

# 거절 조건을 가진 이진 분류기 순차 결합을 통한 얼굴인식

부덕희<sup>0</sup>, 고재필, 변혜란  
연세대학교 컴퓨터과학과  
{dhboo<sup>0</sup>,nonezero, hrbyun}@cs.yonsei.ac.kr

## Face Recognition Using Sequential Fusion of Binary Classifiers with Rejection

Deokhee Boo<sup>0</sup>, Jaepil Ko, Hyeran Byun  
Dept. of Computer Science, Yonsei Univ.

### 요 약

얼굴인식 연구는 얼굴표현에 집중하여 연구가 이루어져 왔으며, 이에 따라 최종 인식단계에서는 간단한 분류기가 적용되었다. 그러나, 보다 강건한 분류 성능을 얻기 위해서는 복잡한 분류기를 도입할 필요성이 있다. 복잡한 분류기를 사용하면서, 얼굴인식의 특성상 온라인 학습을 가능하게 하기 위해서는 이진 분류기 분해/결합 방법을 도입한다. 이러한 기법으로는 OPC와 PWC가 대표적이다. 본 방법의 성능은 기본 이진 분류기의 성능에 좌우되기 때문에 SVM을 적용한다. 본 논문에서는 대표적인 방법인 OPC와 PWC의 장단점을 설명하고, 거절 조건을 갖는 이들 분류기들의 순차 결합 방법을 제안한다. ORL 얼굴데이터 베이스에 대한 실험에서는, 제안하는 방법을 통해 얼굴인식 성능을 향상시킬 수 있음을 보인다.

### 1. 서론

얼굴인식 연구는 얼굴표현에 집중하여 연구가 이루어졌다. PCA, ICA, LDA 등의 얼굴영상 표현 (특징추출) 기법이 제시되어 왔고, 인식단계에서는 K-NN과 같은 단순 분류기를 적용하였다. 분류성능 향상을 위해서는 다층 퍼셉트론과 같은 복잡한 분류기를 사용해야 하나, 새로운 얼굴이 시스템에 등록될 때마다 모든 얼굴을 재학습 시켜야 하는 단점 때문에 온라인 학습이 어렵다. 이러한 경우에, 다중 클래스 문제를 여러 개의 이진 클래스 문제로 분해하고, 다시 이들 이진 분류기들의 결과를 결합하여 최종 결정을 내리는, 출력코딩 (Output Coding) 방법이 효과적이다. 출력코딩 방법을 통해 추가되는 클래스에 대해서 이와 관련된 있는 몇 개의 이진 분류기만을 새로이 학습하면 된다. 이러한 방법은 기반이 되는 이진 분류기의 성능에 크게 의존적이다. 최근 Vapnik [1]이 제안한 SVM (Support Vector Machines)은 최적 이진 분류기로 주목 받고 있으며, 출력코딩 방법의 기반 이진 분류기로서 매우 적합하다. 본 논문에서는, 대표적인 분해 방법인 OPC와 PWC의 장단점을 설명하고 이를 기반으로, 거절조건을 가진 두 방법의 순차 결합 방법을 제안한다. ORL 얼굴 데이터에 대한 실험을 통하여 제안한 방법과 기존 방법의 성능을 비교한다. 본 논문 2장에서는 SVM에 관해 서술하며 3장에서는 기본 분해 방법들을 언급하고 4장에서는 새롭게 제안한 방법들을 소개한다. 5장에서 실험 결과를 보여주며, 6장에서는 결론을 내린다.

