

## 신경망을 이용한 필기체 숫자의 탐지 및 인식에 관한 연구

임길택, 김호연, 남윤석  
한국전자통신연구원 우정기술연구부

### A Study on the Spotting and Recognition of Handwritten Numerals Using Neural Networks

Kil Taek Lim, Ho Yon Kim, Yun Seok Nam  
Postal Technology Development Department, ETRI  
E-mail : ktlim@etri.re.kr

#### Abstract

In this paper, we describe a study on the spotting and recognition of handwritten numerals using neural networks. To recognize a handwritten numeral, two kinds of neural network classifiers are developed. One makes use of the positive samples only, while the other does both of the positive and negative samples. We propose two numeral spotters which discriminate between numerals and non-numerals. Those are also implemented by using neural networks.

From the various experimental results, we found that our methods can be successfully applied to spot and recognize handwritten numerals.

#### I. 서론

지난 수십 년간 문자인식에 관한 수많은 연구가 이루어져왔다[1,2]. 문자인식은 인식대상인 문자에 대한 정보 획득 시간에 따라 온라인과 오프라인 인식으로 나뉘지며, 작성방식에 따라 인쇄체와 필기체 인식으로 분류된다. 특히, 필기체 오프라인 숫자인식은 우편분야 및 금융분야 등 응용분야에서 필요성이 강하게 제기되었고, 또한 클래스의 수가 10개로 적당하고, 실험 세트의 수집도 용이하여 패턴인식 이론의 적용대상 중 가장 많은 주목을 받아왔다고 할 수 있다[3,4].

그러나, 필기체 오프라인 숫자 인식에 대한 연구가

보다 유용하기 위해서는 여러 가지 부가적인 문제에 대한 연구도 병행되어야한다. 가장 중요한 문제 중 하나는 숫자 인식이 숫자가 아닌 기타 입력에 대해서 얼마나 높은 신뢰도로 동작하는가 일 것이다. 예를 들면 우편봉투의 자동인식에 있어서 숫자, 한글, 영어 및 한자 등으로 이뤄진 주소 문자열에서 숫자인식은 숫자를 정확히 탐지하고 인식해야하며, 기타 문자들에 대해서는 인식 기각 결정을 하여야 한다. 만약 한글, 영어, 기호, 숫자를 모두 인식하는 대규모 인식기를 구현하면 별도의 탐지기가 없어도 되지만, 이는 매우 어려운 일이다. 따라서, 문자 종류별로 모듈화된 인식기를 구현하는 것이 보다 효율적인 것이며, 특히 숫자의 탐지는 인식과 더불어 우편주소에 있어서 가장 중요한 과정이다. 숫자의 탐지 및 인식에 대해서는 인쇄체의 경우 연구[5]가 어느 정도 진행되어왔으나, 필기체의 경우 특히 한글 및 영문자가 혼재된 상태에서의 연구는 거의 수행되지 않았다.

본 논문은 필기체 숫자의 탐지 및 인식에 대한 연구로써, 신경망을 이용하여 숫자 탐지기 및 인식기를 구현하고 그 실험결과에 대해서 논한다.

#### II. 입력 특징 및 실험 데이터베이스

##### 2.1 입력 특징

###### (1) 입력 문자영상의 정규화

일반적인 문자인식기는 입력 문자영상을 고정된 크기의 영상으로 정규화 한 후에 특징을 추출하고 인식

하게 된다[6]. 문자영상의 정규화 방법에 따라서 인식률의 차이가 나타날 수 있으나, 본 논문의 실험에서는 간단한 선형 정규화 방법을 사용하였다. 원시 영상이  $f(x, y)$  일때, 크기 정규화된 영상  $f(x', y')$ 는 다음 식과 같이 얻어진다.

$$f(x', y') = f(x, y), \text{ where } x = x'/S_x, y = y'/S_y$$

$$S_x = N/W, S_y = N/H$$

여기서  $S_x$ 와  $S_y$ 는 각각 x축 및 y축의 스케일이다.

(2) 화소 방향 특징

인식기의 성능에 가장 많은 영향을 미칠 수 있는 요인은 사용되는 입력 벡터일 것이다. 따라서, 문자영상을 인식할 때 인식기의 입력으로 영상 자체를 그대로 사용하기도 하지만, 인식을 보다 용이하게 하기 위하여 문자영상이 지닌 고유한 특징을 추출하여 인식기의 입력으로 하는 경우가 보다 일반적이다. 본 논문에서는 인식기의 특징으로 화소의 방향 특징을 이용하였다. 화소의 방향은 방향벡터에 의해서 정의되는데, 방향벡터는  $N \times N$  영역내의 각 흑화소를 벡터로 표현하고 이들의 합을 구함으로써 얻어진다. 그림 1에서는 화소 p에 대한 예로 그림 1-(a)에서의 방향은  $45^\circ$  이며 그림 1-(b)에서는  $63^\circ$  이다.

