

# 국부적 영역에서의 특징 공간 속성을 이용한 다중 인식기 선택

## (Classifier Selection using Feature Space Attributes in Local Region)

신 동 국 <sup>†</sup>      송 혜 정 <sup>\*\*</sup>      김 백 섭 <sup>\*\*\*</sup>  
(Dong-Kuk Shin)    (Hye-Jeong Song)    (BaekSop Kim)

**요 약** 본 논문은 시험 표본 주위의 영역에 대한 속성을 이용한 다중 인식기 선택 방법을 제안한다. 기존의 DCS-LA 동적 인식기 선택 방법은 시험 표본 주위의 학습표본들을 사용해서 각 인식기의 국부적 정확성을 계산하여 인식기를 동적으로 선택하기 때문에 인식 시간이 오래 걸린다. 본 논문에서는 특징공간에서 국부적인 속성을 계산해서 그 속성값에 적합한 인식기를 미리 선정해서 저장해 놓은 후 시험 표본이 들어오면 그 주변의 속성값에 따라 저장된 인식기에서 선택을 하기 때문에 인식시간을 줄일 수 있다.

국부적인 속성으로는 표본 주위의 작은 영역에 대한 엔트로피와 밀도를 계산하여 사용하였으며 이들을 특징공간속성(Feature Space Attribute)라고 하였다. 이들 두 속성으로 이루어지는 속성 공간을 규칙적인 사각형 셀로 나누어, 학습과정에서 각각의 학습표본에 대해 계산된 속성값이 어떤 셀에 속하는지를 구한다. 또한 각 셀에 속하는 학습표본들에 대해 각 인식기의 국부적 정확도를 구하여 셀에 저장한다. 시험 과정에서 시험표본에 대해 속성값 계산을 통해 그 표본이 속하는 셀을 구한 후 그 셀에서 국부적 정확도가 가장 높은 인식기로 인식한다.

Elena 데이터베이스를 사용해서 기존의 방법과 제안된 방법을 비교하였다. 제안된 방법은 기존의 DCS-LA와 거의 같은 인식률을 나타내지만 인식속도는 약 4배 가까이 빨라짐을 실험을 통해 확인할 수 있었다.

**키워드** : 인식기 결합, 다중 인식기 시스템, 동적 인식기 선택, 국부적 정확성, 지역 학습

**Abstract** This paper presents a method for classifier selection that uses distribution information of the training samples in a small region surrounding a sample. The conventional DCS-LA(Dynamic Classifier Selection - Local Accuracy) selects a classifier dynamically by comparing the local accuracy of each classifier at the test time, which inevitably requires long classification time. On the other hand, in the proposed approach, the best classifier in a local region is stored in the FSA(Feature Space Attribute) table during the training time, and the test is done by just referring to the table. Therefore, this approach enables fast classification because classification is not needed during test.

Two feature space attributes are used : entropy and density of k training samples around each sample. Each sample in the feature space is mapped into a point in the attribute space made by two attributes. The attribute space is divided into regular rectangular cells in which the local accuracy of each classifier is appended. The cells with associated local accuracy comprise the FSA table. During test, when a test sample is applied, the cell to which the test sample belongs is determined first by calculating the two attributes, and then, the most accurate classifier is chosen from the FSA table.

To show the effectiveness of the proposed algorithm, it is compared with the conventional DCS-LA using the Elena database. The experiments show that the accuracy of the proposed algorithm is almost same as DCS-LA, but the classification time is about four times faster than that.

**Key words** : Classifier Combination, Multiple Classifier System, Dynamic Classifier selection, local accuracy, local learning

<sup>†</sup> 비 회 원 : 메디슨 DSP팀 연구원  
gonear@medison.com  
<sup>\*\*</sup> 비 회 원 : 한림대학교 정보통신공학부  
hjsong@hallym.ac.kr

<sup>\*\*\*</sup> 종신회원 : 한림대학교 정보통신공학부 교수  
bskim@hallym.ac.kr  
논문접수 : 2004년 5월 21일  
심사완료 : 2004년 9월 23일

