

# Hypercube 영역의 집합으로 표현된 패턴인식 알고리즘의 설계

## (A Design of Pattern Recognition Algorithm as a Collection of Hypercubic Regions)

金 伯 燮\*

(Baek Sop Kim)

### 要 約

본 논문은 패턴인식기의 형태를 hypercube 영역의 집합으로 표현하는 방법을 제안한다. 이 표현의 장점으로 는 사람의 사전지식을 이용할 수 있고 결과를 규칙기반시스템에 사용할 수 있으며 판별속도가 빠르다는 점이다. 이런 형태의 판별기를 설계하는 하나의 방법을 제시하였으며 이 방법에 의해 판별되지 않는 표본들을 기존nearest neighbor 방법으로 다시 판별하더라도 평균 판별시간은 기존의 방법만으로 하는 경우에 비해 적게 걸린다는 것을 실험을 통해 보여 주었다.

### Abstract

In this paper, a method of representing the pattern classifier as a collection of hypercubic regions is proposed. This representation has following advantages over the conventional ones : 1) a simple form of human knowledge can be used in designing the classifier, 2) the form of the classifier is suit for the rule-based system, and 3) this can reduce the classification time. A method of synthesis of the classifier under this representation is also proposed and the experimental result shows that the proposed method is faster than the well-known nearest neighbor classifier.

### I. 서 론

통계적 패턴인식은 표본(sample)의 특징값(feature value)들을 추출해서 이를 기초로 그 표본이 어떤 집단에 속하는가 하는 것을 결정하고자 하는 것이다. 이 결정은 판별규칙(discriminant rule)을 통해서 하게 되는데 판별규칙은 학습표본들(training samples)로부터 얻어진다. 판별규칙을 얻기위한 기존의 통계적 방법은 Bayes의 결정이론(decision theory)에 의거하는데 이는 다음과 같은 단점이 있다.

첫째, 사람이 가지고 있는 사전 판별지식(priori discriminant knowledge)을 이용할 수 없다. 예를들어 키와 몸무게라는 두개의 특징을 사용하여 농구선수와 체조선수를 구별하고자 하는데 “키가 190cm이상이고 몸무게가 70kg이상이면 농구선수이다.”라는 사전 판별지식을 알고 있다고 하자. 이를 판별규칙 설계에 직접이용할 수는 없고 단지 이에 해당하는 표본들을 학습표본에 추가시키는 간접적인 방법만이 가능할 따름이다.

둘째, 얻어진 판별규칙의 형태가 사람이 이해하기 힘든 형태이다. 예를들어 간단한 경우라도 “키에 1.2를 곱한 것과 몸무게를 더한 것이 286보다 크면 농구선수이다.”라는 형태가 된다. 그러므로 판별규칙의 결과를 보고 사람이 필요에 따라 규칙을 수정하는 것이 힘들어 대화형(interactive) 판별기를 만드는데 제약조건이 된다.

\* 正會員, 翰林大學校 電子計算學科  
(Dept. of Comp. Sci., Hallym Univ.)  
接受日字 : 1992年 1月 25日

이러한 단점을 해결하는 하나의 방법으로써 본 논문에서는 판별규칙의 형태를 특징공간에서 hypercube 영역의 집합으로 표현하는 판별기를 제안한다. 이렇게 함으로써 다음과 같은 장점을 얻을 수 있다.

첫째, 간단한 형태의 사전 판별지식의 사용을 가능하게 한다. 여기서 사전 판별지식의 형태는 편의상 특징공간에서 hypercube 영역의 집합으로 표현될 수 있는 것으로 제한한다. 이렇게 제한함으로써 학습표본은 특징공간 내의 점으로써, 그리고 사전 판별지식은 영역의 집합으로써 표현할 수 있다. 이런 형태의 사전 판별규칙은 특징들 혹은 표본들 사이의 관계를 표현할 수는 없지만 판별규칙을 특징공간의 분할문제로 보는 통계적(혹은 비 구조적) 패턴인식 방법에서의 사전 판별지식의 표현 형태로는 적합하다.

둘째, 판별기의 결과를 규칙기반 시스템(rule-based system)의 규칙으로 사용할 수 있다. 이는 판별규칙의 형태가 예를들어 “키가 178cm 이상이거나 키가 170cm 이상 이고 몸무게가 68kg 이상이면 농구선수이다.”와 같이 사람이 이해하기 쉬운 기호(symbol) 형태이기 때문이다.

