

# 앙상블 구성을 이용한 SVM 분류성능의 향상 (Improving SVM Classification by Constructing Ensemble)

제 흥 모<sup>\*</sup>      방 승 양<sup>\*\*</sup>

(Hong-Mo Je)    (Sung-Yang Bang)

**요 약** Support Vector Machine(SVM)은 이론상으로 좋은 일반화 성능을 보이지만, 실제로 구현된 SVM은 이론적인 성능에 미치지 못한다. 주된 이유는 시간, 공간상의 높은 복잡도로 인해 근사화된 알고리즘으로 구현하기 때문이다. 본 논문은 SVM의 분류성능을 향상시키기 위해 Bagging(Bootstrap aggregating)과 Boosting을 이용한 SVM 앙상블 구조의 구성을 제안한다. SVM 앙상블의 학습에서 Bagging은 각각의 SVM의 학습데이터는 전체 데이터 집합에서 임의적으로 일부 추출되며, Boosting은 SVM 분류기의 에러와 연관된 확률분포에 따라 학습데이터를 추출한다. 학습단계를 마치면 다수결(Majority voting), 최소자승추정법(LSE:Least Square estimation), 2단계 계층적 SVM등의 기법에 개개의 SVM들의 출력 값들이 통합되어진다. IRIS 분류, 필기체 숫자인식, 얼굴/비얼굴 분류와 같은 여러 실험들의 결과들은 제안된 SVM 앙상블의 분류성능이 단일 SVM보다 뛰어난 것을 보여준다.

**키워드** : SVM, SVM 앙상블, 얼굴탐지, IRIS 데이터 분류, 필기체 숫자인식

**Abstract** A support vector machine (SVM) is supposed to provide a good generalization performance, but the actual performance of a actually implemented SVM is often far from the theoretically expected level. This is largely because the implementation is based on an approximated algorithm, due to the high complexity of time and space. To improve this limitation, we propose ensemble of SVMs by using Bagging (bootstrap aggregating) and Boosting. By a Bagging stage each individual SVM is trained independently using randomly chosen training samples via a bootstrap technique. By a Boosting stage an individual SVM is trained by choosing training samples according to their probability distribution. The probability distribution is updated by the error of independent classifiers, and the process is iterated. After the training stage, they are aggregated to make a collective decision in several ways, such as majority voting, the LSE(least squares estimation)-based weighting, and double layer hierarchical combining. The simulation results for IRIS data classification, the hand-written digit recognition and Face detection show that the proposed SVM ensembles greatly outperforms a single SVM in terms of classification accuracy.

**Key words** : SVM, SVM ensemble, Face detection, IRIS data classification, the hand-written digit recognition

## 1. 서 론

Support Vector Machine(SVM)은 AT&T Bell연구소의 Vapnik에 의해서 제안된 새로운 분류(classification) 및 회귀(regression) 기술이다[1]. SVM은 마진(margin)을 최대화 하는 구분평면(Separating Hyperplane)을 학

습하는 것이며 이로 인해 좋은 일반화 성능을 보인다[2]. 최근의 이론적 연구는 SVM을 이용한 실제 응용에서의 문제점을 해결하는 것에 주안점을 두고 있으며, 현재까지 얼굴탐지, 필기체 숫자인식, 데이터 마이닝, 생물학 정보칩 분류등 많은 영역에서 성공적으로 적용되고 있다[3,4].

그러나 Support Vector 학습하는 과정은 많은 시간과 저장 공간이 소요된다는 단점이 있다. 실제 구현에서 근사화 된 알고리즘을 적용하여 시간과 저장 공간을 절약할 수 있지만 분류성능 또한 떨어지게 된다.

이러한 단점을 해결하기 위해 SVM 앙상블의 구성을 제안 한다. 각각의 SVM은 전체 학습데이터의 일부분만

\* 비 회 원 : 포항공과대학교 컴퓨터공학과  
invu71@postech.edu

\*\* 종신회원 : 포항공대 컴퓨터공학과 교수  
sybang@postech.edu

논문접수 : 2002년 4월 1일  
심사완료 : 2002년 11월 29일

을 추출하여 독립적으로 학습하며 하나의 SVM이 커버하는 제한된 영역에서는 가장 좋은 분류성능을 보인다. 특정영역에서 가장 좋은 성능을 보이는 여러 개의 SVM을 통합하여 구성된 SVM 앙상블이 하나의 SVM을 사용했을 때 보다 좋은 성능을 나타낼 수 있으리라 기대할 수 있다.

SVM 앙상블의 구성에 관한 기본적인 아이디어는 이미 [7]에서 제안 되었으나, 본 논문에서는 Boosting 과 Bagging을 이용한 SVM 앙상블의 구성과 독립적으로 학습된 여러 개의 SVM을 통합하는 기법으로 다수결(Majority voting), 최소자승추정법(LSE), 2단계 계층적 SVM 등의 방법을 제안한다.

