = \frac{1}{n} \text{로 초기화}$$

2. *for* $t = 1, \ldots T$

   D_t를 사용하여 훈련용 데이터를 재 추출하여 error의 분포가

   $\varepsilon_t = \Pr_{i \sim D_t}[h_t(x_i) \neq y_i]$ 인 분류기 h_t를 생성한다.

3. $a_t = \frac{1}{2} \ln\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right)$ 를 계산한다.

4. $$D_{t+1}(i) = \frac{D_t(i)}{Z_t} \times \begin{cases} e^{-a_t} & \text{if } h_t(x_i) = y_i \\ e^{a_t} & \text{if } h_t(x_i) \neq y_i \end{cases}$$
   $$= \frac{D_t(i)\exp(-a_t y_i h_t(x_i))}{Z_t},$$

   로 최신화 한다. 여기에서 Z_t는 정규화 상수이다.

   end

5. 최종결과 $H(x) = sign\left(\sum_{t=1}^{T} a_t h_t(x)\right)$

4. 모의실험(Simulation)

본 논문의 실험에서는 신경망과 의사결정나무는 SAS Enterprise Miner 4.0을 이용하였고, 이 논문에서 제안하는 부스팅-서포트벡터 분류(SVC on Boosting : SOB)는 MATLAB Release12를 프로그래밍 하였다. 그리고 커널함수는 RBF 커널을 사용하였다. 모의실험 데이터와 실제자료(Wisconsin Breast Cancer Data, German Credit Data, Iris Data)에 대하여 각 알고리즘(신경망, 의사결정나무, SVM, SOB)별로 훈련

오분류율(Training Error)와 평가오분류율(Test Error)를 구하였다. 자료의 수가 크지 않기 때문에 샘플링(Sampling)에 의해 생기는 위험을 줄이기 위하여 50번 반복 수행하였다. SVC에서 사용되는 모수 C는 큰 영향력을 끼치지 않는 것으로 알려져 있으므로 C=10으로 고정하였다. 그리고 SVC의 수행능력에 큰 영향력을 끼치는 커널 모수 σ는 0.05, 0.2 그리고 1로 변화시켜가면서 영향력을 관찰하였다. SOB에서 σ는 0.05, 0.2, 1.2로 변화시켰다. T는 사전실험의 결과 10이하이면 충분할 것으로 사려되어 5, 10으로 고정하여 데이터를 분석하였다.

4.1 모의실험1

모의실험1과 모의실험2의 자료는 Wahba et al.에서 이용된 자료인데[11], 첫 번째 데이터는 $\sin(4\pi x)/4 - y + 0.37 = 0$과 $\sin(4\pi x)/4 - y + 0.63 = 0$을 기반으로하여 생성되었다. $\sin(4\pi x)/4 - y + 0.37 = 0$의 곡선 이하에서 1이 될 확률은 0.95이고, -1이 될 확률은 0.05이다. $\sin(4\pi x)/4 - y + 0.63 = 0$ 곡선 이상에서는 -1이 될 확률은 0.95이고, 1이 될 확률은 0.05이다. 그리고 $\sin(4\pi x)/4 - y + 0.37 = 0$과 $\sin(4\pi x)/4 - y + 0.63 = 0$의 곡선 사이에서는 1과 -1이 될 확률이 0.5로 동일하다. 이러한 분포를 가진 크기가 1000인 표본을 생성하여 실험하였다.

Table 1. 모의실험1 결과

알고리즘	σ	T	Training Error (평균 / 표준편차)	Test Error (평균 / 표준편차)
신경망	-	-	0.2025 / 0.0075	0.1966 / 0.0214
의사결정나무	-	-	0.1425 / 0.0065	0.1886 / 0.0210
SVM	0.05	-	0.1365 / 0.0064	0.1928 / 0.0192
	0.2	-	0.2351 / 0.0073	0.2423 / 0.0227
	1	-	0.4330 / 0.0060	0.4376 / 0.0336
SOB	0.05	5	0.0375 / 0.0079	0.2213 / 0.0216
		10	0.0190 / 0.0052	0.2227 / 0.0215
	0.2	5	0.1510 / 0.0099	0.1829 / 0.0210
		10	0.1502 / 0.0109	0.1807 / 0.0223
	1	5	0.2007 / 0.0126	0.2173 / 0.0234
		10	0.1997 / 0.0126	0.2159 / 0.0290
	1.2	5	0.2030 / 0.0124	0.2189 / 0.0234
		10	0.2027 / 0.0121	0.2190 / 0.0235

모의실험1의 데이터의 경우에는 신경망과 SVC, 의사결정나무, SOB순으로 평가오분류율이 작아지는 것을 볼 수 있다. SVM의 경우에는 σ에 변화에 많은 영향을 받아

오분류율에 큰 차이를 보이지만, SOB에 경우에는 어느 정도 영향을 받지만 σ값과 T값에 큰 영향을 받지 않고 4가지 모형중에 가장 적은 오분류율 0.1807을 σ가 0.2, T가 10일 때 가짐을 알 수 있다.

