신경망을 이용한 필기 숫자 인식에서 부류 분별에 기반한 특징 선택

이 진 선†

요 약

본 논문의 목적은 필기 숫자 인식에서 특징의 부류 분별력을 분석하고, 이를 특징 선택에 활용하는 것이다. 부류 분별력을 측정하기 위하여 Parzen 윈도우를 이용하여 부류 분포를 추정하였고, 서로 다른 부류의 부류 분포간의 거리를 부류 분별로 정의하였다. 이렇게 계산된 부류 분별을 이용하여, 특징 벡터에서 쓸모 없거나 중복성을 갖는 특징을 제거하여 특징 벡터의 차원을 줄인다. 실험은 CENPARMI 필기 숫자에 대해 수행하였으며, 10개 부류 전체 뿐 아니라 2개 부류에 대해서도 수행하였다. 실험 결과 10-부류 필기 숫자 인식에서 256-차원 원래 특징 벡터를 인식률 손실 없이 22% 줄일 수 있어, 부류 분별이 특징 선택을 위한 유용한 도구임을 보였다.

Feature Selection Based on Class Separation in Handwritten Numeral Recognition Using Neural Network

Jin-Seon Lee†

ABSTRACT

The primary purposes in this paper are to analyze the class separation of features in handwritten numeral recognition and to make use of the results in feature selection. Using the Parzen window technique, we compute the class distributions and define the class separation to be the overlapping distance of two class distributions. The dimension of a feature vector is reduced by removing the void or redundant feature cells based on the class separation information. The experiments have been performed on the CENPARMI handwritten numeral database, and partial classification and full classification have been tested. The results show that the class separation is very effective for the feature selection in the 10-class handwritten numeral recognition problem since we could reduce the dimension of the original 256-dimensional feature vector by 22%.

1. 서 론

특징은 최종 인식 성능에 영향을 미치는 주요 요인으로서, 필기 인식에서 매우 중요한 역할을 한다. 이러한 이유로 필기 인식을 위한 다양한 특징들이 개발되었다. 최근 논문들에서 경사-기반 특징[1], 방향 거리 특징[2], 웨이브렛-기반 특징[3], Garbor 필터-기반 특징[4], 화소 거리 특징[5], 그리고 오목성 특징[6] 등이 제안되었다. 특징의 사례 조사를 위해서는 [7, 8]을 참조하기 바란다.

특징은 분별력(discriminating power) 면에서 다른 특징과 구별되는 고유한 특성을 갖는다. 예를 들어 어떤 특징은 특정 부류를 구별하는데 강한 반면, 다른

※ 이 논문은 1998년도 우석대학교 학술연구 조성비에 의하여 연구되었음.
† 정 회 원 : 우석대학교 정보통신 및 컴퓨터공학부 교수
논문접수 : 1998년 10월 8일, 심사완료 : 1998년 12월 7일

부류에 대해서는 비교적 약한 경우가 있다. 이러한 특성을 알아내는 방법 중의 하나는 각각의 특징, 또는 특징 조합에 대하여 분별기(classifier)를 구현하여 실제 인식 실험을 통하여 얻은 혼돈 행렬을 조사하는 것이다. 하지만 이 방법은 매우 힘든 실험을 필요로 하며 인식 성능이 분별기의 매개 변수에 의존하기 때문에 정밀도에 한계가 있다. 따라서 분별기에 영향을 받지 않는, 특징들의 고유한 성질을 분석하는 도구가 필요하다.

부류 분별(class separation)은 수십년동안 특징의 분별력을 분석하기 위해 사용되어 왔다[9, 10]. 두 개의 서로 다른 부류간의 부류 분별은 부류 분포(class distribution)가 특징 공간에서 어떻게 잘 분별되는 가를 보여주는 척도이다. 더 큰 부류 분별은 더 좋은 분별력을 의미한다.

부류 분별은 특징의 부류 분포에 의해 정의할 수 있다. 따라서 충분히 많은 샘플을 가진 훈련 데이터베이스를 사용하여야 한다. 부류 분별을 평가하기 위한 기준 함수 중 하나는 서로 다른 부류간 산포 행렬과 같은 부류내의 산포 행렬을 사용하여 정의된다[9]. 이 기준 함수는 평균 벡터와 공분산 행렬의 통계적 값에 기반을 두며, 정규 분포와 같은 구조에 잘 적용된다. 그러나 필기 패턴은 이러한 조건을 만족하지 않으므로, 이러한 기준 함수는 필기 인식 응용에 제한을 받는다. 또 다른 기준 함수는 확률 분포의 비모수적 추정(non-parametric estimation)을 사용한다[11]. 이 함수는 커널-기반 방법(또는 Parzen 창 방법이라고도 함)을 사용하여 부류 확률 분포를 명시적으로 추정한다. 부류 분별은 이렇게 구한 두 개의 부류 분포 사이의 겹침 거리로 계산된다. 비모수 방법은 분포에 대한 가정을 하지 않으므로, 필기 인식에 적합하다. 이 논문에서는 필기 숫자 인식에서 특징의 부류 분별을 분석하기 위해 비모수 방법을 사용한다.

