공학석사학위논문

하이퍼망 분자컴퓨팅 기반 단어 재인 시뮬레이션

Hypernetwork Based Molecular Computational Simulation of Word Recognition

2006년 8월

서울대학교 대학원
협동과정 인지과학 전공
강 윤 정

하이퍼망 분자컴퓨팅 기반 단어 재인 시뮬레이션

Hypernetwork Based Molecular Computational Simulation of Word Recognition

지도교수 장 병 탁

이 논문을 공학석사 학위논문으로 제출함
2006년 4월

서울대학교 대학원

협동과정 인지과학 전공

강 윤 정

강윤정의 공학석사 학위논문을 인준함
2006년 6월

심사위원장 고 성 룡 (印)
부 위 원 장 장 병 탁 (印)
심 사 위 원 신 효 필 (印)

초 록

본 논문은 단어 완성 문제를 해결함으로써 단어가 두뇌에 저장되는 방식과 인출되는 방식을 설명하고자 하였다. 단어의 저장 방식과 인출 방식에 대한 기존의 연구는 주로 단어의 리스트를 구성한 후 이를 탐색하는 방법으로 이루어진다. 그러나 이러한 방식은 인간의 두뇌라는 제한된 저장 공간에서 이루어지는 기억 인출 및 저장의 효율성과 그 속도를 유연하고 적절하게 반영하지 못하는 측면이 존재한다. 이와 관련하여 인간의 기억 기제를 좀더 충실히 모사하려는 연구가 한 흐름을 형성하고 있는데, 이들의 공통된 특징은 단어 철자를 분산 저장함으로써 그 기제를 이해하려 한다는 점이다.

본 논문에서는 이러한 철자의 분산저장 과정을 hypernetwork 방법론을 통해서 설명함으로써 인간의 단어 표상 과정을 설명한다. Hypernetwork 방법론은 초상호작용주의적 계산 모델로서, 대용량의 higher order 기억 단위를 산정하고 그 기반에서 데이터로부터 다양하고 초병렬적인 연결방식을 이루어낸다.

실험은 알파벳 조합의 저장 단위를 order로 하여 hypernetwork를 통해 단어 완성과제를 수행하였다. 실험 1에서는 분산저장 과정을 통해 학습한 단어에 대한 메모리를 가지고 손상을 입힌 단어에 대한 복원율을 살펴 보았고, 실험 2에서는 학습이 이루어지는 동안에 사용되지 않은 단어의 복원율을 살펴 보았다. 실험 3에서는 손상된 위치를 다양하게 하면서 복원율이 어떻게 바뀌는가를 살펴 보았다.

실험 1과 2에서는 order에 따라 복원율을 달리 보이며 general rule과

specific rule의 적용을 통한 인간의 단어 복원 능력을 모사한다. 실험 3에서는 접두어의 중요성을 설명하는 결과를 얻음으로써 인간의 단어인식에서의 손상위치에 따른 심리학 실험을 잘 설명하고 있다. 기존의 인간의 단어 인출의 연구들과 비교해 봤을 때, 여러 order에서의 hypernetwork를 통한 단어 완성 실험은 hypernetwork가 기존 심리학 연구에서 말하는 친숙성 효과를 잘 반영함을 보여주고, 손상된 위치에 따른 성능의 변화는 기존 심리학 연구에서의 위치 효과를 잘 설명한다.

주제어: 분자컴퓨팅, hypernetwork, 단어 재인, 단어 완성(word fragment completion)

학 번: 2004 - 20106

목 차

그 림 목 차

1. 서 론

1.1 연구 목표 및 범위

본 연구는 단어 재인 시뮬레이션을 분자 컴퓨팅 기반한 hypernetwork에서 구현하였다. 인간은 매우 많은 단어를 알고 있다. 그리고 이 단어들은 학습을 통해서 습득하게 된다. 성인의 경우, 영어권에서는 평균5만개의 단어를 습득한다고 한다.[1] 이것이 정확한 수치는 아닐 수도 있지만, 평균 5만개보다 더 많은 단어를 기억하는 일이 가능한 것은 틀림없다. 우리 머리 속에 이렇게 많은 수의 단어가 기억되어 있다는 사실은 Mental lexicon의 연구를 흥미 있도록 해주는 계기가 된다. 5만개의 전화번호나 지명 이름을 외우는 일과 비교해본다면 더더욱 그렇다.

이렇게 많은 단어가 Mental lexicon에 무작위로 쌓여있지는 않을 것이다. 그 이유는 첫 번째, 위에서 언급한대로 그 수가 매우 많다는 것이다. 둘째, 단어는 우리가 필요할 때 매우 빠르기 인출될 수 있다. 심리학자들은 정보가 구조를 가지는 경우 인간의 기억이 유연할 뿐 아니라 확장이 가능하다는 것을 보여주었다.[2] 이것은 Mental lexicon이 정밀하게 조직된 어휘집이라는 것을 의미한다. 심성 어휘집은 단어를

[1] [Anderson & Freebody, (1981)]
[2] [Gregg, (1986); Gathercole & Baddeley, (1993)]

대용량으로 저장할 수 있어야 하고, 인출의 속도가 빨라야 한다는 두 가지 요구가 있다.

Mental lexicon에서 단어를 인출하는 과정이 의식적으로 자각되는 현상이 아니며, 또한 매우 신속하게 이루어지는 것은 우리 안에 자동화된 처리 기제가 존재하기 때문이라고 생각할 수 있다. 자동화된 처리가 일어나기 위해서는 훈련과 학습이 필요하며, 인간은 읽고 쓰는 과정을 통해서 이러한 훈련과 학습이 이루어진다. 학습은 내용 기반 메모리인 hypernetwork에서도 중요한 요소이다. 인간도 학습과 추론을 통해서 명확한 규칙만으로 설명되지 않는 여러 문제들을 해결할 수 있는 능력을 갖추게 된다. 본 논문에서는 hypernetwork에서 영어 단어를 분산 저장하고 인출시에 복원 과제를 주고 복원 현상을 실험하였다. 학습한 단어의 패턴은 network안에 분산 저장됨으로써 문자(grapheme)를 기반으로 단어 재인이 이루어지며, 인출 단서의 종류에 따라서 인출 정확도에 영향을 줄 수 있다. 더불어 단어를 분산 저장 시 저장 단위의 크기에 따른 평균 단어 복원률을 구해서, 저장 단위와 복원률간의 관계를 논의할 것이다.

1.2 논문의 구성

본 논문에서는 먼저 분자 컴퓨팅 기반 hypernetwork에 대한 전반적인 이론을 살펴보고, 응용 실험으로 단어 재인을 위한

hypernetwork 모델을 구현하고 단어 완성(Word fragment completion)을 수행하였다. 2장에서 분자 컴퓨팅과 hypernetwork이론에 대한 소개와 단어를 기억하는 현상과 더불어 인간의 기억 이론의 하나인 구성적 기억 이론과 hypernetwork의 기본 체계인 병렬 분산 처리와의 연결을 설명한다. 그리고 인간의 머리 속 사전인 심성 어휘집(Mental lexicon)에 대한 설명을 통해서, hypernetwork에 구현된 단어 사전에서 인출 실험 시에 예상되는 심리학적 효과에 대해서 알아보고자 한다. 3장에서는 hypernetwork에서의 단어 재인 시뮬레이션에 대한 구체적인 구현 설명과 알고리즘을 기술하였고, 4장에서는 시행한 시뮬레이션에 사용된 데이터에 대한 설명과 실험 조건에 따른 시뮬레이션들의 결과를 정리하였다. 5장에는 시뮬레이션에 대한 요약과 결론을 기술하고, 향후 과제를 논의하는 것으로 본 논문을 마무리하고자 한다.

2. 관련 연구

2.1 분자컴퓨팅의 정보 처리

인간의 마음에 대한 본질은 무엇인가에 대한 탐구는 지난 수세기 동안 중요한 철학적, 심리학적 주제가 되어왔다. 지난 세기 컴퓨터의 발전은 인간의 마음을 보는 관점에 중요한 틀을 제공하였고, 인간의 마음이 외부 환경에서 오는 자극을 선택하고, 조작해서 재구성하는 기관이라고 생각하게 되었다. 인간의 마음을 정보 처리를 하는 기관으로 생각하고, 인간의 행위를 이 정보 처리의 결과라고 보는 것이 행동 주의 심리학 이후에 자리 잡은 인지 주의 심리학이다. 인지 과학은 이러한 인지 주의 또는 정보 처리 패러다임을 주요한 학문적 틀로 구현하고 있다.

컴퓨터의 발전이 인지 주의의 중요한 관점을 제공하였고, 극단적으로 인간의 마음을 컴퓨터로 비유하기도 하지만, 실제로 마음이 구현되는 인간의 뇌와 컴퓨터는 그 구조상 매우 다르다. 비록, 퍼트남 같은 철학자들은 다중 실현(multiple realizability)의 가능성을 통해서 물리적으로 다른 특성에도 동일한 마음을 구현할 수 있다고 하였고, 이것은 모델링을 하려는 인지 과학자들에게 실제로 뇌를 구현하지 않아도 마음을 모델링 할 수 있다는 희망을 주기도 하였다. 하지만 사실 뇌와 컴퓨터의 정보 처리 기제가 그 물리적인 구조에서 오는 한계를 띄어 넘지 못하고 있다는 점에서 실질적인 대안은 되지 못하고 있다. 실제로 뇌와

컴퓨터의 작동방식은 근본적으로 다른 점이 많다. 뇌는 수많은 신경 뉴런들이 화학적 물질들의 통제아래서 상호간의 매우 많은 연결들을 유지하고 있으며, 가소성(plasticity)이 있기 때문에 손상 시 뇌의 다른 부분으로부터 보완이 되거나 자연 복구가 가능하다. 하지만 컴퓨터는 중앙 처리장치와 그 주변 장치들로 철저히 분업화 되어 있어서, 손상되었을 때 서로 보완할 수 있는 능력이 없다. 본 절에서는 기존의 실리콘 컴퓨터가 제공하지 못하는 물리적 한계인 초 병렬적 상호 연결을 가능하게 하는 분자 컴퓨팅을 이야기할 것이다.

