

신경회로망을 이용한 한글 문자의 인식

°김 상 우 권 윤 호 최 중 호
 서울대학교 공과대학 제어계측공학과

The Recognition of Korean Characters by a Neural Network

Sang-Woo Kim Yun-Ho Jeon Chong-Ho Choi
 Dept. of Control & Instrumentation Eng.,
 Seoul National University

ABSTRACT

A study for the recognition of the Korean characters by a neural network is presented. To reduce the dimension of the input image data, DC components are extracted from each input image and used as input to the neural net. A multi-layer perceptron with one hidden layer was trained with back-error propagation training algorithm. Its performance is tested for 24×24 binary images of Korean characters and the results of several experiments are presented.

1. 서 론

최근에 들어와서 각광을 받고 있는 신경회로망은 '병렬분산 정보처리체제'로 요약할 수 있는데, 구조의 단순성에 비하여 응용분야가 다양하다는 데에 그 매력이 있다. 그러나 이제까지 고안된 신경회로망이 그 성능평가나 응용등에 적용된 경우 그 수준은 대부분 단순작업을 넘지 못하고 있다. 본 논문에서는 신경회로망의 최대 장점인 분류(classification) 기능을 인쇄체 한글의 인식에 이용하여 봄으로써 기존의 방법과는 다른 형태의 효과적인 한글 인식 신경회로망 알고리즘을 제시하려고 한다.

한글은 초성 19개, 중성 21개, 종성 27개의 비교적 적은 수의 기본 자모로 구성되어 있지만, 이들의 결합에 의하여 조합 문자를 이룬다는 특징을 지닌다. 따라서 가능한 문자 갯수는 총 11,172자에 이르지만[1], 이중에 실제로 널리 쓰이는 것은 1,500 - 2,000 자를 넘지 못하는 것으로 알려져 있다. 그럼에도 불구하고 단순자소의 어울림이 이같은 조합 문자를 만들기 때문에 많은 수의 유사 형태의 문자를 생성하여 한글 문자 인식을 어렵게 하는 요인으로 작용한다. 본 논문에서 사용한 문자 체계는 24×24 화소의 이진영상(binary image)으로서 그 글자수는 모두 2,350자이다. 인식을 위하여 사용된 신경회로망의 구조는 다중인식자(multi-layer perceptron)이며 이를 위하여 쓰인 학습방법은 BEP(back-error propagation) 알고리즘의 변형된 형태이다.

그런데 한글처럼 입력이 방대한 경우는 입력 화소를 그대로 신경회로망의 입력으로 사용하는 경우 단순한 다층구조와 BEP 알고리즘만으로는 인식가능한 문자의 갯수에 한계가 있음이 알려져 있다[2]. 이런 난점을 극복하기 위해서는 한글의 문자로부터

기본 자소를 추출(extraction)하는 작업이 필요하여 지는데 이것 자체가 그리 용이한 일이 아니다. 본 논문에서는 한글 영상을 그대로 입력으로 사용하지 않고 전처리과정을 두어서, 한글 영상을 적절한 다른 형태로 변형시킨 것을 신경회로망의 입력으로 사용하였다. 그리고 신경회로망의 구조는 중간층(hidden layer)을 1층을 가지도록 했으며, 입력층의 노드의 갯수는 96개, 중간층의 노드의 갯수는 100개, 출력층의 노드의 갯수는 모두 49개로 하였다.

입력 영상이 표준 위치에 있을 경우에 대하여 먼저 학습을 시킨 후, 문서로부터 스캐너를 통한 문자추출을 가상하여 입력 영상의 좌우이동(left-right shift)에 대해서도 학습을 시도하여 보았다.

본 논문에서는 제안된 신경회로망의 구조와 그 동작 특성, 적용된 학습 알고리즘과 전체 훈련과정, 그리고 실제 데이터를 이용하여 인식 실험한 결과를 보인다. 그럼으로써 한글 인식에 있어서 신경회로망의 적용가능성을 살펴 보기로 한다.

2. 신경회로망 및 전처리과정

신경회로망

인식에 사용한 신경회로망은 그림 1에 나타난 형태의 다중 인식자로서 한층의 중간층 및 입력층과 출력층으로 이루어져 있다.