셋째, 판별시간이 기존의 방법보다 적게 걸린다. 왜냐하면 기존의 통계적 패턴인식 방법들에서 판별함수의 형태는 근본적으로 접합치(matching score)를 계산하는 것이므로 산술 연산이 필요하다. 그러나 제안된 방법은 주어진 표본이 특징공간에서 어떤 hypercube 영역에 속하는지 여부를 판단하는, 비교 연산만으로 가능하기 때문이다.

이렇게 판별규칙의 형태를 미리 결정하고 학습표본으로부터 필요한 모수(parameter)들을 결정하는 방법은 기존 방법중 결정적 방법(deterministic method)들과 비교할 수 있다. 이중 판별규칙의 형태를 hyperplane으로 정한 경우는 Rosenblatt의 퍼셉트론(perceptron), Fisher의 선형 판별 함수(linear discriminant function), Widrow와 Hoff의 델타규칙(delta rule) 등이 잘 알려져 있다.<sup>[1]</sup> 또한 복합형 판별기(compound classifier)<sup>[2]</sup>와 그의 인공 신경망 형태인 제한된 Coulomb에너지망<sup>[3]</sup>의 경우는 판별규칙의 형태를 hypersphere의 집합으로 표현한 것이다.

본 논문은 R.S.Michalsky의 일들에 기초를 둔다. 그는 하나의 가변치 논리체계(variable valued logic system)을 제안하고<sup>[4]</sup> 이를 이산치 변수(discrete variable)를 갖는 경우의 패턴인식기의 설계<sup>[5, 6]</sup>, 연역형 추론기의 설계<sup>[7]</sup>, conceptual clustering<sup>[8]</sup>등에 응용하였다. 이들은 입력 자료의 형태가 이산형 변수<sup>[5, 6]</sup>이거나 논리식 형태<sup>[7]</sup>인데 반하여 본 논문은 이를 연속형 변수(continuous variable) 경우로 확장하여 통계적 패턴인식 분

야에 이용하고자 하는 것이다.

2장에서 기본 개념과 표기법을 제시하고 3장에서 판별규칙의 설계 방법을 설명한다. 4장에서는 기존의 nearest neighbor과 기존의 방법을 실험적으로 비교하고 5장에서 결론을 내린다.

## II. 기본개념과 및 표기법

본 논문에서 사용되는 기호 및 용어는 R.S.Michalsky가 사용한 방법<sup>[5]</sup>을 기본으로 하여 연속변수 경우로 약간 변경하였다. 하나의 표본  $x = (x_1, x_2, \dots, x_d)$ 로 표시되는데 여기서  $x_i$ 는  $x$ 의  $i$  번째 특징값(feature value)이다. 이는 차원이  $d$ 인 특징공간 내에서 하나의 점으로 생각할 수 있다.

$i$  번째 특징값이  $a$  와  $b$  사이에 있는 표본들의 집합을 literal이라고 부르고  $X^b$ 로 표기한다. 즉,

$$X^b = \{ (x_1, x_2, \dots, x_d) \mid a \leq x_i \leq b \} \text{이다.}$$

특히  $a=b$ 인 경우를 기본 리터랄(elementary literal)이라고 부르고  $x_i^a$ 로 표기한다.

Complex는 literal의 집합론적 곱(set-theoretical product)이다 즉, 어떤 complex  $L$ 은

$$L = \bigcap_i X_i, \quad a_i, b_i, \quad i=1, 2, \dots, d \text{이다.}$$

Complex는 특징공간 내에서 hypercube로 표현된다. 이는 점을 유의하기 바란다. 여기서  $d$ 개의 literal들이 모두 기본 literal일 경우 이 complex를 단순(simple) complex라고 부르고 이는 하나의 표본을 나타내며 특징공간 내의 한 점과 대응된다. 단순 complex가 아닌 complex는 특별히 구간(interval) complex라고 부른다.

서로 disjoint한 표본들의 집합  $A = \{x_1, x_2, \dots, x_n\}$  과  $B = \{y_1, y_2, \dots, y_m\}$  이 있다고 하자.  $B$ 에 대한  $A$ 의 cover는  $A$ 에 속하는 모든 표본들을 포함하고  $B$ 에 속하는 어떤 표본도 포함하지 않는 complex들의 집합이다. 이를  $CV(A/B)$ 라 표기하면 이는 다음 조건을 만족하는 complex들의 집합이다.

$$A \subseteq CV(A/B) \subseteq -B.$$

여기서  $-B$ 는  $B$ 에 대한 보집합(complementary set)을 나타낸다.

