

상호활성화모형과 다중해상도경로에 기반한 문자인식을 위한 계산 모형

박창수 김상기 방승양 김대진
포항공과대학교 컴퓨터공학과
{bylkuse, ihistory, sybang, dkim}@postech.ac.kr

A computational model for character recognition based on multi-resolution channels and IAM

Chang-Su Park Sang-Ki Kim Sung-Yang Bang Dae-Jin Kim
Dept. of Computer Science & Engineering, POSTECH

요 약

여기서 우리는 한국어의 글자를 인식하기 위한 계산모형을 제시한다. 이 모형은 상호활성화 모형에 기반하고 있으나, 입력처리부분이 개선되었다. 우리는 기존의 상호활성화모형에서 모든 입력이 규격화된 형태라는 가정을 배제하였다. 또한 우리는 성능향상을 위하여 다중해상도경로 학습을 결합시켰다. 우리는 제안된 모형을 구현하고, 그것을 실제 자료에 적용해 보았다. 우리는 모의실험결과를 통하여 제안된 모형이 실용적인 의미를 갖고 있으며 다중해상도 경로가 실제적으로 인식 성능 향상에 도움을 주고 있음을 확인할 수 있었다.

1. 서론

영어의 시각적 단어인식 모형으로써 상호활성화모형(Interactive Activation Model)이 있다[1]. 이 모형은 원래 단어우월효과(Word Superiority Effect)를 설명하기 위하여 개발되었다. 이 모형은 입력영상에서 오는 정보에 의한 상향식 작용과 이미 학습된 단어의 정보에 의한 하향식 작용을 통해 알파벳을 인식한다. 이 모형의 핵심은 간단하면서도 단어와 알파벳 인식의 여러 가지 측면을 설명함으로써 지금까지 제안된 단어인식의 모형 중에서 가장 확고한 자리를 갖고 있다는 것이다 [2]. 우리는 일찍이 이 모형을 기반으로 한글인식모형을 개발한 바 있다[3]. 개발된 이 한글용 상호활성화 모형은 한글의 글자우월효과(Character Superiority Effect)를 잘 설명할 뿐만 아니라, 실제적인 한글문자인식 시스템에 사용될 수 있는 가능성을 보였다.

물론 이 모형은 위에서 언급했듯이 장점도 많지만, 보다 완전한 사람의 시각적 문자인식 계산 모형이 되기 위해서는 여러 가지 보완이 필요하다[2]. 그러나 여기서 우리의 목적은 사람에 보다 가까운 계산모형을 만드는 것이 아니라 실제 세계에서 유의한 문자인식 계산모형을 만드는 것이다. 이런 관점에서 볼 때, 가장 문제가 되는 것은 입력부분이다. 원래의 상호활성화모형에서 입력 문자는 모두 규격화되어 있고, 그 규격화된 모양에서 특정위치의 선분이 있는지 없는지에 대한 정보가 전달된다. 우리가 전에 제안한 한글용 상호활성화모형에서는 입력신호가 실수값을 가질 수 있게 했고, 입력 글자의 한글 유형을 모르기 때문에 가능한 모든 유형을 가정하여 동시에 처리하게 했다하더라도 입력신호는 여전히 규격화된 선분이었고 그들의 조합으로 자소

가 결정되었다. 글자의 규격화는 필기체가 아니라 인쇄체 문자를 생각하더라도 너무 비현실적이다.

위에서 언급했듯이 우리는 사람의 시각정보처리 원리를 활용하되 실용적인 가치가 있는 글자인식 계산모형을 개발하는 것이 목표이다. 따라서 기존의 상호활성화 모형의 입력부분을 수정하여 입력문자영상의 형태에 대한 제약을 완화하였다. 또한 우리의 모형은 한글용 상호활성화 모형을 기본으로 하면서 여기에 Neocognitron과 다중해상도경로의 개념을 결합하게 된다.

2절에서는 제안된 모형의 기초가 될 한글용 상호활성화 모형을 요약한다. 확장된 입력처리 부분에 관한 설명은 3절에서 한다. 4절에서는 제안된 모형을 써서 실행한 인쇄체 한글인식 초기실험에 대한 보고를 하고, 5절에서는 결론을 제시한다.

2. 한글용 상호활성화모형

여기서는 영어에 적용되었던 상호활성화모형과 한글에 적용가능하도록 수정된 상호활성화모형에 대해서 간단하게 살펴본다. 영어에 적용된 상호활성화모형의 기본적인 구조는 그림 1)의 a)와 같다. 특징계층은 입력되는 시각 자극을 감지하는 부분이고 문자계층은 단어의 특정 위치에서 각 알파벳의 활성화 정도를 나타내는 부분이며 단어계층은 이미 습득된 단어정보와 주어진 시각 자극에 의한 각 단어들의 활성화 정도를 나타내는 부분이다.