DNA분자를 가지고 NP-complete한 문제를 풀었던 Adleman[3] 의 연구 이후로 DNA분자들이 염기 서열간의 정합 반응과 복제가 초 병렬적으로 이루어지고, 그 기제가 화학적 특성을 가진다는 점에 주목하고 있다. DNA 염기서열은 아데닌(adenine), 구아닌(guanine), 시토신(cytosine), 티민(thymine)의 네 가지 뉴클레오타이드(nucleotide)로 이루어져 있고, 아데닌(adenine)과 티민(thymine), 구아닌(guanine)과 시토신(cytosine)의 정합반응으로 염기 서열들이 하나의 이중 나선 구조를 이루고 있다. DNA는 각 생물의 고유한 유전적인 특징을 저장하고 있는 곳이며, DNA컴퓨팅에서도 정보 저장의 기본 단위를

[3] [Adleman, L. (1994)] Molecular computation of solutions to combinatorial problems 참조

아데닌(adenine), 구아닌(guanine), 시토신(cytosine), 티민(thymine)으로 하고 있다.

염기 서열들의 화학적인 상호 작용이 초 병렬적으로 일어나기 때문에 인간의 두뇌 같은 내용 기반 메모리 또는 Associative memory에서 정보 저장과 인출방식을 구현하기 위한 물리적인 기반을 제공한다고 할 수 있다.

2.2 Hypernetwork

Hypernetwork[4]은 초 병렬성을 바탕으로 정보들간의 복잡한 상호 작용을 요구하는 내용 기반의 정보 저장과 인출 또는 associative memory를 구현하기 위한 분자컴퓨팅 기반의 이론적인 모델이다.

Hypernetwork을 이루는 요소는, 어떤 정보를 나타내는 vertice들과 이 vertice들 사이의 상관 관계의 유무를 나타내는 연결들로 이루어진다. 이런 연결들은 vertice들 사이를 방향성이 없이 연결한다. vertice들의 집합이 하나의 hyperedge를 이루며, 이 hyperedge는 가중치를 가지고 있다. Hypernetwork은 이러한 hyperedge들의 집합이라고 할 수 있으며, 요소들의 상관 관계를 확률적으로 표상하고 있는 확률기반 모델이다. 이러한 hypernetwork으로 이루어진 기억 체계는 전체와 부분의 관계가 고정되어 있지 않다. 그래서 기억에 저장된 정보의 전체와 부분을

[4] [Zhang & Kim, (2006)]

유연하게 표상할 수 있는 이점을 가진다. 우리가 외부 세계로부터 다양한 형태의 혹은 다양한 관점에서 발생하는 인출 단서를 사용한다는 점에서 이러한 정보 저장 방식은 유연한 정보 인출을 구현하는데 이점이 될 수 있다.

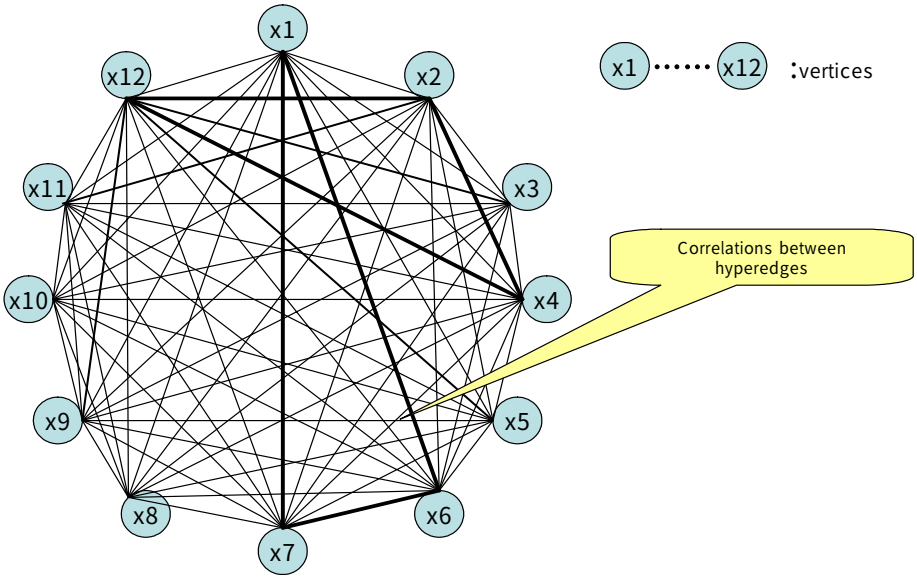


그림 1. Hypernetwork

Hypernetwork는 hyperedges와 그것들 사이의 상관 관계를 하나의 네트워크로 표상하고 있다. Hypernetwork는 가중치를 가진 hyperedge들의 집합이라고 할 수 있으며, hyperedge에서 연결하고 있는 vertice의 개수를 cardinality k라고 한다. 다음 그림에서 그 예를 들어보았다.

$V = \{v_1, v_2, \cdots, v_n\}$ ……vertices

$E = \{E_1, E_2, \cdots, E_m\}$ ……hyperedges, cardinality $k \geq 0$

$W = \{w_1, w_2, \cdots, w_m\}$ ……weight

$G = (V,E)$

$V = \{v_1, v_2, v_3, \cdots, v_7\}$

$E = \{E_1, E_2, E_3, E_4, E_5\}$

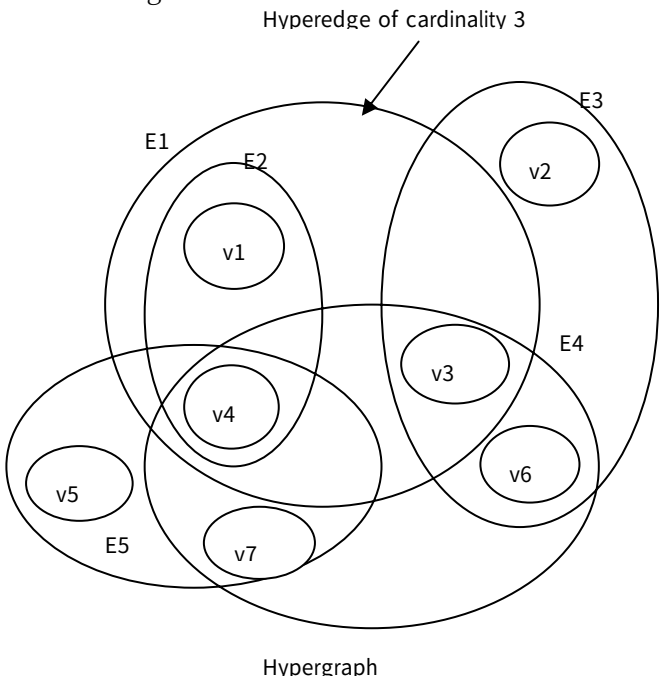
$E_1 = \{v_1, v_3, v_4\}$

$E_2 = \{v_1, v_4\}$

$E_3 = \{v_2, v_3, v_6\}$

$E_4 = \{v_3, v_4, v_6, v_7\}$

$E_5 = \{v_4, v_5, v_7\}$



	v_1	v_2	v_3	v_4	v_5	v_6	v_7
E_1	1		1	1			
E_2	1			1			
E_3		1	1			1	
E_4			1	1		1	1
E_5				1	1		1

그림 2. 7개의 vertice와 5개의 hyperedge로 이루어진 hypergraph의 예

절차는 다음과 같다. Hypernetwork을 구성하기 위해서 우선 학습이 진행될 hyperedge들의 집합인 라이브러리 L을 구성한다. 초기 임의의 라이브러리가 구성되면, 학습시킬 데이터의 패턴 x으로부터 K개의 hyperedge들을 생성한다. 그리고 그 hyperdege들을 라이브러리에 넣어줌으로써 분자들간의 화학 반응을 유도한다. 기대하는 화학 반응이 완료된 hyperedge들은 적정 비율로 복사된 후, 라이브러리에 추가된다. 그리고 다음 학습을 진행시키게 된다.[5]

Hypernetwork이 구현하고자 하는 내용기반 메모리의 기본적인 정의를[6] 살펴보자면, 첫째, 인간의 메모리에서 보이는 현상적인 특징들을 보여 주여야 한다. 이 현상적인 특징들은 아리스토텔레스가 관찰하였던 특징들을 의미하며, Classical Laws of Association이라고 한다.

The Laws of Association

Mental items (ideas, perceptions, sensations or feelings) are connected in memory under the following conditions:

1) If they occur simultaneously (“ spatial contact ”)

2) If they occur in close succession (“ temporal contact ”)

3) If they are similar. (the evoked item might have a high

[5] [장병탁 & 김주경 (2006)] 참조

[6] [Kohonen (1980)]

positive correlation)

4) If they are contrary (the evoked item might have a negative correlation)

둘째, 데이터의 검색은 그것의 위치가 아니라 내용에 기반해서 이루어진다. 이것은 semantic representation of knowledge로 이해되는 좀 더 추상적인 관점으로 데이터들간의 관계 구조를 의미하는 관점이다. 더불어 그 이전의 Hanlon(1966)의 논문에서는 association memory, 또는 content addressable memory에 대한 초기의 정의를 다음과 같이 기술하고 있다.

> “ Associative memories have been generally described as a collection or assemblage of elements having data storage capabilities, and which are accessed simultaneously and in parallel on the basis of data content rather
> than by specific address or location ” and “ the correct meaning of associative refers to interrelationships between data, and not to the storage mechanism ”

위의 정의에 따라, associative memory는 저장된 항목들 간의 모든

비교가 병렬적으로 수행되어야 한다.