판별기를 설계하는데 기존의 방법에서는 표본들에 대

한 정보만을 이용하였으나 여기서는 사전 판별지식이라는 하나의 정보를 추가한다. 사전 판별지식에는 여러가지 형태가 있을 수 있으나 본 논문에서는 complex형태로 제한한다. 즉, “특징  $X_1$ 의 값이 a와 b 사이이고  $X_2$ 가 c보다 크면 집단 A에 속한다.”와 같은 형태이다. 이것은 ( $^aX_1^b$  complex, A)와 같이 (complex, 집단이름)의 쌍으로 표시할 수 있다. (L은 특징이 가질 수 있는 제일 큰 값, S는 제일 작은 값이다.) 이렇게 형태를 제한한 이유는 기존의 통계적 패턴인식에 사전 판별지식을 추가하려고 하였기 때문이다. 통계적 패턴인식에서는 특징 혹은 표본들 사이의 관계를 표현하지 못한다는 점에 유의하기 바란다. 또한, 학습표본들도 “특징  $X_1$ 의 값이 a이고  $X_2$ 의 값이 b이면 집단 A에 속한다.”와 같은 형태로 표현할 수 있으므로 이들도 (complex, 집단이름)의 쌍으로 표시할 수 있다. 결국 본 논문에서 판별기를 설계하는 입력 정보는 모두 (complex, 집단이름)의 쌍으로 주어진다. 여기서 complex들은 학습표본에 대해서는 단순 complex, 사전 판별지식에 대해서는 구간 complex인 점만이 다르다.

이상에서, 두개의 집단 A와 B가 있다고 할 때, 본 논문에서 다루고자 하는 문제는 다음과 같이 정리된다 : (complex, 집단이름)의 쌍들의 집합이 주어졌을 때 cover  $CV(A/B)$ 와  $CV(B/A)$ 를 구하라. cover를 구하는 하나의 방법을 3장에서 설명한다.

### III. Cover 생성과정

이 장에서는 집단 B에 대한 집단 A의 cover  $CV(A/B)$ 를 구하는 방법만을 언급한다.  $CV(B/A)$ 의 경우는 A, B 두 집단의 이름을 서로 바꾸어  $CV(A/B)$ 를 구하면 되므로 과정은 동일하다.  $CV(A/B)$ 를 구하는 과정은 그림 1과 같이 세 부분으로 나누어 지는데 이 그림에서 complex는 구간 complex를 뜻한다. 각각의 과정은 다음과 같다.

#### 1. complex 생성

집단 A내의 각각의 표본에 대해 하나씩의 구간 complex를 생성한다. 이들은 그림 1에서 보듯이 사전 판별지식에 의한 complex들과 함께 complex 선택과정으로 보내지게 된다. 집단 A내의 하나의 표본 x와 집단 B의 표본들이 주어졌을 때 x에 대한 complex 생성 과정은 다음과 같다.

```
function Complex_gen(x, B) : complex;
var L, T : complex;
    i : integer;
begin
```

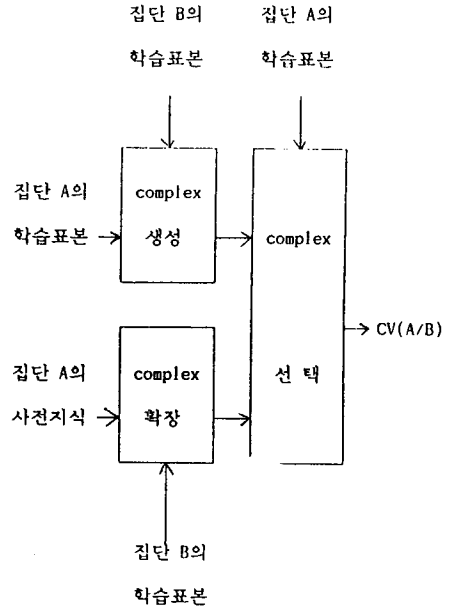


그림 1. Cover 생성 과정  
Fig. 1. Cover generation process.

$$L \leftarrow \bigcap_{i=1}^d S X_i^t : \{ S=Small\ number, L=Large\ number \}$$

for  $i = 1$  to  $m$  do

$begin$   $T \leftarrow extend(x, y_i) ; \{ y_i \text{는 } B \text{의 } i \text{번째 표본} \}$

$L \leftarrow T \cap L ;$

$end ;$

$return(L) ;$

$end ;$

여기서  $extend(x, y)$ 는  $x=(x_1, x_2, \dots, x_d)$ 를 포함하고  $y=(y_1, y_2, \dots, y_d)$ 를 포함하지 않는 하나의 complex를 구하는 함수이다. 이 함수는 다음과 같이 정의된다.

