인공신경망을 이용한 실시간 영문인쇄체 인식

심성균, 정원용
경남대학교 정보통신 공학과

The Real-time Printed Alphabets Recognition using Artificial Neural Networks

Soungkyun Sim, Wonyong Chong
Div. of Information and Communication Eng, KyungNam University
e-mail : ssk_hbej@hanmail.net, wychong@kyungnam.ac.kr

요 약

본 논문은 이미 판서된 오프라인(off-line) 영문 인쇄체를 실시간으로 인식하기 위해 인공신경망의 역전파(Backpropagation) 학습알고리즘을 적용하여 인식 시스템의 성능을 최대화하고, 양질의 특성벡터를 추출함으로서 실시간 처리가 가능하도록 처리시간을 단축시키는 것을 목적으로 하였다.

실시간 영상을 획득하고 처리하기 위한 Genesis 실시간 영상처리 보드와 이 보드를 제어하기 위한 MIL(Matrox Image Library)패키지를 이용하여 실시간 인식시스템을 구현하였고, 인공신경망의 기댓값을 ASCII형태로 변환시켜 출력벡터의 차수를 감소시키는 방법을 제시함으로서 패턴의 학습과 인식처리에 소요되는 시간, 그리고 인식시스템의 성능을 비교해 보았다.

Abstract

The goals of this papper are not only to maximize of performance but also to reduce the response time for the real-time printed alphabets recognition system using the backpropagation algorithm in the artificial neural network.

The Genesis board and MIL(Matrox Image Library) package were used to real-time acquisition, processing and display of images. Through this experiment proved the possibility of real-time recognition processing by comparing response times of the system and proposing the method to reduce of order of the output vectors.

I. 서 론

패턴의 시각적인 특성을 이용한 인식처리기로 잘 알려진 OCR(optical character Recognition) 문자인식에 관한 대부분의 연구들은 인간처럼 지능적이고 정확한 자동서류판독 시스템을 설계하는 것을 목적으로 하고 있다. 1950년 초 컴퓨터의 보급과 전자산업의 발달로 실시간 개념의 적용이 실체화되면서 인식시스템의 성능 개선뿐 아니라 보다 빠른 인식이 가능하도록 처리속도를 개선하는 데 많은 관심을 가지게 되었다.[1]

본 논문에서는 인공신경망을 학습시키는데 학습률변환 방식을 도입한 역전파(Backpropagation) 학습알고리즘의 사용과 출력벡터의 차수를 줄임으로서 학습시간과 인식처리시간을 단축시키고, 특성벡터의 간단한 검출방법을 제시함으로서 시스템의 성능을 최대화하는데 주안을 두어 오프라인(off-line)의 일종으로 영문 인쇄체의 실시간 인식을 목적으로 한다.

Genesis 실시간 영상처리 보드와 이를 제어하는 MIL(Matrox Image Library)패키지의 환경아래 구현한 실험을 통하여 기댓값을 ASCII형태로 변환하여 대응하는 출력벡터의 차수를 감소시키는 방법을 적용함으로서 인공신경망을 구성하는 뉴런(neuron)의 크기가 인식성능과 처리 속도에 미치는 영향에 대해서 비교 · 분석해 보았다.

II. 실시간 인식 시스템

일반적으로 실시간 문자인식은 문자의 기록 및 판서와 동시에 그 패턴의 본질적인 정보를 함유한 특성을 검출하고 대응되는 결과를 즉시 출력하는 온라인 인식을 말한다.[2] 하지만 본 논문에서는 인쇄된 영문체의 오프라인 인식에 적용한 실시간 개념은 그림 1과 같이 입력영상 I_t 가 Δt의 시간 간격으로 획득되어 소정의 영상처리과정을 거친 후, 관련된 결과 영상 O_T가 Δt의 간격으로 출력된다고 가정한다면, 각 입력영상에 관련된 결과영상이 출력되는 데 소요되는 인식 응답시간 ΔT는 $\Delta t \geq \Delta T$의 관계를 만족해야 한다.[3]

따라서 임의의 시간 t에서 영상 I_t가 입력되고, 그

연속영상 $I_{t+\Delta t}$가 입력되기 전에 전반적인 인식처리 과정이 완료되어 영상 I_t에 관련된 결과영상 $O_T \equiv O_{t+\Delta T}$가 출력되어야 한다.