위치 중심적 메모리는 저장된 주소와 저장된 내용이 임의적이지만, 사람을 대상으로 한 경험적인 실험들을 통해서 우리의 뇌는 저장된 이웃 정보들이 서로 상관 관계를 가지고 있음을 보여 주고 있다. 예를 들어서, 점화 효과(priming effect)는 먼저 제시된 단어가 나중에 제시된 단어의 처리에 영향을 주는 현상을 말한다. 이러한 어휘점화효과(lexical priming effect)에서 의미적으로 연상 관계에 있는 단어들 중에서 점화효과가 나타나는 경우를 의미점화효과(semantic priming effect)라고 하고, 단어 형태에 관련한 경우는 형태점화효과(form priming effect), 제시된 횟수와 관련해서는 반복점화효과(repetition priming effect)라고 한다. 의미점화효과의 경우 ‘의사’, ‘간호사’처럼 연상관계가 강한 경우가, ‘버터’,’간호사’보다 표적 단어의 어휘 판단 시간이 빨라진다. 즉 연상관계가 강할수록 점화효과가 커지는 현상을 보여주고 있다. 이것을 인간의 메모리에 저장된 정보들이 임의적 주소에 저장된 것이 아니고, 저장과 인출시의 조건에 따라 저장 장소와 인출 장소가 정해지며, 이 저장된 정보들이 유기적인 연결로 네트워크를 이루면서 그 네트워크 안에서는 가까이 있는 단어들일수록 연상관계가 큰 단어라고 본다. 그래서 네트워크 안에서 한 단어가 활성화가 되면, 그 활성화의 전이가 미치는 범위에 있는 단어는, 활성화의 전이가 미치지 않는 단어보다는 연상관계가 크다. 심성어휘집(mental lexicon)이 단어들간의 연결과 연결 강도로

이루어져 있다고 가정하거나, 의미 네트워크(semantic network)가 우리 두뇌에 있을 것이라고 생각하는 것도 이러한 맥락을 따르기 때문이다.

2.3 병렬 분산 처리 체계와 구성적 기억

우리의 공통적인 일상적 경험과 관련시켜서 볼 때, 시간이 경과함에 따라서 과거에 기억된 내용을 똑같이 기억하지 않을 수 있다는 것을 알 수 있다. 구성적 기억 이론에서는 기억의 재구성과 창조의 과정으로 이런 현상을 설명하고 있다. 구성적 기억 연구는 기억이 정보를 단순히 기억 저장소에 저장하고 회상 시에 그것을 기억 저장소에서 끌어내는 것이 아니라 정보가 들어가서 저장되고 기억되는 시점에서 정보가 구성되는 과정을 겪는다고 이야기 한다.

구성적 기억 연구에서[7] 피험자에게 단어들을 학습시키고, 회상 과제를 주었을 때, 피험자들이 학습하지 않은 단어들을 회상하는 것을 보고하였다. 예를 들면, 학습한 단어들이 '침대', '베개', '침대보'인 경우, 학습하지 않은 단어인 '잠'을 피험자들이 학습한 것으로 잘못 회상하는 경향이 있었다. Deese는 이것을 기억 구성에 대한 연합적 설명(associative account)을 제안하였다. 구성적 기억 연구에서도 그 접근 방향이 다양하였는데, McClelland(1995)는 뇌신경학적 이론을 바탕으로 병렬 분산처리 체계모델(paralled

[7] [Deese (1959)]

distributioned processing model)을 제안하였다. 경험이 뇌의 한 unit에 저장되는 것이 아니라 상호 연결되어 있는 여러 unit에 분산되어 있다고 보았다. 그래서 경험을 표상하는 것은 분산되어 있는 여러 unit들의 활성화 패턴으로 결정된다고 하였다. 위의 Deese의 실험을 병렬 분산처리 체계모델에서는 다음과 같이 설명하였다. ‘침대’, ’베개’와 같이 ‘잠’과 관련된 단어들은 잠이라는 단어가 가지는 활성화 패턴과 매우 닮아 있기 때문에 ‘잠’ 단어가 회상될 수 있다고 하였다. McClelland의 병렬 분산처리 체계모델과 같은 맥락의 뇌신경학적인 구성적 기억 모델은[8] 경험을 패턴으로 표상하고, 패턴의 구성 요소들이 뇌에서 넓게 분산되어 있다고 가정하고, 인출 실패를 패턴 분리(pattern separation)의 실패로 설명한다. 다시 말하면, 기억의 경계를 만들어야 정확한 기억 인출이 가능하지만, 패턴이 다른 패턴과 중첩되는 특징만 기억하게 되면 패턴의 특징이 없어서 다른 패턴과의 분리가 되지 않는 경우를 인출 실패라고 하였다.

기억 저장과 인출에 대한 분산처리 방식은 hypernetwork에서도 같은 맥락을 취한다. 본 논문의 주제인 단어 재인 시뮬레이션에서 단어를 학습할 때 단어 조각들이 hypernetwork에 분산 저장되어 있고, 단어들의 패턴을 학습한 후 단어 완성과제를 통한 인출 실험에서 주어진 인출 단서와 상관도가 높은 패턴을 인출하게 된다.

[8] [(Schacter, Norman & Koustaal, (1998))

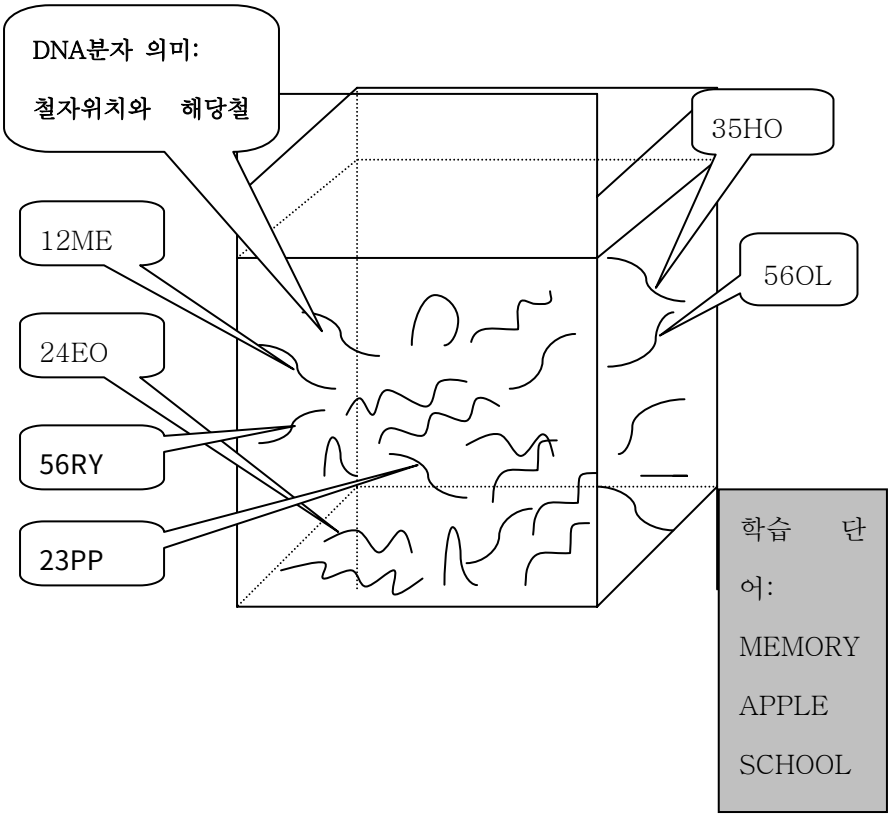


그림 3. Hypernetwork에서 단어 표상

2.4 심성 어휘집(Mental lexicon)

심성 어휘집(Mental lexicon)은 인간의 언어를 처리하기 위한 지식의 집합체라고 생각할 수 있다. 심성 어휘집을 처음에는 단순히 사전의 형태를 띤 단어 모음집일 것이라는 생각에서 발전해서, 현재는 단어의 시각적인 처리, 음운 정보의 처리를 비롯해서 의미 생성처리까지 그 범위가 광범위하다. 인간을 대상으로 한 심리학 실험에서 단어와 관련된

실험으로부터 심성 어휘집이 보이는 몇 가지 현상을 살펴보면, 첫째로 단어 우선 효과를 들 수 있다. 사람들은 단어를 비단어보다 즉각적이고 자동적으로 인식한다. 그리고 익숙한 단어들은 자동적인 응집된 지각 단위이기 때문에 인식을 제어하는 일은 불가능하다. 둘째는 친숙성 효과가 있다. 단어를 응집된 단위로 인식하는 능력은 학습을 통해 습득된다. 대부분의 학습된 능력과 마찬가지로 연습을 통해서 증대된다. 어떤 단어를 더 많이 접촉할수록 이것은 더욱 친숙해지고 더욱 빨리 인식된다. 이러한 현상은 무의미 단어들에 대해서도 성립한다. 친숙성 조건에 따른 하나의 사실은 심성 어휘집이 일반적인 사전과 다르다는 것이다. 종이 책 사전은 찾고자 하는 단어를 찾는 시간이 목록의 길이에 달려있지만, 단어를 인식하는 시간은 우리가 얼마나 많은 단어를 알고 있는가에 달려 있지는 않은 것 같다. 일반적인 사전에서는 일상적인 아닌 단어를 찾는 시간이 익숙한 단어를 찾는 시간보다 길지 않다. 하지만 심성 어휘집에서는 잘 안 쓰는 단어를 찾고자 할 때 시간이 많이 걸릴 것이다. 이것은 심성 어휘집이 어떻게 구성되어 있는가 하는 중요한 문제와 만나게 된다.

