

디더링과 모듈 구조의 다중 MLP를 이용한 무제약 필기체 숫자 인식

임길택(林吉澤)*, 남윤석(南潤奭)*, 진성일(秦成一)**

*한국전자통신연구원 우정기술연구부 우정자동화팀

**경북대학교 전자전기공학부

Unconstrained Numeral Recognition Using Dithering and Multiple Modular MLPs

Kil Taek Lim, Yun Seok Nam, Sung Il Chien

*Postal Technology Deveopment Department, ETRI.

**School of Electronic and Electrical Engineering, Kyungpook National University

E-mail: ktlim@etri.re.kr

Abstract

In this paper, we propose a method of unconstrained handwritten numeral recognition using image dithering and multiple modular MLPs. The set of sample numeral patterns is subdivided into clusters which are extended by their radius. On each extended cluster, we constructed MLP network as the expert recognizer of corresponding cluster. The gating network is also trained by an MLP to weigh the outputs of expert MLPs. In training and test phase of the recognizer, we utilize the multiple dithered numeral images and the combination of the outputs for corresponding dithered images. Experimental results show that our recognition method works very well.

I. 서론

무제약 필기체 숫자의 경우는 매우 복잡한 결정경계를 이루기 때문에 간단한 구조로 높은 인식률을 얻기가 매우 힘들다. 따라서 이러한 경우의 복잡한 인식문제를 해결하기 위해서 최근의 연구들은 다중 특징 및 다중 인식기를 이용하는 복잡한 구조를 가지는 인식시스템에 대한 연구로 진행되어왔다. 이러한 다중 인식기에 대한 연구는 크게 다음과 같이 나눌 수 있다. 첫

째, 다중 특징을 이용하여 각 특징별로 인식기를 구현한 후 인식기들을 결합하여 인식률을 향상시킨다. 이 경우 각 인식기는 각 특징에 대해서 전문가망으로 동작하는데 서로 다른 패턴공간에서 구현된 인식기를 통합하여 인식한다[1,2]. 둘째, 다중 인식기는 서로 다른 알고리즘이나 구조로 구현되어 결합된다[3,4]. 셋째, divide and conquer의 방법을 적용하여 대분류 후 상세분류를 통해 인식을 하는데 이 경우 패턴공간을 보다 작은 공간으로 나누고 각 인식기들은 나누어진 특정한 패턴공간내에서 동작하게 된다[5,6].

본 논문에서는 게이팅망과 모듈화된 MLP 구조를 가지며 디더링(dithering)을 이용하는 새로운 문자인식 방법을 제안한다. 필기체 숫자 인식기의 구현을 위해 전체 패턴 공간을 서로 중첩하는 보다 작은 공간으로 나누고 다중 군집을 구성한다. 초기 다중 군집의 구성은 SOM (self organizing map)[7] 망을 이용하였고 이를 다시 각 군집의 반경에 따라 확장하여 새로운 군집을 만들었다. 구성된 각 군집에 대해서 전문가망으로 동작하는 MLP를 구현하고 이들을 결합시키기 위해 게이팅 기능의 MLP를 구현한다. 학습과 테스트 단계에서는 학습문자를 회전과 이동의 두가지 디더링 과정을 적용하였다. 실제 인식에 있어서 디더링에 의해 발생된 다수의 문자에 대한 인식기의 출력값들을 다양한 방법을 이용하여 결합함으로써 인식률을 향상시켰다.

II. 군집의 생성과 확장

본 논문에서는 학습 패턴들에 대해 SOM 알고리즘을 이용하여 초기 군집을 구성하였다. N_T 개로 이루어진 학습 샘플 집합을 $S = \{X_k, k=1, 2, \dots, N_T\}$ 이라 할 때 이 집합은 M 개의 분리된 부분집합 $S_i, i=1, 2, \dots, M$ 으로 나누어진다. 이때 부분집합 S_i 를 군집 i 라 한다. 이렇게 생성된 군집을 그대로 적용할 경우 실제 S_i 에 속하는 패턴이 S_i 상에서 학습된 인식기에 입력될 때 인식결과는 완전히 잘못될 수 있다. 이에 대처하기 위해 본 논문에서는 군집들이 위치하는 샘플 패턴공간을 서로 중첩되게 나누는 방법으로 군집을 확장하였다. 즉, 어떤 특정 군집 S_i 의 영향권 내에 들어올 수 있는 다른 군집의 패턴들을 군집 S_i 의 원소로 함으로써 군집 S_i 를 확장하는 것이다. SOM의 각 출력노드에 의해 정의될 수 있는 샘플 집합 S_i 는 반경 T_i 에 따라 식 (1)과 같이 \hat{S}_i 로 확장된다.

$$\hat{S}_i = \{X_k \mid \|X_k - W_i\| \leq T_i, k=1, 2, \dots, N_T\} \quad (1)$$