본 논문의 구성은 다음과 같다. 먼저 2절에서 SVM의 이론적 배경을 기술한다. 3절에서는 Bagging 과 Boosting을 이용한 SVM 앙상블의 구성과 3가지 통합방법에 대해서 설명한다. 4절에서는 제안된 SVM 앙상블을 이용한 IRIS 데이터 분류와 필기체 숫자 인식, 얼굴/비얼굴 분류의 실험에 관한 결과를 보이고 분석하였으며, 5절에서 결론과 향후 연구 방향을 기술한다.

## 2. Support Vector Machines

SVM은 최소의 일반화 에러로 나타나게 하는 최적의 분류평면(Separating Hyperplane)을 결정하는 기법이라고 볼 수 있다 [2]. 일반적으로 선형적으로 분류가능한 문제의 분류식은

$$f_{w,b} = \text{sign}(w \cdot x + b), \quad (1)$$

과 같이 나타낼 수 있다.

SVM에서 최적의 분류평면은 서로 다른 클래스들을 구분하는 최대 마진(Margin)사이에 존재한다고 본다. 입력벡터  $x_i$  에 대한 클래스 레이블(label)이  $y_i$ 라고 할 때, 최적의 분류평면은 다음의 제약조건 최소화를 만족해야 한다[2].

$$\text{Min} \frac{1}{2} w^t w \text{ where } y_i(w \cdot x_i + b) \geq 1 \quad (2)$$

선형적으로 구분이 불가능한 경우 위의 최소화 조건은 오분류 데이터를 허용하기 위해 수정되어야 한다. 수정된 식에서 soft margin 분류기가 어느 정도의 에러를 허용하는 대신 제약조건의 위반의 측정치로 새로운 변수인  $c$ 를 포함한다. 그리고  $\alpha_i$ 가 라그랑지(Lagrangian) 계수일 때,

$$\begin{aligned} \text{Min} L(W) &= 1/2 \cdot \langle w, w \rangle - \sum \alpha_i y_i [(\langle w, \varphi(x_i) \rangle + b) - 1] \\ 0 &\leq \alpha_i \leq C \\ \frac{\partial L}{\partial b} &= 0 \\ \frac{\partial L}{\partial w} &= 0 \end{aligned} \quad (3)$$

여기서  $C$ 는  $\xi_i$ 의 가중치이며,  $\varphi(\cdot)$ 는 입력공간을 보다 고차원의 공간으로 매핑하는 비선형함수이다. 식 (3)의 첫 번째 항을 최소화하는 것은 VC 차원(dimension) [1]을 최소화 하는 것과 같은 효과이다[2]. 식(3)을 풀기 위해서 라그랑지 방법을 이용하여 다음 과 같이 변형한다.

$$\begin{aligned} \text{Max: } W(\alpha) &= \sum \alpha_i - 1/2 \cdot \sum \alpha_i \alpha_j y_i y_j K(x_i, x_j) \\ 0 &\leq \alpha_i \leq C \\ \sum \alpha_i y_i &= 0 \end{aligned} \quad (4)$$

이 때  $K(x_i, x_j) = \langle \varphi(x_i), \varphi(x_j) \rangle$ 인 커널함수이다.

두 클래스(binary class)분류에서의 임의의 입력 벡터  $X$ 에 대한 분류함수는 식(1)을 변형한 다음 식과 같다.

$$f(X) = \text{sign}(\sum_{i=1}^l y_i \alpha_i (x_i \cdot X) + b) \quad (5)$$

## 3. Support Vector Machine 앙상블

앙상블의 기본적인 개념은 서로 다른 여러 개의 분류기들의 출력을 통합하여 최종 분류하는 일종의 복수 분류기 시스템(Multiple Classifier System : MSC)라고 볼 수 있다. 분류기들의 앙상블은 테스트 데이터를 분류하기 위해 개별적으로 결정한 여러 개의 분류기들의 출력을 여러 가지 방법으로 결합하는 것이다[8]. 패턴인식 분야에 있어서 앙상블은 개개의 분류기 하나만 사용했을 때 보다 나은 성능을 보인다고 알려져 있다. Haris [16], Giacinto [17], Hansen [9] 그리고 많은 연구자들에 의해서 신경망 앙상블(Neural Network Ensemble)은 연구되어 오고 있었다. 게다가 신경망 앙상블과 유사한 아이디어로 Kim [12]는 Genetic Fuzzy Predictor 앙상블을 제안하였다. 제안된 방법들 모두 이미지 분류, 시계열 예측, 재정흐름 분석 등의 다양한 응용에서 좋은 분류 성능을 보였다.