초기 어휘 처리 모형을 살펴보면, 탐색 기반 모형은 주어진 감각 정보가 심성어휘집의 어휘 군과 일치할 때까지 차례로 탐색된다는 이론으로 실질적인 설득력을 읽은 모형이다. 이후에 탐색이 아닌 활성화 개념이 사용되었는데, 초기 모형에는 Morton(1969)의 로고진 모형(logogen model)이 있다. 로고진 모형은 심성어휘집에서 단어 의미에

대한 모형과 단어 형태에 대한 모형을 구분해야 한다는 가정을 하였다. 그림3는 초기 로고진 모형에 따른 단어재인 모형을 나타낸다. 단어나 형태소를 의미하는 각각의 로고진은 입력 자극의 양이 역치를 넘으면 의미 시스템에 있는 단어의 의미가 활성화되고, 그것의 출력이 반응 버퍼로 나타난다. 각각의 로고진이 가지고 있는 역치는 사용 빈도나 반복 점화(repetition priming)에서 영향을 받는다. 활성화 개념을 사용한 모형은 빈도 효과나 반복 점화를 설명하는데 좀 더 우수하다. 이 방식을 토대로 McClelland와 Rumelhart는 좀더 정교한 어휘처리 모델을 제시하였는데, 이것을 상호작용 활성화 모형(interactive activation model)이라고 한다. 이 모형에서는 마디와 그 마디들의 연결로 이루어져 있으며, 연결에는 흥분성 연결과 억제성 연결이 있다. 이 상호작용 활성화 모형도 로고진 모형처럼 활성화 개념을 사용하긴 하지만, 이 모형에서는 역치는 존재하지 않고, 흥분성 연결과 억제성 연결을 마디마다 설정하고 있다. 로고진 모형은 학습의 빈도가 낮아서 로고진의 역치를 넘지 못하는 단어, 예를 들면 비단어를 처리하지 못하는 단점이 있었다. 하지만, 상호작용 활성화 모형에서는 들어온 입력에서 시작된 활성화는 역치에 따라 차단되거나 통과되는 것이 아니라, 마디의 최상위층까지 활성화가 이루어진다. 다만 억제성 활성화인지 흥분성 활성화 인지에 따라서 문자 수준의 마디와 단어 수준의 마디들 간의 양방향의 활성화가 진행됨에 따라서 단어 재인이 결정된다. 이 모형은 후에 PDP(parallel Distributed

processing)에 영향을 준다. 그림 4는 상호작용 활성화 모형을 나타내고 있다.

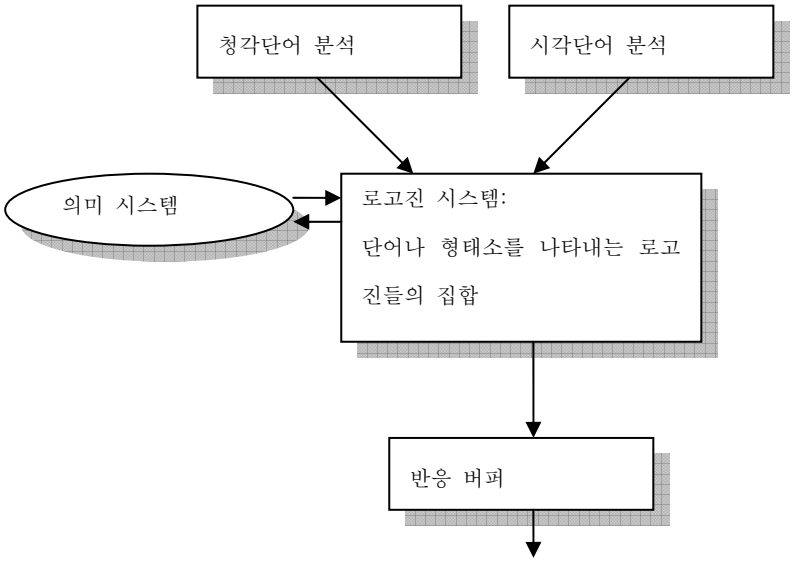


그림 4. Morton의 초기 로고진 모형

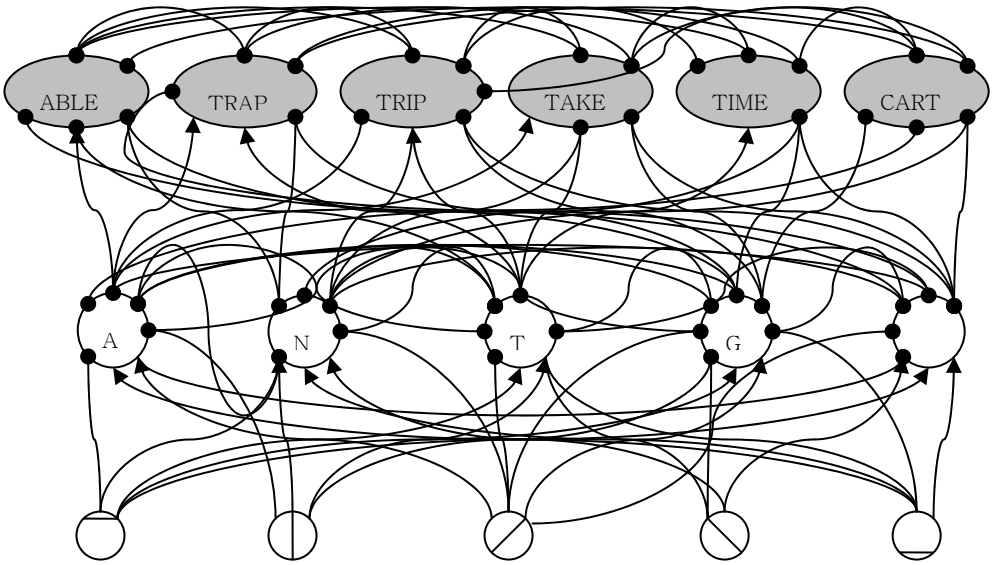


그림 5. McClelland & Rumelhart의 상호작용 활성화 모형

상호작용 활성화 모형은 컴퓨터 시뮬레이션을 신경망 알고리즘을 이용해서 할 수 있다는 장점 때문에 연결주의 모형의 전형이 되어왔다. 이 모형은 단어 빈도 효과뿐 아니라, 단어의 재인이 비단어 재인보다 우월한 경우를 단어 수준에서 문자 수준의 피드백으로 설명하고 있다.

3. 핵심 연구 내용 - 단어 재인을 위한 Hypernetwork 모델

단어 재인을 위한 hypernetwork은 hyperedge들에 단어의 철자 조합을 저장하게 됨으로써, 각 단어를 전체 hypernetwork안에서 분산 저장하고 있다고 할 수 있다. 그림 6에서 보듯이, 한 hyperedge가 4개의 vertice들을 연결하는 경우라면, LOVE란 단어는 위치 정보를 나타내는 vertice 1, vertice 2, vertice 3, vertice 4에 저장되어 있다. MEMORY의 경우는 이 단어에서 가능한 4철자의 조합이 각 위치의 vertice에 분산 저장되어 있다고 할 수 있다. 각 hyperedge들은 DNA분자로 구성된 라이브러리를 학습 시킴으로써 네트워크 내에서 비중이 달라지게 된다. 즉, 같은 형태의 반복적 학습이 많을수록 라이브러리 내에 해당 형태를 표상 하는 DNA분자의 수를 증가시키게 된다. 네트워크로부터의 단어 인출은 단어 완성 과제를 통해서 이루어지며, 주어진 단어와 형태적으로 상관 관계가 높은 철자를 택하게 된다. 그 상관 관계는 hypernetwork을 전체적으로 표상 하는 DNA분자 라이브러리를 학습할 때, 훈련 단어로 주어진 철자 패턴들과 매치되는 DNA분자의 수를 증가시킴으로써 상관 관계의 경중을 네트워크가 저장하게 된다. 인출 시 선택될 확률을 높여주기 때문에 hypernetwork에서는 단어의 저장과 인출 방식이 확률적 모델을 따른다고 할 수 있다.

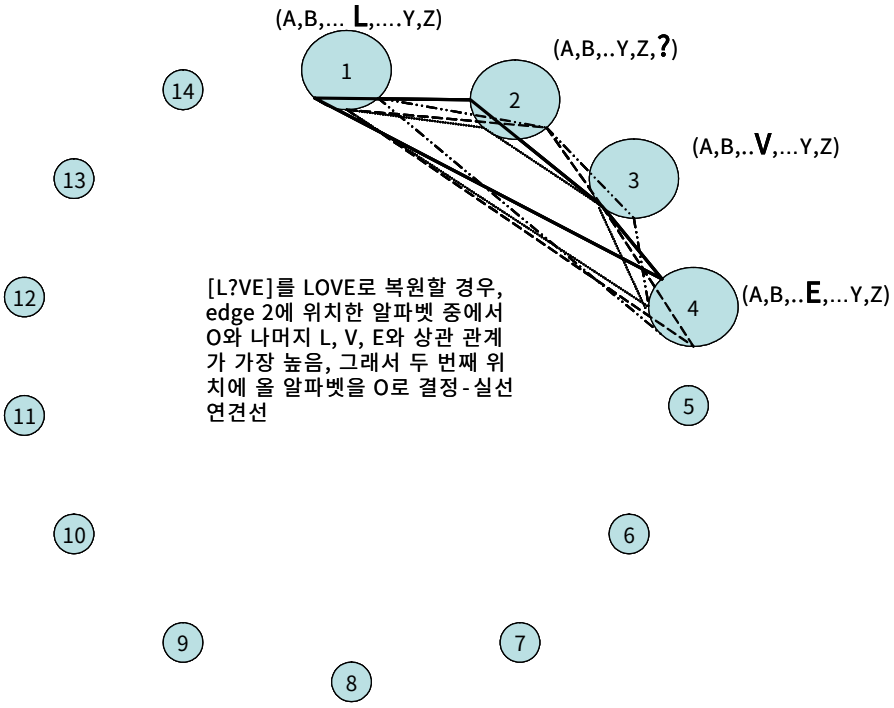


그림 6. Hypernetwork에서 단어 완성

예를 들어서 설명하자면, 각 알파벳은 해당 DNA시퀀스로[9] 코딩 된다.