이 논문의 목적은, 신경망 인식기로 필기 숫자를 인식하는 문제에서 특징들의 부류 분별 능력을 분석하고, 이 분석 결과를 특징 선택에 응용하여 특징 벡터의 차원을 줄이는 것이다. 기본 동기는 부류 분별을 특징 선택(feature selection)에 사용할 수 있다는 데에 있다. 특징 선택은 원래 특징 벡터로부터 우수한 특징들의 부분 집합을 선택하는 문제이다[12]. 낮은 부류 분별을 가진 특징은 분별력이 약한 것이므로 제거하고, 높은 부류 분별을 가진 것을 선택함으로써 좋은 특징들의 부분 집합을 만들 수 있다.

실험을 위해 4,000개의 훈련 샘플(400 샘플/부류)과 2,000개의 검사 샘플(200 샘플/부류)로 구성된 CENPARMI 필기 숫자 데이터베이스를 사용한다. 이 데이터베이스는 실제 우편물로부터 구축되었기 때문에, 다수의 필기자와 필기 도구로 작성된 완전 무제약 샘플들로 구성된다. 따라서 부류 분포를 평가하는 본 논문의 실험에 아주 적합하다고 할 수 있다. 특징 벡터로는 높은 성능을 갖는 경사 특징과 방향 거리 특징을 사용한다.

10개의 숫자 부류 중 단지 2 부류만을 고려하는 2-부류 문제에서 흥미있는 결과 중의 하나는, 3과 8 부류 쌍에 대해 부류 분별값이 가장 높은 특징 셀 하나만 가지고 신경망 인식을 한 결과 94%의 인식을 얻었다는 점이다. 반대로 어떤 특징 셀은 분별력이 거의 없으며 이들을 무효 셀이라 부른다. 256 차원의 특징 벡터에서 가장 좋은 부류 분별력을 갖는 10%만의 특징 부분 집합으로 인식률 손실없이 3-8 쌍을 분별할 수 있었다. 모양이 아주 달라 혼돈이 없는 쌍 6-9에 대해, 단지 5개의 특징 셀 (특징 벡터의 2%)만으로 100% 인식률을 얻었다. 이러한 결과들은 특징 벡터의 일부분이 분별을 위해 필요한 거의 모든 정보를 포함하고 있으며 나머지는 무효이거나 중복된다는 것을 의미한다. 10-부류 분별에서는 인식률 저하없이 약 22%의 특징 셀을 제거할 수 있었다.

2장에서는 용어 정의 및 부류 분별을 측정하기 위한 도구에 대해 기술한다. 3장은 신경망에 의한 숫자 인식에 대해 부분 분류 및 전체 분류에 대해 부류 분별에 기반한 특징 선택에 관한 실험 결과를 보여 준다. 4장은 결론을 기술한다.

2. 부류 분별 추정

이 장에서는 몇 가지 용어 정의와 부류 분별 측정을 위한 수학적 도구에 대해 기술한다.

2.1 용어 정의

ω_i $(0 \le i \le 9)$는 숫자 10개 부류 중 하나를 나타낸다. 부류 ω_i는 훈련 데이터베이스에서 N_i개의 샘플을 갖는다. 특징 벡터 $\underline{X}$는 x_i로 표시되는 특징 셀들의 집합으로 구성된 d-차원 벡터, 즉 $\underline{X}=(x_0,$

$x_1, \ldots, x_{d-1}$)이다. 예를 들어 3장에서 설명할 DDD 특징 벡터는 256-차원 특징벡터, $\underline{X}=(x_0, x_1, \ldots, x_{255})$로 나타낼 수 있다. 부류 ω_i의 샘플 집합은 $Z_i=\{\underline{z_i}^1, \underline{z_i}^2, \ldots, \underline{z_i}^{n_i}\}$로 표시한다. 이때 $\underline{z_i}^k$는 Z_i에서 k번째 위치한 샘플을 의미한다.

필기 숫자 데이터베이스는 g(=10)개의 부류, 부류당 N개의 샘플과 d-차원 특징 벡터를 가진다고 가정하면, g, N, 그리고 d가 각각 한 축을 나타내는 3-차원 공간상의 한 볼륨으로 생각할 수 있다. 이는 거대한 양의 자료이다. 필기 숫자 인식에서 g는 10이고, 특징 벡터는 수 백 차원을 가지며, 훈련 집합에서 샘플 패턴들은 수천 또는 수만 개이다. 4,000 훈련 샘플 (400 샘플/부류)을 갖는 CENPARMI 필기 숫자 데이터베이스의 경우, 총 볼륨은 10*400*256=100만 개의 자료 셀을 포함한다. 이러한 자료 집합을 가시화 하는 것은 매우 어려운 작업이며, 따라서 부류 분별과 같은 수학적 분석 도구들이 필요하다.

두개의 부류가 특징 벡터 $\underline{X}$에 의해 얼마나 잘 분리되는 가를 나타내는 척도인 부류 분별과 관련된 정의를 한다. 척도는 $S^{\infty}(\omega_i, \omega_j, \underline{X})$로 표시되며, 이때 ω_i와 ω_j는 두 개의 부류이고, $\underline{X}$는 하나 이상의 특징 셀로 구성된 특징 벡터이다. 10개 부류를 가진 숫자 인식 경우, $\underline{X}$는 두 개 부류 대신 10개의 부류를 분별해야 한다. 이를 위해 척도 $S^g(\Omega, \underline{X})$를 사용한다. 여기서 $\Omega=\{\omega_0, \omega_1, \ldots, \omega_9\}$이다.