Hansen [9]는 앙상블이 하나의 분류기만을 사용했을 때 보다 좋은 성능을 보이는 이유를 다음과 같이 보였 다.  $n$  개의 분류기  $\{f_1, f_2, \dots, f_n\}$ 로 구성된 앙상블 구조가 있고 테스트 데이터 집합  $x$ 가 있다고 할 때, 어떤 데이터에 대해서 모든 분류기가 같은 결과를 보인다면 앙상블의 성능은 각각의 분류기의 성능과 동일할 것이다. 그러나 분류기들이 서로 다른 에러 값을 보이며 상관도가 낮을 때, 임의의  $f_k(x_i)$ 가 오류(wrong)라고 판정하고, 나머지 대다수의  $f_i(x_i)$ 가 옳음(correct)으로 판정했다면, 앙상블 구조는 다수결(Majority voting)에 의해 그 결과는 옳음(correct)으로 판정한다. 수식으로 자세히 설명하자면, 만약 각각의 분류기들의 에러의 확

를  $p < 1/2$  이고 그 에러 확률들이 독립이면, 다수결 (Majority voting)의 에러확률  $p_E$ 는

$$\sum_{k=n/2}^n p^k (1-p)^{n-k} < \sum_{k=n/2}^n \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{n-k} = \sum_{k=n/2}^n \left(\frac{1}{2}\right)^n \text{와 같이 유도된다.}$$

식에서 알 수 있듯이 분류기의 수가 많을수록 에러의 확률은 점차 줄어들게 되고, 분류기의 개수  $n$ 이 무한대가 되면, 에러확률  $p_E$ 가 0으로 수렴하게 된다.

일반적으로 SVM은 좋은 일반화 성능을 보이며, 전체 최적치(global optimum)를 얻기 위한 파라미터들을 쉽게 학습할 수 있는 것으로 알려져 있다[2]. 그래서 SVM의 앙상블 구성은 분류성능을 크게 향상 시키지는 못할 것이라고 간주 되어질 수 있다. 그러나 SVM의 실제적인 구현은 계산 시간과 기억장소 절감을 위해 근사화 알고리즘을 이용하므로 단일 SVM만으로는 전체 최적치(global optimum)를 위한 정확한 파라미터들을 학습하지 못할 수도 있다. 또한 학습을 통하여 얻어진 Support Vector들이 미지의 테스트 데이터를 완전히 분류하는 데 충분하지 못한 경우도 있다. 그래서 단일 SVM만으로는 모든 테스트 데이터에 대해서 최적의 분류성능을 보장하지 못한다.

이러한 단점을 극복하기 위해, 우리는 SVM의 앙상블 구조를 제안한다. 일반적인 분류기 들의 앙상블 구조가 SVM의 앙상블 구성에도 적용되어질 수 있다. 그림 1은 제안된 SVM 앙상블의 일반적인 구조를 나타낸다. 학습 단계에서 각각의 SVM은 3.1절에서 설명하고 있는 Bagging 혹은 Boosting에 의해 추출되어진 학습 데이터들만을 가지고 독립적으로 학습을 진행하게 된다. 앙상블을 구성하는 모든 SVM들은 3.2절에서 설명하는 여러 가지 통합정책에 의해서 합쳐지게 된다. 테스트 단계에서 테스트 데이터는 앙상블을 구성하는 개개의 SVM에 동시에 적용되어지며 통합정책에 의해 최종적으로 분류 결정을 내리게 된다.

### 3.1 SVM 앙상블의 구성 방법

분류기들의 앙상블 구성을 위한 여러 가지 기법들이 있다. 우리는 SVM 앙상블의 구성을 위해 일반적인 기법을 적용했다. SVM 앙상블 구성에 있어서 주안점은 각각의 SVM이 가능한 서로 다른 학습데이터를 가지게 하는 것에 있다. 각기 다른 SVM에 다른 학습데이터를 할당하기 위해 잘 알려진 있는 기법들 중 대표적인 방법인 Bagging과 Boosting을 적용하였다.

#### 3.1.1 Bagging

먼저, SVM 앙상블 구성을 위한 Bagging 기법[10]을

설명한다. 보통 하나의 전체 학습 데이터 집합  $TR = \{x_i, y_i | i=1, 2, \dots, n\}$ 을 가지게 되는데,  $K$ 개의 SVM들로 구성된 앙상블에서 필요한 데이터 집합은  $K$  개의 독립적인 학습데이터 집합들이 필요하다. 통계적인 특성상 통합한 결과가 더 좋은 성능을 보이게 하기 위해 될 수 있는 한 다른 데이터 집합을 가지게 한다. 따라서 전체 데이터 집합  $TR$ 로부터  $\{TR_k^{Bagging} | k=1, 2, \dots, K\}$ 를 임의로 추출하게 된다. 각각의 데이터는 어느 정도 겹쳐질 수도 있으나, 고른 분포가 되게 하기 위해 특정한 Random Function에 의해 추출한다. 이렇게 추출된  $K$ 개의 데이터 집합이  $K$ 개의 SVM의 개별적인 학습데이터 집합으로 사용되어진다.

