

# 완전결환 신경망을 이용한 무제약 서체 숫자 인식

원상철, 배수정, 최한고  
금오공과대학교, 전자공학부

## Recognition of Unconstrained Handwritten Numerals using Fully-connected RNN

Sang-Choel Won, Su-Jung Bae, Han-Go Choi  
School of Electronic Engineering, Kumoh Nat. Univ. of Tech.  
E-mail: haunter@kmut.kumoh.ac.kr

### Abstract

This paper describes the recognition of totally unconstrained handwritten numerals using neural networks. Neural networks with multiple output nodes have been successfully used to classify complex handwritten numerals. The recognition system consists of the preprocessing stage to extract features using Kirsch mask and the classification stage to recognize the numerals using the fully-connected recurrent neural networks (RNN). Simulation results with the numeral database of Concordia university, Montreal, Canada, are presented. The recognition system proposed in this paper outperforms other recognition systems reported on the same database.

### I. 서론

숫자와 문자와 같은 패턴인식 시스템은 입력신호의 유용한 특징추출을 위한 전처리 과정과 이를 이용하여 정확한 패턴인식을 위한 분류과정으로 나누어 수행된다. 따라서 패턴 인식율을 향상시키기 위해서는 적합한 전처리 과정이 필요하며, 또한 이를 효과적으로 적용할 수 있는 분류 시스템이 동시에 요구된다.

본 논문에서는 무제약 서체 숫자인식을 다루고 있는

데, 서체 숫자는 사람에 따라서 글씨체가 다르며 각 숫자의 크기 또한 다양하다. 그리고 사용되는 필기도구 및 매체에 따라서 숫자의 형태가 매우 다양하므로 이를 인식하기란 쉽지 않다. 이러한 다양성을 갖는 숫자를 직접 사용할 경우 시스템이 매우 복잡하며, 숫자간의 차이가 명확하지 않는 경우 숫자 인식이 어렵다. 따라서 숫자간의 차별성을 강조하기 위하여 각 숫자를 대표할 수 있는 특징을 추출하여 사용하고 있다. 본 연구에서는 기존의 방법 중 Kirsch 마스크를 사용하여 방향성 특징을 추출하는 전처리 과정을 일부 보완한 방법을 사용하고 있다.

분류방법으로서 최근에는 학습능력과 비선형 처리 능력을 갖고 있는 신경망이 숫자 및 문자인식과 같은 복잡한 패턴분류에 널리 사용되고 있으며 여러 논문에서 우수한 인식결과를 보여주고 있다[1-3]. 이러한 신경망들은 네트워크의 구조, 뉴런의 모델 및 활성화 함수, 뉴런간의 연결강도 갱신을 위한 학습법칙 등에 따라서 여러 가지 형태의 신경망이 연구되어 사용되고 있다. 다양하고 복잡한 패턴을 분류하기 위해서는 동특성이 우수한 신경망의 사용이 요구되는데, 이를 구현하기 위해서 신경망에 궤환경로나 지연소자 등을 사용하여 동특성 구현능력을 향상시킴으로써 입력 데이터의 temporal 정보를 효과적으로 처리할 수 있어 복잡한 패턴의 인식율을 향상시킬 수 있다. 본 연구에서는 리커런트 신경망을 사용하고 있는데, 은닉층의 모든 뉴런 출력이 입력층으로 궤환되는 완전연결 리커런트 신경망 (fully-connected RNN)을 사용하여 숫자분

류를 수행하고 있다.

숫자 인식을 위해 실험에 사용된 숫자 데이터는 Concordia 대학의 데이터 베이스로서, 이 데이터 베이스는 각 숫자의 크기가 서로 다른 무제약 (unconstrained) 서체 숫자들로 이루어져 있으며 여러 연구에서 사용되고 있다. 실험에서 신경망의 출력 노드수에 따른 숫자 인식율을 조사하였으며, 실험결과에 의하면 같은 데이터 베이스에 대해 수행되었던 기존의 다른 방법보다 더 우수한 숫자 인식율을 보여줌으로써 제안된 방법이 숫자 패턴분류에 유용함을 확인할 수 있었다.