## 1. 서론

단일 인식기 시스템은 학습 과정에서 하나의 결정구적만을 생성하여 분류 작업에 이용되기 때문에 실제계의 응용을 위한 다양한 특징공간을 분할하기에는 성능의 한계를 가지게 된다[1,2]. 컴퓨터 성능이 발전하고, 다양한 인식 알고리즘이 제안됨에 따라서 이러한 문제점을 극복하기 위한 방법으로 여러 단일 인식기를 결합해서 인식 성능을 향상시키는 인식기 결합 방법들이 개발되어 왔다. 보통 인식기 결합 방법은 학습단계에서 결정되는데 일반적으로 결합에 이용되는 각 인식기가 오류에 독립적이라는 가정으로 개발되었는데 오류에 독립적인 인식기들을 구성하는 것은 매우 어렵다는 문제가 있다[1,2]. 이러한 인식기 결합 방법에 대한 대안으로 제안된 동적 인식기 선택 방법은 인식단계에서 인식기를 선택하는 방법으로 인식기들은 독립적인 필요가 없으며 정확하게 분류되는 하나의 인식기만이 필요하기 때문에 독립가정보다 안정적이며 결합방법을 이용한 다중 인식기 시스템 보다 간단하다[3].

동적 인식기 선택 방법 중 대표적인 것은 Woods가 제안한 DCS LA(Dynamic Classifier Selection-Local Accuracy)이다[4]. 이 방법은 시험표본의 주위의 특징 공간의 국부적 영역에서 각각의 인식기의 정확도(local accuracy)를 추정하여 정확도가 높은 인식기를 동적으로 선택하는 것이다. 그러나 Woods의 방법은 시험표본 주변의 학습표본에 대한 국부적 정확성을 계산하기 위해 많은 시간이 소요되며, 국부적 정확성은 시험표본에 대한 특징 공간상의 구조적 속성을 정확하게 표현해내지 못하는 단점을 가지고 있다.

본 논문에서는 Woods의 동적 인식기 선택방법을 개선하여 시험표본 주변의 국부적 특징 공간의 구조적인 속성을 기반으로 한 다중 인식기 선택 방법을 제안한다. 이 방식은 특징 공간의 구조적인 속성을 나타내는 엔트로피(entropy)와 밀도(density) 등 두 가지 특징을 사용해서 특징 공간을 여러 개의 지역으로 분할하고 각 지역에서 단일 인식기의 정확도를 측정해서 정확도가 가장 높은 단일 인식기를 그 영역을 대표하는 인식기로 선택하는 것이다.

국부적 영역을 결정 하는 데는 두 가지 방법이 있다. 하나는  $k$  최근접 밀도 함수 추정과 같이 학습표본 주위의 표본의 개수를 고정시켜 영역을 결정하는 방법이고, 다른 하나는 Parzen 밀도함수 추정과 같이 영역의 크기를 고정시켜 놓고 그 안의 표본을 사용하는 방법이다. Woods의 논문에서는 전자의 방법을 사용하였다. 그러나 시험 표본에 따라서 학습 표본을 결정하는 것은 영역이 너무 큰 편차를 보이고, 영역이 너무 커져 버린 경

우 국부적 특성을 반영하지 못하는 결과를 초래하게 된다. 본 논문에서는 동등한 국부적 특성을 반영하기 위해 영역의 크기를 고정하였으며 영역 내에 표본이 없는 경우에는 최소한의 표본이 영역 내에 존재하도록 영역 크기를 조정하였다.

본 논문의 구성은 다음과 같다. 2장에서는 다중 인식기 시스템을 구성하는 방법 중 인식기 선택 방법과 관련된 연구를 소개한다. 3장에서 제안된 방법을 설명하고 4장에 실험결과를 제시한 후 5장에서 본 논문의 결론을 내린다.

## 2. 관련 연구

인식기 선택에 대한 연구는 1978년 Dasarathy와 Sheela에 의하여 제안되었다[5]. Dasarathy등은 특징 공간의 지역에 따라 다른 인식기를 사용해서 전체 인식률을 높일 수 있다는 것을 보여 주기 위하여 특징 공간에서 충돌되는 영역에서는  $k$ -NN을 사용하고 나머지 영역에서는 선형 인식기를 사용하였다. 이후 다양한 형태의 인식기 선택 방법들이 제안되었는데 이들은 정적 인식기 선택(static classifier selection)과 동적 인식기 선택(dynamic classifier selection)으로 나누어진다.

- (1) 정적 인식기 선택은 학습단계에서 영역을 나누어 각 영역에 가장 적합한 인식기를 선택하는 방법으로 시험 단계에서 입력된 미지의 값은 가까운 영역을 선택하고 해당 영역을 책임지는 인식기에 의해 결과를 출력하게 된다.
- (2) 동적 인식기 선택은 시험 단계에서 입력된 값과 관련된 영역에 대해 여러 인식기들의 성능을 평가하여 가장 적합한 인식기를 선택하는 방법이다.