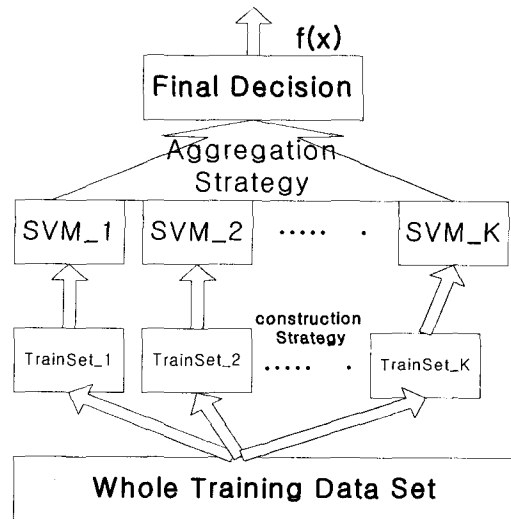


그림 1 SVM 앙상블의 일반적인 구조

#### 3.1.2 Boosting

대표적인 Boosting 알고리즘은 AdaBoosting 알고리즘 [11]이다. Bagging과 마찬가지로 각각의 SVM은 서로 다른 학습데이터 집합을 가지게 된다.  $l$ 번째 순서에서 모든 학습데이터  $x$ 에 대한 확률 분포 함수  $p_l(x)$ 가 있고,  $l$ 번째 순서를 수행하면  $l$ 번째 SVM의 학습데이터가 분포함수  $p_l(x)$ 에 의해서 추출되어지며, 이 데이터를 학습단계에 적용하게 된다.  $l$ 번째 SVM의 학습데이터에 대한 에러 윌로 계산되어진  $\epsilon_l$ 은  $(l+1)$ 번째 순서의 확률 분포 함수  $p_{l+1}(x)$ 을 변경에 영향을 미치게 된다. 그림 2에서 간단한 AdaBoosting 알고리즘을 서술하였다.

Input : a set  $S$  of  $m$  labeled examples :  
 $S = \{(x_i, y_i), i=1, 2, \dots, m\}$   
 Labels  $y_i \in Y = \{1, \dots, K\}$   
 Learn (Using general learning algorithm)  
 a constant variable  $L$

Initialize for all  $i$  :  $w_i(i) = \frac{1}{m}$

for  $l = 1$  to  $L$

for  $i = 1$  to  $m$

$$p(i) = w(i) / (\sum_k w_k(i))$$

end

$h_l = \text{Learn}(p)$

$$\epsilon_l = \sum_i p(i) | \{h_l(x_i) \neq y_i\} |$$

if  $\epsilon_l > 1/2$  then

$L = l - 1$

break

$$\beta_l = \frac{\epsilon_l}{(1 - \epsilon_l)}$$

for  $i = 1$  to  $m$

$$w_{l+1} = \begin{cases} w_l(i) + \beta_l & |h_l(x_i) \neq y_i \\ w_l(i) & |h_l(x_i) = y_i \end{cases}$$

end

end

그림 2 AdaBoosting 알고리즘

### 3.2 SVM 앙상블 통합전략

학습이 끝나면, 독립적으로 학습된 여러 개의 SVM들을 적절한 결합방법에 의해 통합하는 과정이 필요하다. 우리는 선형 혹은 비선형적인 결합방법으로 나누어서 고려해 보았다. 선형적인 결합방법은 복수개의 SVM들을 다수결(Majority voting)과 최소자승추정법(LSE)등과 같은 선형적인 기법에 의해 통합하는 방법이다. 이 두 방법들은 Bagging과 Boosting에 자주 사용되어지는 방법이다. 비선형적인 방법으로는 하위 레벨에 있는 SVM들을 통합하면서 최종결정을 위해 상위레벨에서 다른 SVM을 두어 2단계 계층적 SVM을 구성하는 것이다.

#### 3.2.1 다수결(Majority voting)

Majority voting은 여러 개의 SVM을 결합하는 방법들 중 가장 간단한 방법이다.

SVM 앙상블을 구성하는  $k$  개의 SVM들로부터 출력된 결정함수들이  $f_k(k=1, 2, \dots, K)$ 이고,  $j$  클래스의 라벨이  $C_j(j=1, 2, \dots, C)$ 일 때,  $N_j = \text{NumberOf}(h_{f_k}(c) = C_j)$ 는  $j$  클래스로 인식한 SVM의 수이다. 주어진 입력벡터  $x$ 에 대한 최종결정 단계에서의 다수결(Majority voting) 결정함수  $f_{maj}(x)$ 는 다음과 같이 표시할 수 있다.

$$f_{maj}(x) = \arg \text{Max} N_j \quad (6)$$

즉 앙상블을 구성하는 SVM들이 가장 많이 결정된 클래스를 최종적으로 결정된 클래스라고 본다.

#### 3.2.2 최소자승추정법(LSE:Least Square Estimation)에 의한 가중치

LSE 기반의 가중치 부여방법은 앙상블을 구성하는 각 SVM들에게 다른 가중치를 부여하게 된다. 때로는 분류기들의 정확도의 비율에 따라 가중치를 부여하기도 하지만 [2], 제안된 방법은 LSE 기반으로 하여 가중치를 학습하게 된다.

앙상블을 구성하는  $k$ 개의 SVM들의 결정함수들이  $f_k(k=1, 2, \dots, L)$ 이고, 앙상블 구성방법에 의해 구성된 학습데이터 집합이  $TR_k = \{(x_i; y_i) | i=1, 2, \dots, L\}$ 와 같이 정의 될 때, 가중치 벡터  $w$ 는 다음과 같이 구해 질 수 있다.

$w_E = A^{-1}y$ 이며, 여기서  $A = (f_i(x_j))_{K \times L}$ 이고  $y = (y_i)_{1 \times L}$ 이다.