각 계층안에서는 여러 노드들이 있고, 각 계층의 노드들은 같은 계층의 다른 노드들 내지는 다른 계층의 노드들과 상호작용을 한다. 노드사이의 상호작용에는 노드의 활성화 정도를 크게하는 자극작용과 반대로 활성

화 정도를 감소시키는 억제작용이 있다. 이렇게 계층들 사이와 각 계층안에서의 상호작용이 반복적으로 이루어지면 결국 문자계층이나 단어계층에서 그 활성화 값이 다른 노드의 활성화 값보다 월등히 크게되는 노드가 나타나게 될 것이다. 우리는 그러한 노드에 해당하는 값으로 주어진 자극을 인지하게 되는 것이다.

이렇게 영어의 단어 인식에 맞도록 제안된 모형을 한글에 적용하기 위해서는 무엇보다 각 계층에 포함된 각 노드가 갖는 의미를 한글에 맞게 다시 정의할 필요가 있다. 영어의 가장 기본적인 인식단위는 알파벳인 반면, 한글에서는 자음과 모음(자소)이다. 그리고 바로 다음 단계의 구성요소는 영어의 경우 단어가 되지만, 한글의 경우는 조금 다르게 글자가 된다. 이러한 것들을 고려해 볼 때, 영어에서의 단어계층은 한글에서는 글자계층에 대응 되어야한다. 이러한 변경을 통해 한글에 적용된 상호활성화 모형의 기본구조가 그림 1)의 b)에 나와 있다. 또한 한글에서는 글자가 가질 수 있는 6가지 유형이 있는데, 이들 각각을 동시에 동작시키기 위한 구조적인 변경이 필요하다. 그러나 모형의 전체적인 작동방법은 한글에 적용된다고 하더라도 크게 변하지 않는다.

3. 입력처리의 확장

3.1 기본방향

앞서 설명한 상호활성화모형의 경우 영어와 한글 모두 입력을 실제 영상의 형태가 아닌 규격화된 선분들의 유무를 이용하고 있으므로 실제적인 인식문제에 적용하기에 적합하지 않다. 그래서 우리는 앞에서 살펴본 상호활성화모형의 입력부분을 다음과 같이 보완하였다. 우선 기존의 모형에서 규격화된 형태로 제한되었던 입력 문자영상에 아무가정도 두지 않았다. 그래서 기존의 모형에서는 인식에 사용되는 특징의 추출이나 감지가 수동적으로 이루어진 반면, 보완된 모형에서는 특징의 추출과 감지가 자동적으로 이루어진다. 이것을 위해서 우리는 신경망을 이용하기로 하였다.

또한 우리는 인식을 좀 더 효과적으로 수행하기 위해서 여러 다른 해상도에서 추출된 특징을 사용하기로 하였다. 저해상도의 특징은 입력의 전체적인 모형을 추출하는데 사용되고, 그 정보가 빠르게 전달된다. 우리는 그 정보를 인식처리의 상위단계 즉, 단어계층에 전달하고

그것을 후보 문자집합을 줄이는데 사용한다. 고해상도의 특징은 자소를 정확하게 확인하기 위해서 입력에서 좀 더 자세한 정보를 추출하게 하는데 사용된다. 어느 경우이든 입력으로부터 정보를 추출하는 것은 마찬가지로지만, 상호활성화모형의 입장에서 보면 저해상도 경로는 하향식처리에 도움을 주고, 고해상도 경로는 상향식처리에 도움을 준다고 할 수 있다.

제안된 모형에서 우리는 저해상도, 중해상도, 고해상도의 세가지 경로를 사용한다. 다음에서 각각의 경로에 사용되는 특징과 구조에 대해서 설명한다.

3.2 저해상도 경로

저해상도 경로의 역할에 대해서는 여러 가지로 생각해 볼 수 있으나 우리는 우선 글자의 대략적인 모양으로 인식의 대상이 되는 글자의 후보 범위를 축소시키는 것으로 하였다. 우리는 저해상도에서 추출된 특징을 이용하여 전체 글자를 미리 군집화해 냈다가, 입력이 들어오면 입력과 유사한 군집에 속한 글자들은 자극 신호를 받게 된다.

저해상도 경로의 특징을 얻기위해서 입력영상을 겹침없이 5x 5의 윈도우로 나누고 해당 윈도우안에서 1의 값을 갖는 화소의 비율을 얻어낸다. 이것은 입력이 30x 30의 해상도를 갖는 이미지라면 36차원의 특징벡터가 추출됨을 의미한다. 우리는 앞서 설명한 상호활성화모형의 글자계층에 포함되는 모든 글자를 자기조직화지도(SOM)를 이용해 군집화하였다.