정적 인식기 선택에는 학습 전략에 따라 미리 영역을 분할하여 각각의 영역을 대표하는 인식기를 할당하는 방법[6]과, 전체 학습 데이터 집합에 대해 각 인식기가 최선인 영역을 찾는 방법[7] 등 두 가지로 나누어진다. 동적 인식기 선택 방법들은 다음과 같이 국부적 정확성에 기반 한 방법, 군집화와 선택에 의한 방법, 그리고 다중 네트워크 신경망을 이용한 방법 등으로 나눌 수 있다.

- (1) 국부적 정확성 방법은 시험표본과 국부적으로 가까운 학습표본들을 이용하여 분류된 결과가 가장 우수한 인식기를 선택하는 방법이다. Woods[4]는 단일 인식기들의 정확한 인식률을 인식기 성능 측정방법으로 이용하였고, Giacinto[3,8]는 단일 인식기로부터 국부적 영역 안으로 정확하게 분류될 확률을 이용하여 인식기를 선택하였다.
- (2) 군집화와 선택방법[9,10]은 특징 공간의 각 영역에

서 가장 분류 성능이 좋은 인식기를 선택하고자 두 단계로 나누어 처리한다. 군집화 단계에서는 학습 데이터를 특징공간상에서 서로 배타적인 집합으로 구성하기 위한 부분으로 군집화를 통하여 얻어진 각 클러스터는 하나의 클래스와 대응된다. 선택단계에서 시험 표본이 입력되면 가장 가까운 클러스터를 찾고 해당 클러스터의 후보 인식기들 중에서 성능이 가장 좋은 인식기로 최종 결과를 출력한다.

- (3) 다중 네트워크 신경망을 이용한 방법[11,12]은 국부적으로 전문적인 네트워크들의 앙상블로 다중 네트워크 시스템을 구성하여 각 네트워크의 출력 값을 결합하거나 하나의 네트워크를 선택하는 방법이다. 다중 네트워크 시스템은 하나의 네트워크를 이용하여 전체 학습 데이터를 학습하는 경우에 발생하는, 학습 시간이 길고 일반화 능력이 떨어진다는 문제를 해결하기 위한 목적으로 제안되었으며 하나의 작업을 부 작업으로 나누어 각각의 부 작업에 전문적인 네트워크를 구성하여 단일 네트워크가 해결하기 힘든 문제를 쉽게 해결할 수 있도록 하였다.

### 3. 국부적 속성을 이용한 다중 인식기 선택

본 논문은 기존의 대표적인 동적 인식기 선택 방법인 Woods[4]의 DCS-LA 구조를 개선하고자 하는 것이다. 이 장에서는 DCS-LA 구조를 설명한 후 본 논문에서 제안하는 특징 공간 속성을 이용한 인식기 선택 방법에 대해 설명한다.

#### 3.1 Woods의 DCS-LA 구조

Woods의 DCS-LA 구조를 그림 1에 보인다. DCS-LA는 특징공간에서 시험표본 주위의 국부적 영역에서 각각의 인식기의 정확도를 측정 하여 가장 국부적 정확성이 높은 인식기를 동적으로 결정하여 이 인식기로 인식을 한다. 여기서 국부적 영역은 k-NN 인식기와 같이 학습표본을 중심으로 k번째 가까운 학습표본까지의 거리를 반지름으로 하는 원(hypersphere)으로 정의된다.

학습표본으로 n개의 단일 인식기들( $D_1, \dots, D_n$ )을 미리 학습해 놓으면 인식단계에서 그 중 하나를 선택하

게 되는데, 구체적인 알고리즘은 다음과 같다.

- (1) 시험표본  $x$ 에 대해 모든 단일 인식기가 동일한 클래스 레이블을 출력한다면 그 레이블을 인식결과로 출력하고 종료한다.
- (2) 만약 다르다면,  $x$ 주위의 국부적 영역 안에 있는 k개의 학습표본들을 이용하여 각 인식기의 국부적 정확도를 평가한다. 정확성을 평가하는 방식에 따라 두 가지 방식으로 나누어 진다.
  - (가) Overall Local Accuracy -  $D_i$ 로 k개의 학습표본을 인식한다. 인식결과가 학습표본의 클래스 레이블과 일치하는 비율을  $D_i$ 의 국부적 정확도로 한다.
  - (나) Local Class Accuracy -  $D_i$ 로  $x$ 를 먼저 인식한다. 인식결과 클래스 레이블이  $s$ 라면, k개의 학습표본 중에 클래스 레이블이  $s$ 인 것을 골라  $D_i$ 가 이들을 올바르게 인식하는 비율을  $D_i$ 의 국부적 정확도로 한다.
- (3) 정확도가 가장 높은 인식기로  $x$ 를 인식하여 인식결과를 출력한다.