### 2. Support Vector Machines

SVM은 구조적 에러를 최소화하는 분류기이다. 두 클래스를 선형 분리하는 결정평면과 가장 가까운 샘플을 SV라 하고, SV와 결정평면간의 거리를 '마진' 이라고 정의하였을 때, 마진을 최대화하는 결정평면을 구해 좋은 일반화 성능을 가져 온다.  $x_i \in \mathbb{R}^n$ ,  $y_i \in \{-1, +1\}$ 인 학습집합  $(x_1, y_1), \dots, (x_b, y_b)$ 이 주어졌을 때, 결정평면은  $w \cdot x + b = 0$  이다.  $x$ 는 입력벡터,  $w$ 는 가중치벡터,  $b$ 는 바이어스이다. 마진을 최대화하는 비용함수를 식(1)과 같이 정의할 수 있다.

$$\phi(w) = 1/2 \cdot \|w\|^2 \quad (1)$$

선형적으로 분리 가능하지 않은 경우는 오류에 대한 패널티( $\epsilon$ )를 적용하여 식(2)의 비용함수를 통해 최적화 시킨다. 여기서  $C$ 는 마진의 최대화와 분류 에러를 조절하는 파라미터이다.

$$\phi(w) = 1/2 \cdot \|w\|^2 + C \sum_{i=1}^b \epsilon_i \quad (2)$$

분리면이 비선형인 경우 커널함수를 도입하며 대표적인 커널로 다항함수, RBF, 2층 퍼셉트론이 있다. 다중클래스 문제에 대한 분해방법들에서 SVM 알고리즘은 문제의 복잡도를 적절하게 할 수 있는 이진기들을 생성하며 분해 방법들로 생성된 이분자들이 두 클래스 분류를 유도할 수 있을 때 문제의 복잡도에 알맞게 SV의 수를 선택하는 등의 장점이 있다 [6].

3. 다중 클래스 문제 해결을 위한 기본 분해 방법  
분해 방법으로 구현되는 학습 기계들은 분해(인코딩)와

재결합(디코딩)의 두 단계로 구성된다.

분해 단계에서는  $f_1, \dots, f_L$ 인 L개의 이분기들을 학습시키기 위해 K클래스들을 열거한 분해 행렬  $D: \{-1, 0, +1\}^{L \times K}$ 을 생성한다. 이분기  $f_i$ 은 행  $D(i, \cdot)$ 에 따라 학습된다.  $D(i, k)=+1$ 이면, k클래스의 모든 예들이 positive이며  $D(i, k)=-1$ 이면, k클래스의 모든 예들이 negative이고  $D(i, k)=0$ 이면,  $f_i$ 의 학습에서 k클래스의 어떤 예들도 포함하지 않음을 의미한다. D의 열을 코드워드(codeword)[7]라고 부른다.

재결합 단계에서는 단순히 NN규칙이 일반적으로 사용되며 클래스 결과는  $f(x)$ 와 열  $D(\cdot, k)$ 간 왜곡된 측정값  $d: \mathbf{R}^L \times \{-1, 0, +1\}^K \rightarrow [0, \infty]$ 을 최소화하는 것을 선택한다.

$$class\_output = \operatorname{argmin}_k d(f(x), D(\cdot, k)) \quad (3)$$

측정값이 마진기반으로 정의되는 경우 마진기반 디코딩이라 부르며 분류기가  $h(x) \in \{-1, 1\}$ 로 정확한 이진 출력할 때, 해밍 거리를 이용한 방법은 해밍 디코딩이라 부른다.

### 3.1 OPC (One Per Class)

각 이분기  $f_i$ 는 한 클래스를 나머지 다른 클래스들과 분리하며 이는 클래스 수만큼의 이분기가 필요하다. 재결합에서 max-min 디코딩이 이 방법에 흔히 사용된다. 즉, 새 입력  $x$ 가  $f_i$ 에서 가장 높은 값을 갖는다면 j 클래스로 분류된다. 한번에 모든 클래스들을 학습하는 것이 큰 장점이며 학습 샘플이 얼굴 인식과 같이 각 클래스에 대해 적을 때 유용하다. 그러나 다양한 클래스들을 하나로 그룹화하므로 복잡한 이분기들을 생성할 수 있다.