WORD

W - - CGCGGCGATG

O - - ACATATCGGT

[9] 부록 참조

R - - CTGTCGAGTG

D - - CAGTTATTTC

위의 형식으로 단어들이 코딩 된 후, 초기 라이브러리를 구성한다. 그 라이브러리의 형태는 영어 단어의 위치 정보와 해당 철자로 이루어져 있고, 가능한 철자 조합은 2개(order 2)부터 14개(order 14)까지 이다. 초기 라이브러리에는 학습할 단어에 대한 연속된 철자 조합이 모두 포함되어 있다. 만약 WORD라는 단어로 2개 철자 조합으로 구성된 라이브러리를 학습시킨다면, 12WO, 23OR, 34RD가 학습할 철자로 주어진다. 이 연속한 철자들과 정합 반응을 한 초기 라이브러리의 DNA분자 시퀀스는 개수를 증가시킴으로써 라이브러리 내의 DNA분자 시퀀스의 농도에 변화를 주어 라이브러리를 업데이트 하게 된다. 이렇게 일련의 단어들이 순차적으로 제공되면서 학습이 끝날 때까지 라이브러리를 업데이트 하게 된다. 그림 7이 학습 과정을 설명한 그림이다.

단어 완성 과제를 통해서 hypernetwork으로부터의 단어 인출 방식은 다음과 같다. 예를 들어서, 123456ME?ORY (MEMORY)가 테스트 단어라면, order2(연속한 철자 조합이 두 개)인 경우에 23E?, 34?O와 같이 손상된 철자 조합이 라이브러리에 인출 단서로 주어진다. 주어진 인출 단서들과 반응을 한 23EA ~23EZ, 34AO~ 34ZO의 DNA시퀀스를 추출한다. 추출된 DNA시퀀스(23EA~23EZ)에서 농도가 가장 높은 것을

선택하여 후보 철자를 결정한다. 다음으로 추출된 DNA시퀀스(34AO~34ZO)에서 농도가 가장 높은 것을 선택하여 후보 철자를 결정한다. 만약, 두 경우에서 후보 철자가 다른 경우, DNA시퀀스의 농도가 더 큰 것으로 결정하게 된다.

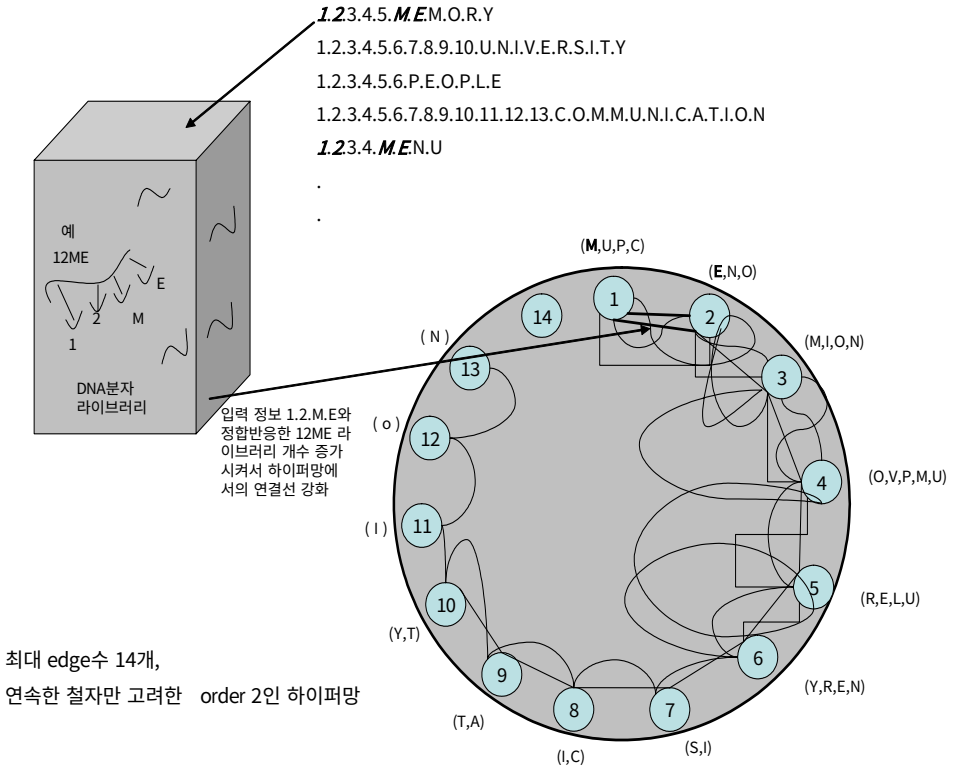


그림 7. 라이브러리 학습 과정

4. 실험 및 결과

브라운 대학의 쿠체라Kučera와 프랜시스Francis는 1,014,232어절의 미국 문어 샘플인 브라운 코퍼스에서 50,406가지의 단어를 발견하였다. 본 실험에서 선택한 단어는 대략 5만개의 단어에서 길이가 4글자부터 14글자까지의 단어만 사용하였고 최종적으로 39,311개의 단어를 선택하였다. 다음은 단어 길이에 따른 분포를 나타낸다.

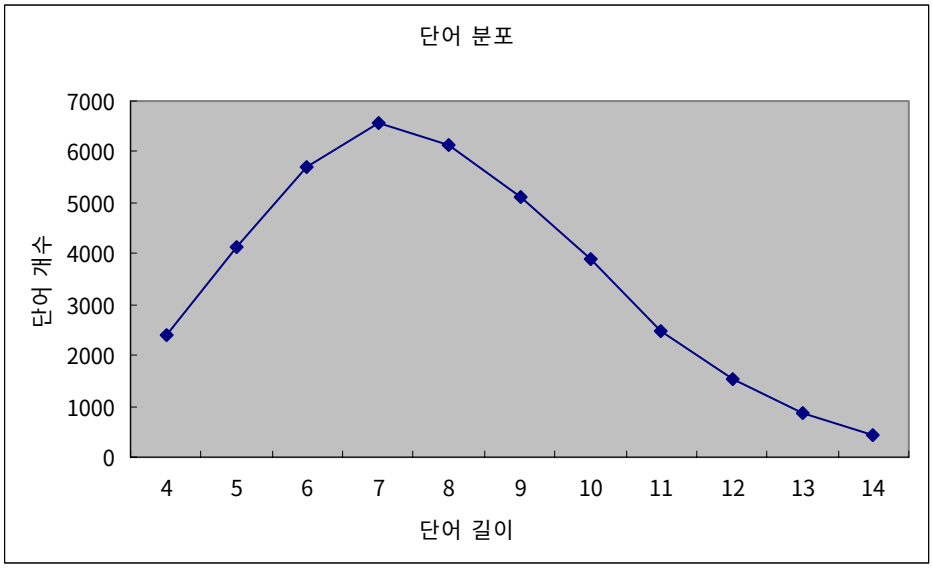


그림 8. 실험에 사용한 단어의 길이 따른 분포

다음 두 실험에서, 연속한 철자 조합(order)의 크기에 따른 단어 복원률을 알아 볼 것이다. 이때, 과제의 난이도를 측정하기 위해서 정보

이론에서의 entropy와 perplexity를 측정하였다. 그것의 공식과 측정된 Perplexity는 다음과 같다.

$$p(x) = \frac{\text{각 조합의 횟수}}{\text{훈련 데이터에서 모든 철자 조합의 개수}}$$

$$\text{Entropy } H(X) = -\sum_{x \in X} p(x) \log_2 p(x)$$

$$\text{Perplexity} = 2^H$$

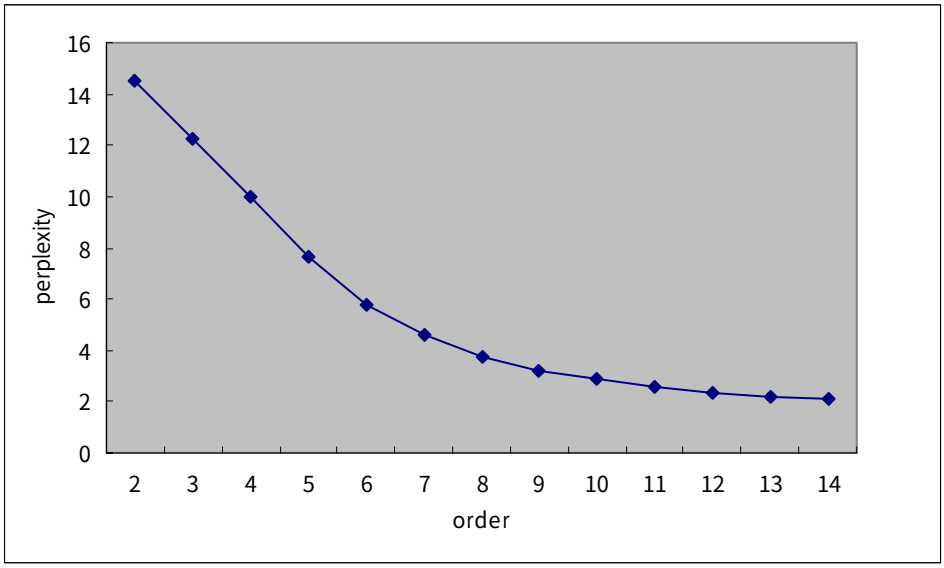


그림 9. 실험에 사용한 단어의 perplexity

Perplexity가 작아질수록 문제의 난이도가 쉬워지는 것을 의미한다.