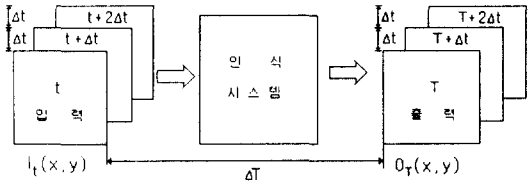


그림 1. 실시간 인식처리 개념
Fig 1. The concept of real-time recognition

III. 인쇄된 영문자 인식 시스템

개개의 독립적인 객체들로 형성된 집단 **P**가 하위집단 $\mathbf{S}_1, \cdots, \mathbf{S}_5$로 구성되었을 때, 임의의 하위집단으로부터 한 표본객체에 대한 인식처리를 수행하는 인식시스템의 전형적인 기본 구성을 그림 2에 표현한다.[2]

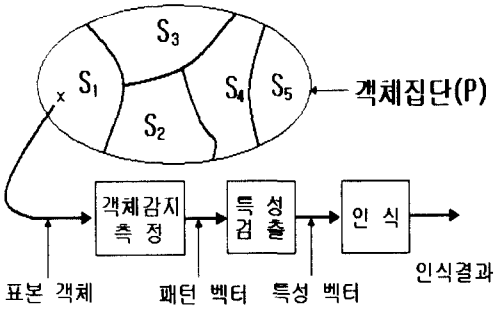


그림 2. 일반적인 인식시스템
Fig 2. General recognition system

인식하고자 하는 객체가 가진 속성들은 입력디바이스와 처리보드에 의해 감지되고 측정되어져 패턴 벡터 $\mathbf{M} = [m_1, m_2, \cdots, m_p]$으로 변환되고, 패턴벡터 M은 특성 검출기의 변환함수 φ에 의해 N ($N \ll P$)차원의 특성벡터 $\mathbf{x} = [x_1, \cdots, x_N] = \varphi(\mathbf{M})$로 특성화된다. 특성 벡터 **x**는 인식 시스템 또는 인식시스템 학습알고리즘의 입력으로 주어짐으로서 관련된 유일한 결과를 얻기 위해 연산되어진다.

IV. 특성벡터 검출

실질적으로 대부분 인식 시스템의 성능은 입력 패턴으로부터 검출되는 특성에 많이 의존하고 있는 만큼 특성의 검출과정은 패턴인식에서 아주 큰 비중을 차지하는 과정이기도 하다. 문자패턴들은 그 크기, 위치 그리고 모양 등의 많은 속성을 포함하고 있는데 이 속성들이 특성벡터로 변환된다. 이렇게 산출된 패턴벡터는 하나 이상의 다른 부류의 패턴들로부터 명백하게 분류될 수 있어야 하지만 반드시 인간에 의해 이해 가능한 의미를 포함할 필요는 없다.[2]

1. 거리 특성

거리 특성은 이진 영상에서 각각의 흰 픽셀들은 가장 인접하는 검은 픽셀과의 거리를 가진다는 성질을 이용한 것이다. [4]

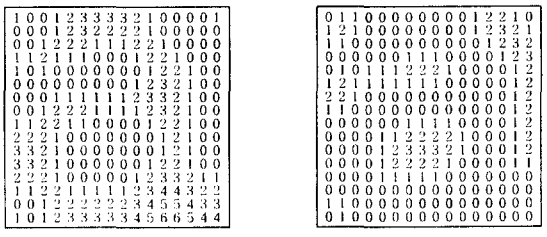
일반적인 거리함수 d는 Euclidean 공간에서 식 (1)과 같이N차원 패턴 $\mathbf{x} = (x_1, \cdots, x_N)^T$과 $\mathbf{y} = (y_1, \cdots, y_N)^T$의 사이에서 양의 실수 함수로 나타낼 수 있다.[2,4]

$$d(\mathbf{x}, \mathbf{y}) = \left[\sum_{i=0}^{N} |x_i - y_i|^n \right]^{\frac{1}{n}} \quad (1)$$