각층의 노드의 연결은 전방향으로만 이루어지는데 입력층을 제외한 중간층과 출력층의 각 노드의 실제 출력값은

$$f(x) = [1 + e^{-(x+\theta)}]^{-1}$$

여기서

$$x = \sum \omega_i o_i$$

o_i : x에 연결된 이전 층의 실제 출력값

ω_i : o_i 와 x를 연결하는 가중치값

θ : x의 bias값

형태의 sigmoid 함수로 주어진다.

출력층의 노드의 갯수는 49개로서 그림 2에 주어져 있다. 출력을 코드화(decoding)하지 않고 이렇게 넓게 퍼뜨려 놓은 것

(encoding)은 각 층간의 연결선의 증가를 통하여 정보의 분산을 퍼하기 위한데 이 경우 혼련시켜야 할 가중치 갯수의 증대로 인한 혼련 시간의 증가가 문제로 남지만, 신경회로망이 포용할 수 있는 정보량이 효과적으로 늘어남을 확인할 수 있었다.

중간층 노드의 갯수는 100개를 설정하였다. 중간층 노드의 갯수에 관한 정확한 기준은 알려진 것이 없기 때문에 출력층 노드의 갯수의 이배수 정도를 중간층 노드의 갯수로 삼았다.

전처리과정

신경회로망의 입력층의 노드의 갯수는 모두 96개로서 이는 입력 영상인 24 × 24 화소의 이진 데이터를 바로 입력으로 사용하지 않고 적절한 전처리과정을 두었기 때문이다. 이렇게 한 이유는 크게 두가지를 들 수 있는데, 먼저 24 × 24 화소를 그대로 입력으로 사용하는 경우에는 입력 노드의 갯수가 모두 576개로서 이는 중간층 노드와 어울릴 경우에 신경회로망의 전체의 가중치의 갯수가 수만에서 수십만 개에 이르게 되어 결과적으로 막대한 혼련 시간을 요구하게 되는데 본 논문에서 도입한 전처리과정은 입력의 갯수를 96개로 줄임으로써 가중치의 갯수를 감소시키고 이것은 결국 혼련 시간을 줄일 수 있도록 하여 준다. 다음으로는 전처리과정을 거친 후의 입력자료가 원래의 한글 화소 영상보다 한글 자체가 지니는 특징을 나타내는데 좀 더 효과적일 수도 있다는 점을 들 수 있다. 즉 576개의 이진 화소는 많은 양의 중요도가 떨어지는 정보를 포함할 수 있는데 전처리과정을 통하여 이런 정보들을 어느 정도 걸러낼 수 있을 것으로 생각된다.

본 논문에서 시도한 전처리과정은 그림 3에 나타나 있다. 처음 24개의 데이터는 화소를 가로방향으로 합한 값이고, 다음 24개의 데이터는 화소를 세로방향으로 합한 값이다. 세번째 24개의 데이터는 +45도 방향으로 두축씩 화소를 더한 것을 2로 나눈 값이다. 마지막의 24개의 데이터는 +135도 방향으로 두축씩 화소를 더한 것을 2로 나눈 값이다.

이상의 96개의 데이터는 0, 45, 90, 135도 방향의 DC 성분을 뺀 후 샘플로서 실제로 한글을 2차원 FFT 하여 보면 그림 4에서 알 수 있듯이 DC 성분이 그 글자의 주요 특징을 이루고 있음을 알 수 있다. 이렇게 구한 DC 값들은 0에서 24까지의 정수값들로서 이것을 24로 나누어서 0에서 1까지의 값으로 변환시킨 것(normalized value)을 노드의 실제 입력값으로 삼았다.

3. 혼련 과정

신경회로망을 혼련시키기 위하여 사용된 학습 알고리즘은 Rumelhart 등이 제안한 BEP 알고리즘[3]의 변형으로서 이는 지도학습이론(supervised learning)의 대표적인 방법이다. BEP 알고리즘의 가중치를 변형시키는 식은

$$\Delta\omega_{ij}(n+1) = \eta\delta_{pj}\omega_{pi} + \alpha\Delta\omega_{ij}(n)$$

이다.