$function\ extend(x, y) : complex ;$

$var\ d, t : real ;$

$i\ integer ;$

$begin$

$$L \leftarrow \bigcap_{i=1}^d S X_i^t : \{ S=Small\ number, L=Large\ number \}$$

$d \leftarrow square\_dist(x, y) ;$

$c \leftarrow farthest\_coord(x, y) ;$

$t \leftarrow (x_c + y_c) / 2.0 ;$

$if\ (x_c < y_c)\ then\ return\ (L \cap S X_c^t)$

$else\ return\ (L \cap ^t X_c^t) ;$

$end ;$

여기서,  $square\_dist(x, y) = \max_k |x_k - y_k|$ 이고  $far$

the<sub>st</sub>\_coord(x, y)는  $|x_k - y_k|$ 가 최대가 되는 좌표축을 되돌려 주는 함수이다.

그림 2에 위 방법으로 생성된 complex들을 보인다. 여기서 동그라미로 표시된 것들은 집단 A의 표본들이고 네모로 표시된 것들은 B의 표본들이다.  $L_1, L_2, L_3$ 는 각각 집단 A의 표본들  $x_1, x_2, x_3$ 에 대해 생성된 complex들을 나타낸다. 이 그림에서 보듯이 집단 A의 표본에 대한 complex는 그 표본을 집단 B에 대해 일반화(generalize)시킨 것이라고 생각할 수 있다. 또한 하나의 complex(예를들어  $L_2$ )내의 어떤 점에서도 그 complex를 생성한 표본( $x_2$ )과의 거리가 집단 B의 어떤 표본과의 거리보다 가깝다는 사실에 주의하기 바란다. 여기서의 거리는 사각거리(square distance)를 뜻한다.

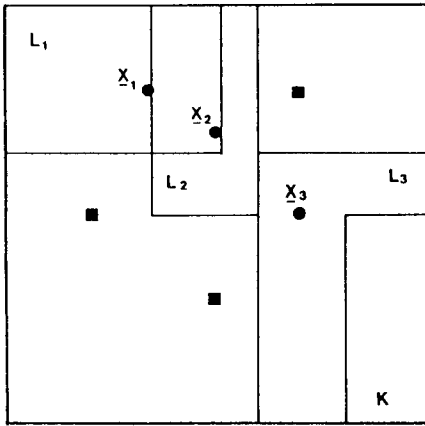


그림 2. 학습표본, 사전 판별지식 및 생성된 complex들  
Fig. 2. Training samples, priori discriminant knowledge and complexes generated from the training samples.

2. complex 확장

집단 A에 대한 사전 판별지식이 주어지면 이를 일반화시킨 후 complex선택과정으로 가게된다. 사전 판별지식을 일반화 시키는 과정은 1절과 비슷하게 주어진 사전 판별지식을 집단 B의 표본들에 대해 확장하는 것이다. 사전 판별지식은 앞에서 설명한 바와 같이 complex들의 집합인데 하나의 complex를 확장하는 과정은 다음과 같다.

- 1) complex와 대응하는 hypercube의 중점을 구한다. 이를  $x_i$ 이라고 하자.
- 2) 집단 B에 대한  $x_i$ 의 cover,  $CV(x_i/B)$ 를 1절의 방법으로 구한다.
- 3)  $CV(x_i/B)$ 에 속하는 complex중 사전 판별지식 complex를 포함하고 면적이 가장 큰 것을 골라 이것으로 사전 판별지식 complex를 대체한다.

그림 2에서 complex K는 사전 판별지식 complex를 나타내고  $L_3$ 가 위 과정에 따라 K를 확장한 complex이다.

3. complex 선택

앞의 1, 2절에서 각각 집단 A의 표본들에 대한 complex들과 사전 판별지식의 확장에 의한 complex들을 구하였다. 전자를  $\{LS_1, LS_2, \dots\}$ , 후자를  $\{LR_1, LR_2, \dots\}$ 라고 하자. 이  $L_i$  complex들 중에서 집단 A의 표본들을 모두 포함하고 complex 갯수가 최소가 되는  $CV(A/B)$ 를 얻고자 한다. 이들 complex들의 모든 조합을 고려해서 최적해를 구할 수는 있지만 이 경우 계산시간이 지수함수적으로 증가하기 때문에 현실적으로 곤란하다. 그러므로 다음과 같은 단계적 방법을 사용한다.