이 방법은 시험표본이 들어올 때마다 적합한 단일 인식기를 선택하는 동적 선택 방법을 사용하고 있다. 동적 선택을 하는 과정에서 최소  $n*k$  번 인식을 해야 하기 때문에 인식 시간이 많이 걸린다는 단점이 있다.

#### 3.2 제안된 다중 인식기 선택 방법

본 논문에서 제안하는 다중 인식기 선택방법을 그림 2에 보인다. 그림에서 실선으로 표시된 부분은 학습단계를 나타내고 점선으로 표시된 부분은 인식단계를 나타낸다. 제안된 방법은 인식단계에서 FSA(Feature Space Attribute) 테이블을 참조해서 인식기를 정적으로 선택한다는 점이 DCS-LA와 크게 다르다. FSA 테이블은 학습단계에서 생성되는데 이 과정에서 DCS-LA에서와 같이 국부적 정확도를 이용한다.

Bottou와 Vapnik[13]은 특징공간에서 결정경계(decision boundary)가 복잡한 영역에서는 복잡도(complexity)가 높은 인식기를 사용하는 것이 적합하고, 그렇지 않은 영역에서는 복잡도가 낮은 인식기를 사용하는 것이 전체적인 인식오차를 줄일 수 있다고 하였다. 이는 영역의 속성에 따라 다른 인식기를 선택하는 것이 바람직하다고 해석할 수 있다. 본 논문에서는 특징공간속성(Feature Space Attribute) 계산을 통해 특징공간을 속성공간(Attribute Space)으로 변환한다. 속성공간을 규칙적인 작은 셀로 분할하여 각 셀에 인식기를 할당하여 FSA 테이블에 저장한다. 인식과정에서는 단지 학습표본을 속성공간으로 변환하여 어떤 셀에 속하는지를 판단한 후 FSA 테이블에서 인식기를 얻어 그 인식기로

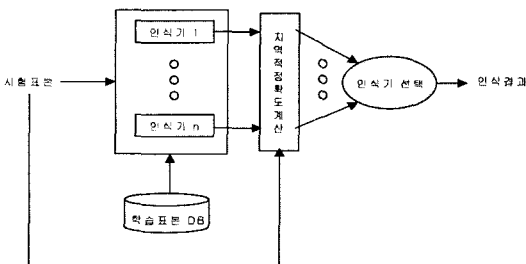


그림 1 DCS-LA 구조

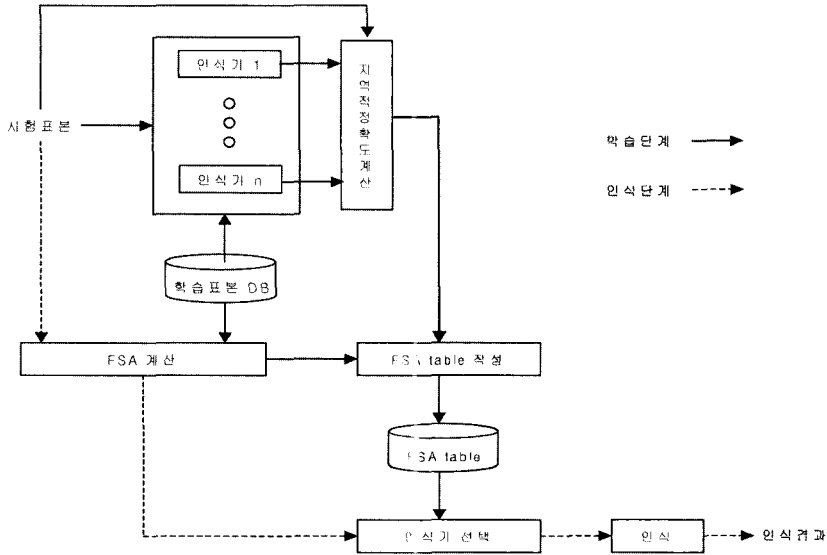


그림 2 제안된 다중 인식기 선택 방법

인식한다. 그러므로 인식과정에서 단지 1번의 인식으로 결과를 얻을 수 있어 기존의 DCS-LA에서  $n \times k$ 번 인식에 비해 빠르게 된다. 제안한 방법의 각 블록에 대한 구체적인 설명은 아래와 같다.