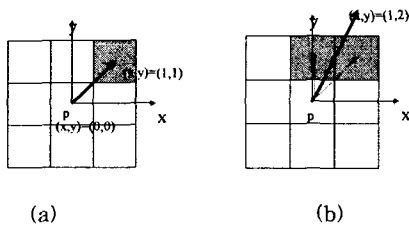


그림 1. 화소 p의 방향벡터

각 화소 특징은  $24 \times 24$ 의 정규화 영상에서  $4 \times 4$  크기의 부분영역 내의 각 화소들에 대한 방향을  $45^\circ$  간격으로 나누어 저장한다. 따라서,  $6 \times 6$ 개의 부분영역에 대해 8개(360/45)의 방향들이 정의되므로 전체 288 차원의 특징 벡터가 얻어진다.

2.2 실험 데이터베이스

본 논문에서 사용한 실험 문자영상 데이터베이스는 4세트로 되어있다. 2세트는 숫자들만 있으며, 나머지 2세트는 우편 봉투에서 직접 추출한 숫자 이외의 영상으로 되어있다. 본 논문에서 구현하는 인식기는 숫자를 인식대상으로 하므로, 숫자 영상을 정샘플 (positive sample) 이라 하고 숫자가 아닌 영상을 부샘플

(negative sample) 이라 한다. 첫 번째 정샘플 데이터베이스는 경북대학교의 숫자데이터베이스로 11690자로 이루어져있으며 본 논문에서는 AP 라 명한다. 다음으로 연세대학교에서 수집한 숫자 데이터베이스로 전체 41193자로 되어있고, BP 라 명한다. 부샘플 데이터베이스는 AN과 BN이 있는데, AN은 9113개이며 BN은 4709개로 되어있다. AN의 부샘플의 예를 그림 2에 나타내었는데, 우편 봉투에 씌어진 '대전'라는 필기 한글에서 인식을 위해 분리해낸 후보 문자영상들이다.



그림 2. 부샘플의 예

III. 숫자 탐지 및 인식

3.1 신경망을 이용한 숫자의 탐지 및 인식

본 논문에서는 숫자의 탐지 및 인식을 위하여 다층 퍼셉트론 신경망[7]을 이용한다.

인식기의 실험은 두가지로 동일한 구조의 신경망 인식기가 정샘플만을 이용했을 때의 성능과 정샘플과 부샘플을 모두 이용했을 때의 성능을 비교해보고자 한다. 정샘플 데이터베이스 AP만을 이용하여 인식기, Reco-P와 정샘플 AP와 부샘플 AN을 모두 이용하는 인식기, Reco-PN을 한 개의 중간층을 가지는 다층퍼셉트론 신경망으로 구현하였다. 각 인식기의 입력으로는 288 차원의 화소 방향특징을 이용하여서, 입력노드는 288개이며, 입력 숫자 클래스마다 1개의 출력노드를 할당하므로, 10개의 출력노드로 되어있다.

인식기의 학습단계에서 출력 목표 값은 입력 영상의 클래스에 해당하는 노드만 '1'이며 다른 노드는 '0'으로 두었다. 인식기의 중간층 개수는 한 개이며, 중간층의 노드 수는 30, 50, 70, 90개로 변화시켜서 실험하였다. 인식기의 학습을 위한 파라미터는 학습률을 0.1로 하였고, 관성항을 0.7로 하였다. 학습반복횟수는 학습률을 충분히 크게 하였으므로 50회만을 수행하였다.

표 1에 인식기의 인식 실험결과를 나타내었는데, 충분히 예측 가능한 결과로써 Reco-PN이 Reco-P와 유사한 인식률을 얻기 위해서는 보다 더 많은 중간층 노드가 필요함을 알 수 있다. 그러나 이 결과는 어떠한 기각도 없을 때의 경우이다. 두 인식기의 성능을 보다 더 살펴보기 위해서 기각 임계값을 변화시켜가며 인식기의 신뢰도와 오류율, 기각률, 인식률, 그리고 부샘플

데이터 처리율의 변화 추이를 그림 3과 4에 나타내었다. 각 그래프에서 신뢰도와 오류율, 기각률, 인식률은 정샘플로 실험한 결과이고, 부샘플 처리율은 부샘플에 대한 기각률을 나타낸 것이다. 기각 방법은 최고출력 노드 값이 기각 임계값보다 작을 때 기각하였다.