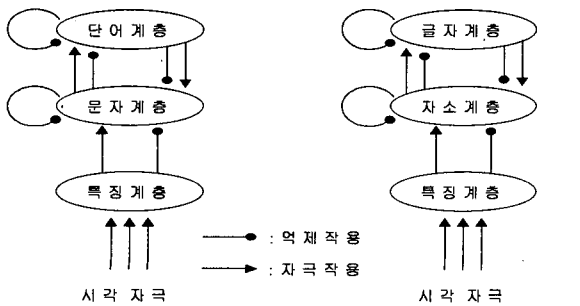
3.3 중해상도 경로

우리는 중해상도 경로를 입력된 글자의 유형을 결정하는데 도움을 주기위해 사용하였다. 입력된 글자의 유형은 입력된 글자가 무엇인지 인식하기 전에는 정확히 알 수 있는 방법이 없다. 그러므로 우리는 단지 입력된 글자의 대략적인 모양을 이용해서 글자의 유형을 추측해 볼 수 있을 뿐이다.

중해상도 경로의 특징을 추출하기 위해서 우리는 입력 영상에 겹침을 허용하면서 5x 5 크기의 윈도우를 띄우고, 각각의 윈도우로부터 저해상도와 같은 방법으로 특징을 추출하게 된다. 그리고 이 경로의 학습을 위해 다중계층퍼셉트론(MLP) 신경망을 사용한다. 신경망의 출력노드가 갖는 의미는 글자의 각 유형이 된다. 훈련자료를 이용해서 신경망을 미리 학습시킨후, 입력이 주어지면, 이를 이용해 신경망의 출력노드의 출력값을 계산하고, 각 출력 노드에 해당하는 유형의 글자들에 해당 노드의 출력값에 비례하게 자극 신호를 보내게 된다.

3.4 고해상도 경로

고해상도 경로에서 얻어지는 정보는 글자의 각 위치에 해당하는 자소의 종류이다. 고해상도 경로는 실제적으로 글자의 각 유형에 대응하는 6개의 분리된 경로로 이루어지고, 이들은 병렬적으로 동작한다. 각각의 경로는 입력이 자신에 해당하는 유형이라고 가정하고 입력을 처리한다. 또한 한글에서 각 글자는 유형별로 2~4개의 자소로 이루어지므로, 각 유형의 경로내에서도 2~4개의 자소인식경로가 존재한다. 결국 고해상도 경로내에는 병렬적으로 작동하는 17개의 각기 다른 인식기가 존재하는 것이다.



a) 영어에서의 상호활성화모형 b) 한글에서의 상호활성화모형

그림 1) 상호활성화모형의 기본 구조

우리는 각각의 자소 인식기에 Neocognitron형태의 피드포워드 신경망을 사용한다. 각각의 자소인식기의 구조나 동작은 입력의 숫자나 출력의 숫자가 다르다는 것을 제외하고는 모두 동일하다. 각각의 자소인식기는 독립적으로 작동하기 때문에 우리는 이들을 개별적으로 학습시킬수 있다.

각 자소인식기는 두 개의 은닉계층으로 구성되고, 각 계층은 단순세포층과 복합세포층의 쌍으로 구성된다. 단순세포층의 노드는 인간시각영역의 단순세포와 비슷한 기능을 한다. 즉, 단순세포계층의 노드는 주어진 수용영역안에서 미리 정해진 특징을 감지한다. 그리고 복합세포층의 노드는 그 전 계층의 수용영역안에서 하나의 노드라도 발화하면 자신도 발화함으로써 어느 정도의 왜곡을 흡수하게 된다. 그러므로 단순세포층에서 복합세포층으로 가면서 차원의 감소가 이루어진다.

각 자소 인식기의 입력계층의 크기는 해당 위치에 나타나는 모든 자소를 모두 포함할 수 있는 정도의 크기로 결정된다. 각각의 은닉계층은 여러개의 플레인(같은 특징을 감지하는 노드들의 집합)으로 구성된다. 첫 번째 단순세포층의 플레인들은 입력으로부터 미리 정해진 미세특징을 감지한다. 각 미세특징은 특정 방향의 짧은 선성분이다. 두 번째 단순세포층의 플레인은 전단계에서 감지된 미세특징을 조합해서 이루질 수 있는 좀더 큰 특징을 감지한다. 출력계층의 노드들은 각각의 자소 인식기에 해당하는 위치에 나타날 수 있는 자소의 집합에 대응된다.