3.2.1 특징공간 속성(FSA)

특징공간 속성은 어떤 표본(시험표본 혹은 학습표본)  $x$  주위의  $k$ 개의 학습표본이 어떻게 분포하고 있는가를 나타낸다. 속성으로는 밀도(density)와 엔트로피(entropy)를 사용하며 각각은 다음과 같이 계산된다.

$x$ 와  $k$ 번째 가까운 학습표본과의 거리를  $r$ 이라고 하고 특징공간의 차원을  $d$ 라고 하자. 밀도는 표본들이 얼마나 밀집되었는가의 정도를 말하는 것으로 식 (1)과 같이 정의 된다.

$$\text{Density}(x) = \frac{1}{r^d} \tag{1}$$

표본이  $C_1, C_2, \dots, C_m$  등  $m$ 개의 클래스 레이블을 갖는다고 하고  $x$  주위의  $k$ 개의 학습표본중에서 클래스 레이블이  $C_i$ 인 학습표본이  $s_i$ 개 있다고 하면 엔트로피는 다음과 같이 정의 된다.

$$\text{Entropy}(s_1, s_2, \dots, s_m) = - \sum_{i=1}^m p_i \log_2(p_i) \tag{2}$$

여기서  $p_i$ 는 클래스  $C_i$ 에 속하는 표본의 확률로서  $s_i/k$ 로 추정한다.

특징공간에서 하나의 표본에 대해 이와 같이 두개의 속성을 계산할 수 있다. 두개의 속성으로 구성된 공간을 속성공간이라고 할 때, 특징공간에서 하나의 표본은 속성공간에서 한 점으로 투사된다.

3.2.2 FSA 테이블

FSA 테이블은 속성공간에 각 인식기의 정확도를 할당하기 위해 작성된다. 먼저 밀도와 엔트로피 값을 각각 규칙적으로 10개의 구간으로 나누어 속성공간을 100개의 셀(cell)로 분할한다. 하나의 셀에는 선택개수(number of hits)와 각 인식기의 국부적 정확도 등  $n+1$ 개의 슬롯(slot)이 부여된다. 여기서  $n$ 은 단일 인식기의 개수이다.

FSA 테이블은 학습표본을 인가해서 셀을 선택하고 선택된 셀에 대한 슬롯 값을 바꾸어서 작성된다. 구체적인 과정은 다음과 같다.

- (0) 속성공간에서 모든 셀의 모든 슬롯을 0으로 초기화한다.
- (1) 학습표본을 순차적으로 선택한다. 선택된 학습표본을  $x$ 라고 하자.
- (2)  $x$ 에 대하여 엔트로피와 밀도를 계산한다.
- (3) 속성공간에서 계산된 엔트로피와 밀도에 해당하는 셀을 찾아 선택개수를 증가한다.
- (4) DCS-LA 방법으로  $x$  주위에 대해 각 인식기의 국부적 정확도를 계산한다.  $x$ 를 정확하게 인식하지 못한 인식기는 국부적 정확도를 0으로 한다.
- (5) (3)에서 구한 셀의 각 인식기의 국부적 정확도 슬롯에 대해 (4)에서 구한 값을 누적한다.
- (6) 다음 학습표본을 선택해서  $x$ 라고 하고 (2)로 간다. 더 이상 선택할 학습표본이 없으면 종료한다.

3.2.3 인식기 선택

학습과정에서 생성된 FSA 테이블은 인식 단계에서

인식기 선택에 이용된다. 기본적으로 시험표본이 속한 셀을 찾아 그 셀에서 국부적 정확도가 가장 높은 인식기를 선택하게 되지만 셀의 선택개수가 너무 작으면 기존의 DCS-LA 방법으로 인식기를 선택하게 된다. 보다 구체적인 과정은 다음과 같다.

- (1) 입력된 시험표본  $x$ 에 대해 주위의 학습표본을 사용해서 밀도와 엔트로피를 계산한다.
- (2) 속성공간에서 계산된 엔트로피와 밀도에 해당하는 셀을 찾는다.
- (3) 위에서 구한 셀의 선택개수 슬롯값이 임계치보다 작으면, DCS-LA 방법으로 인식기를 선택하여 인식결과를 출력하고 종료한다.
- (4) (2)에서 구한 셀의 선택개수 슬롯값이 임계치보다 크면, 해당 셀의 국부적 정확도 슬롯값이 가장 큰 인식기를 선택해서 그 인식기로  $x$ 를 인식하여 인식결과를 출력하고 종료한다.