## II. 숫자 인식 시스템

### 2.1 데이터 베이스

이 논문에서 사용된 데이터 베이스는 미국 전역에 있는 우체국에서 발송되지 못한 편지 봉투의 우편번호를 캐나다의 콘코디아 대학에서 데이터 베이스화시킨 것이다[4]. 이 데이터 베이스 숫자들은 모두 Run Length Coding으로 압축되어 있으며 전체 숫자는 6000개로서 4000개의 학습 데이터와 2000개의 실험 데이터로 구성되어 있다. 그림 1은 데이터 베이스에서 발췌된 학습 및 시험 샘플 데이터의 예를 보여주고 있다. 그림에서 확인할 수 있듯이 각 숫자들의 크기, 폭, 기울기, 그리고 형태 등이 글씨를 쓰는 스타일에 따라서 서로 다름을 알 수 있다.

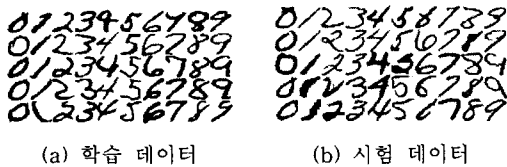


그림 1. 샘플 데이터

Fig. 1. Sample data

### 2.2 데이터 특징 추출

신경망의 입력 데이터로서 숫자 영상을 그대로 사용할 경우 시스템이 복잡해지므로 우수한 인식율을 기대할 수 없다. 따라서 숫자 영상에 대한 사전 정보를 이용하면, 즉, 입력 영상으로부터 특징을 추출하여 신경망의 입력으로 사용한다면 신경망의 인식성능의 향상을 기대할 수 있다. 이를 위해서는 인식에 필요한 중요한 정보의 손실을 최소화하는 특징 추출과정을 통해 특징 벡터를 추출할 필요가 있다.

일반적으로 숫자는 짧은 라인 세그먼트로 구성된 것

으로 볼 수 있으므로 라인 에지 탐지를 위해 Kirsch 마스크를 사용한 영상 에지 추출방법을 사용하였다. 이 Kirsch 마스크는 마스크 연산 시 주위 8개 이웃 화소(pixel) 모두를 고려함으로써 다른 마스크에 비하여 4방향에 대해 보다 정확한 에지를 검출하므로 숫자 영상의 에지 검출에 적합하다.

영상으로부터 에지를 검출하기 위한 Kirsch 방법은 다음과 같이 정의된다[5].

$$G(i, j) = \max_{k=0}^7 [ |5S_k - 3T_k| ] \quad (1)$$

$$S_k = A_k + A_{k+1} + A_{k+2} \quad (2)$$

$$T_k = A_{k+3} + A_{k+4} + A_{k+5} + A_{k+6} + A_{k+7} \quad (3)$$

여기서  $G(i, j)$ 는  $i$ 번째 행과  $j$ 번째 열의 화소의 기울기 값이고,  $A$ 는 8진수로 계산된 나머지 값이며,  $A_k (k=0, 1, \dots, 7)$ 는 특정 화소  $(i, j)$ 에 이웃하는 8개 화소들이다. 각 화소에 곱해지는 숫자의 의미는 방향성 기울기의 scale factor를 나타낸다.