주어진 입력 벡터  $x$ 에 대한 LSE 기반 가중치 부여법에 의한 최종 결정은 식(7)과 같다.

$$f_{LSE}(x) = \text{sign}(w_E \cdot [(f_i(x))_{K \times L}]) \quad (7)$$

#### 3.2.3 2단계 계층적 SVM

SVM 앙상블을 구성하는 여러 개의 SVM들의 출력들을 통합하기 위해 상위 레벨에서 다른 하나의 SVM을 두어 상위레벨의 SVM에서 최종결정을 하도록 하는 방법이다. 이 결합에서는 하위레벨의 SVM들의 출력벡터가 상위레벨의 SVM의 입력벡터가 된다. 이 구조는 M. Jordan에 의해 소개되어진 MOE(Mixture of Expert) [13]과 유사한 개념이다.

앙상블을 구성하는  $k$  개의 SVM들의 결정함수들은  $f_k(k=1, 2, \dots, L)$ 이고, 함수  $F$ 는 상위레벨 SVM의 결정함수이다. 주어진 입력 벡터  $x$ 에 대한 2단계 계층적 SVM 결합에 의한 최종결정은 다음과 같다

$$f_{SVM}(x) = F(f_1(x), f_2(x), \dots, f_k(x)). \quad (8)$$

## 4. 실험 및 결과

Bagging과 Boosting을 이용한 제안된 SVM 앙상블

의 효율성을 측정하기 위해서 세 종류의 분류문제 즉 IRIS 데이터 분류, 필기체 숫자인식, 얼굴탐지에 적용해 보았다.

Bagging과 Boosting으로 각각 다르게 구성된 앙상블에 대해 세 가지 통합 정책, 다수결(Majority voting), LSE기반 가중치, 2단계 계층적 SVM의 모든 경우에 대해서 실험하였으며 또한 단일 SVM에 대한 성능향상을 보이기 위해 SVM 하나만 사용했을 경우의 결과도 보인다. SVM 앙상블과 다른 패턴분류 기법과의 비교를 위해 KNN(K-nearest-neighbor)알고리즘[19] 혹은 신경망(MLP)[20]을 적용했을 때의 결과와 비교하여 분석한다.

실험을 위한 SVM 학습 알고리즘은 SMO(Sequential Minimal Optimization)[7]로 구현하였다.

4.1 IRIS 데이터 분류

IRIS 데이터 [14,15]는 패턴인식 분야에서 잘 알려진 데이터베이스들 중 하나이다. IRIS 데이터는 3개의 클래스로 나뉘어져 있으며 각각의 클래스는 50개의 데이터 집합을 가진다.

이들 중 하나의 클래스는 다른 클래스와 선형적으로 분리가 가능하지만, 다른 두 클래스는 서로 선형적으로 분리가 불가능하다. 우리는 두 클래스(binary class) 구분 SVM 앙상블을 다 클래스 문제로 확장하였고, 그 구성은 일대일(one-against-one) [4] 방법을 취하였다. 5개의 다 클래스 구분 SVM이 있는데 각각의 다 클래스 구분 SVM은 다시 3개의 SVM 요소로 구성된다.

즉  $SVM^{1..5} = (SVM_{c_1c_2}, SVM_{c_2c_3}, SVM_{c_3c_1})^{1..5}$ 가 된다.

$SVM^1, SVM^2, \dots, SVM^5$ 는 제안된 SVM 앙상블을 구성하는 기본 단위의 다 클래스 SVM이며, 이들은 다시 3개의 일대일(one-against-one) SVM를 결합하여 다 클래스 구분으로 확장하였다. 다 클래스 SVM의 클래스 결정은 최다승선택(Max Win) [4]에 의한 방법을 택하였고, 앙상블 전체의 최종 결정을 위해서 5개의 다 클래스 SVM들을 제안한 통합정책에 의해 결합하였다. 기본 단위 SVM들은 모두 2차 다항식형(2-d polynomial) 커널 함수로 학습하였다. 우리는 전체 150개의 데이터들 중에 90개를 임의로 추출하여 학습을 위한 데이터 집합으로 지정하였다. Bagging에서는 이들 중 60개(클래스당 20개)를 하나의 SVM의 학습을 위해서 임의로 추출하였고, Boosting에서는 매 순서마다 60개를 여러 확률분포에 의해서 추출하였다. 테스트를 위한 데이터는 학습데이터와 중복되지 않는 60개(클래스당 20개)로 구성되었다.

표 1은 IRIS 데이터 분류를 위한 실험결과를 보여준다.

K-NN은 k = 1일 때 K-nearest-neighbor 알고리즘을 적용한 결과이며, MLP는 10개의 은닉노드를 가지는 신경망을 역전파 알고리즘(BP : Back Propagation)[20]으로 학습한 것이다. 제안된 SVM 앙상블 구조에서의 결과는 Bagging, Boosting 각각에 대해 다른 통합정책을 적용한 결과를 나타내고 있다. 결과에서 알 수 있듯이 SVM 앙상블을 Boosting으로 구성하고 통합을 LSE기반 가중치 혹은 2단계 계층적 SVM으로 하였을 때 97.33%로 가장 좋은 분류성능을 보였다.