4.3 모의실험2

모의실험2의 데이터는 $x^2+y^2=0.85^2$과 $x^2+y^2=0.7^2$을 기반으로 하여 생성되었다. $x^2+y^2=0.85^2$의 원 바깥쪽에서 1이 될 확률은 0.95이고, -1이 될 확률은 0.05이다. $x^2+y^2=0.7^2$의 원 안쪽에서는 -1이 될 확률은 0.95이고, 1이 될 확률은 0.05이다. 그리고 $x^2+y^2=0.85^2$과 $x^2+y^2=0.7^2$의 원 사이에서는 1과 -1이 될 확률이 0.5로 동일하다.

Table 2. 모의실험2 결과

알고리즘	σ	T	Training Error (평균 / 표준편차)	Test Error (평균 / 표준편차)
신경망	-	-	0.1433 / 0.0056	0.1566 / 0.0195
의사결정나무	-	-	0.1350 / 0.0025	0.1866 / 0.0200
SVC	0.05	-	0.0703 / 0.0063	0.2093 / 0.0214
	0.2	-	0.1955 / 0.0057	0.2113 / 0.0181
	1	-	0.3815 / 0.0070	0.3871 / 0.0302
SOB	0.05	5	0.0014 / 0.0011	0.2237 / 0.0214
		10	0.0001 / 0.0004	0.2278 / 0.0228
	0.2	5	0.1192 / 0.0107	0.1894 / 0.0214
		10	0.1059 / 0.0104	0.1893 / 0.0194
	1	5	0.1461 / 0.0074	0.1547 / 0.0139
		10	0.1462 / 0.0075	0.1548 / 0.0139
	1.2	5	0.1469 / 0.0084	0.1555 / 0.0149
		10	0.1468 / 0.0086	0.1555 / 0.0148

모의실험2의 데이터의 경우에는 SVC와 의사결정나무, 신경망 그리고 SOB순으로 평가오분류율이 작아지는 것을 볼 수 있다. 모의실험1과 마찬가지로 SVC의 경우에는 σ에 변화에 많은 영향을 받아 오분류율에 큰 차이를 보이지만, SOB에 경우에는 어느 정도 영향을 받지만 σ값과 T값에 큰 영향을 받지 않고 4가지 모형 중에 가장 적은 오분류율 0.1547을 σ가 1, T가 5일 때 가짐을 알 수 있다.

4.3 Wisconsin Breast Cancer Data

Wisconsin Breast Cancer Data는 30개의 독립변수로 유방종양의 상태가 양성인지 악성인지 파악하는 문제이다. 30개의 독립변수는 이미지 정보인데 그 이미지의 세포핵의 특징을 나타내고 있다. 데이터의 개수는 569이며, 30개의 독립변수는 10가지의 특징을 3가지로 분리해서 측정한 것이다. 10가지의 특징은 다음과 같다.

반지름, 구조, 주변길이, 면적, 부드러운 정도(반지름에서 지역적인 변동), 밀집도(주변길이 제곱/면적-1.0), 오목함(외곽의 오목한 부분의 정도), 오목한 지점, 대칭성, 분열된 차원이다.

30개의 독립변수를 가지고 있기 때문에 계산시간, 다중공선성 문제 때문에 변수를 선택하였다. 필요한 변수를 선택하기 위해서 SAS E-miner에서 제공하는 Variable selection노드를 이용할 수 있다. 카이제곱 통계량을 이용한 결과 W_perimeter(최소-최대주변길이), W_concave_point(최대-최소오목지점), M_concave_points(평균 오목지점), W_texture(최대-최소 조직구조)4개의 변수가 나왔다. 또한, 상관관계 방법을 이용한 결과 W_perimeter, W_concave_point, M_concave_points, W_texture, W_area(최대-최소 면적) 5개의 변수가 선택되었다. 두 방법 중 상관관계를 이용한 모형이 우수하여 5개의 독립변수를 사용하여 데이터를 분석하였다.