### 4. 실험 및 결과 분석

#### 4.1 실험환경

실험은 ELENA Project[14]로부터 얻은 데이터베이스 중 Phoneme\_CR과 Satimage\_CR 데이터 집합을 사용하였다. 여기서 Satimage\_CR은 6개의 클래스를 가지고 있는데, 이 논문의 실험에서는 4번과 6번 클래스만을 뽑아 두 개의 클래스 문제로 변형하여 사용하였다. Phoneme\_CR은 5개의 특징을 가지며 4,144 표본이 있고, Satimage\_CR은 36개의 특징을 가진 2134개의 표본으로 구성되어 있다. 인식률은 divide by half 방법으로 계산하였다. 즉 전체 데이터베이스에서 임의로 반을 나누어 학습표본 집합으로 사용하고 나머지 반은 시험표본 집합으로 사용하였다. 이 과정을 10번 반복해서 각각의 인식률의 평균을 전체 인식률로 하였다. 단일 인식기로는 k-NN, 선형 커널을 사용한 SVM(SVM-linear), RBF(Radial Basis Function)을 사용한 SVM(SVM-RBF) 등 세 가지를 사용하였다.

실험환경은 펜티엄4 2.0Ghz, 메모리 256Mbytes이고, Windows 2000상에서 Matlab 프로그램을 사용하였다.

### 4.2 실험결과

#### 4.2.1 인식률

표 1은 단일 인식기, 기존의 DCS-LA 방법, 그리고 제안한 방법 각각의 인식률을 보인 것이다.

표 1 Woods의 DCS-LA와 제안한 방법에 대한 인식률

method of Classification		Phoneme	Satimage
단일인식기	K-Nearest Neighbor	86.41%	88.61%
	SVM_linear	77.37%	82.39%
	SVM_RBF	87.02%	90.61%
다중인식기 선택방법	DCS-LA - Overall	86.98%	90.61%
	DCS-LA - Local	87.06%	90.67%
	제안한 방법	87.41%	90.78%

표 1에서 인식률이 가장 높은 것은 굵은 글씨로 표시하였다. 제안한 방법이 기존의 DCS-LA보다 약간 높은 인식률을 보이고 있으나 차이가 미미하여 기존의 국부적 정확도를 측정하는 DCS-LA 방법과 비슷한 정도의 인식률을 보이고 있다고 할 수 있다.

그림 3은 제안한 방법이 속성공간의 각 셀에서 어떤 인식기를 선호하였는가를 보여주고 있다. 그림에서 각 셀의 선택개수 슬롯값이 임계치보다 작으면 0이라고 표시하였고 그렇지 않은 경우는 국부적 정확도가 가장 높은 인식기를 표시하였다.

그림에서 전반적으로 엔트로피가 높은 경우에는 SVM-RBF가 선택되고 엔트로피가 낮은 경우에는 k-NN이 선택되고 경향을 보이고 있다. k-NN은 아주 간단한 결정함수를 사용하는 국부적 인식기(local classifier)인데 반해 SVM-RBF는 복잡한 결정함수를 사용하는 전역적 인식기(global classifier)이다. 엔트로피가 낮은 영역, 즉 시험표본의 주위의 학습표본들 대부분이 동일한 클래스 레이블을 가지는 경우는 간단한 결정함수

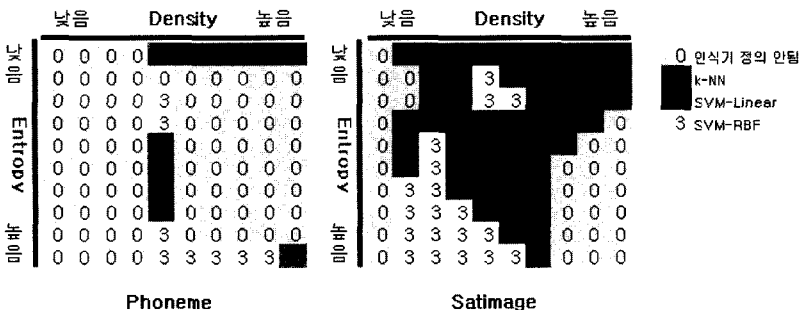


그림 3 속성공간의 각 셀에서 인식기 선호도

표 2 인식방법별 학습시간 (단위 : 초)