본 연구에서는 원래의 숫자 이미지의 가로와 세로 크기가 약 20에서 80픽셀 분포의 다양한 형태로 구성되어 있으므로 크기를 정규화하고 방향성 특징벡터를 생성하여 신경망의 입력으로 사용하였다. 이를 위한 전처리 단계로 원 숫자 이미지를 256X256의 이미지로 변환하여 모든 숫자를 같은 크기로 변환시키고, 변환된 이미지를 64X64로 압축한 후 다시 16X16의 압축된 이미지를 생성하였다. 그리고 각 이미지는 4가지의 방향성 이미지 벡터, 즉, 수평, 수직, 좌, 그리고 우 방향에 대한 특징벡터를 추출하기 위해 다음 식을 이용하여 방향정보를 갖는 벡터를 계산하였다.

$$G(i, j)_H = \max ( |5S_0 - 3T_0|, |5S_4 - 3T_4| ) \quad (4)$$

$$G(i, j)_V = \max ( |5S_2 - 3T_2|, |5S_6 - 3T_6| ) \quad (5)$$

$$G(i, j)_R = \max ( |5S_1 - 3T_1|, |5S_5 - 3T_5| ) \quad (6)$$

$$G(i, j)_L = \max ( |5S_3 - 3T_3|, |5S_7 - 3T_7| ) \quad (7)$$

4쌍으로 이루어진 Kirsch 마스크를 사용하여 구한 이미지 벡터는 평균화 연산을 통하여 다시 4X4의 벡터로 압축하였으며 이 벡터를 각 숫자의 국부적(local) 특징을 갖는 입력벡터로 사용하였다. 따라서 각 숫자에 대한 국부적 특징벡터는 4X4X4의 방향정보 벡터로 구성된다. 무제약 필기체 숫자인식에 있어서 이러한 국부적인 특징만을 이용할 경우, 문자의 전역적인(global) 특징을 고려할 수 없으므로 정규화된 16X16 이미지를 직접 4X4 이미지로 단순 압축하여 이를 전역적인 특징벡터로 사용하였다. 따라서 본 연구에서는 각 숫자에 대해 모두 5개의 4X4 특징벡터 (4X4X4 개

의 국부적 특징벡터와 1X4X4 개의 전역적 특징벡터)가 추출되어 신경망의 입력으로 사용하였다. 그림 2는 숫자 2에 대한 특징벡터의 이미지 추출결과를 보여주고 있다.

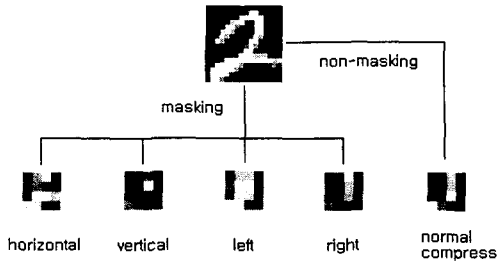


그림 2. 특징 벡터 이미지 추출과,   
 Fig. 2. Processing of extracting image features

### 2.3 완전 연결 RNN

그림 3은 3층으로 구성된 완전케환 리커런트 신경망 구조를 보여주고 있다. 은닉층의 모든 뉴런이 입력층으로 케환된 구조로서 신경망의 동적특성을 향상시킨다. 신경망은 오차 역전달 알고리즘에 의해 층간의 연결계수  $w_{ji}(n)$ 와  $w_{kj}(n)$ 가 다음과 같이 갱신된다.

$$w_{kj}(n+1) = w_{kj}(n) - \eta \frac{\partial E(n)}{\partial w_{kj}(n)} + a\Delta w_{kj}(n-1) \quad (9)$$

$$w_{ji}(n+1) = w_{ji}(n) - \eta \frac{\partial E(n)}{\partial w_{ji}(n)} + a\Delta w_{ji}(n-1) \quad (10)$$

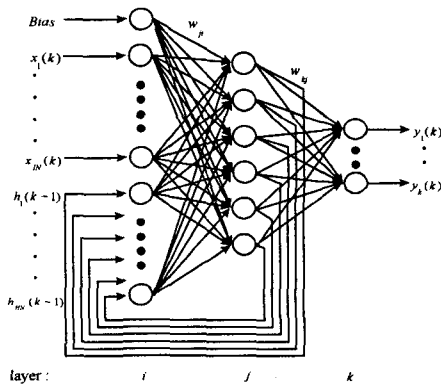


그림 3. 완전 케환 RNN 구조   
 Fig. 3. Structure of RNN

여기서  $\eta$ 와  $a$ 는 각각 학습율과 모멘텀을 나타내며 오차함수  $E(n)$ 는 원하는 출력과 신경망 출력의 차의 제곱으로서 오차가 최소가 되도록 연결계수가 갱신된다. 본 신경망의 각 뉴런의 활성화함수로는 시그모이드

함수를 사용하였다.