4.1 실험 1: 학습한 단어의 단어 완성 실험

브라운 코퍼스에서 선택한 39,311개의 단어를 학습시키고, 그 중 20%를 테스트 단어로 사용하였다. 테스트 단어에 대한 손상 위치는 랜덤하며, 손상 개수는 3개까지 실험하였다. 그리고 복원한 단어의 오답 여부는 손상시키기 전 단어와 비교해서 철자가 모두 맞아야 정단어로 인정하였다. 다음은 연속 철자 조합(order 2~order14)으로 이루어진 각각의 hypernetwork에서의 단어 복원률이다.

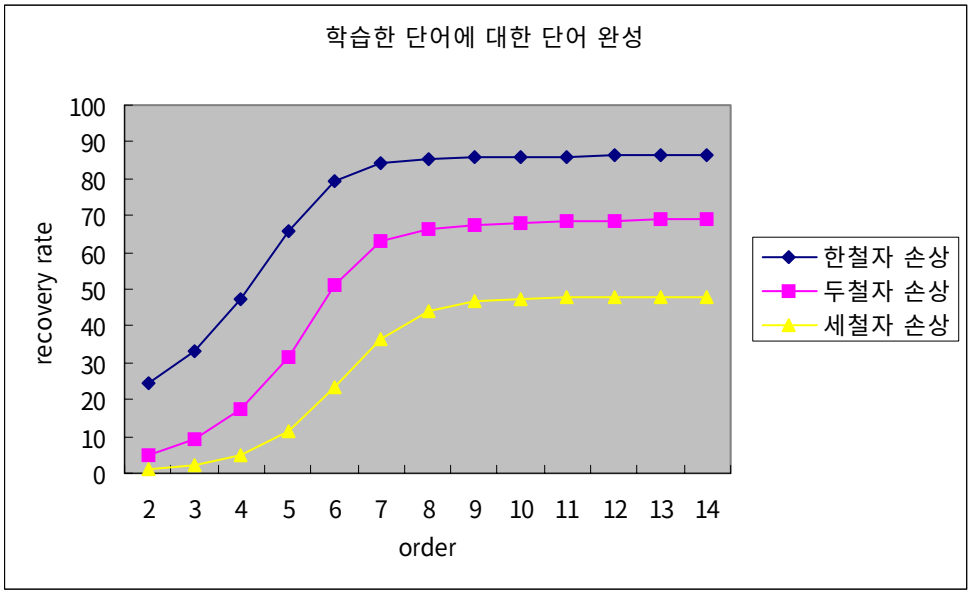


그림 10. 학습한 단어에 대한 단어 복원률

손상된 철자 수가 늘어날수록 단어를 복원하는데 필요한 정보의 감소로 복원률이 떨어지는 것을 볼 수 있다. 반면에 order가 커지면서 복원률이 증가하는 것을 관찰 할 수 있는데, 이것은 ?에 올 수 있는 A부터 Z까지 후보 철자가 연속된 철자정보를 많이 알수록 올바른 철자가 복원될 확률이 커지기 때문이다. 예를 들어서 2개의 연속 철자 조합의 hypernetwork라면, ME?ORY를 복원하는 문제에서 E?, ?O에서 M이 복원되어야 하겠지만, EA, EB, 혹은 TO, BO과 같은 빈도가 높은 철자 조합이 있을 경우, M이 복원되기는 힘들다. 하지만 Order3의 경우에서는 ME?, E?O, ?OR에서 M이 복원될 확률은 Order2의 경우보다 커진다. 다시 말하면, Order가 하나씩 커질수록 간섭 확률이 1/26씩 줄어들기 때문이다.

전체적인 복원 경향은 order 4에서 order 5으로 증가할 때 가장 큰 폭으로 복원률이 좋아지고, order7이후는 복원률의 증가 폭이 매우 작고 거의 복원률이 일정 값으로 유지되고 있다고 볼 수 있다. 더 이상 복원률이 올라가지 않는 이유는 단어 길이가 해당 order와 같은 경우는 모두 복원 되지만 order보다 짧은 단어인 경우 복원 되지 않는 경우가 있기 때문에 평균 복원률이 계속 증가하지 않는다.

4.2 실험 2: 새로운 단어의 단어 완성 실험

브라운 코퍼스에서 선택된 39,311개의 단어에서 훈련 단어로 31,448개의 단어를 사용하고, 테스트 단어로 7,863개의 단어를 사용하였다. 테스트 단어에서 손상 위치는 역시 랜덤하게 정했으며, 손상 개수는 3개까지 주었다. 복원된 단어의 정단어 여부는 손상시키기 전 테스트 단어와 비교해서 결정하였다. 복원률은 다음과 같다.

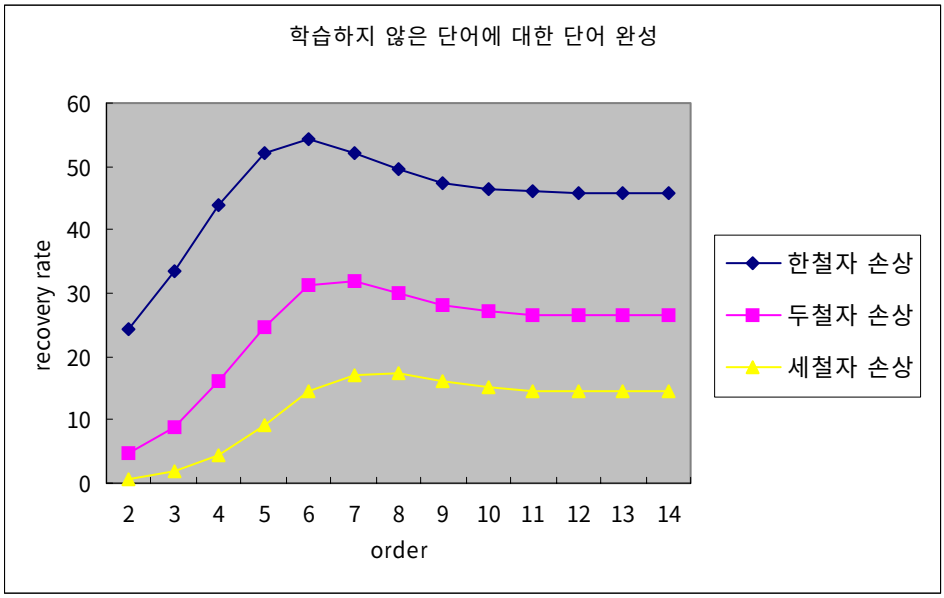


그림 11. 학습하지 않은 단에에 대한 단어 복원률

새로운 단어에 대한 복원률은 order 6에서 가장 높고, order가 커질수록 복원률이 다시 떨어지는 것을 볼 수 있다. 이것은 낮은 order에서는

이전에 학습했던 비슷한 철자들로부터 복원 정보를 얻을 수 있지만, order가 어느 지점 이후로 커지게 되면 복원해야 할 단어에 대한 전체 철자 조합 정보는 hypernetwork이 모르기 때문에, 복원률이 다시 떨어지게 된다. 그리고 학습한 단어와 비슷한 새로운 단어라도 단어길이가 길어지면 비슷한 철자 구간이 상대적으로 짧아지기 때문에 손상 위치가 비슷한 구간이 아닌 곳에 정해질 경우도 많아지게 된다. 하지만 정점에서부터 복원률이 계속 떨어지지 않고 어느 정도 복원률을 유지하는 이유는 비슷한 단어로부터 복원 정보를 얻고 있기 때문이다. 다음은 비슷한 단어의 복원 예를 보였다.

예)

학습한 단어

새로운 단어

새로운 단어의 손상

MEMORY	SCHOOL
MEMORABILIA	SCHOOLING
?EMORABILIA	SCHOO?ING

CARRYING	RESORT
CARRYOVERS	RESORTING

CARR?OVERS	?ESORTING
MINDFUL	DRESSING
MINDANAO	DRESSED
MI?DANAO	D?ESSED
CITIZEN	TRAVEL
CITIZENRY	TRAVELLED
CI?IZENRY	TRA?ELLED
LEGAL	
LEGALIZED	
LEG?LIZED	
…	
…	

4.3 실험 3: 위치 별 손상에 따른 단어 완성

실험 3에서는 hypernetwork이 학습한 패턴에 관해서 좀 더 분석하기 위한 실험을 하였다. 실험 1,2와는 달리 테스트 단어를 손상

시키는 위치를 단어의 첫 번째 위치부터 단어의 끝 위치까지 순차적으로 정하였다. 손상시킨 철자의 수는 두 개까지이고, order의 크기는 5까지 제한하였다. 이유는 이 실험에서 고려한 영어 단어의 특정 형태가 5개의 철자 조합을 넘지 않기 때문이다. 테스트로 사용한 단어는 단어길이 8과 단어길이 11을 사용하였다. 왜냐하면 브라운 코퍼스에서 선택한 영어 단어에서 존재하는 특정한 형태(예, tion, ing, ed, ly, able, tive, ment)를 포함하는 비율을 조사했을 때, 단어길이 8은 33%로 가장 적게 포함하였고, 단어길이 11은 81%로 가장 많이 특정 형태를 포함하고 있기 때문이다. 그래서 단어 길이 8과 11인 경우만 대표로 실험하였다.