여기서 $\Delta\omega$ 는 가중치의 변화량이고, δ 는 역방향으로 전파된 오차의 분담값이며 ω 는 노드의 출력값이다. 다음 상수 η 는 학습률로서 이 값이 클수록 가중치의 변화량이 커지게 된다. 이 값이 커지게 되면 학습 속도가 증가되는 장점이 있지만 수렴하지

못하고 진동할 우려가 있으며 너무 작으면 진동은 막아지지만 학습 속도가 너무 느려지는 단점이 생긴다. 다음 상수 α 는 관성항(momentum)으로서 전번 단계의 계산에서 구해진 가중치의 변화 경향을 다음 단계에 반영시키는 역할을 하는데 이렇게 함으로써 수렴속도의 증가를 유발하게 된다[3].

본 논문에서는 학습률은 0.5, 관성항은 0.8을 사용하였다.

4. 결과 및 분석

신경회로망의 구현 및 시뮬레이션은 IBM PC-386 상에서 이루어졌다. 신경회로망의 규모때문에 많은 양의 시간이 소모되기는 했지만 몇 가지 경우에 있어서 바람직한 결과를 얻을 수 있었다. 전처리과정을 거치지 전의 한글 영상 화소는 명초채 24 × 24 의 이진 영상으로서 실험에 사용된 전체 글자수는 2,350자였다.

그림 5는 시뮬레이션이 이루어지고 있는 화면을 옮겨 놓은 것으로서 혼련중인 문자가 표시되고, 그 상태에서의 출력오차, 현재의 sweep 횟수, 그전 sweep 까지의 인식된 문자의 갯수, 그전 sweep 까지의 전체오차 등을 표시하도록 되어 있다.

처음 시도한 것은 2,350 자 전체에 대한 학습으로서 2,337 자 정도가 학습됨을 확인할 수 있었다.

다음은 2,350 자의 25%에 해당하는 588자를 임의로 선택한 후, 이것을 혼련 세트로 사용하여 신경회로망을 혼련시킨 후, 이렇게 혼련된 가중치 집합을 이용하여 전체 글자 집합에 대하여 어떻게 반응하는가를 살펴 보았는데 1,949자를 인식하였다. 이는 전체의 25%를 혼련에 이용하여 전체의 83% 정도를 인식해낸 것으로서 혼련된 신경회로망이 입력범위에 대하여 어느 정도 일반화(generalization)를 하고 있음을 알 수 있다.

588자의 문자 세트에 대해서는 중간층 노드의 갯수를 30개로 줄여서 학습을 시도하여 보았는데 학습속도에는 별 영향이 없었지만 2,350자 전체 세트에 대한 인식의 정도는 아래 표와 같다.

중간층 노드의 갯수	인식 정도
100	1,949 / 2,350
30	1,631 / 2,350

표에서 알 수 있듯이 중간층 노드의 갯수가 적어지면 신경회로망이 포함할 수 있는 정보량이 적어지고 따라서 전체 문자 세트에 대한 일반화도 제대로 이루어지고 있지 못함을 볼 수 있다.

이상에서는 입력이 매우 이상적인 형태라고 가정하였으나 궁극 목표인 스캐너에 의한 문자 인식을 가상하기 위하여 좌우로 이동된 문자에 대해서 인식정도를 실험하여 보았다. 혼련에 쓰인 것은 588자 및 그것들이 좌우로 각각 한 비트씩 이동된 후의 1,764자로서 혼련과정에서의 인식정도는 1,740자 정도였다. 이렇게 혼련된 신경회로망을 전체 2,350자가 좌우로 이동된 후인 7,050자에 대하여 적용시켰을 경우의 인식정도는 4,500자 정도로서 이는 전체의 25%를 혼련에 사용하여 전체의 65% 정도를 인식해 냈음을 의미한다.

좀더 악화된 환경을 고려하기 위하여 기본자체가 좌우로 각 두 비트씩 이동된 것까지 포함된 2,940자에 대하여 훈련시켜 보았는데 그 훈련과정은 그림 6과 같다.

2,940자 중 인식과정에서 인식한 최대의 글자수는 952자이고 그것을 고비로 하여 점차 인식정도가 떨어짐을 알 수 있는데, 이는 변환된 글자 세트 전체를 수용하기에는 신경회로망의 규모가 작기 때문인 것으로 생각되며 인식 정도가 떨어지는 현상은 앞으로 좀 더 규명하여야 할 것이다.