표 3 IRIS 데이터 분류 실험결과

	Bagging	Boosting
K-NN		93.38%
MLP		95.8%
Single SVM		95.94%
Majority Voting	96.00%	96.60%
LSE 가중치	97.03%	97.33%
계층적 SVM	97.25%	97.33%

4.2 필기체 숫자 인식

그림 3에서 일부 예를 보이고 있는 UCI 필기체 숫자 인식[15]에도 제안된 SVM 앙상블을 적용하였다. 이 데이터베이스는 3823개의 학습 데이터와 1723개의 테스트 데이터로 나뉘어 있다. 원래 이미지의 크기는 32\*32인데, 8\*8 단위의 block 마다 특징을 추출하여, 숫자 하나에 대한 특징벡터의 크기는 64\*1이 된다.

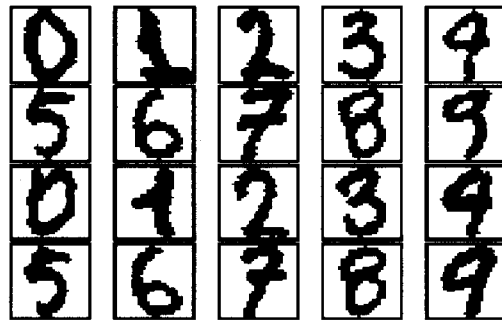


그림 3 UCI 데이터베이스의 필기체 숫자들의 예

우리는 이 문제에서도 다 클래스 확장을 위해서 일대일(one-against-one)방법을 채택하였다. 각각의 다 클래스 SVM은 최다승(Max Win)으로 클래스를 결정하며, 11개의 다 클래스 구분 SVM으로 앙상블을 구성하였다. 모든 SVM들은 2-D Polynomial 커널함수로 학습하였다.

일반적으로 C클래스를 one-against-one방법으로 분류하기 위해서는  $\frac{C(C-1)}{2}$  개의 분류기가 필요하다. 필기체 숫자인식은 10클래스 문제이므로 하나의 다 클래스 SVM은 45개의 분류기로 구성된다. 즉  $SVM^{1..11} = (SVM_{0..1}, SVM_{0..2}, \dots, SVM_{0..9}, SVM_{1..2}, \dots, SVM_{9..0})^{1..11}$ 와 같게 된다. Bagging에서는 하나의 SVM이 전체 학습 데이터 집합 3823개에서 1000개를 임의 추출하여 학습하였고, Boosting 에서도 마찬가지로 매 iteration 마다 여러 확률 분포 함수에 의해 1000개씩을 추출하여 학습하였다. 표 2는 필기체 숫자 인식 실험의 결과를 보인다. K-NN의 k = 3 일 때, MLP는 은닉노드 100개를 가지는 신경망이다. 테스트 데이터 집합 1723개를 모두 적용한 결과이다. 실험결과 SVM 앙상블의 Boosting 구성에서 2 단계 계층적 SVM으로 통합하였을 때 97.83%로 가장 좋은 성능을 보였다.

표 4 UCI 데이터베이스 필기숫자 인식 실험결과

	Bagging	Boosting
K-NN	95.32%	
MLP	96.5%	
Single SVM	96.02%	
Majority Voting	96.85%	97.15%
LSE 가중치	97.27%	97.61%
계층적 SVM	97.38%	97.83%

4.3 얼굴/비얼굴 분류

얼굴탐지를 다음과 같이 정의 할 수 있다. 입력으로 임의의 이미지 혹은 비디오신호가 들어오면 그 곳에 사람의 얼굴이 있는지 없는지 판단하고 결과로 얼굴의 위치가 어디에 있는지 나타내는 과정이다. 얼굴탐지에 있어서 중요한 것은 다른 물체를 탐지하는 것과 마찬가지로 사람의 얼굴인 것과 그렇지 않은 것을 구분하는 것이 중요하다. 그런 연유로 얼굴탐지는 근본적으로 서로 다른 두 종류의 패턴을 구분하는 두 클래스(binary class) 분류과정으로 고려되어질 수 있다. SVM은 두 클래스(binary class)를 가장 잘 분류하기 위한 구분영역을 구하기 위한 학습방법이므로 얼굴탐지를 위한 좋은 분류기로 사용되어 질 수 있다.

SVM을 이용한 얼굴탐지는 Osuna[8]에 의해서 이미 제안되었지만, 본 논문에서는 탐지성능을 향상시키기 위해 SVM 앙상블을 적용하였다. 얼굴탐지에 있어서 가장 중요한 기능이라고 볼 수 있는 얼굴/비얼굴의 분류성능을 검증하기 위해서, 잘 알려진 얼굴 데이터베이스를 대

상으로 실험한 결과들을 보이고 분석한다.

얼굴/비얼굴 분류실험에서 SVM 앙상블의 성능을 측정하기 위해서 CMU+IM 얼굴 데이터베이스[18]을 사용하였다. 이것은 카네기 멜론 대학에서 공개용으로 배포하고 있는 CMU DB와 포항공대에서 제작한 얼굴데이터 베이스를 통합한 것이다. 학습을 위해서 데이터베이스로부터 다양한 조명변화가 적용된 얼굴 데이터 1000개와 비얼굴 데이터 1000개를 추출하여 사용하였으며, 테스트 데이터는 학습데이터와 겹치지 않게 같은 수량의 데이터집합을 추출하였다. 그림 3과 그림 4는 각각 실험에 사용한 얼굴 데이터, 비얼굴 데이터의 예를 보인다.