인식 집합	인식 방법	인식기 학습	FSA 테이블 작성	합 계
Phoneme	DCS LA	264.087	0	264.087
	제한한 방법	264.087	14.756	278.843
Satimage	DCS LA	186.834	0	186.834
	제한한 방법	186.834	11.213	198.048

표 3 인식방법별 시험시간 (단위 : 초)

인식 집합	인식 방법	kNN 탐색	인식기 선택	인 식	합 계
Phoneme	DCS LA	5.966	17.814	0.059	23.840
	제한한 방법	5.968	0.0001	0.00006	5.968
Satimage	DCS LA	4.371	10.464	0.0046	14.840
	제한한 방법	4.377	0.0001	0.00002	4.377

로 인식하는 것이 시험오차의 분산(variance)을 줄일 수 있기 때문에 k-NN이 주로 선택된 것으로 판단된다. 반면 엔트로피가 높은 영역에서는 간단한 결정함수로는 시험오차의 편차(bias)가 커지기 때문에 복잡한 결정함수를 선호한 것으로 판단된다. SVM-RBF와 SVM-linear는 속성공간에서 비슷한 위치에서 선택되는데, SVM-RBF가 선택되는 빈도가 월등히 많다. 이는 표 1에서 단일 인식기에 대한 인식을 비교에서도 알 수 있듯이 시험 대상이 되는 데이터베이스가 SVM-linear로 인식하기에는 결정경계가 복잡하여 복잡도가 높은 SVM-RBF를 선호한 것으로 추정된다.

4.2.2 계산시간

각각의 데이터 집합에 대해서 학습 시간과 인식 시간으로 나누어서 측정하였다. 학습 시간은 인식기 결합을 위해 각각의 인식기를 학습하는 시간과 제안한 논문에서 필요한 FSA 테이블 작성 시간으로 나누어진다. 인식 시간은 시험표본이 들어 왔을 때 그것으로부터 가까운 k개의 학습표본을 선택하는데 걸리는 시간(kNN 탐색시간), 시험표본의 특징공간속성을 계산하여 FSA테이블에서 인식기를 선택하는데 걸리는 시간(인식기 선택시간), 마지막으로 선택한 인식기로 인식하여 인식결과를 얻는 시간(인식시간)으로 나누어진다. 표 2는 학습시간을 측정할 결과를 보여주고, 표 3은 인식 시간을 측정할 결과를 보여 주고 있다.

제한한 방법은 기존의 DCS-LA에 비해 학습 단계에서 FSA테이블 작성 때문에 약간의 시간이 추가로 필요하나 단일 인식기들을 학습하는데 걸리는 시간에 비해 매우 작아 전체 시간은 기존의 방법과 큰 차이를 보이지 않는다. 그러나 인식 단계에서는 인식기를 선택하는 데는 기존 방법에 비해 현저하게 시간이 줄어들어 전체 시간이 크게 단축됨을 알 수 있다. 이는 기존 방법이 동적으로 인식기를 선택하는 반면 제안한 방법은 FSA 테이블을 사용하여 단순한 계산을 통해 인식기를 선택할

수 있기 때문이다. 전체 인식 시간은 기존의 방법에 비해 4배 가까이 빨라졌다. 여기서 kNN 탐색은 간단히 전역탐색(full search) 알고리즘으로 구현하였지만 k-d tree 같은 고속 탐색방법으로 구현하면 탐색시간을 더 줄일 수 있어 전체 인식시간의 차이가 더 커질 것이다.

5. 결론

본 논문에서는 Woods의 국부적 정확도를 사용한 동적 인식기 선택방법(DCS-LA)의 인식시간을 줄이는 방법을 제안하였다. 표본 주변의 학습표본들의 분포를 분석하여 특징공간을 속성공간으로 변환하고 속성공간을 규칙적인 셀로 나누어 각 셀에 각 단일 인식기의 국부적 정확도를 저장하였다. 인식단계에서는 시험표본이 속한 셀을 속성공간에서 구해 그 셀에서 인식기를 빠르게 선택하여 인식시간을 줄였다. 시험결과 기존의 방법에 비해 거의 비슷한 인식률을 보이지만 인식시간은 4배 가까이 단축됨을 알 수 있었다.