2.2 비모수 방법

특징 벡터 $\underline{X}$에 의한 한 부류의 확률 밀도 분포는 다음과 같이 추정한다. $\underline{X}$는 특징 셀의 집합으로 구성된 특징벡터, 즉 $\underline{X}=(x_0, x_1, \ldots, x_{d-1})$이다. 단위 계단 함수는 다음과 같다.

$$\Phi(\underline{X})=\begin{cases}1, \text{if } |x_i|<\frac{1}{2}, i=0,1,\ldots,d-1\\ 0, \text{otherwise}\end{cases}$$

$\Phi(\underline{X})$는 전체 d-차원 공간 R_d 상에서 전체 합이 1이 되는 함수로서, 커널 함수라 부른다. $\Phi(\underline{X})$의 이동과 크기 변환된 함수를 다음과 같이 정의한다.

$$\Phi_n(\underline{X})=\Phi((\underline{X}-\underline{X_i})/h_n)$$

위에서, h_n은 분포 추정에 사용하는 평활화(smoothing) 인자이다. 큰 h_n값은 높은 정도의 평활화를 의미한다.

부류 ω_i의 확률 분포는 다음 식으로 추정한다.

$$P_n^{\omega_i}(\underline{X})=\frac{1}{N_i}\sum_{k=1}^{N_i}\frac{\Phi((\underline{X}-\underline{z_i}^k)/h_n)}{V_n}$$

이때 $V_n=(h_n)^d$이다.

추정된 부류 분포를 사용해서, 두 부류 ω_i와 ω_j의 부류 분별은 다음과 같이 정의한다.

$$S^{\infty}(\omega_i, \omega_j, \underline{X})=\int_{R_d}|P_n^{\omega_i}(\underline{X})-P_n^{\omega_j}(\underline{X})|\,d\underline{X}$$

이때 R_d는 d-차원 실수 공간이며, $P_n^{\omega_i}(\underline{X})$와 $P_n^{\omega_j}(\underline{X})$는 각각 부류 ω_i와 ω_j에 대한 부류 분포이다. 이 수식은 두 분포간의 겹쳐진 거리 정도를 측정한다. 겹치지 않은 극단적인 경우에, S^{∞} 는 최대값 2.0을 가진다. 완전히 겹친 경우는 0 값을 가지며, 특징 벡터 $\underline{X}$는 ω_i와 ω_j의 분별에 쓸모가 없으며, 특징 벡터 $\underline{X}$는 무효이다.

부류 집합에 대한 부류 분별은 S^{∞} 를 사용해서 다음과 같이 정의한다.

$$S^g(\Omega, \underline{X})=\sum_{\omega_i\in\Omega}\sum_{\omega_j\in\Omega, j\neq i}S^{\infty}(\omega_i, \omega_j, \underline{X})$$

이때 $\Omega=\{\omega_0, \omega_1, \ldots, \omega_9\}$ 이다.

$\underline{X}$의 차원이 클 때, 실제적인 계산이 불가능한 고 차원 문제(curse of dimensionality)가 발생한다. 추정된 분포에 대한 샘플링 비율이 r, 차원이 d, 추정에 사용한 샘플의 수가 N일 때, 하나의 분포를 추정하기 위한 계산 복잡도는 $O(N*r^d)$ 이다. 만일 m개의 분포를 평가할 경우, 복잡도는 $O(m*N*r^d)$ 이다. 따라서 $\underline{X}$ 의 차원을 작게 제한을 가해야 한다. 이 논문에서는 1-차원(즉, $\underline{X}=(x_i)$)과 2-차원(즉, $\underline{X}=(x_i,x_j)$)에 대해 실험하였다.

3. 부류 분별에 기반한 특징 선택

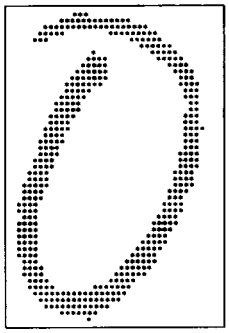
3.1 특징 벡터 추출

이 절에서는 실험에서 사용할 두 가지 특징에 대해 설명한다. 이들은 필기 숫자에 대해 좋은 인식 성능을 갖는 것으로 증명된 특징들이다[1, 2].

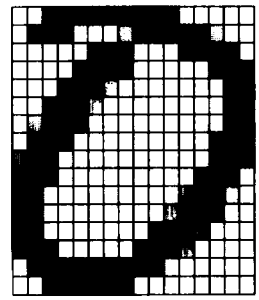
입력 패턴 P_{m*n} 은 먼저 16*16 메쉬, R_{16*16} 로 크기 정규화된다. R은 P 메쉬상에 16*16 격자 메쉬를 씌움

으로써 계산된다. P의 몇 개 격자들이 서로 다른 비율로 R의 한 개의 격자와 겹쳐지게 된다. R의 한 격자의 값은 P의 겹쳐진 모든 격자들을 사용해서 계산한다. R 격자의 값은 P의 겹치는 화소들 값에 가중치(겹치는 정도)를 곱한 값의 합으로 계산한다. R의 모든 격자 값이 계산된 후, 이 값들은 0.0과 1.0사이의 값으로 정규화된다. 모든 겹치는 격자들을 겹침 정도에 따라 고려하였으므로, 크기 정규화 과정에서 정보 손실이 최소화되었다고 말할 수 있다.