### III. 실험 결과

숫자 인식율을 조사하기 위해 3가지 구조의 리커런트 신경망에 대해서 실험하였다. 첫 번째 구조는 신경망의 입력층, 은닉층, 그리고 출력층의 뉴런수가 각각 160개 (80개 특징벡터의 외부입력과 은닉층으로 부터의 80개 케환 벡터), 80개, 그리고 4개이다. 두 번째 구조는 200개 (80개 외부입력과 120개 케환입력), 120개, 4개이며, 세 번째 구조는 200개, 120개, 10개인 구조를 사용하여 실험하였다. 출력이 10개이고 은닉층의 뉴런수가 80개일 경우 숫자 인식이 제대로 이루어지지 않았으므로 은닉층의 개수를 120개로 늘려서 실험하였다.

출력층의 뉴런수가 4개인 경우 숫자 '0'에서 '9'에 대해 출력 값이 0000에서 1001로 각각 대응시켰고, 뉴런수가 10개인 경우에는 해당 출력뉴런이 1이 되게 하고 나머지 뉴런은 모두 0이 되도록 학습하였다. 그리고 1과 0의 결정은 뉴런의 출력 값에 임계값과 비교함으로써 이루어지는데, 본 실험에서는 4가지 임계값 경우, 즉, 0.9/0.1, 0.8/0.2, 0.7/0.3, 그리고 0.6/0.4에 대해서 각각 실험하였다. 학습율과 모멘텀 율은 4개 출력층 뉴런의 경우 0.15와 0.7을 사용하였고, 10개 출력층 뉴런의 경우 0.1과 0.5를 사용하였다. 이러한 신경망의 구조와 변수 값들은 실험에 의해서 결정되었다.

학습 데이터에 대한 신경망의 학습횟수는 학습 데이터의 모든 숫자가 인식이 완전히 이루어 질 때까지 수행되었는데, 각 신경망 구조에 대해 대부분 100 epoch 보다 적은 반복횟수에서 학습이 이루어졌다. 또 다른 실험은 학습이 성공적으로 수행되었음에도 불구하고 반복횟수가 200 epoch 될 때까지 계속하였다. 그리고 이러한 실험은 4가지 임계값에 따라서 각각 달리 실험하였다.

표 1은 시험 데이터에 대해 5번 수행한 후 평균 인식율을 보여주고 있는데 대부분의 경우 인식율은 98% 이상을 나타내고 있다. 이러한 결과는 주어진 숫자 데이터 베이스에 대해 매우 양호한 인식결과를 보여주고 있다. 표 1에서 인식율#1은 학습 데이터에 대해 모든 숫자의 인식이 이루어질 때 학습이 종료된 경우이며, 인식율#2는 200 epoch 까지 학습이 수행된 것으로서, 첫 번째 경우(4개 출력층 노드, 80개 은닉층 노드, 임계값 0.9/0.1)를 제외하고는 인식율#1 보다 약간 향상되었다.