표 1. 인식기 인식 실험결과

	데이터 세트							
	AP				BP			
중간층 노드	30	50	70	90	30	50	70	90
Reco-P	99.83	99.98	99.99	100	93.56	94.52	94.96	95.27
Reco-PN	98.96	99.50	99.76	99.91	91.95	93.66	94.25	94.57

실험은 모두 학습에 사용하지 않은 테스트 데이터를 이용하였다. 그림 3은 정샘플만을 이용하여 학습한 인식기, Reco-P에 대한 실험결과이고, 그림 4는 정샘플과 부샘플을 모두 이용하여 학습한 인식기, Reco-PN으로 실험한 결과이다. 예상할 수 있는 바와 같이 두 그래프를 보면 모두 기각 임계값을 높임에 따라서 신뢰도는 점차로 상승하고 있으며 오류율은 감소하고 있다. 또한, 기각된 데이터와 오인식 데이터를 포함하고 있는 인식률은 점차로 하락함을 알 수 있다.

그림 3과 4의 두 그래프에서 알 수 있는 두 인식기의 가장 큰 차이는 부샘플에 대한 처리율이다. 그래프에서 알 수 있듯이 부샘플로 학습한 인식기는 기각 임계값이 낮을 때부터 매우 높은 기각률을 보이는데 반해서 정샘플로만 학습한 인식기는 기각 임계값을 상당히 높인 후에도 처리율이 낮음을 알 수 있다. 즉, 정샘플로만 학습한 인식기는 문자가 아닌 입력 데이터에 대해서도 높은 출력값의 결과를 내고 있는 것이다. 이러한 부샘플 처리율은 정샘플만을 인식대상으로 하는 낱자 인식에서는 중요치 않으나, 인식기를 단어인식이나 문자탐지에 적용할 때는 매우 중요한 역할을 하게 된다. 부샘플 처리율이 높으면 잘못 분할된 데이터나 엉뚱한 데이터가 입력되었을 때 이를 신뢰성 있게 처리할 수 있기 때문이다. 이 실험을 통하여 부샘플을 학습 데이터에 추가하면 정샘플에 대한 인식률에서는 미세한 손실이 있을 수 있으나 부샘플이 입력되었을 때 이를 신뢰성 있게 처리할 수 있기 때문에 인식기의 신뢰도를 크게 향상시킬 수 있다는 것을 알 수 있다.

숫자의 탐지를 위해서 탐지기는 두 가지 방법으로 구현되는데 그림 5에 그 구조를 나타내었다. Spot-PN2는 두 개의 출력층 노드를 가지며, 각각 정

샘플과 부샘플에 대응된다. Spot-PN1에는 11개의 출력층 노드가 있으며, 10개는 각 숫자 클래스를 위한 것이며, 마지막 노드는 부샘플에 대응된다. Spot-PN1의 탐지 결정은 최고 출력값이 상위 10개의 노드에서 나타나면, 입력을 정샘플이라 결정하고, N에 해당하는 노드 출력값이 가장 크면 부샘플로 결정한다.