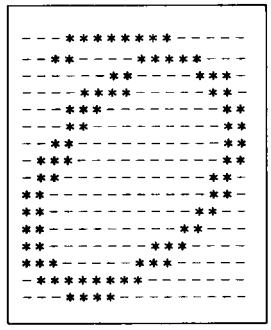
(그림 1(a), 1(b), 1(c))는 이진 입력 패턴, 크기 정규화된 R 메쉬, 그리고 R의 이진화 결과를 보여 준다.



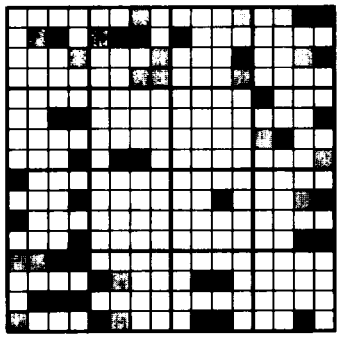
(a) 입력 패턴, P_{36*51}



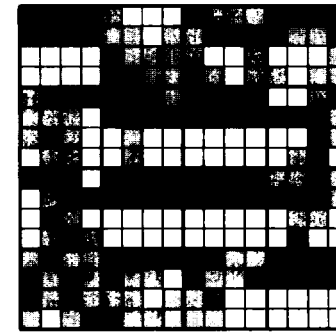
(b) 크기 정규화된 메쉬, R_{16*16}



(c) 이진 메쉬



(d) CGD 특징 벡터



(e) DDD 특징 벡터

(그림 1) 숫자 패턴과 추출된 특징 벡터의 예
(Fig. 1) An example numeral pattern and the extracted feature vectors

(1) CGD (Contour-based Gradient Distribution, 외곽-기반 경사 분포) 특징

경사 특징을 얻기 위해 (그림 1(b))의 정규화 메쉬 R에 Sobel 에지 연산자를 적용하여, 경사 강도 맵, M_{16*16}과 경사 방향 맵, D_{16*16}를 얻는다[13]. 0° 와 360° 사이값을 갖는 경사 방향은 16단계(22.5°/단계)로 구간화된다. 경사 크기 맵은 전체 맵상의 평균치를 임계치로 사용하여 이진 맵으로 변환된다. 따라서 이진 맵은 입력 패턴의 외곽에 있는 화소들만 1 값을 갖는다.

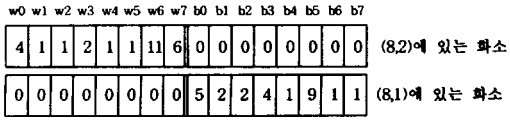
경사 방향 맵 D_{16*16}을 4*4 블록 메쉬로 나눈다. 즉, D의 4*4 부 영역이 하나의 블록이 된다. 각 블록에 대해 16개 구간을 갖는 하나의 히스토그램을 계산한다. 히스토그램 누적시 이진 경사크기 맵에서 1 값을 갖는 화소만 참여한다. 그러므로 입력 패턴의 외곽에 있는 화소들만 히스토그램 구성에 영향을 미친다. 16개 블록 모두에 대해 히스토그램 계산을 마친 후, 히스토그램 값을 0.0과 1.0사이의 값으로 정규화한다. 보다 자세한 알고리즘을 위해서는 [1]을 참조하라.

(그림 1(d))는 최종 CGD 특징 벡터를 보여준다. 4*4 블록 메쉬 및 각 블록 메쉬에 대해 16개 구간의 히스토그램을 볼 수 있다. 어두울수록 높은 값을 의미한다. 이 특징 벡터는 256개의 특징 셀(16개 블록과 16 구간/블록)을 가지며, 따라서 CGD는 256 차원을 갖는다. 0에서 255 사이의 일련 번호로 CGD의 셀들을 구별한다. CGD 셀 i는 $\lfloor i/16 \rfloor$ 번째 블록과 (i mod 16) 번째 구간에 있는 셀을 의미한다.

(2) DDD (Directional Distance Distribution, 방향 거리 분포) 특징

DDD 특징은 (그림 1(c))에 있는 이진 맵으로부터 계산한다. 이 특징은 거리 정보에 기반을 둔다. 입력 패턴의 각 화소에 W (흰) 집합과 B (검은) 집합이라 부르는 두 개의 8바이트 집합을 (그림 2)처럼 할당한다. 흰 화소는 8 방향에 대해 가장 가까운 검은 화소까지의 거리값을 W 집합에 기록한다. B 집합은 거리 계산없이 단순히 0으로 채운다. 같은 방식으로 검은 화소는 8방향에 대해 가장 가까운 흰 화소까지의 거리값을 B 집합에 기록한다. W 집합은 0으로 채운다. (8-방향 코드는 0(동), 1(북동), 2(북), 3(북서), 4(서), 5(남서), 6(남), 7(남동)이다.)

