인지과학, 제28권 제2호
Korean Journal of Cognitive Science
2017, Vol. 28, No. 2, 111～131.
http://dx.doi.org/10.19066/cogsci.2017.28.2.002

# RBM을 이용한 언어의 분산 표상화*

유 희 조 남 기 춘 남 호 성†
고려대학교 고려대학교 고려대학교
심리학과 심리학과 영어영문학과

연결주의 모델은 계산주의적 관점에서 언어 처리를 연구하는 한 가지 접근법이다. 그리고 연결주의 모델 연구를 진행하는데 있어서 표상(representation)을 구축하는 것은, 모델의 학습 수준 및 수행 능력을 결정한다는 점에서 모델의 구조를 만드는 것만큼이나 중요한 일이다. 연결주의 모델은 크게 지역 표상(localist representation)과 분산 표상(distributed representation)이라는 두 가지 서로 다른 방식으로 표상을 구축해 왔다. 하지만 종래 연구들에서 사용된 지역 표상은 드문 목표 활성화 값을 갖고 있는 출력층의 유닛이 불활성화하는 제한점을, 그리고 과거의 분산 표상은 표상된 정보의 불투명성에 의한 결과 확인의 어려움이라는 제한점을 갖고 있었으며 이는 연결주의 모델 연구 전반의 제한점이 되어 왔다. 본 연구는 이와 같은 과거의 표상 구축의 제한점에 대하여, 제한된 볼츠만 머신(restricted Boltzmann machine)이 갖고 있는 특징인 정보의 추상화를 활용하여 지역 표상을 가지고 분산 표상을 유도하는 새로운 방안을 제시하였다. 결과적으로 본 연구가 제안한 방법은 정보의 압축과 분산 표상을 지역 표상으로 역변환하는 방안을 활용하여 종래의 표상 구축 방법이 갖고 있는 문제를 효과적으로 해결함을 보였다.

주제어 : 연결주의 모델링, 지역 표상, 분산 표상, 제한된 볼츠만 머신

---

* 이 논문 또는 저서는 2016년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2016R1A2B4007876).
이 논문은 2016년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2016S1A5A8019370)
† 교신저자: 남호성, 고려대학교 영어영문학과, (02841) 서울특별시 성북구 안암로 145
연구 분야: 음성학
E-mail: hnam@korea.ac.kr

# 서 론

연결주의 모델은 철자, 발음, 그리고 의미 정보 등의 상호관계를 학습시켜 인간의 언어 처리 과정을 모사하고, 언어심리학적 현상을 계산주의적으로 표현하고자 하는 방법이다. 이와 같은 연결주의 모델을 구성하는데 있어서 연구자는 크게 두 가지, 모델의 구조와 표상(representation)의 구조를 결정할 필요성이 있다.

모델의 구조는 말 그대로 모델이 어떻게 언어처리과정을 표현하는가에 대한 이야기이다. 즉 연구자는 '어떤 정보를 입력 받아서, 어떤 과정을 거쳐서 최종적으로 어떤 정보를 출력해낸다'를 결정함으로써 모델의 구조를 결정한다. 비단 연결주의 모델 뿐 만이 아니라, 다른 모든 모델들 또한 모델의 구조를 제시한다.

반면 표상의 구조는 처리되는 정보가 어떤 형태로 이루어져 있는가에 대한 이야기이다. 연결주의 모델을 구성할 때 연구자는 모델을 구성함에 있어서 '어떤 정보가 어떤 구조를 갖고 있는가' 를 결정해야만 한다. 하지만 많은 경우 이와 같은 구조에 대해서는 깊은 논의가 이루어지지 못하는 것이 사실이다.

## 1. 연결주의 모델에서의 층(layer)

연결주의 모델의 구조 중 정보를 표상하는 위치는 주로 층(layer)으로 정의되며, 층은 여러 개의 유닛(unit)의 집합으로 구성된다. 그리고 표상은 이들 각 유닛이 어떤 값을 갖는가에 의해 결정된다.

많은 경우 유닛의 값은 0에서 1 사이의 값을 갖고 있으며 시그모이드(sigmoid), 소프트맥스(softmax)와 같이 유닛의 활성화 값을 계산하는 함수들 또한 이 범위 내에서 값을 구하도록 되어 있는 경우가 많다. 이에 대해서는 몇 가지 이유가 있는데 첫째로 정보의 표상 방식을 0(없음)과 1(있음)로 간략화 할 수 있기 때문이다. 이 표현방식은 동시에 '0.8의 활성화 값을 80%의 확률로 추정'하는 것과 같이 0과 1 사이의 값을 확률로서 기능할 수 있도록 한다.