식 (2)는 위 식에서 n=1일 때, 멘헤턴(Manhattan) 거리 또는 검정 픽셀에 대해 4방위의 인접거리를 고려하는 city block 거리(d_4)를 나타낸 것이다.

$$d_4[(x_1, y_1), (x_2, y_2)] = |x_2 - x_1| + |y_2 - y_1| \quad (2)$$

그림 3은 식(2)를 이용한 16×16크기 문자 패턴 "a"와 반전된 "a"의 거리특성 결과를 나타낸 것이다.



a. 패턴 "a" b. 반전된 패턴 "a"

그림 3. 패턴 "a"의 City block거리 특성
Fig 3. City block distance feature of pattern "a"s

2. 정규화 특성

패턴 벡터로부터 가장 쉽게 얻을 수 있는 특성벡터 중 하나가 정규화 특성벡터이다. 이 정규화 특성벡터를 획득하기 위해서는 먼저, 입력패턴을 그물모양의 동일한 부영역으로 나눈 후, 각각의 부영역에 대하여 정보를 가진 픽셀의 총 수를 계산한다. 그리고 각 부영역에

서 계산된 검정픽셀의 총 수는 그 부영역의 면적으로 정규화 된다. 일반적으로 패턴의 본질적인 정보를 회손하지 않는 범위에서 정규화 특성벡터의 크기를 결정하게 된다.

그림 4는 문자패턴 "a"에 대해 각각 8×8, 16×16, 크기로 정규화 처리결과를 나타낸 것이다.

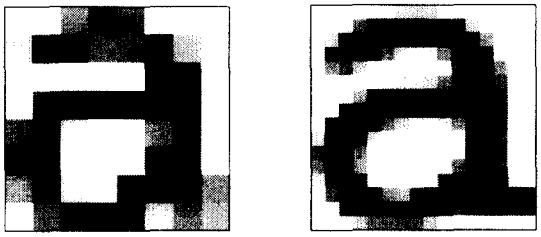


그림 4. 패턴 "a"의 정규화 특성
Fig 4. Normalized features of pattern "a"

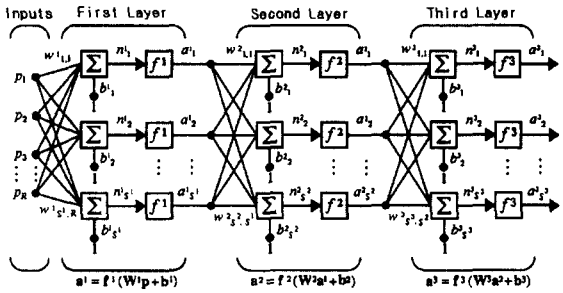


그림 5. 전형적인 3계층 인공신경망 구성도
Fig 5. Three layers artificial neural network

V. 인공신경망

패턴의 본질적인 정보를 함유하고있는 특성벡터를 분류기에 통과시킴으로 실질적인 분류가 이루어진다.

본 논문에 적용한 다층 인공신경망은 그림 5와 같이 입력층과 출력층 사이에 하나 이상의 은닉층(hidden layer)을 가지는 신경망의 형태로 은닉층의 개수가 증가할수록 신경망의 결정구역이 더욱 다양화된다.[5]

각 계층은 다수의 뉴런으로 구성되는 데, m번째 계층의 뉴런 수를 s_m이라 할 때 R차원의 입력벡터 $\mathbf{p}=(p_1,\cdots,p_R)$와 관련된 첫 번째 계층의 출력벡터 $\mathbf{a}=(a_1,\cdots,a_{s_m})$사이의 관계는 아래 식 (3)의 형태로 일반화 할 수 있다.

$$\mathbf{a}^{m+1} = \mathbf{f}^{m+1}(\mathbf{W}^{m+1}\mathbf{a}^{m} + \mathbf{b}^{m+1}) \quad (3)$$

각 계층의 전달함수 f는 그 계층에 따라 독립적으로 적용된다.