(그림 1(c))에 있는 예제 패턴에서, (8,2)의 화소는 (그림 2)의 위에 있는 WB 인코딩을 갖는다. 이 화소는 흰색이므로 B 집합은 0값을 갖는다. W 집합의 8 방향 거리를 계산하기 위해 이 화소는 각 방향으로 광선을 쏘며, 이 광선은 검은 화소와 충돌할 때까지 진행한다. 충돌하면, 진행 거리를 광선 방향에 해당하는 바이트에 기록한다. 한 예로서, 방향 0의 광선은 (8,2)W->(9,2)W->(10,2)W->(11,2)W->(12,2)B의 경로를 밟는다. 결국 진행 거리 4를 W 집합의 첫 번째 바이트에 기록한다.



(그림 2) WB 인코딩 예
(Fig. 2) Example WB encoding

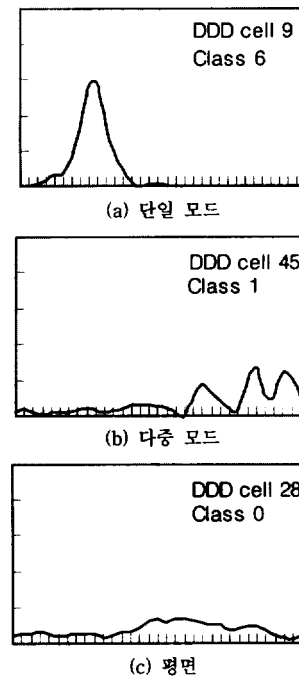
(그림 2)의 아래에 있는 WB는 (8,1)에 있는 검은 화소에 대한 WB 인코딩 예를 보여준다. 이 예에서는 흰 화소와 충돌없이 맵의 경계에 도달하는 경우가 발생한다. 이 경우에는 배열을 원형으로 간주한다. 예를 들어, 방향 1을 갖는 광선은 (8,1)B->(9,0)B->(10,15)W의 경로를 밟으며, 진행 거리 2를 B 집합의 두 번째 바이트에 기록한다.

256개 모든 화소에 대해 WB 코딩을 계산한 후, R을 4*4 블록 메쉬로 나눈다. 한 블록은 4*4 화소로 구성된 부 영역이다. 각 블록은 16개 WB 코딩의 평균을 구한다. 마지막으로 값을 0.0과 1.0 사이로 정규화한다. 보다 자세한 알고리즘을 위해서는 [2]를 참조하기 바란다.

최종적으로 256-차원(16 블록과 16 값/블록)의 DDD 특징 벡터를 얻는다. (그림 1(e))는 4*4 블록 메쉬와 16개의 값/블록을 가진 DDD 특징 벡터를 보여준다. DDD의 셀 번호는 CGD와 유사하게 붙여진다. DDD 셀 i는 $\lfloor i/16 \rfloor$ 번째 블록과 (i mod 16) 번째 바이트에 있는 셀을 의미한다.

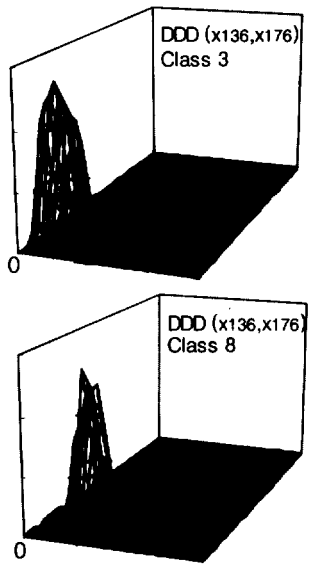
3.2 부류 분포

이 절에서는 CGD와 DDD 특징에 대하여, 필기 숫자 데이터베이스로 추정한 부류 분포를 보인다. (그림 3)은 추정된 1-차원 분포들을 보여준다. (그림 3(a))는 DDD 특징 셀 $\underline{X}=(x_9)$에 의해 추정된 분포로서, CENPARMI 필기 숫자 데이터베이스의 훈련 집합에서 부류 6에 대한 400개의 샘플을 사용하였다. 평활화 인자 h_n 값으로 1/32 를 주었다. 이것은 전형적인 단일 모드 분포를 보여준다. 다중 모드 구조와 평면 구조를 갖는 분포들이 (그림 3(b))와 (그림 3(c))에 있다. (그림 4)는 DDD 특징 벡터에서 2-차원 특징 벡터 $\underline{X}=(x_{136}, x_{176})$를 사용하여 부류 3과 8에 대해 추정한 2-차원 분포를 보여준다.