그리고 이러한 변환에 무관(invariant)하기 위해서는 다른 형태의 신경회로망을 고려해야 할 것으로 생각된다.

5. 토의 및 결론

본 논문에서는 한글자체가 지니는 여러 방향의 DC 성분을 바탕으로 신경회로망을 도입하여 입력문자를 전처리과정을 두어서 인식하는 한글 인식용 신경회로망을 구현하였다.

정위치에 있는 문자에 대해서는 DC 성분을 입력 데이터로 이용하여 좋은 성과를 얻을 수 있었지만, 좌우 변환이 심한 경우에 대하여는 그다지 만족스러운 결과를 얻지 못하였다. 그에 대한 대책으로는 우선 중간층 노드 갯수의 증가를 생각할 수 있지만 이 경우 훈련시간의 연장이라는 문제를 불러 일으킨다. 다음으로는 DC 성분만으로는 문자의 이동에 효과적으로 대응할 수 없다고 보고, 전처리과정에서 이것의에 또다른 것을 입력 데이터로 추출해 보는 것이다. 이를 위하여 고려하고 있는 것은 Hu 가 제안한 관성함수(momentum function)를 이용하는 방법 [4,5]으로서 적절히 활용한다면 상당히 좋은 효과를 얻을 수 있을 것으로 생각된다.

또 여기서는 거의 변형이 없는 다층인식자 구조를 사용하였는데 이것에 약간의 구조 변형을 가한다면 인식률을 증가시키는 데 도움을 줄 수 있을 것으로 생각된다.

그밖의 다른 방법으로는 한글 자체가 지니는 구조 특성을 인식에 이용하는 것으로서 [2] 전체 한글 글자 세트를 몇 개로 분할함으로써 훈련 시간의 감소를 가져 올 수 있고, 인식률의 향상을 추구할 수 있다는 면에서 유용한 방법으로 생각된다.

다층인식자는 입력이 24 x 24 인 한글 문자에 대해서 전처리과정을 통하면 좋은 인식 능력을 가지고 있으므로 한글 문자 인식에 유력한 신경회로망 구조라 할 수 있다.

참 고 문 헌

- [1] 이주근, "한글 문자의 인식에 관한 연구(IV)," 전자공학회지, 제9권, 1972, 9월, pp.25-32.
- [2] 고병기의 3 인, "경쟁하는 신경망 조직을 이용한 인쇄체 한글 문자의 인식," 제 1회 영상처리 및 이해에 관한 워크샵 발표 논문집, 1989, pp.127-134.
- [3] D. E. Rumelhart and J. L. McClelland, Parallel distributed processing, vol.1, vol.2, 1986, MIT Press.
- [4] A. Khotanzad and J. H. Lu, "Distortion invariant character recognition by a multi-layer perceptron and back-propagation learning," IEEE 2nd Int'l. Conf. on Neural Network, vol.1, pp.625-632, 1988.
- [5] S. A. Dudai, K. J. Breeding and R. B. Mcghee, "Aircraft identification by moment invariants," IEEE Trans. on Computers, vol. C-26, pp.39-46, 1977.