3. 5 모형의 동작

앞서 설명한대로 세 개의 다른 경로로부터의 출력신호를 통합하기 위한 기본적인 구조는 상호활성화모형이다. 원래의 상호활성화모형에서 입력으로부터 특징을 감지하는 특징계층을 고해상도 경로가 대신하게 된다. 저해상도 경로와 중해상도 경로로부터의 출력은 하향식 처리와 관련되기 때문에 문자계층으로 바로 입력된다.

4. 모의 실험

우리는 우리가 제안한 모형을 구현하고 실제자료를 사용해서 모의실험을 행하였다. 우리가 행한 모의실험은 단지 제안된 모형의 가능성을 검증하기 위한 것이었으므로 이것은 예비적인 모의실험의 성격을 띠고 있다. 제안된 모형의 실제 가능성을 분석하고 장· 단점을 찾내기 위해서는 다양한 자료를 사용한 더 많은 모의실험이 필요하다.

이 모의실험에서 우리는 한글의 상용문자 2,350자중에서 가장 자주 쓰이는 600자만을 대상으로 하였다. 또한 우리는 그 중에서 유형3과 6의 글자를 제외시켰다. 이러한 대상글자의 감소는 적당한 시간내에 모의실험을 행하기 위해서 필요한 것이었다. 그리고 시스템이 성공적으로 동작하기 위한 여러 매개변수들의 조정을 위해 한 폰트의 인쇄체 문자를 사용하기로 결정하였다.

입력은 30×30크기의 이진 화소 영상이고, 세 개의 다른 경로를 학습시키기 위해서 잡음이 포함되지 않은 문자집합과 5%의 무작위 잡음을 첨가된 5개의 문자집합을 준비하였다. 또한 평가를 위해서 5%의 무작위잡음이 첨가된 또 다른 5개의 문자집합을 준비하였다. 모의실험의 결과가 표 1)에 있다.

표 1) 모의실험 결과

경로	훈련자료	평가자료
고해상도 경로	97.8%	93.4%
저· 고해상도 경로	98.6%	95.1%
중· 고해상도 경로	98.9%	98.0%
저· 중· 고해상도 경로	99.2%	98.3%

5. 결론

여기서 제안된 모형은 인지모형에 기반한 실제적인 인식시스템의 개발을 위한 시도이다. 그러기 위해서 우리는 한글형 상호활성화모형에서 출발하였다. 그리고 규격화된 형태가 아닌 실제 영상을 입력으로 받아들이기 위해 입력부분을 개선하였다. 더 나아가 인간 시각 처리의 다중해상도 경로 학설의 잇점을 취하기 위하여 세 가지 다른 해상도 경로를 결합하였다. 각 경로는 여러 가지 신경망 구조로 구현되었다.

제안된 모형의 성능과 가능성을 검증하기 위하여 실제적인 한글자료를 이용하여 모의실험을 행하였다. 모의실험을 통하여 우리는 시스템이 예상했던대로 동작함을 확인했다. 우리는 저· 중해상도 경로가 고해상도 경로를 보충해주면서 시스템의 전체적인 성능을 향상시키는데 도움을 준다는 것을 확인하였다.

이것은 단지 예비적인 모의실험결과이다. 앞으로 다양한 자료를 이용한 더 많은 모의실험과 분석이 행해져야 할 것이다.

6. 참고문헌

- [1] McClelland J. L. & Rumelhart D.E., An Interactive Activation Model of Context Effects in Letter Perception: Part 1. An Account of Basic Findings, *Psychological Review*, 88(5):375-407, 1981.
- [2] Glyn W. Humphreys & Vicki Bruce, *Visual Cognition:Computational, Experimental, and Neuropsychological Perspectives*, Lawrence Erlbaum Associates Ltd., Publishers, pages 241-247, 1989.
- [3] Park C. S. & Bang S. Y., Modeling Character Superiority Effect in Korean Characters by Using IAM, *Proceedings of Biologically Motivated Computer Vision*, 1:316-325, 2000.
- [4] Kunihiro Fukushima, Sei Miyake & Takayuki Ito, Neocognitron:A Neural Network Model for a Mechanism of Visual Pattern Recognition, *IEEE Transaction on systems, man and cybernetics*, 13(5):826-834, 1983.
- [5] Toshio Inui, A Model of Human Visual Memory:Data Compression with Multi-resolution, *Scandinavian Conference on Image Analysis*, 6:325-332, 1989.

Acknowledgement

본 연구는 BK21 사업을 통하여 포항공과대학교 전자·컴퓨터공학부에 주어진 교육부의 재정지원과 과기부가 후원하고 있는 뇌과학연구개발사업의 재정지원을 통해 이루어진 것입니다.