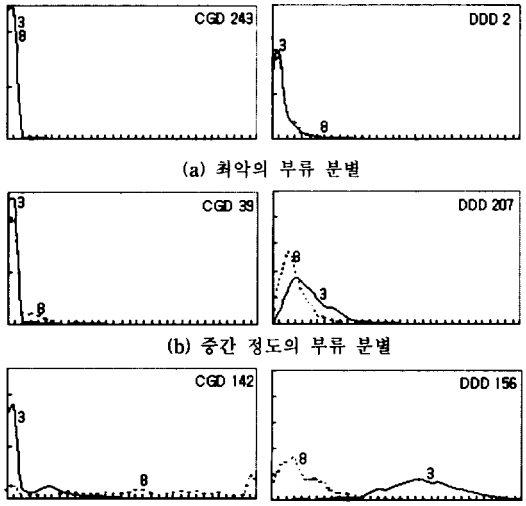
(a) 단일 모드
(b) 다중 모드
(c) 평면

(그림 3) Parzen 창에 의해 추정된 1차원 부류 분포
(Fig. 3) 1-D class distributions estimated by Parzen window



(그림 4) Parzen 창에 의해 추정된 2-D 부류 분포
(Fig. 4) 2-D class distributions estimated by Parzen window

(그림 5)는 CGD와 DDD에 대해, 3과 8 부류의 부류 분별을 보여 준다. (그림 5(a))의 CGD 셀 243과 DDD 셀 2는 3과 8 부류의 분포가 거의 겹치며, 이는 이 특징 셀들은 분별력을 갖지 않음을 의미한다. 반대로, (그림 5(c))의 최상의 특징 셀들은 잘 분리된 분포를 보이며, 이는 좋은 분별력을 의미한다.



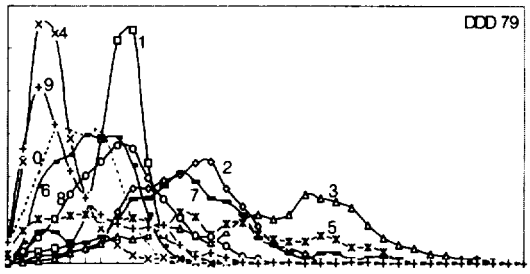
(a) 최악의 부류 분별

(b) 중간 정도의 부류 분별

(c) 최상의 부류 분별

(그림 5) 다양한 부류 분별
(Fig. 5) Distributions for various kinds of class separation

10개 부류를 위해서는, 각 특징 셀에 대해 10개의 부류 분포를 사용한다. (그림 6)에서 DDD의 한 특징 셀인 x_{79} 에 대해 추정된 10개 부류들의 부류 분포를 보여준다. 10-분별의 경우, 10개 부류 모두를 한 그룹으로 간주하는 기준 함수를 사용해야 한다. 이를 위해 2.2절에서 이미 정의한 기준 함수 $S^g(\Omega, \underline{X})$를 사용한다. <표 1>은 (그림 5) 분포들의 부류 분별 S^{cc}와 S^g 를 보여준다. 3과 4 부류 쌍은 가장 큰 부류 분별 값 1.75를 갖는다. (그림 5)의 그래프에서 부류 3과 4의 부류 분포의 겹침 정도가 작음을 볼 수 있다. 반대로, 부류 2와 7은 가장 작은 부류 분별, 0.31을 갖는데, 이들의 분포는 거의 겹침을 볼 수 있다



(그림 6) 10개의 숫자 부류에 대한 부류 분포
(Fig. 6) Class distributions for 10 numeral classes

3.3 특징 벡터의 차원 축소

지금까지 개개의 특징 셀은 분별력 면에서 다른 셀과 구별되는 특성을 가짐을 확인했다. 이러한 부류 분별 정보를 사용하여 인식률의 손실없이 특징 벡터의 차원을 축소할 수 있다.

이를 위해 개개 특징 셀에 대해 부류 분별을 측정한다. 여기서 개개 특징 셀은 1-차원 특징 벡터, $\underline{X}=(x_i)$ (i는 0에서 d-1 사이)를 의미한다. 다음 알고리즘을 이용하여 특징 셀들을 부류 분별의 값에 따라 정렬한다. 다음은 두 부류 ω_p와 ω_q에 대한 특징 정렬 알고리즘이다.

<u>알고리즘-1D</u>:

1. R = empty list, and P = $\{x_i \mid 0 \le i \le d-1\}$.
2. If P is empty, stop.
3. Choose $x_k \subseteq P$ such that for all $x_{k'} \subseteq P$ and $k' \neq k$,

$$S^{cc}(\omega_p, \omega_q, \underline{X}=(x_k)) \geq$$

〈표 1〉 DDD의 특징 셀 x_{79}에 대한 부류 분별
〈Table 1〉 Class separations for a feature cell, x_{79} of DDD

	S^{cc}										합
	0	1	2	3	4	5	6	7	8	9	
0	0.00	1.07	1.54	1.66	0.70	1.07	0.33	1.38	0.81	0.57	9.13
1	1.07	0.00	1.24	1.53	1.55	1.20	0.77	1.20	0.64	1.23	10.44
2	1.54	1.24	0.00	1.04	1.73	0.79	1.38	0.31	0.92	1.39	10.33
3	1.66	1.53	1.04	0.00	1.75	0.75	1.63	0.96	1.33	1.48	12.13
4	0.70	1.55	1.73	1.75	0.00	1.26	0.93	1.55	1.29	0.44	11.21
5	1.07	1.20	0.79	0.75	1.26	0.00	1.05	0.54	0.77	0.90	8.34
6	0.33	0.77	1.38	1.63	0.93	1.05	0.00	1.25	0.53	0.73	8.59
7	1.38	1.20	0.31	0.96	1.55	0.54	1.25	0.00	0.81	1.20	9.20
8	0.81	0.64	0.92	1.33	1.29	0.77	0.53	0.81	0.00	0.92	8.02
9	0.57	1.23	1.39	1.48	0.44	0.90	0.73	1.20	0.92	0.00	8.88
										S^g	96.27