여기서, $T_i = \max \|X_k - W_i\|$, for $X_k \in S_i$ 이다. 이와 같은 과정을 그림 1에 개념적으로 나타내었다. 그림 1의 (a)에서 흰 사각형 패턴은 군집 i 의 대표벡터 W_i 에 가깝고 검은 사각형 패턴은 군집 j 의 대표벡터 W_j 에 가깝게 위치한다. 따라서 군집 S_i 는 흰 사각형 패턴들을 포함하고 S_j 는 검은 사각형 패턴을 포함한다. 그림 1의 (a)를 보면 S_i 에 있는 5개의 패턴들이 T_i 에 의해 정의되는 흰 사각형 패턴 영역에 위치하므로 그림 1의 (b)와 같이 \hat{S}_i 에 새롭게 추가된다.

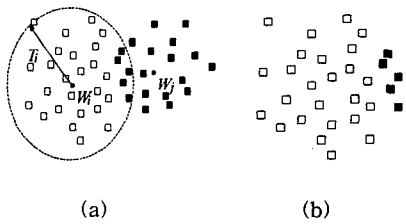


Figure 1. S_i is extended to \hat{S}_i .

III. MLP 모듈 구조의 문자인식기와 디더링을 이용한 인식

본 논문에서는 확장된 각 군집마다 MLP신경망을 적용하는 모듈구조의 문자 인식기로 인식을 한다. M

개의 확장된 특정 군집에 대해서 M 개의 MLP들이 학습되어 해당 군집에 대해서 상세분류기로서 동작한다. 즉, 각 MLP는 해당 군집에 대한 전문가망으로 동작한다. 입력된 패턴에 대해 군집을 선택하는 대분류를 위해서는 군집의 평균값이나 SOM 출력노드를 사용하기도 하지만 이들의 특성이 근본적으로 비교사학습이므로 클래스 분별력이 떨어질 가능성이 많다. 그러므로 우리는 상세분류기, 즉, 각 군집에 할당된 MLP들을 신뢰도 높게 제어하기 위해서 MLP를 이용하여 게이팅망을 구현한다. 이때 게이팅망의 출력노드의 개수는 전체 군집 전문가 MLP의 수와 같다. 이러한 인식기를 그림 2에 나타내었다. 그림 2의 인식기에서 어떤 특정 클래스 i 에 대한 인식기의 출력값은

$$o_i = \sum_{k=1}^M a^k f_i^k \quad (2)$$

와 같다. 여기서 o_i 는 클래스 i 에 대한 출력값이고, a^k 는 k 번째 군집에 대한 게이팅망의 출력값으로써 입력 패턴이 k 번째 군집에 속할 가능성을 나타낸다. f_i^k 는 k 번째 MLP의 클래스 i 에 대한 출력값을 나타낸다. 그림 2의 인식기의 결정규칙은 o_m 이 식 (3)와 같이 계산될 때 입력패턴의 클래스는 m 으로 결정된다.

$$o_m = \max o_i, i=1, 2, \dots, L \quad (3)$$

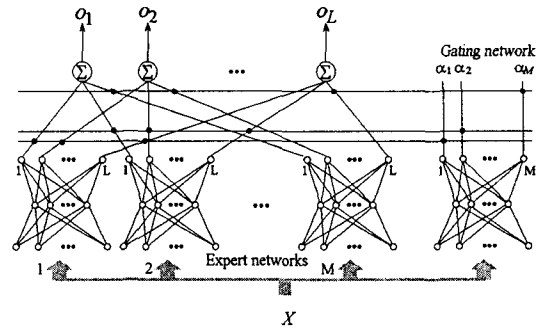


Figure 2. Proposed classification scheme: Gating network weighs the output values of individual MLPs specializing in a cluster.

본 논문에서는 필기체 숫자 영상 대해 회전과 이동 연산을 통한 디더링을 함으로써 확보된 숫자 집합을 확장하고 인식 성능을 높이고자 하였다. 입력 패턴에 대해 K 개의 디더링 문자들이 생성되어 인식기에 입력된다. 원시 입력 문자와 각각의 디더링 문자들은 그림 3과 같이 인식기에 입력되어 그 출력값들이 출력 배열에 저장된다. 이러한 $K+1$ 개의 출력 벡터 값들을 이용

하여 최종 인식 클래스를 결정하기 위해서 다양한 방법들을 실험하여보았다

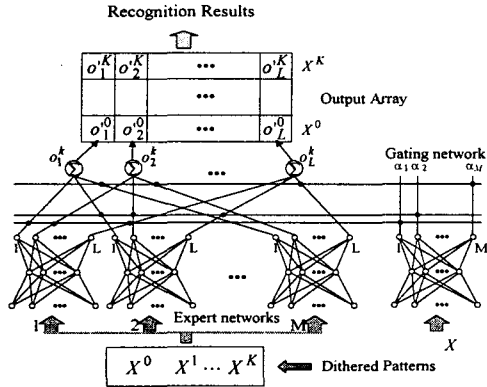


Figure 3. The proposed recognition scheme.