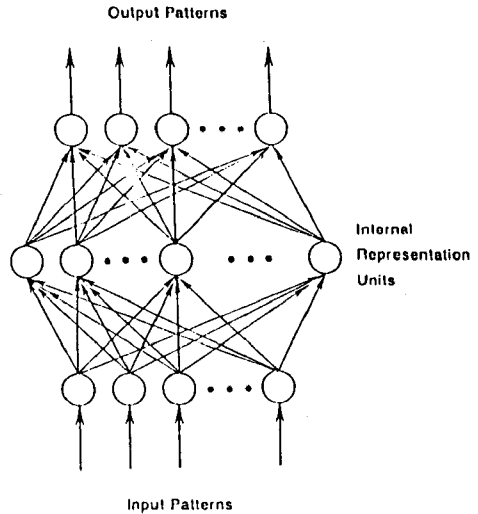


그림 1. Multi layer perceptron

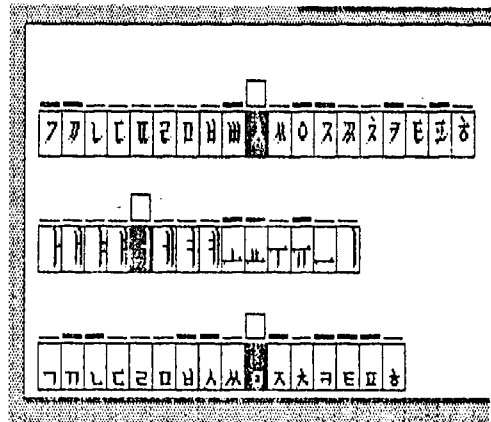


그림 2. 출력 노드

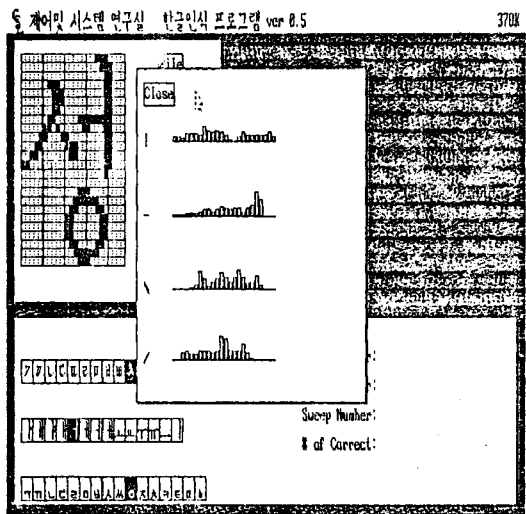


그림 3. 전처리과정 - 96 개의 DC 성분

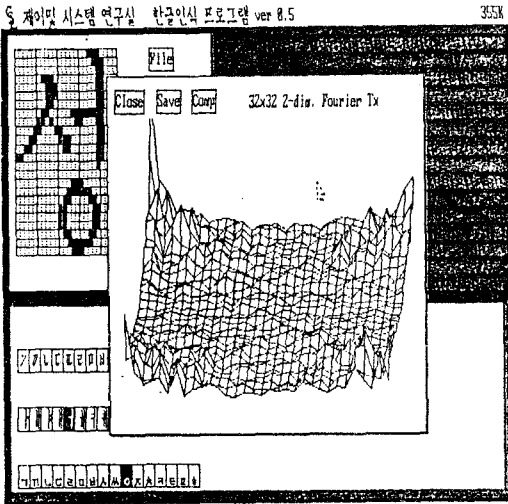


그림 4. '성'자에 대한 2차원 FFT

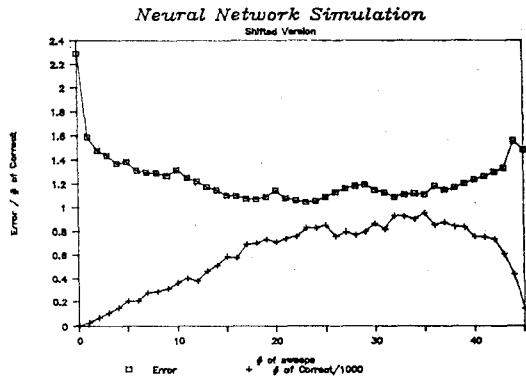


그림 6. 두 비트 좌우이동에 대한 시뮬레이션 과정

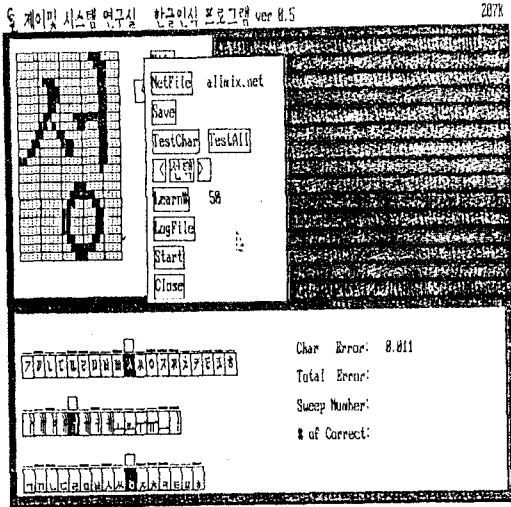


그림 5. 시뮬레이션 과정