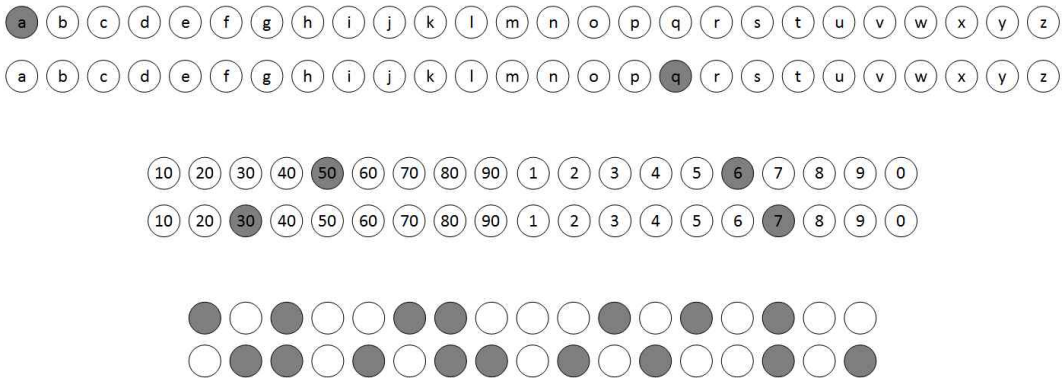
둘째로 오버플로(overflow)를 줄일 수 있기 때문이다. 연결주의 모델은 학습을 진행하면서 몇 번의 연속된 곱셈 연산을 수행하게 된다. 이 때 0과 1사이의 값은 곱한다 할지라도 값이 증가하지 않으므로 문제를 줄일 수 있게 된다. 차후에 설명하게 될 '제한된 볼츠만 머신(Restricted Boltzmann Machine, 이하 RBM)' 과 같은 일부 학습 알고리즘들은 0과 1을 넘어갈 경우 정상적으로 학습이 일어나지 못하고, 오버플로가 발생하게 된다.

셋째로 신경학적 타당성을 들 수 있다. 연결주의 모델의 기반이 되는 신경망(neural network)은 인간의 신경의 전파방식을 모사한 알고리즘이다. 인간의 신경을 구성하는 뉴런(neuron)은 역치(threshold)에 따라 그 상태가 정해지게 되며, 역치(threshold)를 넘지 않을 경우 뉴런이 활성화되지

않고, 역치를 넘어갈 경우 급격하게 활성화가 일어나 연결된 다른 뉴런에 전기적, 화학적 방식으로 활성화를 전달한다. 신경망의 표상 정보에서 0을 비활성화 1을 활성화로 정의할 경우 이와 같은 인간의 신경 전파와 유사성을 확보할 수 있게 된다.

## 2. 지역 표상(localist representation)과 분산 표상(distributed representation)

연결주의 모델에서 이용되는 표상 구조들은 크게 두 가지 형태, 지역표상(localist)과 분산표상(distributed)으로 요약될 수 있다. 지역표상은 단일한 유닛이 하나의 정보를 표상하는 구조를 의미한다. 가령 알파벳 26자를 지역표상 구조로 표현한다면 1번 유닛에 'a', 2번 유닛에 'b'와 같은 구조가 나타나게 된다(그림 1의 위와 중간). 세부적으로 보면 지역표상은 다시 두 부류로 나눠질 수 있다. 하나는 한 패턴에 오직 하나의 유닛만이 활성화되는 단일 활성화(one-hot) 구조(그림 1의 위)이다. 다른 하나는 분산표상과는 달리 지역주의 표상의 특징으로 각 유닛이 특정한 정보를 보유하지만, 층 전체의 부분집합이 특정 기능 혹은 속성을 표상하여 내부 구조를 가지게 되어, 하나 이상의 유닛이 활성화되는 속성(feature) 구조이다(그림 1의 중간).




(그림 1) 표상 방식에 따른 활성화 차이 위) 지역 표상 중 단일 활성화 표상의 예시. 각 유닛에 하나의 철자가 할당되어 있으며, 한 표상은 오직 하나의 유닛만이 활성화된다.
가운데) 지역 표상 중 속성 구조의 예시. 예시의 각 유닛에는 10진수 숫자가 할당되어 있다. 두 속성의 조합으로 0에서 100까지의 숫자 중 하나의 숫자를 표상해낸다.
아래) 분산 표상의 예시. 각 유닛에는 할당된 바가 없다. 전체 유닛의 활성화 패턴이 하나의 표상을 만든다.