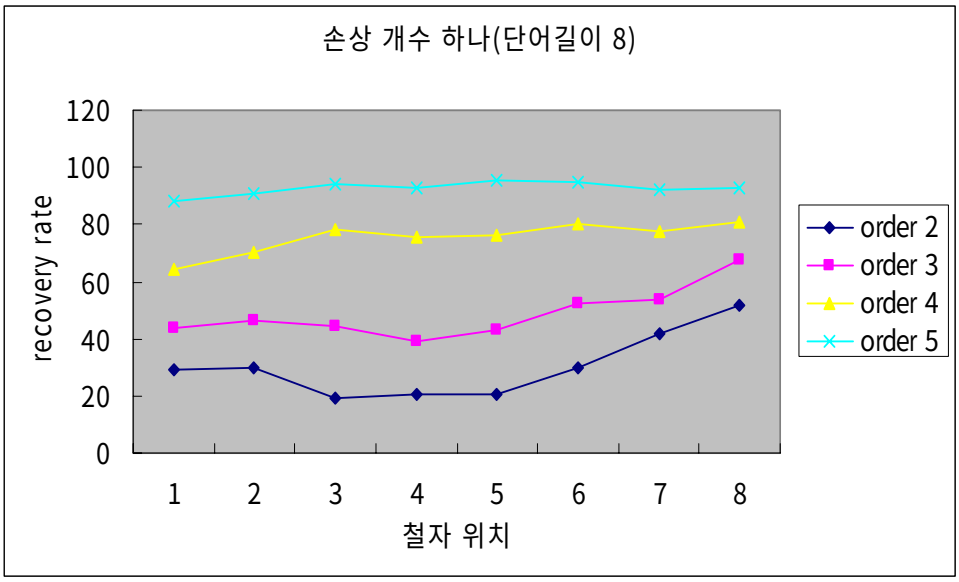


그림 12. 손상 위치에 따른 복원률 (손상 개수 = 1, 단어 길이 = 8)

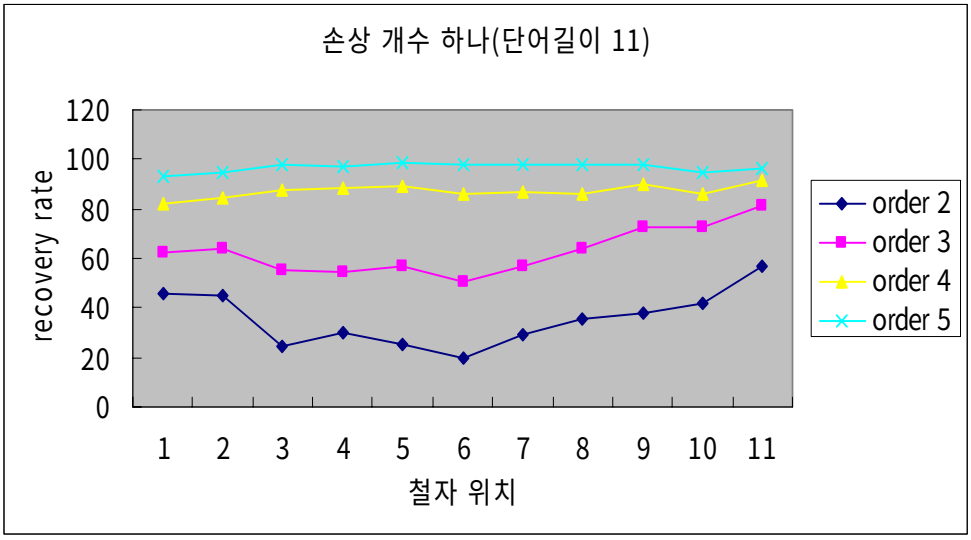


그림 13. 손상 위치에 따른 복원률 (손상 개수 = 1, 단어 길이 = 11)

손상 개수가 하나인 경우, order 2에서는 손상 위치가 앞이나 뒤인 경우 recovery rate가 좋아지는 것을 볼 수 있다. 단어는 접두어와 접미어가 존재하기 때문에 단어 앞과 뒤에서 같은 형태가 반복되는 경우가 많다. 그래서 hypernetwork이 이러한 접두어, 접미어의 패턴을 학습하였다고 볼 수 있다. 하지만 order가 커질수록 손상 위치에 따른 복원률이 단어 처음부터 끝까지 비슷한 것을 보이는데, 이것은 order가 커지면 단서가 많아지기 때문에 손상 위치에 크게 영향을 받지 않는다고 할 수 있다. 다음은 철자를 2개 손상 시켰을 경우의 결과이다.

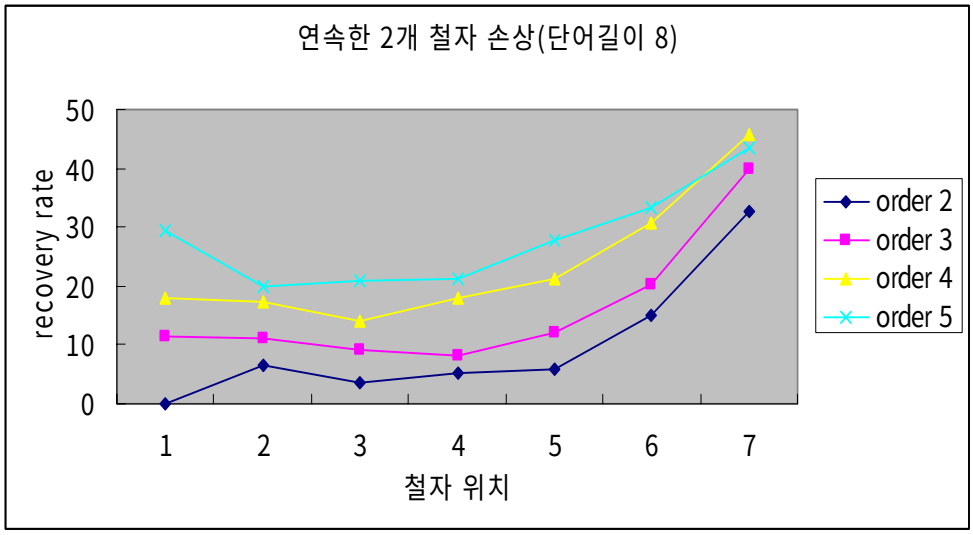


그림 14. 손상 위치에 따른 복원률 (손상 개수 = 2(연속철자), 단어 길이 = 8)

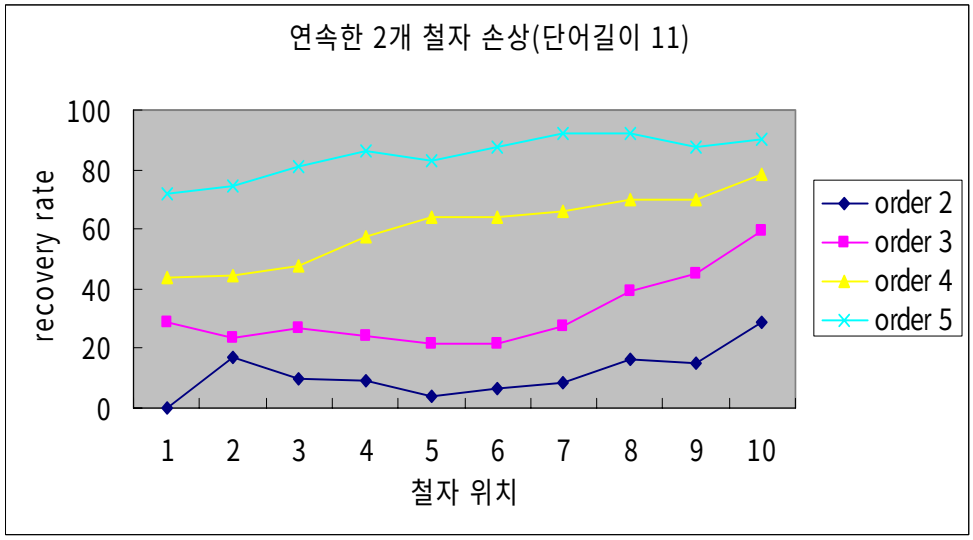


그림 15. 손상 위치에 따른 복원률 (손상 개수 = 2(연속철자), 단어 길이 = 11)

손상 개수가 증가 하였기 때문에, 복원률이 떨어졌고 특히 order2인

경우, 손상 위치가 앞일수록 복원률이 저조 함을 볼 수 있는데, 이것은 단어에서 앞쪽 정보를 주지 않았을 경우, 사람들의 단어 재인 능력이 떨어진다는 심리학적 실험 결과와 비슷하다고 볼 수 있다. 하지만 order가 커질수록 복원 단서가 충분히 주어지기 때문에, 이러한 위치 효과는 거의 나타나지 않았다. 전체 실험 3의 그래프에서 보듯이, hypernetwork이 단어의 접두어와 접미어의 패턴을 다른 철자 조합 패턴보다 더 잘 학습하였다고 볼 수 있다.

5. 결 론

5.1 연구 결과 요약 및 정리

본 논문에서 실험한 주요한 내용은 DNA분자로 구현된 hypernetwork에 단어를 분산 저장하고, 분산 저장된 철자 조합들로부터의 단어 재인을 시뮬레이션 한 것이다. 단어 재인을 위해서 단어 완성 과제를 수행하였으며, 단어를 이루는 연속한 철자 조합의 크기(order2~order14)에 따른 복원률을 알아보았다. 그리고, 손상 위치에 따른 단어 완성 과제를 시뮬레이션 하였다.

첫 번째 실험에서 학습한 단어를 가지고 단어 완성 과제를 했을 때, order가 커질수록 증가 폭은 줄어들지만, 복원률은 계속 증가하는 것을 볼 수 있다. 새로운 단어의 단어 완성 과제에서는 복원률의 패턴이 order가 커지면서 좋아지다가 order가 계속 커지면 다시 감소하였다. 어느 정도 order가 증가하면 간섭 확률은 줄어들지만, 학습하지 않은 단어인 경우에는 order가 길어질수록 해당 단어가 hypernetwork에 저장되어 있지 않기 때문에, 복원률이 감소하게 된다. hypernetwork의 학습은 단어종류가 많을수록, 그리고 같은 단어를 학습할 기회가 많을수록 복원률이 좋아진다. Order가 낮은 경우는 단어를 복원시에 hypernetwork의 좀 더 많은 hyperedge들이 복원에 참여하게 되지만, order가 높아지면서 주어진 단어를 복원하는데 적용될 수 있는

hyperedge들의 종류가 줄어들게 된다. 다시 말하면, 낮은 order인 경우 단어 복원에 일반적으로 적용되며, order가 높아지면서 특정 단어 복원에 적용되는 경우라고 할 수 있다.