1. 디더링 문자에 대한 MLP 출력값 결합 방법

표 1에 본 논문에서는 사용한 디더링 된 패턴들에 대한 인식기 출력값들을 결합하는 다양한 방법들을 나타내었다. 인식기의 출력값들을 결합하는 각 방법들을 설명하기 위한 심볼들은 다음과 같다. 디더링 문자 X^k 가 입력될 때 클래스 i 에서의 출력값은

$$o_i^k = \sum_{m=1}^M \alpha^m f_i^m(X^k) \quad (4)$$

와 같다. 여기서 M 은 전체 MLP의 수와 같다. 디더링 된 각 문자들에 대한 각 MLP의 출력값들은 출력 배열에 변환되어 저장되는데 이들은 결합방법에 따라 결합된다. o_i^k 의 변환된 값을 o_i^k 로 두고, $\Lambda = \{1, 2, \dots, L\}$ 을 L 개의 클래스 레이블의 집합, $\Gamma = \{0, 1, \dots, K\}$ 를 모든 디더링 인덱스의 집합으로 정의한다. $o_j^k = \max_{j \in \Lambda} o_j^k$ 으로 X^k 에 대한 L 개의 출력값 중 최대값이며 $o_{j^2}^k = \max_{j \in \Lambda_2} o_j^k$ 는 두 번째 큰 값으로 j^* 는 최대값의 인덱스이며 j^2 는 두 번째 최대값의 인덱스이다. 이때 $\Lambda_2 = \Lambda - \{j^*\}$ 이다. 몇몇 방법에서 출력값 o_i^k 는 정규화되어 결합된다. 즉,

$$s^k o_i^k = o_i^k \quad (5)$$

가 된다. 여기서 $s^k = \sum_{i=1}^L o_i^k$ 로 k 번째 디더링 된 패턴에 대한 스케일로서 $\frac{1}{s^k} \sum_i o_i^k = \sum_i o_i^k / s^k = 1$ 이 된다.

입력 문자 X 를 디더링한 후 생성된 모든 문자들에 대한 출력값들을 결합한 후의 인식 클래스를 $D(X)$ 라

하며

$$D(X) = j, \quad (6)$$

$$j = \arg \max_{i \in \Lambda} \sum_{k=0}^K o_i^k$$

와 같이 정의한다. 방법 I의 경우는 o_i^k 의 모든 정보를 이용하는 것이며 방법 II의 경우는 각 문자영상에 대한 인식기의 신뢰도 값을 이용하는 것이다. 방법 III의 경우는 보우팅과 그변형에 대한 것이다.

Table 1. Summary for methods of output combination.

Methods	o_i^k	Decision
I-1	o_i^k	$D(X)$
I-2	o_i^k/s^k	
II-1	$(o_j^k - o_{j^2}^k)\delta_{ij}$	
II-2	$((o_j^k - o_{j^2}^k)/s^k)\delta_{ij}$	
II-3	$(o_j^k/s^{k+1} + (1 - o_{j^2}^k/o_{j^2}^k))\delta_{ij}$	
III-1	δ_{ij}	
III-2	$o_j^k \delta_{ij}$	
III-3	$o_j^k \delta_{ij} / s^k$	

VI. 인식실험

1. 인식기의 학습

본 논문의 실험에 사용한 데이터는 Concordia대학교의 필기체 숫자 데이터베이스이다. 이 데이터베이스는 6000자로 구성되어있는데, 상위 4000자는 학습에 사용하였고 하위 2000자는 테스트에 사용하였다. 입력 특징으로는 144차원으로 된 숫자영상의 기울기 벡터[8]에서 추출된 주성분인데 군집생성과 게이팅망의 경우는 54차원이며 각 MLP들은 144차원을 가진다. SOM 학습에 사용된 파라미터는 초기 학습률은 0.85이며 학습 반복 횟수는 100회이다. 이웃의 구조는 사각형으로 하였으며 6x6의 출력구조를 가지고있다. 확장된 군집들에서 학습되는 MLP들은 모두 동일한 구조를 가진다. 입력 노드의 개수는 144이며 출력 노드의 개수는 10개이며 은닉층의 노드 수는 30개이다. 학습률은 0.1로 하였고 관성항은 0.7 최대 학습 반복 횟수는 100회로 제한하였다. 입력 문자 영상을 디더링할 때 회전각 θ 는 5°, 10°로 두었다. 이동의 경우는 이동변이값, dx 와 dy 를 2로 제한하였다. 게이팅망의 은닉층의 노드 수는 25개로 하였고 출력층의 노드는 군집의 개수인 36개가 된다. 학습률은 0.1이고 관성항은 0.7로 하였다.