- 1) 다음과 같이 초기화 한다.  
 $S \leftarrow \{LS_1, LS_2, \dots\}$   
 $DC \leftarrow \cup_i LR_i$   
 $A \leftarrow A - \{x \mid x \in CV\}$
- 2) S중 A에 속하는 표본들을 가장 많이 포함하는 complex를 구한다.  
 이를  $LS_k$ 라고 하자. 만약 동물이 발생하면 그들중 면적이 큰 complex를 선택한다.
- 3) 다음과 같이 값들을 변경한다.  
 $S \leftarrow S - LS_k$   
 $CV \leftarrow CV \cup LS_k$   
 $A \leftarrow A - \{x \mid x \in CV\}$
- 4) 만약 A가 공집합이면 끝나고 아니면 1)로 가라.

이 과정을 통하여 그림 2의 complex들중  $L_1, L_3$  두개의 complex들이 선택되는데 이들을 학습표본들과 함께 그림 3에 보인다. 이들 두개의 complex들이 집단 B에 대

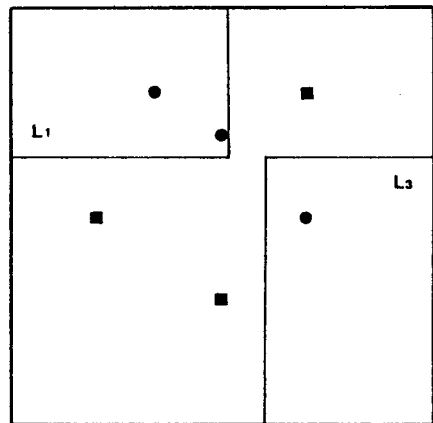


그림 3. 학습표본 및 계산된 cover  $CV(A/B)$   
Fig. 3. Training samples and calculated cover  $CV(A/B)$

한 집단 A의 cover, 즉  $CV(A/B)$ 가 된다. 그림 2와 마찬가지로 동그라미는 집단 A의 표본들을, 네모는 집단 B의 표본들을 나타낸다.

IV. 실험 결과

그림 4에 하나의 실험 예를 보인다. 그림 4(a)는 판별기 설계의 입력으로써 집단 A에 대한 사전 판별지식이 두 개의 complex들( $K_0, K_1$ )로 주어지고 집단 A, B 각각의 학습표본이 50개씩 주어져 있다. 동그라미는 집단 A의 표본들이고 X는 집단 B의 표본들이다. 이로부터 3장의 방법으로 구한  $CV(A/B)$ 는 그림 4(b)에 볼수 있듯이  $L_1$ 부터  $L_4$ 까지의 4개의 complex들로 구성되었다. 표본들은 평균이  $\mu_A=(0.7, 0.3), \mu_B=(0.4, 0.6)$  공분산 행렬이

$$\Sigma_A = \begin{bmatrix} 0.00625 & 0.00375 \\ 0.00375 & 0.00625 \end{bmatrix} \quad \Sigma_B = \begin{bmatrix} 0.00625 & -0.00375 \\ -0.00375 & 0.00625 \end{bmatrix}$$

인 정규분포에서 추출한 것이다.

공분산행렬과 집단 A의 평균은 그대로 두고 집단 B의 평균을( $x_c, 1-x_c$ )로 높고  $x_c$ 를 0.1부터 0.5까지 0.05씩 증가해 가며 실험한 결과를 그림 5, 6, 7에 보인다. 이들 실험에서는 각 집단마다 100개씩의 표본을 추출하여 그중 반은 학습표본으로, 나머지 반은 시험표본으로 사용하였다. 실험은 매 경우마다 10번씩 반복하여 평균을 구하였다. 이들 실험은 기존의 nearest neighbor 방법과 비교하기 위한 것이었으므로 사전 판별지식은 없는 것으로 하였다.