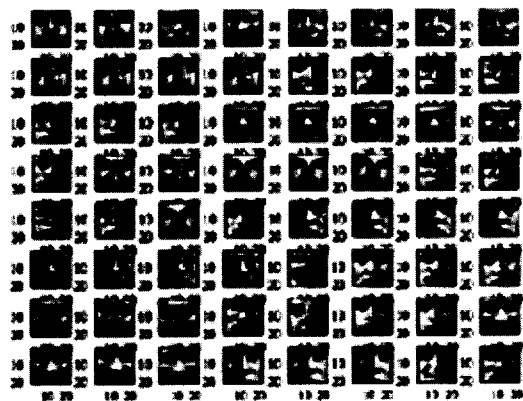


그림 3 얼굴 데이터 예

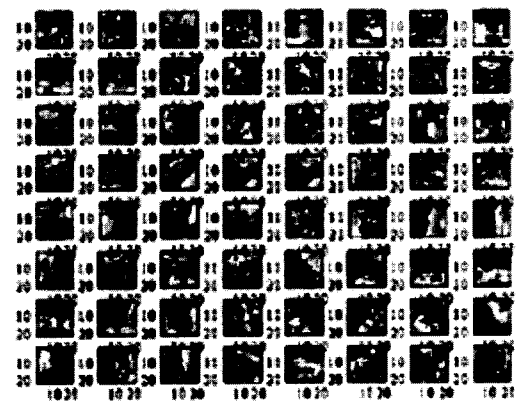


그림 4 비얼굴 데이터 예

표 3은 얼굴/비얼굴 분류 실험에 대한 결과를 나타낸다. SVM\_S는 단일 SVM에 대한 실험결과이며, EN\_5, EN\_11은 각각 SVM\_S를 5개, 11개를 앙상블로 구성한 것을 나타낸다, MLP는 Rowley[7]의 방법으로써

잘 알려진 얼굴탐지 기법과 비교하기 위해 실험하였다. 모든 SVM의 커널함수는 2차다항형(2D-Polynomial)이며 양상블의 구성은 EN\_5, EN\_11 공히 Bagging, 통합전략은 다수결(Majority voting)을 사용하였다. 일반적으로 SVM 양상블에서 SVM의 수가 증가할수록 분류성능은 점차적으로 향상 된다고 생각할 수 있다.

"FALSE"는 비얼굴을 얼굴로 오인식 하는 것을 말하며, "MISS"는 얼굴을 비얼굴로 분 인식하여 분류에 실패한 것을 말한다. "Correct"는 FALSE와 MISS를 제외한 나머지, 즉 얼굴데이터를 얼굴로 인식하고, 비얼굴을 비얼굴로 올바르게 인식한 분류 정확도를 백분율로 표시한 것이다.

단일 SVM보다는 EN\_5가 EN\_5보다 EN\_11의 성능이 좋음을 알 수 있고, 실험결과가 보여 주듯이 EN\_11은 MLP보다 좋은 분류성능을 나타내고 있다.

표 5 얼굴 / 비얼굴 분류 실험 결과

	SVM_S	EN_5	EN_11	MLP
FALSE	75	42	38	52
MISS	39	20	8	8
Correct	89.6%	93.8%	95.4%	94.0%

### 5. 결론 및 향후 연구 과제

일반적인 경우 SVM의 실제적인 구현은 계산시간과 저장 공간의 절약을 위해 근사화된 알고리즘을 이용하게 된다. 그 결과 SVM의 분류성능이 이론적인 기댓값에 미치지 못하는 경우가 발생하게 된다.

이러한 문제점을 극복하고자, 본 논문에서는 독립적으로 학습된 여러 개의 SVM으로 구성된 SVM 양상블에 대해 논의하였다. 양상블의 구성에 있어서 통계적 학습 방법에서는 잘 알려진 Bagging과 Boosting을 도입하여 개개의 SVM들이 될 수 있는 한 서로 다른 학습데이터 집합을 가질 수 있게 하였다. 이렇게 구성된 SVM 양상블은 다수결(Majority voting), LSE 기반 가중치, 2단계 계층적 SVM등의 통합정책에 의해 결과들이 합쳐지며 최종결정을 내리게 된다.

제안된 SVM양상블의 분류성능을 측정하기 위해서 IRIS 데이터 분류, 필기체 숫자 인식, 얼굴/비얼굴 분류 문제 등에 적용해 보았다. 모든 실험의 결과를 통해서 SVM 양상블이 SVM 단독으로 학습했을 때는 물론이고, 다른 패턴인식 기법들 보다 향상된 분류성능을 보이고 있음을 알 수 있다. 양상블의 통합정책으로 제안한 세 가지 방법들의 성능비교에서는 2단계 계층적 SVM

이 가장 좋은 성능을 보였고, LSE 가중치, 다수결(Majority voting)의 순으로 나타났다.