이러한 속성구조의 대부분은 여러 개의 단일 활성화 구조를 복수 배치한 구조로 볼 수 있다. 가령 (그림 1)의 위와 같이, 각각 알파벳을 표상하는 26 유닛들을 4배로 배치하고 각 유닛에 1,2,3,4의 위치를 지정한다면, 이 표상 구조는 위치에 따른 26개 유닛 내에선 하나의 유닛만 1이 되고 나머지는 0이 되는 단일 활성화 구조지만, 전체 104개 유닛은 4개 유닛이 동시에 1이 되고

나머지는 0이 되어 4문자열로 구성된 단어를 표상하는 속성 구조가 만들어지게 된다.

그렇다면 분산표상은 어떤 점에서 지역 표상과 다른가? 분산표상은 하나 이상의 활성화된 유닛을 갖는 다는 점에서 지역표상 중 속성 구조와 유사하다. 하지만, 분산 표상의 경우 어떤 정보가 하나의 유닛에 존재하지 않는다. 즉, 하나의 유닛 활성화가 여러 정보 구성에 관여하며, 동시에 하나의 정보를 위해서 여러 유닛이 활성화되는 구조이다(그림 1의 아래). 이 차이를 명확하게 드러내는 설명은, 유닛의 손상으로 인한 정보의 손실 측면이다. 이러한 차이는 속성 구조를 포함한 지역 표상은 비록 층 전체의 활성화 패턴을 통해 어떤 표상을 이뤄내지만, 세부적으로 본다면 단일 유닛에 특정한 정보가 기록되어 있기 때문이다. 가령 문자열을 표상하는 지역 표상에서 '1번 위치의 a'에 대한 정보를 갖는 유닛의 활성화를 0으로 강제하여 비활성화 시킴으로써 손상(lesion)을 입히면, 해당 층은 '1번 위치의 a'에 대한 모든 정보를 소실한다. 하지만 분산표상은 '1번 위치의 a' 라는 정보를 여러 유닛이 나눠서 보관하고 있기 때문에 일부 유닛에 문제가 생겨 0으로 고정한다 할지라도 다른 유닛에 보관되어 있는 '1번 위치의 a'에 대한 부분적인 정보에 의해 '1번 위치의 a'가 불완전하게나마 표현될 수 있다. 이러한 분산 표상의 장점은 단점이 되기도 한다. 강제 비활성화를 적용했을 때, 지역 표상은 특정 정보의 손상만 가져오지만, 분산표상은 한 유닛의 활성화가 여러 개의 정보와 관련되어 있기 때문에 다른 정보에도 손상을 줄 수 있다. 예를 들어, 지역 표상의 경우'1번 위치의 a'에 대한 유닛 의 손상이 '2번 위치의 f'에 대한 정보에 영향을 주지 않지만, 분산표상에서 한 유닛의 손상은 '1번 위치의 a'와 같은 특정 정보만이 아니라 다른 정보들까지도 어느 정도의 손상을 가져 오게 된다. 한 유닛의 손상이 얼마나 많은 정보를 얼마나 심각하게 손상시키는지는 층 전체의 크기와, 정보 하나에 대한 분산의 정도에 따라 크게 달라지지만, 이 '정보의 분배'와 '손상의 분배'는 지역표상과 분산표상을 가르는 중요한 차이점이다.

### 3. 과거 연결주의 모델 연구에서의 지역 표상과 분산 표상

지역표상 구조는 다양한 연결주의 모델에서 이용되어 왔는데, 단일 활성화 구조보다는 속성 구조를 더 많이 이용해 왔다. Hinton과 Shallice(1991)은 연결주의 모델을 통해 후천적 난독증(acquired dyslexia) 환자의 증상을 모사하고자 철자 정보를 입력 받아 의미 정보를 표상하는 모델을 제안하였다. 이 연구에서 철자 정보는 각 유닛이 특정 위치의 특정 자소(letter)를 나타내는 속성 구조로 표상되었고, 의미 정보는 각 유닛 별로 특정 속성을 보유하고, 이를 조합하는 속성 구조로 표상되었다(Hinton & Shallice, 1991).