그림 에서 동그라미는 nearest neighbor 방법을 사용한 경우의 판별오차를 나타내고 네모와 세모는 각각 제안된 방법을 사용한 경우의 판별오차와 미판별(unclassified)을 나타낸다. 그림 4에서 알 수 있듯이 제안된 방법의 cover는 특 정공간의 모든 영역을 포함하지 못한다. 단지 학습표본들을 포함하는 일반화된 영역을 구할 따름이다. 그러므로 시험표본중에는  $CV(A/B)$ 나  $CV(B/A)$  어디에도 포함되지 않는 미판별표본들이 발생한다.

그림 6은 미판별 표본들을 다시 neighbor 재판별한 결과를 보인다. 이것은 미리 예상되었던 것이다. 왜냐하면  $CV(A/B)$ 의 영역내의 어떤 점도 사각거리(squaredistance)가 가장 가까운 학습표본은 집단 A에 속하는 것이 되도록 complex 생성을 하였기 때문이다.

그러나 기존의 nearest neighbor 방법을 사용한 경우와 제안된 방법에 재판별을 거치는 경우의 판별시간에는 차이가 있다. 전체 학습표본의 갯수를  $N_s$ , 전체 complex 갯수를  $N_c$ , 미판별 확률을  $\alpha$ 라고 하자. 어떤 시험표본이

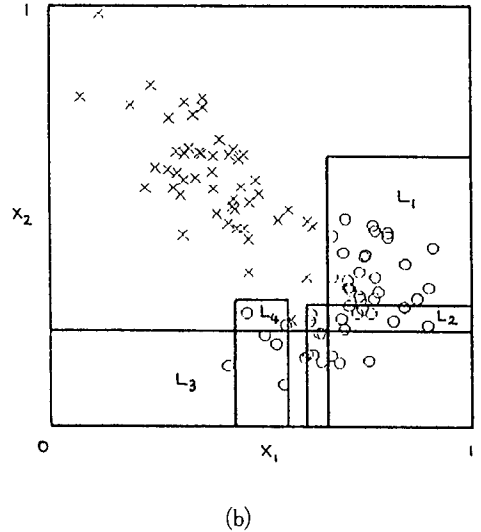
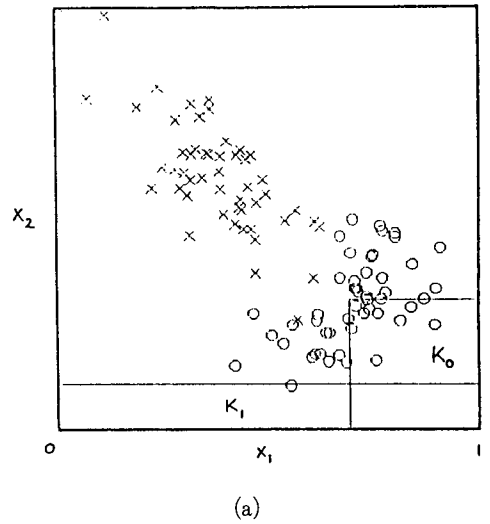


그림 4. (a) 학습표본 및 사전 판별지식 (b) 학습표본 및 계산된 cover  $CV(A/B)$

Fig. 4. (a) Training samples and priori discriminant knowledge, (b) training samples and calculated cover  $CV(A/B)$ .

cover에 포함된다면 이를 판별하는데는 최악 경우  $N_c$ 개 만큼의 complex들에 대해 포함시험(inclusion test)를 거치게 된다. 포함시험은 시험표본이 하나의 complex에 포함되는가를 시험하는 것으로써 이는 2d번의 비교연산으로 가능하다. Nearest neighbor 방법으로 하나의 시험표본을 판별하는데는  $N_c$ 번의 Euclidean 거리계산이 필요한데, 이는 2d번의 부동점 곱셈과 덧셈이 필요하므로 이는 포함시험보다 오래 걸린다. 계산 편의상 포함시험과 거리계산에 걸리는 시간을

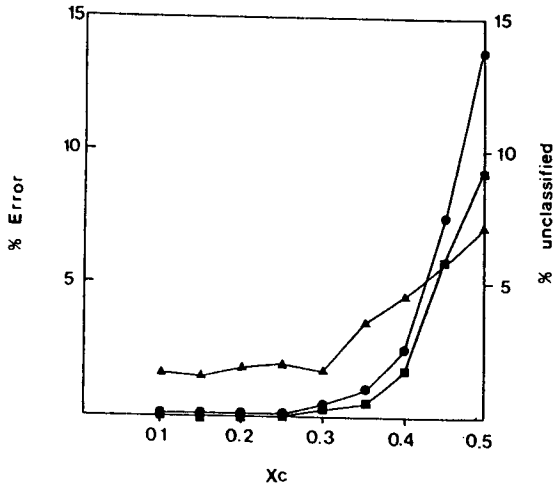


그림 5. Nearest Neighbor 방법의 판별오차 (●), 제안된 방법의 판별오차 (■), 및 미판별율 (▲),  
Fig. 5. Classification errors of Neighbor method (●), those of proposed method (■) and ratio of unclassified samples (▲).