표 1. 시험 데이터에 대한 평균 인식율

Table 1. Average recognition rate of test data

no. of output nodes	no. of hidden nodes	threshold (1/0)	인식율#1 (%)	인식율#2 (%)
4	80	0.9/0.1	97.86	97.56
		0.8/0.2	98.22	98.67
		0.7/0.3	98.38	99.14
		0.6/0.4	99.13	99.44
	120	0.9/0.1	97.65	98.12
		0.8/0.2	98.74	99.23
		0.7/0.3	98.98	99.55
10	120	0.6/0.4	99.44	99.53
		0.9/0.1	98.24	98.82
		0.8/0.2	98.64	99.23
		0.7/0.3	99.05	99.63
		0.6/0.4	99.46	99.69

표 2. 숫자별 인식율

Table 2. Confusion matrix of numerals

	0	1	2	3	4	5	6	7	8	9	미분류	오판율 (%)	인식율 (%)
0	200											0.0	100.0
1		200										0.0	100.0
2	4		193	1			1				1	3.5	96.5
3	1	2		194				2			1	3.0	97.0
4					200							0.0	100.0
5						200						0.0	100.0
6			2				197				1	1.5	98.5
7				2				198				1.0	99.0
8	1								197		2	1.5	98.5
9										200		0.0	100.0
											평균치	1.05	98.95

표 3. 다른 인식방법들의 인식율

Table 3. Recognition rate of other methods

Methods	Recognized	Substituted	Rejected	Reliability
Kim	95.85	4.15	0.00	95.85
Krzyzak	94.85	5.15	0.00	94.85
Lam	93.10	2.95	3.95	96.98
Legault	93.90	1.60	4.50	98.32
Mai	92.95	2.15	4.90	97.74
Nadal	86.05	2.25	11.70	97.45
Suen	93.05	0.00	6.95	100.0
Cho	96.05	3.95	0.00	96.05
Lee	97.30	2.70	0.00	97.30

표 2는 출력층 노드 4개, 은닉층 노드 80개, 임계값이 0.6/0.4인 경우에 대한 숫자별 인식내용을 보여주고

있는데, 표에서의 미분류는 해당 임계값으로 결정할 수 없는 경우이다. 표 3은 같은 데이터 베이스에 대해 지금까지 발표된 다른 시스템의 인식결과들을 보여주고 있는데[6], 본 연구에서의 인식결과가 기존 방법들에 의한 인식결과 보다 더 우수함을 알 수 있다.

#### IV. 결론

본 논문에서는 리커런트 신경망을 이용한 무제약 서체 숫자인식을 다루고 있다. 전처리 과정으로는 기존의 방향성 특징 추출과정을 보완하였으며 완전 연결된 리커런트 신경망을 이용하여 숫자 인식을 수행하였다. 실험결과에 의하면 숫자 인식결과가 매우 양호한 것으로 나타났으며, 같은 데이터 베이스에 대해 수행되었던 기존 숫자 인식 시스템의 인식결과 보다 더 우수한 인식율을 보여줌으로써 본 논문에서 사용한 서체 숫자 인식방법의 유용성을 확인하였다.

#### 참고문헌(또는 Reference)

- [1] Y. Le Cun et al., "Constrained neural-network for unconstrained handwritten digit recognition", in Proc. 1st Int. Wkshp. Frontiers Handwriting Recognition, pp. 145-154, Montreal, Canada, 1990.
- [2] S. W. Lee, H. H. Song, "A New Recurrent Neural-Network Architecture for Visual Pattern Recognition", *IEEE Trans. on Neural Networks*, Vol. 8, No. 2, Mar. 1997.
- [3] R. Urbanczik, "A recurrent neural network inverting a deformable template model of handwritten digits", in Proc. Int. Conf. Artificial Neural Networks, pp.961-964. Sorrento, Italy, 1994
- [4] C. Y. Suen, C. Nadal, T. Mai, R. Legault, and L. Lam, "Recognition of handwritten numerals based on the concept of multiple experts," in *proc. 1st Int. Wkshp. Frontiers Handwriting Recognition*, Montreal, Canada, pp. 131-144, 1990.
- [5] W. K. Pratt, *Digital Image Processing*, New York: Wiley, 1978.
- [6] S. B. Cho "Neural-Network Classifiers for Recognizing Totally Unconstrained Handwritten Numerals", *IEEE Trans. on Neural Networks*, Vol. 8, No. 1, Jan. 1997.