

# k 근방 원형상에서 최근접 결정법을 이용한 패턴식별법

\*김응규, 이수종

한밭대학교 정보통신공학전공, 한국전자통신연구원  
 e-mail : kimeung@hanbat.ac.kr, sjletri@etri.re.kr

## A Pattern Classification Method using Closest Decision Method in k Nearest Neighbor Prototypes

\*Eung-Kyeu Kim, Soo-Jong Lee

\*Dept. of Infor. & Commu. Eng., Hanbat Nat'l University  
 Speech/Language Information Research Center, ETRI

### Abstract

In this paper, a pattern classification method using closest decision method based on the mean of norm in the closet prototype from an input pattern and its k nearest neighbor prototypes is presented to do accurate classification in arbitrary distributed patterns when the number of patterns is very low. Also this method can be used to classify input pattern precisely when the number patterns is very low because this method considers the weight by the difference of variance in prototypes around the discrimination boundary.

### I. 서론

선형분리가능한 패턴에 대한 패턴식별 방법중 k 근방 결정법(k-NN method)에 기초한 방법이 있다 [1]-[4]. 이 방법은 우선, 분류된 클래스 패턴(class pattern)인 원형상(prototype)과 입력패턴과의 거리에 따라 그 거리가 작은 순서로 k개의 원형상을 선택한다. 다음으로, 이 k개의 원형상중에서 클래스별 다수결을 취해 최대가 된 클래스를 입력패턴이 속하는 클래스로 한다. 각 클래스의 원형상 분포가 집중될 최근접 결정법 k=1에 의해 고정밀도 식별이 가능하게 된다. 또한, 클래스별 최근접 원형상 근방의 원형상 조밀도의 차이가 각 클래스에서 작게 됨으로써 k 근방 결정법 k≥2에 의해 그림 1과 같이 고정밀도 식별이 가능하게 된다. 그러나, 각 클래스의 원형상 분포가 희박한 경우,

클래스 별 최근접 원형상에 의한 식별경계가 정확한 해의 식별경계로부터 크게 이탈하기 때문에 최근접 결정법 k=1에서는 식별 정밀도가 저하될 우려가 있다.

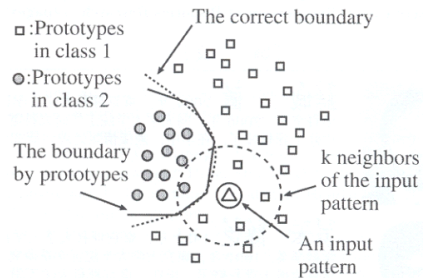


그림 1 각 클래스의 원형상 분포가 집중된 경우, 원형상에 의한 식별경계와 입력패턴근방의 원형상

또한, 클래스별 최근접 원형상 근방의 원형상 조밀도의 차이가 각 클래스에서 크게됨으로써 k 근방 결정법 k≥2에서는 식별 정밀도가 저하될 우려가 있다. 이에 본 연구에서는 클래스별 원형상(prototype) 분포가 선형분리 불가능하고 동시에 분포가 서로 다르면서 빈약한 분포의 원형상에서도 입력패턴의 고정밀도 식별을 행하는 것을 목적으로 한다. 다시말해서, 입력패턴과 클래스별 최근접 원형상의 거리를, 클래스별 최근접 원형상과 그 k 근방의 원형상에서 평균벡터와 그 평균벡터의 산출에 이용한 각 원형상과의 노름 (norm) 평균에 기초한 가중치 부가를 행하여 최근접 결정법을 이용한 식별을 행하고자한다.

이하에 각각 제안한 방법의 알고리즘 설명 및 인공적인 데이터와 실제 데이터에 대해서 기존의 방법인 k-NN법, 매해러노비스 거리[5]-[8], CAP[9], KCAP [10], SVM[11]의 각각에 기초한 방법과 제안하는 방법에 대해 기술한다.

## II. 제안방법 알고리즘 및 유효성 평가

### 2.1 제안방법 알고리즘

각 단계별 제안 알고리즘을 아래에 나타낸다.

[단계 1] 클래스별 최근접 원형상(prototype)의 결정.

[단계 2] 클래스별 최근접 원형상(prototype)과 그 k 근방 원형상 집합의 결정.

[단계 3] 클래스별 최근접 원형상(prototype)과 그 k 근방 원형상에 있어서 평균과 노름(norm) 평균의 산출.

[단계 4] 입력패턴과 클래스별 최근접 원형상과 노름 평균에 의한 가중치를 부가한 거리의 산출

[단계 5] 산출한 거리에 기초한 입력패턴의 식별.