Hinton과 Shallice(1991)의 연구를 확장한 Plaut와 Shallice(1993)은, 연결주의 모델을 통해 난독증의 기전을 설명하고자 하였고, 종래 Hinton과 Shallice(1991)가 제안한 모델 구조에 의미층(semantic layer)에서 음운층(phonological layer)로 가는 경로가 추가된 모델을 제안하였다. 해당 모델에서 발

음 정보는 각 유닛이 특정 위치의 특정 음소(phoneme)를 나타내는 속성 구조였다(Plaut & Shallice, 1993).

Kawamoto(1993)가 제시한 모델 역시 일부 표상은 지역 표상 구조를 사용하였다. Kawamoto (1993)은 단어가 갖고 있는 복수의 정보를 사용하여 의미 중의성을 해결하는 방안을 연결주의 모델을 통해 보이고자 하였다. 이를 철자, 발음, 품사, 그리고 의미까지 총 4가지의 정보가 상호작용하는 구조의 연결주의 모델을 제안하였다. 연구에서 지역 표상 구조가 사용된 부분은 철자 표상으로, McClelland와 Rumelhart(1981)의 모델에서 이용된 지역표상 방식을 참고하여 이용하였다. 이 구조는 직사각형 안에 테두리와 안의 대각선 등의 각 선을 나타내는 12개 유닛을 이용하여 디지털시계와 유사한 방식으로 알파벳을 표상하는 속성 구조로서, Kawamoto(1993)의 모델은 4문자로 구성된 문자열을 표현하기 위해 48개 유닛을 이용하였다(Kawamoto, 1993; McClelland & Rumelhart, 1981).

Plaut 등(1996)은 영어 동사의 규칙/불규칙에 관한 단일경로 모형을 제안한 Seidenberg와 McClelland(1989)의 연구를 확장을 시도하였다. 이를 위해 영어 동사가 부분적으로 규칙성을 보유하고 있는 케이스인 준 규칙성(quasi-regularity) 개념을 연결주의 모델을 통해 보이고, 후천적 표면성 난독증(acquired surface dyslexia) 환자의 현상을 모사함으로써 단일경로모델을 주장하였다. 이 연구는 철자 정보와 발음 정보 모두를 지역 표상 구조로 표상하였으며, 각 유닛은 단일한 위치와 철자 혹은 발음을 나타내는 속성 구조였다. 이와 같은 구조는 후속 연구에서도 적용되어 연결주의 모델에 의미 정보를 포함시켜 단어의 읽기와 어휘판단과제를 모사하는 방안을 제안한 Plaut(1997)의 모델 또한 동일한 철자층과 음운층의 표상 구조를 사용하였다(Plaut, 1997; Plaut, McClelland, Seidenberg, & Patterson, 1996).

Harm과 Seidenberg(1999)는 발음의 획득과 난독증의 발생 메커니즘을 연결주의 모델을 통해 제안하였다. 해당 모델에서 철자층은 과거 연구들과 유사하게 각 유닛마다 위치와 철자가 할당되어 최대 8자의 문자를 입력 받을 수 있도록 하는 속성 구조였다. 동시에 음운층은 각 유닛이 단일 음소 정보를 갖는 표상 구조를 이용한 과거 연구들과 다르게, 조음 기관(articulator)을 표상하는 구조를 이용하였고, 이에 따라 몇 가지 조음 기관이 조합되어 하나의 발음 정보가 표상되는 속성 구조였다(Harm & Seidenberg, 1999).

Rogers 등(2004)는 객체의 지각적 표상과 단어 간의 상호작용에 의해 의미가 활성화된다는 고전적 주장을 검증하기 위해 연결주의 모델을 사용하였으며, 이를 위해 이름 그리고 지각적(perceptual), 기능적(functional), 사전적(encyclopedic) 정보와 같은 객체에 대한 서술 정보(verbal descriptor)와 시각적 속성(visual feature)들이 양방향으로 연결되어 있는 속성 구조를 제안하였다. 이 연구에선 표상 구조를 형성하는데 두 단계를 거쳤다. 먼저 참여자들로부터 각 카테고리에 대해 서술과 그림으로 표현하도록 함으로써, 각 카테고리가 보유하고 있는 서술 정보와 시각적 속성을 수집, 이를 통해 서술 정보와 시각적 속성 관한 원형(prototype)을 제작하였다. 그리고 이 원

형을 바탕으로 하여 각 유닛에 이진 무작위할당(0 혹은 1을 할당하는 방식)을 진행, 8개의 본보기(exemplar)를 제작하였다. 단 본보기를 제작할 때 원형이 보유한 속성에 따라 0.8, 0.2, 0으로 해당 속성이 본보기에 적용될 확률을 다르게 설정함으로써 대부분 보유하는 속성과 드물게 보유하는 속성, 보유하지 않는 속성을 구분하였다(Rogers et al., 2004).