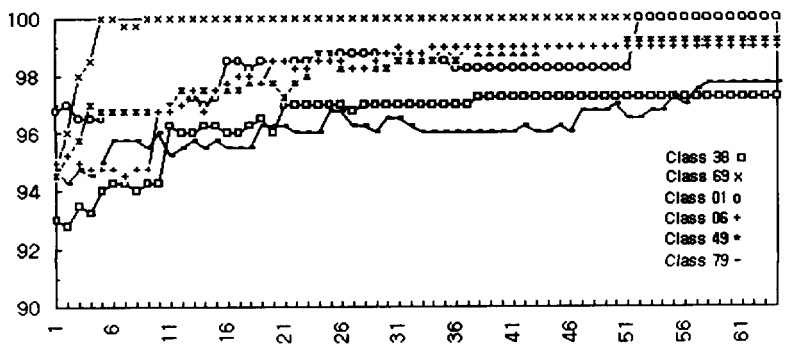
$$S^{cc}(\omega_p, \omega_q, \underline{X} = (x_{k'}))$$

4. R <- x_k, and P = P-x_k.
5. Goto step 2.

인식 실험을 위해서는 전향 다층 퍼셉트론(Feed-forward multiple layer perceptron) 신경망을 사용하였다. 신경망은 10개의 부 망으로 구성되고, 각 부 망은 10개 부류 중 하나를 책임지는 모듈러 구조를 갖는다. 모듈러 신경망을 사용한 이유는 비 모듈러 구조에 비해 인식률이 높기 때문인데, 모듈러 신경망에 대한 구체적인 사항은 [14]를 참조하기 바란다. 실험에서는 부분 분별기와 전체 10-분별기 둘 다 사용한다. 10보다 적은 수 k에 대한, k-분별기는 10개 부류 중 단지 k개만을 고려하는 분별기이다. 실험에서는 2-분별기를 사용한다. (실제 응용에서, 2-분별기는 두 부류의 혼돈 해결기로 사용할 수 있다.) k-분별기의 신경망 구조는 단지 k개의 출력 노드만을 갖는다는 점을 제외하고는 10-분별기와 동일하며, 훈련시 훈련 집합의 해당 k개 부류에 속하는 샘플들만을 사용한다.

(1) 2-분류기

위의 알고리즘으로 정렬한 특징의 순서 리스트에서 상위 k개의 특징 셀을 사용해서 신경망 인식기로 인식 실험을 수행하였다. 이 실험으로 최고 인식 성능을 얻기 위해 필요한 부분 특징 벡터를 알아낼 수 있다. (그림 7)에서 6가지 종류의 2-분류기에 대한 DDD 특징

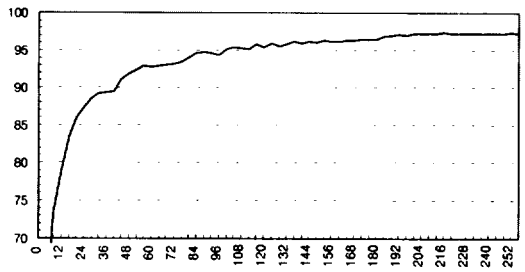


(그림 7) 2-분류기에 대한 상위 k개 (x축) 셀들의 인식률 (y축)
(Fig. 7) Recognition rate (y axis) of the top k (x axis) cells for several 2-classifiers

벡터를 사용한 실험 결과를 볼 수 있다. 3-8 부류 쌍의 경우, 최고 성능 97.25%가 k=38 근처에서 얻어졌다. 이는 256차원 DDD 특징 벡터의 약 15%(38/256) 정도로 최고 성능을 얻을 수 있음을 의미한다. 다른 셀들은 무효이거나 중복되며, 인식 성능 손실없이 이들을 제거할 수 있다. 모양이 달라 혼돈 가능성이 적은 6-9 쌍에 대해서는, 단지 5개의 특징 셀만으로 100% 인식률을 달성하였다. 또 다른 쌍 0-1에 대해서는 k=52에서 100% 인식률을 얻었다. 혼돈 가능성이 높은 쌍 0-6에 대해서는 최고 인식률 98.75%가 k=32에서 얻어졌다. 혼돈 가능성이 높은 쌍 4-9에 대해서는 최고 인식률 99%가 k=50에서 얻어졌다. 그리고 혼돈 가능성 높은 쌍 7-9에 대해서는 k=60에서 최고 인식률 97.75%를 얻었다.

(2) 10-분류기

10-분류기에 대한 실험도 수행하였다. (그림 8)이 실험 결과를 보여준다. 상위의 한 개의 특징 셀은 25.65%의 인식률을 보였다. 상위 4개의 특징 셀은 59.70%의 인식률을 보였다. 그리고 특징 셀을 하나씩 추가함에 따라, 인식률은 처음에 급속히 증가하다가 점차적으로 둔화되었다. 그리고 k=200 주위에서 97.3%의 최고 성능에 도달하였다. 이는 남아있는 56개 특징 셀을 제거할 수 있음을 의미한다. 결과적으로 인식률 손실없이 DDD 특징 벡터 차원을 22% 줄일 수 있음을 의미한다.