1. 역전파(Backpropagation) 학습알고리즘

다층 인공신경망을 학습시키는 역전파 학습알고리즘은 LMS(Least Mean Square)알고리즘의 비선형적인 확장형태로 성능지표면을 따라 수렴하는 최급강하 알고리즘이다. 역전파 학습알고리즘의 성능지표로 사용되는 MSE(Mean Square Error)를 반복횟수 k에 대한 근사식으로 표현하면 식 (4)와 같다.[5]

$$\hat{F}(\mathbf{x}) = (\mathbf{t}(k)-\mathbf{a}(k))^T(\mathbf{t}(k)-\mathbf{a}(k)) = \mathbf{e}^T(k)\,\mathbf{e}(k) \quad (4)$$

m번째 계층의 전달함수의 i번째 입력에 대한 $\hat{F}$의 변화성분의 s_i^m과 위 식을 근거로 인공신경망의 최급강하 알고리즘의 관계식을 표현하면 다음과 같다.

$$\mathbf{W}^m(k+1) = \mathbf{W}^m(k) - \alpha\,\mathbf{s}^m(\mathbf{a}^{m-1})^T \quad (5)$$

$$\mathbf{b}^m(k+1) = \mathbf{b}^m(k) - \alpha\,\mathbf{s}^m \quad (6)$$

α가 학습률을 나타낼 때, 반복횟수 k에 대한 강도와 바이어스의 변화가 없도록 학습이 진행되어야 한다.

임계값 s은 식 (7)에 표현한 것처럼 $m+1$번째 층으로부터 m층 방향으로 역전파되는 회기적 연산이 수행된다.

$$s^M \rightarrow s^{M-1} \rightarrow \cdots \rightarrow s^1 \rightarrow s^0 \quad (7)$$

식 (8)은 m+1계층의 임계값과 m계층의 임계값의 관계를 나타낸 것이고, 식 (9)는 출력층의 임계값 s^M를 일반화한 것이다.

$$\mathbf{s}^m = \dot{\mathbf{F}}^m(\mathbf{n}^m)(\mathbf{W}^{m+1})^T\mathbf{s}^{m+1} \quad (8)$$

$$\mathbf{s}^M = -2\,\dot{\mathbf{F}}^M(\mathbf{n}^M)(\mathbf{t}-\mathbf{a}) \quad (9)$$

VI. 실험 및 결과

본 논문의 실험에서 영문자 인쇄체의 실시간 인식을 위해 관련 패턴들을 획득하고 처리하는 하드웨어로 영상의 획득, 처리, 출력 등의 기능을 포함한 주보드(Main board)와 그와 직접적으로 인터페이스 함으로서 실시간으로서 신호처리능력을 극대화하는 처리보드(Processor board)로 구성된 Genesis 실시간 영상처리 보드를 사용하였고, 이 보드를 제어하기 위한 소프트웨어로 Visual C++ 또는 Visual Basic프로그램과 빠른 호환성을 가지며 영상의 표현 및 처리 등을 위한 최적화된 명령어들과 그에 상응하는 함수들을 포함한 MIL(Matrox Image Library)패키지를 사용하였고 버전 6.0의 Visual C++프로그램 환경에서 구현해 보았다.

실험에서 인공신경망의 전달함수로 미분 가능한 식 (10)에 표현한 시그모이드(sigmoid) 함수를 사용하였다.

시그모이드 전달함수는 특성상 0과 1의 값을 출력하기 위해서는 극한의 입력값을 요구하고, 성능지표의 원점은 성능지표면이 변화하지 않는 saddle point가 되거나 평탄한 모양일 가능성이 높다.

$$a = \frac{1}{1+e^{-n}} \qquad a: output, \quad n: input \quad (10)$$

이러한 이유 때문에 인공신경망의 원활한 학습이 진행되도록 출력의 범위를 0.001에서 0.99사이로 제안하고, 연결강도와 바이어스를 -0.5와 0.5사이의 랜덤한 값으로 초기화하였다.

표 1은 영문소문자 26개를 대상으로 출력벡터의 차수에 대한 출력 형태를 표현한 것이다. 즉 출력벡터의 차수가 5일 때는 ASCII형태로 출력하고, 반면 크기가 25일 때는 각 패턴의 순서에 해당하는 자리만 1(약 0.99) 값을 가지고 나머지는 0(약 0.001)의 값을 기대하였다.