Table3. Wisconsin Breast Cancer Data 실험 결과

Algorithm	σ	T	Training Error (평균 / 표준편차)	Test Error (평균 / 표준편차)
신경망	-	-	0.0263 / 0.0021	0.0352 / 0.025
의사결정나무	-	-	0.0131 / 0.0052	0.0529 / 0.024
SVC	0.05	-	0.0104 / 0.0030	0.0451 / 0.018
	0.2	-	0.1071 / 0.0057	0.1014 / 0.0224
	1	-	0.3702 / 0.0027	0.3489 / 0.0064
SOB	0.05	5	0.0000 / 0.0000	0.0406 / 0.0149
		10	0.0000 / 0.0000	0.0444 / 0.0160
	0.2	5	0.0001 / 0.0006	0.0471 / 0.0139
		10	0.0000 / 0.0000	0.0465 / 0.0139
	1	5	0.0296 / 0.0060	0.0332 / 0.0118
		10	0.0285 / 0.0056	0.0330 / 0.0122
	1.2	5	0.0313 / 0.0062	0.0343 / 0.0130
		10	0.0308 / 0.0071	0.0337 / 0.0127

Wisconsin Breast Cancer Data의 경우 마찬가지로 SOB가 4가지모형 중에 가장 적은 평가오분류율 0.0332를 σ가 1, T가 5일 때 가짐을 알 수 있다.

4.4 German Credit Data

German Credit Data의 수는 1000개이며, 20개의 독립변수로 이루어져 있다. 20개의 독립변수 중 7개는 연속형이고, 13개는 이산형 변수이다. 목표변수는 신용의 좋고 나쁨을 나타내는 이산형 변수이다. 독립변수는 다음과 같다. 당좌예금구좌(질적), 거래를 한 달 수(연속), 신용상태(질적), 대출용도(질적), 신용점수(연속), 보통예금(질적), 일한 기간(질적), 비상금(연속), 성별 및 결혼여부(질적), 거주형태, 계좌 개수(연속), 직업(질적), 부양가족(연속), 전화등록여부(질적), 외국인 직장인여부(질적)이다. 결측치는 없으며 신용 좋음의 빈도수는 700이며, 신용 나쁨의 빈도수는 300이다.

German Credit Data의 경우에도 Wisconsin Breast Cancer Data 의 경우와 마찬가지로 변수선택과정을 거쳐 11개의 독립변수를 선택하여 분석을 하였다.

Table 4. German Credit Data 실험 결과

알고리즘	σ	T	Training Error (평균 / 표준편차)	Test Error (평균 / 표준편차)
신경망	-	-	0.2175 / 0.0012	0.2624 / 0.0392
의사결정나무	-	-	0.2330 / 0.0032	0.3024 / 0.0423
SVC	0.05	-	0.0000 / 0.0000	0.3105 / 0.0169
	0.2	-	0.0630 / 0.0052	0.2863 / 0.0118
	1	-	0.3000 / 0.0000	0.3000 / 0.0000
SOB	0.05	5	0.0001 / 0.0005	0.3281 / 0.0225
		10	0.0000 / 0.0000	0.3349 / 0.0217
	0.2	5	0.0000 / 0.0002	0.3193 / 0.0196
		10	0.0000 / 0.0000	0.3238 / 0.0186
	1	5	0.1541 / 0.0116	0.2972 / 0.0271
		10	0.1248 / 0.0122	0.2933 / 0.0215
	1.2	5	0.1832 / 0.0121	0.2888 / 0.0230
		10	0.1610 / 0.0116	0.2841 / 0.0223

4가지 모형 중에 신경망이 가장 적은 평가오분류율 0.2624을 가지고 있다. 다음으로는 SOB가 평가오분류율이 0.2841고 나타났다.

4.5 Iris Data

Iris Data의 수는 150개이며, 4개의 독립변수로 이루어져있다. 4개의 독립변수들은 모두 연속형 데이터이다. 목표변수는 붓꽃의 3가지 종류(Setosa, Versilcolor, Viginica)이다. 독립변수는 꽃받침의 길이(sepal length), 꽃받침의 두께(sepal width), 꽃잎의 길이(petal length) 그리고 꽃잎의 두께(petal width)이다. 단위는 모두 cm로 측정되었

다. 결측치는 없으며 각각의 붓꽃의 부류에 대해 모두 50개씩의 자료가 있다. 본 논문에서는 이원분류에 대해서만 고려하므로 Setota와 Versilcolor를 한 분류로 생각해서 분석을 시도했다.