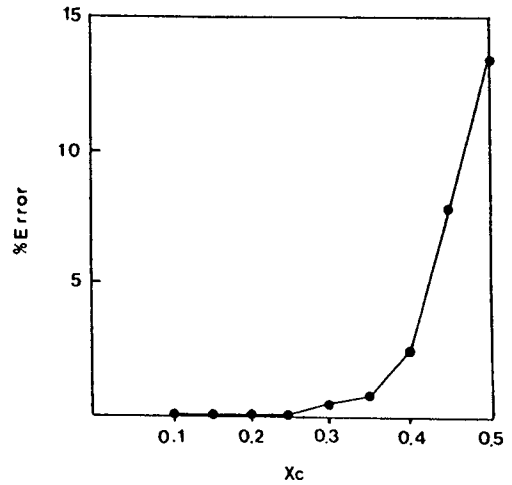


그림 6. 미판별표본을 Nearest Neighbor 방법으로 판별할 경우 오차  
Fig. 6. Classification errors when the unclassified samples are re-classified by the nearest neighbor method.

같다고 하고 이를 단위시간으로 잡자. 그러면 제안된 방법의 경우 시험표본이 cover에 포함되지 않는 경우, 즉 미판별인 경우 판별시간은  $N_c + N_s$  이므로 판별시간의 기대치는  $(1 - \alpha) N_c + \alpha (N_c + N_s) = N_c + \alpha N_s$ 가 된다. nearest neighbor 경우의 판별시간은  $N_c$ 라고 하였으므로 둘 사이의 비는  $N_c / N_s + \alpha$ 가 됨을 알 수 있다. 여기서 앞항을 구별하려는 두 집단의 separability에 대한 것이고 뒷항은 3-1절에서 생성된 complex가 학습표본을 얼마나 일반화(generalize)하였는가를 나타내는 항이다. 두 항은 모두 판별오차가 증가함에 따라 커진다.

그림 7에  $x_c$ 에 따른 complex 갯수, 즉  $N_c$ 의 변화를 보인다.  $x_c$ 가 0.4인 경우  $N_c$ 는 7이고 이 경우 그림 5에서  $\alpha$ 는 4.5%이므로 제안된 방법은 nearest neighbor에 비해 판별시간이 단지  $7/100 + 0.0045 = 11.5\%$  만큼만 소요된다. 이 경우 제안된 방법이나 nearest neighbor 방법 모두 판별오차는 2.5%이었다.

## V. 결 론

본 논문에서는 판별기의 형태를 특징공간에서 hypercube 영역의 집합으로 미리 결정하여 패턴인식을 설계하는 하나의 방법을 제시하였다. 판별기의 형태를 어떻게 정하면 학습표본의 정보뿐 아니라 사람이 가지고 있는 사전 판별지식의 사용을 가능하게 하고, 얻어진 판별

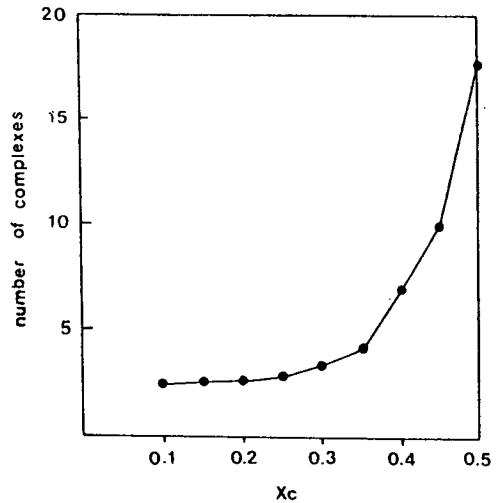


그림 7.  $x_c$ 에 따른 complex 갯수의 변화  
Fig. 7. Change of unnumber of complexes as varying  $x_c$ .