### 3.2 PWC (PairWise Coupling)

이진분류기  $f_{ij}$ 는 클래스들의 모든 가능한 쌍에 대해 클래스i로부터 클래스j를 분리하므로, 한 이분기는 오직 두 클래스와 관련된 샘플들만으로 학습한다. 이는 간단한 이분기들을 만드는 장점을 갖지만 클래스 i도 j에도 속하지 않는 입력  $x$ 가  $f_{ij}$ 에 넣어진다면, 무의미한 결과로 인해 디코딩 과정을 고려할 때 큰 단점이 된다. 클래스 수가 늘어날 때마다, 무의미한 결과들로 성능은 떨어지므로 좋은 성능을 얻기 위해 관계 있는 분류기만을 디코드 할 필요가 있다. 이분기의 수는  $K C_2 = K(K-1)/2$ 개이며 해밍 디코딩이 이 방법에 흔히 사용된다. K=4에 대한 OPC와 PWC 방법의 분해 행렬이 그림1 (a), (b)에서 보여준다.

class 1 2 3 4	class 1 2 3 4
svm1 $\begin{pmatrix} +1 & -1 & -1 & -1 \\ -1 & +1 & -1 & -1 \\ -1 & -1 & +1 & -1 \\ -1 & -1 & -1 & +1 \end{pmatrix}$	svm1_2 $\begin{pmatrix} +1 & -1 & 0 & 0 \\ +1 & 0 & -1 & 0 \\ +1 & 0 & 0 & -1 \\ 0 & +1 & -1 & 0 \end{pmatrix}$
svm2 $\begin{pmatrix} -1 & +1 & -1 & -1 \\ -1 & -1 & +1 & -1 \\ -1 & -1 & -1 & +1 \end{pmatrix}$	svm1_3 $\begin{pmatrix} +1 & 0 & -1 & 0 \\ +1 & 0 & 0 & -1 \\ 0 & +1 & -1 & 0 \\ 0 & +1 & 0 & -1 \end{pmatrix}$
svm3 $\begin{pmatrix} -1 & -1 & +1 & -1 \\ -1 & -1 & -1 & +1 \end{pmatrix}$	svm2_3 $\begin{pmatrix} 0 & +1 & -1 & 0 \\ 0 & +1 & 0 & -1 \\ 0 & 0 & +1 & -1 \end{pmatrix}$
	svm2_4 $\begin{pmatrix} 0 & +1 & 0 & -1 \\ 0 & 0 & +1 & -1 \end{pmatrix}$

(a) OPC

(b) PWC

그림 1. 분해 행렬들. 각 행은 이분기, 각 열은 클래스

### 4. 거절 조건을 통한 두 분해 방법의 결합

OPC와 PWC는 다중 클래스 문제에 대한 분해 방법으로 흔히 사용되나, 전자는 복잡한 이분기들을 갖고, 후자는 관계되지 않은 이분기들에서 나온 무의미한 결과들이 최종 결정에 영향을 미치는 단점을 가진다. OPC와 PWC의 학습 예리

바운드는 [8]에 자세히 볼 수 있다.

$$\epsilon_K = K\epsilon_b \quad (4)$$

$$\epsilon_K \leq (K-1)\epsilon_b \quad (5)$$

$\epsilon_b$ 는 잘못 분류된 학습 예들의 평균 이분기들의 에러율이며 K는 클래스 수이다. 각 PWC 이분기는 두 클래스들만을 가지고 학습하므로 OPC보다 낮은 에러율을 가져 (5)가 (4)보다 낮은 사실에 일치한다. 이에 기반하여, 거절을 통한 두 분해방법의 결합을 제안한다. 인식과정은 OPC로 수행되며 OPC가 주어진 입력에서 낮은 신뢰를 가진 인식 결과를 출력하면 OPC의 디코딩 과정을 거절하여 이를 적당한 PWC의 입력으로 보낸다.