표 1. 출력벡터 차수에 대한 기댓값 형태
Table 1. The forms for order of outputs

순서	문자 패턴	출력벡터의 차수 : 5	출력벡터의 차수 : 26
1	a	10000	10000000000000000000000000
2	b	01000	01000000000000000000000000
3	c	11000	00100000000000000000000000
⋮	⋮	⋮	⋮

표 2. 출력벡터의 크기에 대한 학습 / 인식시간
Table 2. Result for output vectors size

출력벡터 차수	5		10		26	
	학습	비학습	학습	비학습	학습	비학습
평균MSE ($\times 10^{-3}$)	2.0	10.1	3.1	9.6	0.77	40.0
학습시간(분)	약 38		약 50		약 54	
처리시간(초)	0.306		0.315		0.33	

표 2는 각 패턴들의 초기학습률을 지역최소점에 수렴하지 않도록 충분히 작은 0.125로 고정하고, 한 패턴에 대한 학습 횟수는 1200회로하고, 모든 패턴에 대한 학습횟수를 20회로 초기화 한 상태에서 160-160-N(N:출력벡터의 크기)로 구성된 인공신경망을 통한 영문인쇄체의 고립단어에 대한 학습·인식 처리시간을 측정 비교한 것이다.

평균 MSE는 출력벡터와 기댓값 사이의 MSE값으로 영소문자 26패턴에 대한 평균값을 나타낸 것이다.

표 2에서 출력벡터의 크기가 5일 때 학습시간과 인식처리시간이 약 38분과 0.3초로 가장 적고, 10차원의 ASCII형태와 일반적인 26차원 형태의 학습시간은 비슷하지만 인식처리 시간은 차수의 크기에 의존함을 알 수 있다. 반면에 평균 MSE값은 학습된 패턴에 대해서는 출력벡터가 26차원일 때 0.00077의 값으로 더 좋은 결과를 산출하였다. 실제적으로 입력되는 특성벡터의 크기에 비해 출력벡터의 크기가 큰 26차원 형태는 카메라의 조리개의 열림 또는 블러링(blurring) 현상과 같은 외부환경에 영향 받은 비학습 패턴의 경우에 MSE값은 증가하지만 오히려 더 좋은 인식성능을 발휘하였다.

VII. 결 론

본 논문에서는 Genesis 실시간 영상처리보드와 이를 제어하는 MIL패키지를 사용하여 실시간 영문 인쇄체 인식 시스템을 구현해 보았다.

실험을 통해서 실시간으로 패턴의 학습과 인식처리에 소요되는 시간은 출력벡터의 크기가 증가할수록 더 많은 학습시간과 인식처리시간이 요구됨을 확인하였고, 특성벡터의 크기에 비해 출력벡터의 크기가 클수록 외부환경의 영향을 받은 비학습 패턴에 효율적인 결과를 얻을 수 있었다.

비록 한 고립단어를 인식하는 데 약 0.3초로 MPEG-2에서 기준속도에는 미치지 못하지만 입력패턴과 병렬처리로 수행되는 실시간 영문자 인식시스템의 성능에는 아무런 문제가 없었다.

이러한 이유로 더욱 지능적이고 정확한 실시간 인식 시스템의 설계를 위해서는 보다 빠른 속도로 수렴하는 알고리즘의 고안과 적은 특성벡터의 차수를 사용하여 모든 패턴을 인식할 수 있는 양질의 특성벡터의 개발이 요구된다.

참고문헌

[1] S.Singh and A. Amin, "Neural Network Recognition of Hand-printed Chacters", Neural Comput & Applic 1999

[2] B.D.Ripley, "Pattern Recognition and Neural Networks", Cambridge,1996

[3] Phillip A.Lapiante and Alescnder D. Stoyenko, "Real-Time Imaging", IEEEPress 1996

[4] Il-Seok Oh, DhinY. Suen , "Distance features for neural network-based recognition of hendwritten characters", Springer-Velag 1998

[5] Martin T,Hagan, Howard B. Demuth and Mark Beale,"Neural Network Design" ,PWS Publishimg Company, 1996