제한한 방법은 기존의 방법과 마찬가지로 국부적 정확도를 사용하여 인식기를 선택하고 있다. 국부적 정확도를 추정하는 과정에서 표본 주위의 k개의 학습표본을 사용하는데 정확도를 추정하는데 표본 개수가 적어 추정 오차가 크다는 근본적인 문제를 해결하지는 못했으며 이 때문에 인식률이 향상되지 않은 것으로 판단된다. 인식기를 평가하는데 기존의 국부적 정확도를 계산하는 방법을 개선하면 인식률을 향상할 수 있을 것으로 기대되어 향후 과제로 남긴다.

참고 문헌

[1] L. Xu, A. Krzyzak, and C. Y. Suen, "Methods of combining multiple classifiers and their applications to handwritten character recognition," *IEEE Trans. Systems, Man, and Cybernetics*, Vol. 22, No. 3, pp. 418-435, 1992.  
 [2] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas,

"On combining classifiers," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 20, No. 3, pp. 226-239, 1998.

- [3] G. Giacinto and F. Roli, "Methods for Dynamic Classifier Selection," ICIAP '99, 10th International Conference on Image Analysis and Processing, Venice, Italy, pp. 659-664, 1999.
- [4] K. Woods, W. P. Kegelmeyer Jr., K. Bowyer, "Combination of multiple classifiers using local accuracy estimates," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 19, pp. 405-410, April 1997.
- [5] B. V. Dasarathy and B. V. Sheela, "A composite classifier system design : concepts and methodology," Proceeding of IEEE, Vol. 67, pp. 708-713, 1978.
- [6] A. Verikas, A. Lipnickas, K. Malmqvist, M. Bacauskiene, and A. Gelzinis, "Soft combination of neural classifiers: A comparative study," Pattern Recognition Letters, Vol 20, pp. 429-444, 1999.
- [7] L. A. Rastrigin and R.H. Erenstein. "Method of Collective Recognition," Energoizdat, Moscow, 1981.
- [8] G. Giacinto and F. Roli, "Dynamic classifier selection based on multiple classifier behaviour," Pattern Recognition, Vol. 34, Iss. 9, pp. 1879-1881, September 2001.
- [9] L. I. Kuncheva, "Cluster-and-Selection method for classifier combination," Proc. 4th International Conference on Knowledge-Based Intelligent Engineering Systems & Allied Technologies (KES '2000), pp. 185-188, Brighton, UK, 2000.
- [10] Y. S Hwang and C. Y. Suen, "a method of combination multiple experts for the recognition of unconstrained handwritten numerals," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 17, No. 1, pp. 90-94, January 1995.
- [11] A. J. C. Sharkey, "Multi-net systems, in Combining Artificial Neural nets : Ensemble and Modular Multi-Net Systems," pp. 1-30, Springer-Verlag, London, 1999.
- [12] J. X. Dong, A. Krzyzak, and C. Y. Suen, "A Multi-Net Local Learning Framework for Pattern Recognition," In Proc. of 7th ICDAR, pp. 328-332, 2001.
- [13] L. Bottou and V. Vapnik, "Local learning algorithms," Neural Computation, vol. 4, no. 6, pp. 888-900, 1992.
- [14] <http://www.dice.ucl.ac.be/neural-nets/Reserach/Projects/ELENA/>



신 동 국

2002년 한림대학교 학사 졸업(공학사)  
2004년 한림대학교 석사 졸업(공학석사)  
2004년~현재 메디슨 DSP팀 연구원. 관  
심분야는 영상처리, 패턴인식 등



송 혜 정

1989년 한림대학교 전자계산학과(이학  
사). 1998년 한림대학교 컴퓨터공학과(공  
학석사). 2003년 한림대학교 컴퓨터공학  
과(공학박사). 2000년~현재 한림대학교  
정보통신공학부 연구전임강사. 관심분야  
는 컴퓨터비전, 패턴인식 등



김 백 섭

1978년 한양대학교 전자공학과(공학사)  
1980년 한국과학기술원 산업전자공학과  
(공학석사). 1987년 한국과학기술원 전기  
및 전자공학과(공학박사). 1984년~1986  
년 한양대학교 의과대학 제왕의학교실  
조교수. 1989년~1990년 Syracuse 대학  
CASE center 객원연구원. 1987년~현재 한림대학교 정보  
통신공학부 교수. 관심분야는 컴퓨터 비전, 영상처리, 패턴  
인식, 신경망, 생체인식 등임