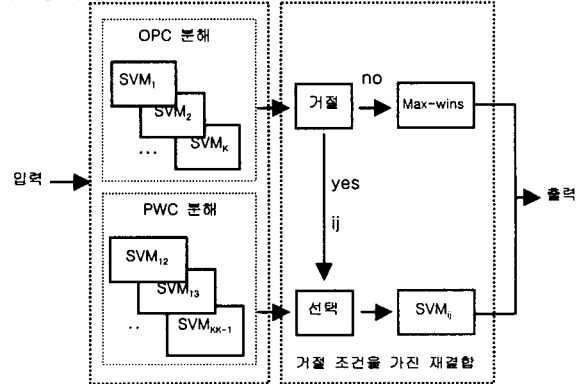


그림 2. 거절 조건을 가진 이진 분류기 순차 결합 구조

max-win 디코딩을 하는 OPC분해에서는, 다음과 같은 거절 조건을 세울 수 있다: 클래스  $c_i = \operatorname{arg}\{s_i = f_i(x)\}$ 이고  $c_j = \operatorname{arg}\{s_j = f_j(x)\}$ 이며  $i=c_j, j=c_i$ 로 정의한  $\{f_i(x), \dots, f_j(x)\}$  사이에  $s_i$ 은 가장 큰 결과값이고  $s_j$ 는 두번째로 가장 큰 결과값이다.

$$reject = \begin{cases} true & \text{if } |s_1 - s_2| \leq \theta \\ false & \text{else} \end{cases} \quad (6)$$

$$class\_output = \begin{cases} c_i & \text{if } reject = false \\ c_i & \text{if } reject = true \text{ and } SVM_{ij} \geq 0 \\ c_{ij} & \text{if } reject = true \text{ and } SVM_{ij} < 0 \end{cases} \quad (7)$$

$SVM_{ij}$ 는 클래스 i를 positive로써, 클래스 j를 negative로써 학습한 이진 분류기이다.

## 5. 실험 결과

### 5.1 얼굴 데이터베이스

40명으로 구성되어 사람당 10개 영상이 있는 ORL데이터를 사용한다. 사람마다 빛, 표정, 포즈가 다르며 회전, 스케일링, 변형의 문제들을 고정하는 전처리 과정이 눈 위치를 매뉴얼하게 찍어주면서 수행된다. 히스토그램 평활화를 적용하여 마스크를 씌운 후 PCA변형을 수행한다. 다양한 커널과 C를 가지고 SMOBR SVM을 이용하며 실험 데이터들에 leave-one-out cross validation 방법을 적용하였다.

5.2 랭크에 의한 애러율 변화

제안된 방법은 가장 높은 후보와 두번째로 가장 높은 후보만을 고려한다. 이는 적당한 답이 두 후보들 중에 하나임을 의미한다. 그림3는 랭크 변화에 따른 애러율을 보여준다.  $\sigma=1$ ,  $C=4$ 인 RBF 커널이 사용된다. 그림3.(a)에서 보는 바와 같이 랭크가 증가함에 따라 애러율이 감소함을 볼 수 있다. 게다가, 그림3.(b)에서 두 랭크간의 애러율 차는 랭크1과 랭크2에 있을 때 다른 쌍들과 확연히 구분됨을 볼 수 있다.

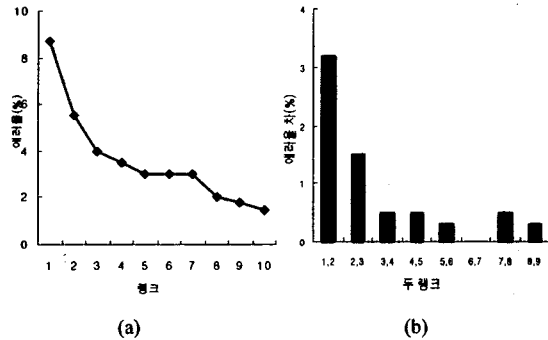


그림 3. (a) 랭크에 따른 애러율, (b) 랭크간 애러율 차

5.3 거절된 샘플 테스트

그림 4(a)는 임계 값에 의해 거절된 샘플들의 수를 보여준다.