Cree, McNorgan, 그리고 McRae(2006)가 제한한 모델의 의미층 또한 속성 구조가 사용되었다. Cree, McNorgan와 McRae(2006)는 연결주의 모델을 통해 의미표상을 처리하는데 있어 대다수의 속성보다 특히 중요한 역할을 갖는 특정 속성들이 존재함을 확인하고자 하였다. 이 연구에서 의미층은 하나의 유닛이 하나의 속성을 갖고 있으며, 이들 속성의 조합으로 의미를 표상하는 속성 구조를 사용하였다(Cree, McNorgan, & McRae, 2006).

Laszlo와 Plaut(2012)가 제안한 모델에서의 철자 정보 또한 속성 구조를 이용하였다. Laszlo와 Plaut(2012)는 연결주의 모델을 통해 사건 관련 전위(event-related potential, ERP)의 N400패턴 모사를 시도하기 위해, 철자 정보와 의미 정보를 사용하였으며, 이 중 철자 정보가 속성 구조를 이용하였다. 구체적으로는 철자 정보를 위한 입력층에서 각 유닛은 특정한 철자 위치의 특정 철자를 표상하는 형태의 속성 구조가 사용되었다(Laszlo & Plaut, 2012).

지역표상 만큼이나 분산 표상 역시 다양한 모델에서 표상 구조로서 이용되어 왔다. 위의 Plaut 등 (1996)에서 잠깐 언급되었던 Seidenberg와 McClelland(1989)는 분산 표상을 이용한 대표적인 연결주의 모델링 연구이다. Seidenberg와 McClelland(1989)는 영어 동사의 규칙/불규칙 처리에 관한 전통적인 이론인 이중경로 구조에 대해 반박하며, 단일한 경로로 해당 처리가 가능함을 보이고자 연결주의 모델을 제안하였다. 이 연구에서 분산 표상을 구성하기 위해서 사용된 개념은 트라이그램(trigram)이었다. 트라이그램은 3개의 알파벳 문자로 구성된 문자열을 지칭하는 말로서, 철자정보라면 3개의 철자를, 발음 정보라면 3개의 음소를 조합하여 구성한다. 해당 연구에서는 철자층의 각 유닛에는 철자에 대한 트라이그램 10개를 무작위로 할당하였고, 음운층의 각 유닛에는 음소에 대한 트라이그램 10개를 무작위로 할당하였다. 가령, 1번 유닛에 'hav, _ga, str, tri, rik, gum, tr, al_, pho, dal'와 같은 트라이그램들이 할당되었다면('_'는 단어의 처음과 끝을 나타내는 기호), 'have'나 'try'와 같은 단어를 표상할 때 이들 단어가 포함하고 있는 'hav', 'tr'에 의해 이 유닛이 활성화 되게 된다. 동시에 각 트라이그램은 복수의 unit에 할당될 수 있도록 하였다. 즉 위의 예시의 트라이그램들은 1번 유닛에 할당되었다 하더라도 이에 상관없이 다른 유닛에도 할당될 수 있도록 함으로써 하나의 트라이그램이 여러 유닛을 활성화 가능하게 하였다(Seidenberg & McClelland, 1989).

Kawamoto(1993) 또한 철자 (지역 표상)를 제외한 나머지에 대해 분산표상 구조를 이용하였다. 발음 표상은 48개 유닛의 각 12개 유닛이 특정 위치에 관한 정보라는 점에서는 지역표상에 가깝지만, 위치 내에서의 정보는 이진 무작위 할당으로 전체의 패턴의 조합으로 표상을 형성하는 분산표상 구조였다. 또한 품사와 의미 표상은 완전하게 이진 무작위 할당으로 무작위 할당된 분

산표상 구조였다(Kawamoto, 1993).

Plaut(1997)에 의해 제안된 모델의 의미 정보 또한 분산표상 구조를 사용되었다. 위에 언급된 다른 연구들과 마찬가지로 분산표상 구축에는 무작위 할당을 이용하였으며, 원형과 본보기 개념을 사용하였다. 이는 각 의미 표상에 대한 원형을 유닛 별로 이진 무작위 할당을 통해 구성하고, 원형 표상에 추가로 일부 유닛에 이진 무작위 할당을 적용하여 본보기를 구축하는 방식으로서, 모델은 원형을 직접 학습하는 것이 아니라 구체적 표상인 본보기를 학습에 사용하였다(Plaut, 1997).