(그림 8) 10-분류기에 대한 상위 k개 (x축) 셀들의 인식률 (y축)

(Fig. 8) Recognition rate (y axis) of the top k (x axis) cells for 10-classifier

특징 차원의 축소는 두가지 면에서 유리함을 제공한다. 첫째는 메모리 공간의 효율성이다. 우리가 사용한 10-분류 신경망의 경우 수 만개의 연결 강도를 갖는데 이를 22% 줄일 수 있다. 특히 2-분류 경우 위의 실험 결과를 살펴보면 훨씬 효과적임을 알 수 있다. 둘째는 계산 효율이다. 신경망 인식기의 계산 시간은 연결 강도 개수에 대략 비례하므로 10-분류 경우 약 22%의 계산 시간 단축 효과를 가져온다.

4. 결 론

필기 숫자를 신경망으로 인식하는 문제에서, 부류 분별을 측정하여 특징 선택을 수행하고 이를 이용하여 특징 벡터의 차원을 줄일 수 있었다. 실험 결과는 비모수 방법이 필기 숫자 데이터베이스에 대해 잘 적용됨을 보였다. 또한 숫자의 부분 분류와 전체 분류에 대해 특징 차원 감소를 효과적으로 달성할 수 있음을 보였다.

본 논문에서는 특징 셀들간의 종속성(correlation)을 고려하지 않았다. 하지만 하나의 특징 벡터에 속하는 특징 셀들 간의 종속성은 매우 높기 때문에 보다 효과적인 특징 선택을 위해서는 종속성을 고려하는 알고리즘이 개발되어야 한다. 최근 각광받고 있는 유전자 알고리즘 등을 도입하여, 개개 특징의 분별력, 특징들간의 상호 의존성, 특징들 간의 보완성(complementarity)을 모두 고려한 방법 개발 등이 향후 연구 과제이다.

참 고 문 헌

[1] G. Srikantan, S. W. Lam, and S. N. Srihari, "Gradient-based contour encoding for character recognition," Pattern Recognition, Vol.29, No.7, pp.1147-1160, 1996.

[2] 오일석, Ching Y. Suen, "광학 문자 인식을 위한 거리 특징", 정보과학회 논문지(B), 제25권, 제7호, pp.1028-1043, 1998.

[3] S. W. Lee, C. H. Kim, H. Ma, and Y. Y. Tang, "Multiresolution recognition of unconstrained handwritten numerals with wavelet transform and multilayer cluster neural network," Pattern Recognition, Vol.29, pp.1953-1961, 1996.

[4] Y. Hamamoto, S. Uchimura, M. Watanabe, and T. Yasuda, "Recognition of handwritten numerals using Garbor features," Proceedings of ICPR'96, pp.250-253, Vienna, Austria, 1996.

[5] N. W. Strathy and C. Y. Suen, "A new sys-

tem for reading handwritten ZIP codes," Proceedings of ICDAR, pp.74-77, Montreal, Canada, 1995.

[6] J. T. Favata, G. Srikantan, and S. N. Srihari, "Handprinted character/digit recognition using a multiple feature/resolution philosophy," Proceedings of IWFHR'94, pp.57-66, Taiwan, 1994.

[7] O. D. Trier, A. K. Jain, and T. Taxt, "Feature extraction methods for character recognition- a survey," Pattern Recognition, Vol.29, No.4, pp. 641-662, 1996.

[8] I. S. Oh and C. Y. Suen, "Distance features for neural network-based recognition of handwritten characters," International Journal on Document Analysis and Recognition, Vol.1, No.2, pp.73-88, 1998.

[9] R. O. Duda and P. E. Hart, 'Pattern Classification and Scene Analysis,' A Wiley-interscience Publication, 1973.

[10] J. Schurmann, 'Pattern Classification : A Unified View of Statistical and Neural Approaches,' John Wiley and Sons, Inc., 1996.

[11] E. A. Patrick and F. P. Fischer II, "Nonparametric feature selection," IEEE Transactions on Information Theory, Vol.15, No.5, pp.577-584, September 1969.

[12] J. Kittler, "Feature Selection and Extraction," in Handbook of Pattern Recognition and Image Processing (Edited by T. Y. Young and K. S. Fu), Academic Press, 1986.

[13] R. M. Haralick and L. G. Shapiro, 'Computer and Robot Vision,' Addison-wesley publishing company, 1992.

[14] I. S. Oh, J. S. Lee, K. C. Hong, and S. M. Choi, "Class-expert approach to *handwritten numeral* recognition," Proceedings of the Fifth IWFHR, Essex, England, pp.35-40, 1996.

이 진 선

e-mail : jslee@core.woosuk.ac.kr

1985년 2월 전북대학교 전산통계학과 졸업(학사)

1988년 2월 전북대학교 대학원 전산통계학과(이학석사)

1988년~1992년 2월 한국전자통신연구소 연구원

1995년 8월 전북대학교 대학원 컴퓨터공학과(공학박사)

1995년~현재 우석대학교 정보통신 및 컴퓨터공학부 조교수

관심분야 : 패턴인식, GIS, 멀티미디어