하지만 모든 응용에 2단계 계층적 SVM 통합정책이 적합함을 의미하지는 않는다. 다수결(Majority voting)은 아주 간단한 아이디어의 통합정책으로 계산량이 많지 않으므로, 다른 기법에 비해 분류성능이 크게 뒤 떨어지지 않는다면 실시간 응용에 SVM양상블을 적용할 때 적극적으로 고려해 볼 만한 방법이다.

본 논문에서는 SVM 양상블의 구성을 학습단계에서 개개의 SVM이 서로 다른 데이터 집합을 가지도록 하는데 주안점을 두었다. 그러나 데이터가 충분치 않을 때에는 데이터의 분할이 의미가 없어질 수도 있다. 데이터의 수가 적은 응용에서 SVM양상블을 구성하기 위해서는 동일한 데이터에 대해서 학습과정에서 다른 종류의 커널함수를 적용하거나 파라미터를 변경하여 SVM들이 서로 다른 일반화 성능을 나타낼 수 있도록 특징짓는 Support Vector들을 생성하도록 하는 구성방법에 대한 연구가 필요하다.

### 참 고 문 헌

- [1] C. Cortes and V. Vapnik, "Support vector network", *Machine Learning*, vol. 20, 273-297, 1995.
- [2] Christopher J.C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition", *Data Mining and Knowledge Discovery*, vol. 2, no. 2, 121-167, 1998.
- [3] T. Joachims, "Making large-scale support vector machine learning practical", *Advances in Kernel Methods: Support Vector Machines*, MIT Press, Cambridge, MA, 1999.
- [4] John Platt, "Fast training of support vector machines using sequential minimal optimization", *Advances in Kernel Methods: Support Vector Machines*, MIT Press, Cambridge, MA, 1999.
- [5] L. Bottou, et. al., "Comparison of classifier methods: a case study in handwriting digit recognition", *Proceedings of International Conference on Pattern Recognition*, IEEE Computer Society Press, pp. 77-87, 1994.
- [6] S. Knerr, L. Personnaz, G. Dreyfus, "Single-layer learning revisited: a stepwise procedure for building and training a neural network", J. Fogelman, editor, *Neurocomputing: Algorithms, Architectures and Application*, Springer-Verlag, 1990.
- [7] Valdimir N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, 1999.
- [8] Thomas G. Dietterich, "Machine Learning Research: Four Current Directions", *The AI*

- Magazine*, vol. 18, no. 4, 97-136, 1998.
- [9] L. Hansen, P. Salamon, "Neural network ensembles", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, 993-1001, 1990.
- [10] L. Breiman, "Bagging predictors", *Machine Learning*, vol. 24, no. 2, 123-140, 1996.
- [11] Robert E. Schapire, "A Brief Introduction to Boosting", *proc. 6th Int Joint Conf. On AI*, pp.1-6, 1999.
- [12] Daijin Kim and Chul-Hyun Kim, "Forecasting time series with genetic fuzzy predictor ensemble", *IEEE Transaction on Fuzzy Systems*, vol. 5, no. 4, 523-535, 1997.
- [13] M. Jordan and R. Jacobs, "Hierarchical mixtures of experts and the EM algorithm", *Neural Computation*, vol. 6, no. 5, 181-214, 1994.
- [14] R.A. Fisher, "The use of multiple measurements in taxonomic problems", *Annual Eugenics*, 7, Part II, 179-188, 1936.
- [15] B. D. Bay, The UCI KDD Archive [<http://kdd.ics.uci.edu>]. Irvine, CA: University of California, Department of Information and Computer Science, 1999.
- [16] M. Haris, V. Ganapathy, "Neural network ensemble for financial trend prediction", *IEEE TENCON 2000, Proceedings*, Volume: 3, 157-161, 2000.
- [17] G.Giacinto, F. Roli, L. Bruzzone. "Combination of neural and statistical algorithms". *Pattern Recognition Letter 21*, pp. 385-397, 2000.
- [18] <http://nova.postech.ac.kr>.
- [19] Richard O. Duda, *Pattern Classification 2nd ED.*, 1997.
- [20] Christopher M. Bishop, *Neural Networks for Pattern Recognition*, Clarendon Press. Oxford, 1995.



방 승 양

1966년 일본 Kyoto대학 전기공학과 학사. 1969년 서울대학교 전기공학과 석사. 1974년 University of Texas 전산학 박사. 1974년~1975년 Wayne State University, Assistant Professor. 1979년~1981년 Bell 연구소 근무. 1981년~1984년 한국전자기술연구소 시스템부 부장. 1984년~1986년 (주)유니온시스템 전무. 1986년~현재 포항공과대 컴퓨터공학과 교수. 뇌연구센터 소장



제 홍 도

1999년 8월 영남대학교 컴퓨터공학과(공학사). 2002년 2월 포항공과대학교 컴퓨터공학과(공학석사). 2002년 2월~2003년 2월 삼성SDS정보기술연구소 선임연구원. 2003년~현재 포항공과대학교 컴퓨터공학과 박사과정