### 2.2 제안방법의 유효성 평가

#### 2.2.1 인공패턴에 대한 실험

우선, 인공패턴에 대한 각 방법의 특성평가를 행한다. 여기에서는 두 클래스 패턴의 식별을 행한다. 두 클래스의 선형분리 불가능한 인공패턴을 생성하기 위해, 우선, 다음과 같은 초반구(super-hemisphere)를 사 고한다.

$$x_n = \sqrt{r_c^2 - x_1^2 - x_2^2 - \dots - x_{n-1}^2} \quad (1)$$

각 클래스에 있어서 패턴의 벡터는 다음의 확률분포 를 만족하는  $d$  를 각각 부가하여 생성한다.

$$p(d) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{d^2}{2\sigma^2}\right\} \quad (2)$$

또한, 식별률  $C_r$  은 다음 식으로 표시한다.

$$C_r[\%] = \frac{1}{C_{Num}} \sum_{i=1}^{C_{Num}} \left[ \frac{1}{P_{Num}} \left\{ \sum_{j=1}^{P_{Num}} D(p_j^{(i)}) \right\} \times 100 \right] \quad (3)$$

먼저, 각 클래스의 원형상(prototype) 수에 대한 k-NN 에 의한 식별률을 나타내고, 다음으로 차원수에 대한 각 방법의 식별률을 나타낸다. 차원수에 대한 각 방법 에 대한 식별률 실험결과, 각 클래스에서 원형상의 수는 20개, 입력패턴은 각 클래스별 100개로 한 경우, 제 안하는 방법, SVM에서는 차원수가 증가하더라도 식별 률은 고정밀도로 유지되고 있음을 알게 되었다.

#### 2.2.2 실제 패턴에 대한 실험

실제패턴에 대한 각 방법의 특성평가를 행하였다. 실험에 사용한 데이터는 문자식별용 데이터베이스인 미니스티(MINIST)를 이용했다[8]. 이것은 문자가 0-9 까지 10개의 클래스, 각각 28×28 픽셀 사이즈 (pixel size), 8[bit/pixel] 계조영상에 대해 라스터 스캔(raster scan)에 의한 벡터화를 취하여 실험을 행하였다. 또한 원형상의 수 785로부터 1,200개까지 각 방법에 있어서 식별률 실험을 행하였다. 그 결과, 타 방법에 비해 제 안하는 방법(Prop.)에서 약간 높은 식별률을 나타냈지 만, 매해라노비스(Mahalanobis)외의 방법은 거의 동등 한 정도의 식별률을 나타냈다.

다음으로, 각 클래스의 원형상 수를 10개로 한 경우 각 클래스의 최근접 원형상의 수와 그 k 근방 k=2의 패턴을 그림 2에 나타낸 것으로, 각 클래스의 원형상 수가 많은 경우 각 클래스의 최근접 원형상과 그 k 근

방 원형상의 유사성이 높다는 것을 알 수 있다. 또한 표 1로부터 각 클 래스의 원형상 수가 10개인 경 우 각 클래스에 있어서 노름 평균이 높음을 알 수 있다.

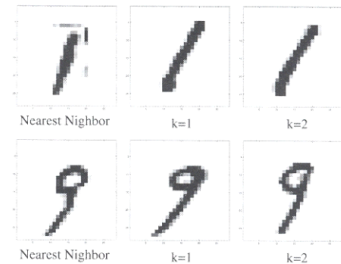


그림 2 원형상의 수 10개에 있어서 각 클래스의 최 근접 원형상과 그 k 근방의 패턴

표 1 원형상 수 10개에 있어서 각 클래스별 노름평균

방 법	노름 평균
클래스 1	702.0
클래스 2	814.0

## III. 결론 및 향후 연구방향

클래스(class)별 원형상(prototype)의 분포가 선형 분리 불가능하고 동시에 분산이 서로 다른 희박한 분포의 원형상에서도 입력패턴에 대한 고정밀도의 식별을 행하기위해 클래스별 최근접 원형상과 그 k 근방 원형상에서 노름(norm) 평균에 기초한 최근접 결정법을 이용한 패턴식별방법을 제안했다. 제안하 는 방법의 유효성을 평가하기위해 인공적인 패턴과 실제 패턴에 대해 일반적인 k-NN법, CAP 등 각각 에 기초한 방법과 제안하는 방법을 적용하여 식별률 에 의한 평가를 행하였다. 그 결과, 인공적인 패턴에 대한 실험에 있어서 제안하는 방법은 타 방법들과 비교하여 원형상의 수를 변화시킨 경우 와 차원수를 변화시킨 경우 양자 모두 식별률이 향상되었고, 원형상의 분포가 희박한 경우에 있어서 특히 유효 했 다. 금후의 과제로서 노름 평균에 대한 가중치 부여 에 대한 검토 등 식별률을 보다 향상시킬 수 있는 방안을 모색 해야할 것이다.

### 참고문헌

- [6] N. Kato, M. Suzuki, S. Omachi, H. Aso and Y. Nemoto, "A Hand-written Character Recognition System Using Directional Element Feature and Asymmetric Mahalanobis Distance", IEEE Trans. on PAMI, Vol.21, No.3, pp.258-262, 1999.
- [8] Y. Mitani and Y. Hamamoto, "Classifier design based on the use of nearest neighbor samples", Proc. 15<sup>th</sup> Int. Conf. Pattern Recognition, Vol.2, pp.773-776, 2000.
- [9] Y. LeCun, L. Bottou, Y. Bengio and P. Haffner, "Gradient-based learning applied to document recognition", Proceedings of IEEE, Vol.86, No.11, pp.2278-2324, 2005.