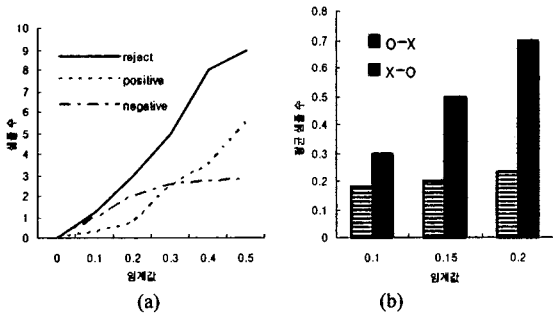


그림 4. (a) 임계 값 변화에 따른 거절된 샘플 수, (b) 임계 값에 의해 거절된 후 PWC의 결과

positive는 거절 없이 OPC에 의해 바로 분류될 수 있지만 거절 조건으로 거절된 샘플이며 negative는 거절 없이 OPC에 의해 잘못 분류되지만 거절 조건으로 거절된 샘플을 나타낸다. Reject는 거절된 샘플들의 총 수를 나타낸다. 그림4.(a)에서 임계값 0.1과 0.3사이에 Negative의 샘플 수가 Positive의 수보다 많음을 볼 수 있다. 각 거절된 샘플에 대한 PWC의 결과는 그림4.(b)에서 보여준다. o-x는 적당한 PWC를 통해 잘못 분류됐지만 거절 전에는 바로 분류된 샘플이다. x-o는 그 반대이다. x-o 샘플들의 평균 수는 o-x의 수보다 많다.

5.4 성능 비교

표1의 비교된 애러율을 보면 제안된 방법이 PWC와 OPC보다 나은 성능을 보여준다.

표 1. 애러율 결과 비교

방법	PWC (다항2차)	OPC (RBF)	제안된 방법 (다항2차)
C			
3	9.75	8.75	7.5
4	9.5	8.75	7.25
5	9.5	8.75	7.25

6. 결론

본 논문에서는 대표적인 분해방법 OPC와 PWC의 장단점을 보았다. 분류기로 SVM을 이용하며 거절과정을 통해 두 분해방법의 결합을 소개하였다. 이는 OPC로부터 잘못된 결과를 거절함으로써 OPC의 단점을 극복할 수 있다. 거절된 샘플은 PWC의 입력으로 보내진다. PWC의 무의미한 결과 문제는 OPC에 의해 가장 높은 두 랭크의 클래스들만을 고려함으로써 해결된다. ORL얼굴 데이터로 시험한 결과 제안된 방법이 실 데이터 집합 하에서 애러율이 감소함을 볼 수 있다.

□ This work was supported in part by the Biometrics Engineering Research Center, (KOSEF)

참고문헌

- [1] V.N. Vapnik, "Statistical Learning Theory", John Wiler & Sons, NY, 1998
- [2] G.Guo, S.Z.Li, and K.L.Chan,"Support vector machine for face recognition", Image and Vision Computing, Vol.19, pp.631-638, 2001
- [3] B.Heisele, P.Ho, and T.Poggio, "Face recognition with support vector machines: Global versus Component-based Approach", Proc. of IEEE ICCV, pp688-694, 2001
- [4] T.Hastie and R.Tibshirani, "Classification by Pairwise Coupling", Advanced in Neural Information Processing Systems, Vol.10, MIT Press, 1998
- [5] T.G.Dietterich and G.Bakiri, "Solving Multiclass Learning Problems via Error-Correcting Output Codes", Journal of Artificial Intelligence Research, Vol.2, pp.263-286, 1995
- [6] G.Valentini, "Upper bounds on the training error of ECOC-SVM ensembles", Technical Report TR-00-17, DISI Univ. of Genova, 2000
- [7] E.Alpaydm abd E.Mayoraz, "Learning Error-Correcting Outputs Codes from Data", Proc. of ICANN, 1999
- [8] A.Klautau, N.Jevtic and A.Orlisky, "Combined Binary Classifiers with Applications to Speech Recognition", Proc. of ICSLP, pp2469-2472, 2002