Iris Data의 경우에는 의사결정나무, 신경망과 SVC, SOB순으로 평가오분류율이 작아지는 것을 볼 수 있다. SVC인 경우에는 σ에 변화에 많은 영향을 받아 오분류율에 큰 차이를 보이지만, SOB에 경우에는 어느 정도 영향 받지만 σ값과 T값에 큰 영향을 받지 않고 4가지 모형 중에 가장 적은 평가오분류율 0.0396을 σ가 1.2, T가 10일 때 가진다.

Table 5. Iris Data 실험 결과

알고리즘	σ	T	Training Error (평균 / 표준편차)	Test Error (평균 / 표준편차)
신경망	-	-	0.0242 / 0.0130	0.0460 / 0.0211
의사결정나무	-	-	0.0256 / 0.0110	0.0600 / 0.0321
SVC	0.05	-	0.0000 / 0.0000	0.0431 / 0.0299
	0.2	-	0.1139 / 0.0119	0.1427 / 0.0531
	1	-	0.2973 / 0.0162	0.3147 / 0.0690
SOB	0.05	5	0.0000 / 0.0000	0.0476 / 0.0299
		10	0.0000 / 0.0000	0.0480 / 0.0247
	0.2	5	0.0000 / 0.0000	0.0578 / 0.0314
		10	0.0000 / 0.0000	0.0560 / 0.0267
	1	5	0.0173 / 0.0110	0.0524 / 0.0290
		10	0.0109 / 0.0105	0.0483 / 0.0290
	1.2	5	0.0225 / 0.0134	0.0472 / 0.0313
		10	0.0152 / 0.0122	0.0396 / 0.0295

이상의 결과에서 SOB가 다른 방법에 비해 우수한 결과를 보여줄 수 있음을 알 수 있다. 그리고 SVC에서 중요한 역할을 하는 커널모수 σ에 대해서 큰 영향 받지 않는 것으로 나타났다. 그리고 SOB의 경우에는 σ가 1이상의 큰 값을 가질 때 좋은 결과를 주는 것으로 나타났다.

5. 결론 및 향후과제

본 논문에서 실험한 SON는 모의실험자료와 실제자료(Wisconsin Breast Cancer Data, German Credit Data, Iris Data)에서 우수한 성능을 보여주었다. SOB는 σ값에

큰 영향을 받지 않고 유의한 결과를 보여주었다. 또한 이상치가 존재할 경우 평가오류율이 0이 되지 않는 사실로부터 이상치의 존재 가능성을 예측 할 수 있다. 그러나 다음과 같은 향후과제도 남아있다.

SOB는 T값(부스팅의 반복회수)을 결정하여야 한다. 따라서 T값을 찾아내는 효과적인 알고리즘 개발이 필요하다. 그리고 SOB가 σ값에 크게 영향을 많이 받지 않는다고 해도 여전히 σ의 선택의 문제가 남아있다. 기존의 CV(Cross Validation)방법의 단점을 극복할 수 있는 효과적인 알고리즘 개발이 시급하다.

6. 인용문헌

1. C. Burges(1998). *A Tutorial on Support Vector Machines for Pattern Recognition*, In Data Mining and Knowledge discovery 2, Kluwer Academic Publishers, Boston.
2. N. Cristianini and J. Shawe-Taylor(2000). *An Introduction to Support Vector Machines*, Cambridge.
3. Y. Freund and R. E. Schapire(1997). A Decision-Theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Science*, 55(1), 119-139.
4. Y. Freund and R. E. Schapire(1999). A Brief Introduction to Boosting, *Proceeding of the Sixteenth International Joint Conference on Artificial Intelligence.*
5. S. Gunn(1998). Support Vector Machines for Classification and Regression, *ISIS Technical Report*, U. of Southampton.
6. E. Osunam R. Freund and F. Girosi(1997). Support Vector Machines : Training and Applications, *MIT AI Lab., Technical Report.*
7. A. J. Smola, B. Scholkopf(1998). A Tutorial in Support Vector Regression, *NeuroCOLT2, Technical Report, NeuroCOLT.*
8. L. G. Valiant(1984). A Theory of the Learnable, *Communication of the ACM*, 27, 1134-1142.
9. V. Vapnik(1985). *The Nature of Statistical Learning Theory*, Springer.
10. V. Vapnik(1998). *Statistical Learning Theory*, Springer.
11. Wahba, G., Lin, Y. and Zhang, H.(2000). Generalized Approximate Cross Validation for Support Vector Machines, or, Another Way to Look at Margin-Like Quantities, *TR 1006*, April 1999. Expanded version of TR1006 posted here February 1999. (With revisions) in *Advances in Large Margin Classifiers*, Smola, Bartlett, Scholkopf and Schurmans, eds., MIT Press(2000), 297-309.

[2002년 7월 접수, 2002년 8월 채택]