규칙을 사람이 이해하기 쉬운 논리식 형태가 되므로 규칙의 일부를 편집할 수 있으며 규칙기반 시스템과의 접속이 용이하다는 장점이 있다. 또한 이러한 형태의 판별규칙을 사용하여 어떤 표본을 인식하는데는 기존의 산술연산 대신에 비교연산만으로 가능하므로 인식시간이 적게 걸린다는 장점도 아울러 얻어진다.

사전 판별지식이 없는 경우에 기존의 nearest neighbor방법과 실험을 통해 비교한 결과 예상했던 대로 기존의 방법에 비해 인식오차는 비슷하며 인식시간이 적게 걸린다는 것을 알 수 있었다. 인식오차가 비슷한 이유는 판별하려는 표본이 어떤 hypercube 영역에 속한다면 그 표본과 거리가 가장 가까운 학습표본도 역시 이 hypercube 영역에 속하게 되도록 설계하였기 때문이다. 즉 기존의 방법은 표본과 가장 가까운 학습표본을 구하여 그 학습표본이 속하는 집단으로 표본을 인식하는데 반하여 제안된 방법은 표본과 가장 가까운 학습표본이 속하는 hypercube를 구하여 그 hypercube에 대응하는 집단으로 표본을 인식하려는 것이다. Hypercube에 대응하는 집단이름은 그 안에 속하는 학습표본이 속하는 집단이름과 같으므로 결국 제안된 방법으로 인식된 집단이름은 기존의 nearest neighbor방법에서 얻어진 집단이름과 같게 된다. 기존의 방법은 Euclidean 거리를 사용하는데 비해 제안된 방법은 사각거리를 사용했다는 점이 다르나 nearest neighbor방법의 asymptotic property는 사용된 거리척도(distance metric)에 무관하므로 학습표본이 매우 많을 경우 두 방법의 인식오차는 같아질 것이다.

그러나 제안된 방법은 판별하려는 집단들 사이의 separability가 적은 경우나 특징들 사이에 상관도(correlation)가 클 경우 생성되는 hypercube의 수가 증가하게 되어 효율도가 떨어진다는 단점이 있다. 이 문제는 판별기의 형태를 hypercube 영역의 집합으로 미리 정했기 때문에 피할 수 없는 것이다. 그러므로 이 방법은 특징들 사이의 상관도가 비교적 적고 집단들 사이의 separability가 큰 경우에 사용하는 것이 바람직할 것이다.

### 參 考 文 獻

[1] R.O. Duda and P.E. Hart, *Pattern Classifica-*

*tion and Scene Analysis*, Wiley-Interscience Publ., '73.

- [2] B.G. Batchelor, *Practical Approach to Pattern Classification*, Plenum Press, '74.
- [3] D.P. Morgan and C.L. Scofield, *Neural Networks and Speech Processing*, Kluwer Academic Publishers, '91.
- [4] R.S. Michalsky, "A geometrical model for the synthesis of interval covers," Technical Report no. 461, Dept. of CS, Univ. of Illinois, Urbana, June '71.
- [5] R.S. Michalsky, "A variable-valued logic system as applied to picture description and recognition," in *Graphics Languages*, F. Nake and a. Rosenfeld, Eds, North-Holland, pp. 20-47, '72.
- [6] R.S. Michalsky, "AQVAL/1- computer implementation of a variable-valued logic system and the application to pattern recognition," *Proc. 1st Joint Conf. on Pattern Recognition*, Washington DC, pp. 3-17, '73.
- [7] R.S. Michalsky, "Pattern recognition as rule-guided inductive inference," *IEEE Trans. on Pattern Anal. Machine Intell.*, vol. PAMI-2, no. 4, pp. 349-361, '80.
- [8] R.S. Michalsky and R.E. stepp, "Automated construction of classification: conceptual clustering versus numerical taxonomy," *IEEE Trans. on Pattern Anal. Machine Intell.*, vol. PAMI-5, no. 4, pp. 396-410, '83.

### 著 者 紹 介



金 伯 燮(正會員)

1955年 11月 4日生. 1978年 2月 한양대학교 전자공학과 졸업(공학사). 1980年 2月 한국과학기술원 산업전자공학과 졸업(공학석사). 1985年 2月 한국과학기술원 전기및 전자공학과 졸업(공학박사). 1985年~1987年 한양대학교 계량학교실 조교수. 1989年~1990年 Syracuse대학 CASE Center 객원 연구원. 1987年~현재 한림대학교 전자계산학과 부교수. 주관심분야는 패턴인식, 인공신경망, 문자인식 등임.