2. 인식결과

표 1에 있는 각 방법에 대한 인식실험 결과를 표2에 나타내었다. 실험에서 회전각 $\theta=(+10,-10,+5,-5)$ 로 4종류로 하고 이동의 경우 (dx,dy) 는 $(-2, -2),(+2, +2)$ 의 2가지로 하였다. 그러므로 7개의 영상($k=0,1, \dots, 6; K=6$)이 생성된다. 표2의 인식 결과는 동일한 데이터베이스에 대한 기존의 방법들에 비해 매우 높은 인식률을 보여준다.

Table 2. The results of output combination.

Methods	Recognition rate (%)
method I-1	98.30
method I-2	98.30
method II-1	98.15
method II-2	98.20
method II-3	98.25
method III-1	98.05
method III-2	98.05
method III-3	98.15

Table 3. The experimental results for assessing validity of dithering.

Methods	Recognition rate	Remark
LN-TN	96.20	
LN-TD	96.30	
LD-TN	97.95	
LD-TD I-1	98.30	method I-1

표 3에서는 디더링 방법의 효용성을 보여준다. 여기에서 'L'은 학습을 나타내며 'T'는 테스트를 나타낸다. 또한 'D'는 디더링, 'N'은 디더링을 하지않았다는 것을 나타낸다. 예를 들어 LD-TN은 학습에는 디더링을 하였고 테스트할 때에는 디더링을 하지않았다는 것을 뜻한다. 이 표를 살펴보면 학습과 테스트 단계 모두에서 디더링을 한 경우의 인식률이 보다 높음을 알 수 있다. 또한 테스트 단계에서 디더링을 하는 경우보다 학습단계에서 디더링을 하는 것이 인식률이 보다 높음을 알 수 있다.

VII. 결론

이 논문에서는 게이팅망과 모듈화된 MLP 구조를 가지는 신경망 기반 인식기와 디더링을 이용한 인식방법을 제안하였다. 초기 다중 군집의 구성은 SOM 망을 이용하였는데 각 군집의 반경에 따라 서로 중첩하게 하여 새로운 군집을 만들었다. 구성된 각 군집에 대해서 전문가망으로 동작하는 MLP를 구현하고 이들을 결합시키기 위해 게이팅 기능의 MLP를 구현하였다. 이때 각 MLP의 학습에 있어서는 학습 집합을 확장하

고자 회전과 이동의 2가지 디더링 과정을 거쳐 생성된 패턴들을 포함하였다. 인식에 있어서는 학습과 마찬가지로 입력 패턴에 대해서 회전과 이동 연산을 통한 디더링을 함으로써 발생된 새로운 패턴들을 발생시키고 이들에 대해 각각의 신뢰도값을 얻어 다양한 방법들을 적용하여 결합함으로써 보다 높은 인식률을 얻고자하였다. 인식실험 결과 제안한 인식방법이 높은 인식률을 보임을 알 수 있었다. 특히, 인식시스템의 디더링 효용성에 관한 실험에서는 학습 단계에서의 디더링이 테스트단계에서 보다 더 효과적이며 두 단계 모두에 디더링을 적용한 경우가 가장 좋은 성능을 보임을 알 수 있었다.

References

- [1] 류강수, 진성일, 모듈화된 신경회로망 중간층 출력의 재학습에 의한 필기체 숫자 인식, *정보과학회 논문지-B*, 제23권 9호, pp. 931-940, 1996.
- [2] S. N. Srihari, "Recognition of handwritten and machine printed text of postal address interpretation," *Pattern Recognition Letters*, vol. 14, pp. 291-302, 1993.
- [3] L. Xu, A. Krzyzak, and C. Y. Suen, "Method of combining multiple classifiers and their application to handwritten numeral recognition," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 22, pp. 418-435, 1992.
- [4] F. Kimura and M. Shridhar, "Handwritten numeral recognition based on multiple algorithms," *Pattern Recognition*, vol. 24, no. 10, pp. 976-983, 1991.
- [5] 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.
- [6] B. S. Kang, S. I. Chien, and K. T. Lim, "Large scale pattern recognition system using hierarchical neural network and false alarming nodes," *Proc. of ICTAI*, Newport, California, pp. 196-203, Nov. 1997.
- [7] T. Kohonen, "The self-organizing map," *Proc. of IEEE*, vol. 78, no. 9, pp. 1464-1480, 1990.
- [8] G. Srikanthan, S. W. Lam, and S. N. Srihari, "Gradient-based contour encoding for character recognition," *Pattern Recognition*, vol. 29, pp. 1147-1160, 1996.