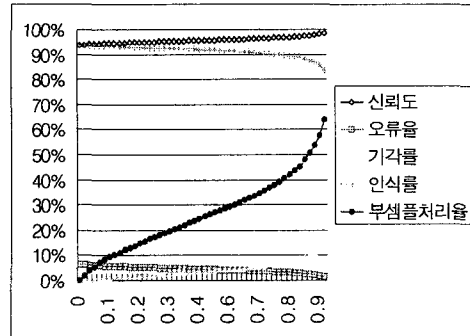


그림 3. 정샘플로만 학습한 Reco-P의 특성

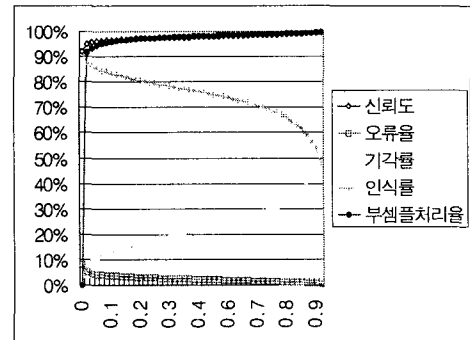


그림 4. 정샘플과 부샘플로 학습한 Reco-PN의 특성

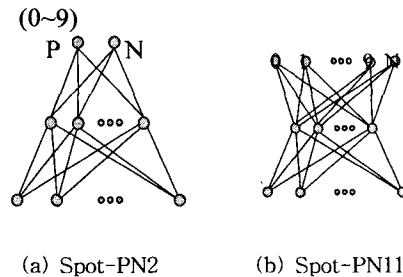


그림 5. 탐지기의 구조

정샘플들에 대한 탐지율을 나타내는 표 2를 보면 보다 많은 출력 노드를 가진 Spot-PN1의 탐지율이 훨

선 높음을 알 수 있다. 이는 정샘플과 부샘플에 대한 출력값들의 공간이 보다 넓게 분포하여 분별이 보다 용이하기 때문인 것으로 판단될 수 있다. 부샘플에 대한 탐지기들의 탐지율을 나타낸 표 3을 보면 각 탐지기들의 탐지율이 매우 높게 나타나서 그 차이가 크게 두드러지지 않았다. 이는 많은 부샘플들이 하나의 부샘플 노드에 집중되어 부샘플에 대한 학습이 충분히 되어 탐지율이 높게 나타났다고 볼 수 있다. 그러나, 표 2의 결과를 다시 보면, Spot-PN2의 정샘플 탐지율은 마찬가지로 하나의 노드에 많은 정샘플들이 학습되지만, 탐지율은 낮게 나타난 것으로 보아 부샘플들에 비해 정샘플들의 각 클래스 분포공간이 보다 명확하게 차이가 나서, 하나의 노드로 이들을 수용하기 어려웠던 것으로 판단된다.

표 2. 정샘플에 대한 탐지기의 탐지 실험결과

	데이터 세트							
	AP				BP			
중간층 노드	30	50	70	90	30	50	70	90
Spot-PN2	87.45	88.73	87.27	86.83	69.23	71.69	69.71	68.56
Spot-PN11	87.52	93.70	97.13	97.42	68.76	78.29	83.29	83.74

표 3. 부샘플에 대한 탐지기의 탐지 실험결과

	데이터 세트							
	AN				BN			
중간층 노드	30	50	70	90	30	50	70	90
Spot-PN2	99.33	99.39	99.31	99.60	98.24	98.26	98.43	98.47
Spot-PN11	99.59	99.60	99.53	99.85	99.13	98.43	98.19	98.43

#### IV. 결론

본 논문에서는 필기체 숫자의 탐지 및 인식을 위하여 신경망을 이용하여 탐지기와 인식기를 각각 구현하였다. 실험 결과 숫자 인식기에 있어서 부샘플을 학습 데이터에 추가하면 정샘플에 대한 인식률에서는 미세한 손실이 있을 수 있으나 부샘플이 입력되었을 때 이를 신뢰성 있게 처리할 수 있기 때문에 인식기의 신뢰

도를 크게 향상시킬 수 있다는 것을 알 수 있었다. 또한 숫자의 탐지에서는 정샘플과 부샘플에 각각 대응되는 2개의 출력노드를 가지는 탐지기보다는 정샘플 클래스 10개와 부샘플 클래스 1개에 각각 대응되는 11개의 출력노드를 가지는 탐지기가 보다 우수한 성능을 보여준다는 것을 알 수 있었다.

본 논문의 실험은 신경망을 적용한 숫자 탐지기와 인식기의 가능성에 대한 연구이므로, 구현된 탐지기와 인식기를 최적화하고 조합하여 최적의 탐지율 및 인식율을 얻어, 실제 우편 봉투상의 주소 문자열에 대해 성공적으로 적용하는 것이 향후의 연구과제로 남아있다.

#### 참고문헌

- [1] S. Cho, "Neural-network classifiers for recognizing totally unconstrained handwritten numerals," *IEEE Transactions on Neural Networks*, vol. 8, pp. 43-53, 1997.
- [2] D. Lee and S. N. Srihari, "A theory of classifier combination: the neural network approach," *Proc. of the Third ICDAR*, Montreal, Canada, pp. 42-45, 1995.
- [3] J. Cao, M. Ahmadi, and M. Shridhar, "Recognition of handwritten numerals with multiple feature and multistage classifier," *Pattern Recognition*, vol. 28, no. 2, pp. 153-160, 1995.
- [4] C. Y. Suen, C. Nadal, R. Legault, T. A. Mai, and L. Lam, "Computer recognition of unconstrained handwritten numerals," *Proc. of IEEE*, vol. 80, no. 7, pp. 1162-1180, 1992.
- [5] J. H. Kim, K. Kim, and S. Chien, "Korean and English character recognition system using hierarchical classification neural network," *Proc. of ICSMC*, Vancouver, vol. 1, pp. 759-764, Oct. 1995.
- [6] S. Lee and J. Park, "Nonlinear shape normalization methods for the recognition of large-set handwritten characters," *Pattern Recognition*, vol. 27, no. 7, pp. 895-902, 1994.
- [7] D. E. Rumelhart, G.E. Hinton, and R.J. Williams, "Learning Internal Representations by Error Propagation," *Parallel Distributed Processing*, vol. 1, Eds. D. E. Rumelhart and J. L. McClelland, pp. 319-362, 1986.