위치 별 손상에 따른 단어 완성 실험에서는 손상 개수가 하나일 때, 낮은 order에서는 손상 위치가 단어의 중간 보다는 앞과 뒤일 때 복원률이 좋음을 볼 수 있다. 이것은 단어의 형태가 앞과 뒤에 반복되는 형태가 많기 때문인데, 어근과 어미정보를 hypernetwork이 저장하고 있다고 볼 수 있다. 손상 개수가 연속된 위치에서 두개일 때는 대체로 어근은 어미 정보보다 길이가 짧기 때문에, 어근 정보를 잃어 버리게 되어 복원률이 낮았다. 이것은 심리학 실험에서 사람들에게 단어의 앞쪽 정보가 중요함을 보여주는 위치 효과라고 할 수 있다. 하지만 order가 커질수록 위치 효과는 보이지 않으며, 그 이유는 복원에 필요한 인출 단서가 많아졌기 때문이라고 할 수 있다.

정리하자면, hypernetwork은 단어를 분산 저장하고, 분산된 철자 조합들로부터 전체 단어가 조합되는 모델이며, 단어를 인출할 때 인출할 철자를 이웃한 철자들과의 상관 관계에서 확률적으로 정한다. 단어 완성 과제를 수행할 때, 낮은 order에서는 hyperedge들이 좀 더 일반적으로 단어 복원에 적용되는 general rule에 기반한 단어 복원이 이루어지고, 높은 order에서는 특정 단어를 복원하기 위한 specific rule에 기반한다고 할 수 있다. 이것은 hypernetwork이 상황에 따라 규칙들을

유연하게 적용시킬 수 있는 모델이라고 할 수 있다.

심성 어휘집에서의 단어 인출은 리스트 방식이 아니다. 다시 말해서, 친숙한 단어와 형태적으로나 의미적으로 연관 있는 단어들이 그렇지 않은 단어보다 빨리 재인되는 현상을 보여준다. 실험에서 보여주는 단어 복원 현상은 학습의 정도와 인출 단서에 따라서 단어 복원이 달라짐을 보여줌으로써 심성 어휘집에서의 단어 인출 현상과 관련이 깊다고 할 수 있다. 또한 이러한 실험 결과는 인간의 메모리 같은 내용 기반 메모리에서의 기억 현상을 보여준다고 할 수 있다.

5.2 향후 과제

먼저 분산되어 저장된 철자 조합들이 언어학적으로나 심리학적으로 어떤 정보를 포함하는가에 대한 연구가 필요하다. 본 논문에서는 철자 조합들의 의미적인 정보는 구체적으로 언급하지 않고, 형태적으로 다양한 크기의 철자 조합들로부터 단어를 인출하는 실험을 하였다. 단어의 인출에 영향을 주는 형태적인 패턴뿐만 아니라 궁극적으로는 철자 너머에 내포되어 있는 여러 가지 의미에 관한 고려가 이루어져야 hypernetwork에서의 단어 표상을 좀 더 포괄적으로 이야기 할 수 있을 것이다.

본 실험에서는 단어를 많이 학습할수록, 그리고 자주 접할수록 복원률이 좋아지는데, 인간이 단어를 학습할 때 발생하는 여러 가지

현상적인 기억 이론을 적용시켜 시뮬레이션 모델을 학습시킨다면, 좀 더 구체적인 심리학적인 데이터를 적용할 수 있을 것이다. 그리고 실험에서 각각의 라이브러리에서 order를 구성하였지만 이것을 하나의 라이브러리에 구현해서 general rule과 specific rule이 동시에 적용되는 실험을 한다면 심성 어휘집에서 일어나는 언어 현상에 관하여 다양하게 시뮬레이션 할 수 있을 것이다.

참 고 문 헌

Adleman, L.M. (1994). Molecular computation of solutions to combinatorial problems, Science, 266, 1021 - 1024.

Anderson, J. (1985). Cognitive psychology and its implications. New York: W.H. Freeman

Baddeley, A. (1990). Human Memory: Theory and practice. Hillsdale, NJ, pp. 263 - 291

Ezziane, Z. (2006) DNA computing: applications and challenges, Nanotechnology 17, R27 - R39.

Gardenfors, P. (2000) Conceptual Spaces: The Geometry of Thought, MIT Press

Goldstein, E.B. (2001). Sensation and perception (6th ed.). Pacific Grove, CA: Brooks/Cole.

Kim, Ji-Soo. (2005). Emotional Information Categorization by Molecular Computing, Master Thesis, School of Computer Science and Engineering, Seoul National University.

Kohonen, T. (1980). Content-Addressable Memories, Springer-Verlag, Berlin

Lee, Eun-Seok. (2003). DNAgram: Molecular Simulation of Anagram Problem Solving, Master Thesis, Interdisciplinary Program in Cognitive Science, Seoul National University.

MacLeod, C.M. & Kampe, K.E. (1996). Word Frequency Effects on Recall, Recognition, and Word Fragment Completion Tests, Journal of experimental Psychology, Learning, Memory and Cognition, Vol.22, No.1, 132-142

Massaro, D.W., Taylor, G.A., Venezky, R.L., Jastrzembski, J.E. and Lucas, P.A. (1980). Letter And Word Perception: Orthographic Structure and Visual Processing in Reading, N.H.C, NY

Miyake, A. and Shah, P. (1999). Model of Working Memory. Models of Working Memory: Mechanisms of Active Maintenance and Executive Control, New York: Cambridge University Press, pp. 375 - 411

Raja Sekhar, G.P. DNA Computing - Graph Algorithms, Department of Mathematics, Indian Institute of Technology

Rumelhart, D.E., and McClelland, J.L. (1986). Parallel Distributed Processing, vols. 1 and 2, Cambrideg, MA: MIT Press

Seidenberg, M.S. (1995). Visual word recognition: An overview. In J. L. Miller, and P. D. Eimas, Eds., Speech, Language, and Communication. San Diego: Academic Press.

Seidenberg, M.S and McClelland, J.L (1989). A Distributed, Developmental model of Word Recognition and Naming, Psychological Review,Vol.96, No.4, 523 - 568

Son, Jeong - Wook. (2005). Molecular computational modeling of recognizing emotion in text, Master Thesis, Interdisciplinary Program in Cognitive Science, Seoul National University.

Ullman, M.T. (2004). Contributions of memory circuits to language: the declarative/procedural model, Cognition, Vol. 92, No. 1 - 2. pp. 231 - 270.

Zhang, B. - T. and Jang, H. - Y. (2005). A Bayesian algorithm for in vitro molecular evolution of pattern classifiers, Lecture Notes in Computer Science, 3384:458 - 467

Zhang, B. - T. and Jang, H. - Y. (2005). Molecular learning of wDNF formulae, Lecture Notes in Computer Science*, DNA 11*, 3892:427 - 437

Zhang, B. - T. and Jang, H. - Y. (2005). Molecular learning of wDNF formulae, Preliminary Proceedings of the Eleventh International Meeting on DNA Computing (DNA 11), pp. 185 - 195

Zhang, B.-T. and Kim, J.-K. DNA Hypernetworks for Information Storage and Retrieval, (to be submitted)

인지심리학 (개정판). (2003). 이정모, 외 16인 공저 학지사.

장병탁, (2005). 나노바이오지능 분자 컴퓨터: 컴퓨터 공학과 바이오 공학, 나노기술, 인지뇌과학의 만남, 한국정보과학회지 바이오정보기술 특집호

부 록

DNA시퀀스(이은석, 2003)

A:GAAATGAGTT	B:TGATGCTACA
C:GGTTGTGGCG	D:CAGTTATTTC
E:GGCTACGATG	F:GAGCACTGCG
G:AGACGACGAC	H:GTAGAGGTAC
I:TCTCAGAGAT	J:TCGATTTGGA
K:TAGGCAGATG	L:ACAGCACTAC
M:TCGGATAGGA	N:TGATTATGTC
O:ACATATCGGT	P:CTCTTCAACG
Q:AGAAGAAACT	R:CTGTCGAGTG
S:ACATTTCACC	T:ATCACGTAAT
U:CAGAACGGAT	V:TGTGTAGTGC
W:CGCGGCGATG	X:CTATGCTGGG
Y:AACTTGGCGC	Z:GCATGATCAT
?: ATCTGGCCCT	

Abstract

This paper presents a graph-based model to explain how words are memorized and retrieved in the human brain by solving word completion problems. Previous studies have mainly focused on the process of building up and searching a word list. However, they cannot appropriately explain the retrieval process of the human brain. Recent works suggest breaking up words into small segments that explain word storage and retrieval.

Here, we try to solve the memorizing mechanism by using a hypernetwork model. The proposed method is a hyper-interaction computational model employing higher-order memory chunks, which are densely connected to each other. In this paper, we show that the hypernetwork architecture can be a possible candidate for explaining the human cognitive process.

The experiments are performed by setting the hypernetwork order that indicates the storage unit of alphabets. Experiment 1 examines the recovery rate of damaged words, and experiment 2 examines the recovery rate of new words which are not shown in the learning process. Experiment 3 tests how the recovery rate varies according to the damaged location. The experimental results show that the performance changes over the orders represent the word completion capability of the brain through general rule and specific rule. Also, the damaged location, especially in prefixes

explains the psychological word retrieval effect. The hypernetwork order and the damaged location elucidate the ‘familiarity effects’ and the ‘position effects’, respectively.

Keywords: moleculer computing, hypernetwork, word recognition, word fragment completion

Student Number : 2004 - 20106