Cree, McNorgan, 그리고 McRae(2006)가 제안한 모델의 의미층은 속성 구조의 지역 표상 구조였으나, 단어층은 분산 표상을 사용하였다. 해당 연구에선 단어층의 전체 30개의 유닛 중 무작위로 3개를 선정해 1을 할당하는 방식으로 분산표상을 구성하였다(Cree 등, 2006).

Dilkina, McClelland, 그리고 Plaut(2008, 2010)은 의미적 치매(semantic dementia) 환자들 간에 나타나는 개인차를 모사하고, 나아가 언어처리에서 전통적으로 고려되어 왔던 심상어휘집(mental lexicon)이 의미 정보와 별도로 존재할 필요성이 있는가에 대해 확인하고자, 의미 정보를 표상하는 행동/사전 정보 층, 철자 층, 음운 층, 시각 층 등이 통합적으로 존재하는 연결주의 모델을 제안하였다. 해당 연구는 시각층은 이진 무작위 할당을 통한 분산표상 구조를, 의미층은 카테고리 별 원형과 세부 속성을 추가하는 본보기를 활용한 분산표상 구조를, 철자와 발음에 대해서는 유닛의 위치는 할당되어 있으나 위치 별 패턴은 이진 무작위 할당인 분산표상 구조를 사용하였다(Dilkina, McClelland, & Plaut, 2008, 2010).

Laszlo와 Plaut(2012) 역시 제안한 모델의 의미층은 이진 무작위할당을 이용해 분산 표상을 구성하였다. 다만 해당 연구는 일부 표상은 50개 중 3개가 일부 표상은 50개 중 7개를 1로 할당하는 방식을 이용함으로써, 의미 정보의 구체성(concreteness)에 차이를 두고자 하였다.

### 4. 지역 표상과 분산 표상의 장단점

먼저 지역 표상 구조의 경우는 두 가지 장점이 존재한다. 첫째로, 지역표상 구조는 구조를 만들고 이해하는 것 양 쪽 모두가 편하다. Plaut 등(1996)이나 Harm와 Seidenberg(1999)의 예시와 같이 지역표상 구조를 사용한 연구의 경우 규칙성이 명확하므로 표상 정보는 실제 사람이 사용하는 정보와 직접적으로 대응된다. 이는 연구자가 표상 정보를 만들어내는데 많은 노력이 필요하지 않으며, 동시에 모델을 이해하고자 하는 사람들 또한 규칙성에 따라 쉽게 표상 구조를 이해할 수 있다(Harm & Seidenberg, 1999; Plaut et al., 1996).

둘째로, 모델의 학습 결과를 확인하는 것 또한 쉽다. 가령 모델이 산출한 어떤 값에 대해 정오판단을 시도할 때, 지역표상이라면 활성화/비활성화에 대한 기준을 정하고(ex. 0.8과 0.2) 목표로 한 유닛들만 활성화되고, 나머지 유닛들은 그보다 낮다는 것을 확인하는 것만으로도 해당 시

도가 정말 원하는 것을 올바르게 표상하였지 여부를 확인할 수 있다. 하지만 분산표상에서 이를 검증하기 위해선 출력 값을 모든 가능한 목표 패턴들과 비교하여 유사성을 계산한 뒤 본래의 목표 패턴이 가장 유사하다는 것을 확인해야 한다. 학습된 패턴이 작을 경우 이 비용은 큰 문제가 되지 않지만, 패턴이 증가하면 증가할수록 한번 정오판단을 시도하는데 들어가는 비용은 기하급수적으로 증가하게 된다. 이와 같이 결과 확인이 쉬운 것은 지역 표상의 매우 큰 장점이다.

하지만 동시에 지역표상 구조는 몇 가지 단점도 존재한다. 첫째로, 표상하고자 하는 정보가 커질수록 사용되어야 하는 유닛도 증가한다. 위에서 언급한 알파벳을 예시로 든다면, 1문자를 표현하는데 26개의 유닛이 필요하며, 최대 10자로 된 단어를 표현하기 위해서는 $26 \times 10 = 260$개의 유닛이 필요하게 된다. 그리고 연결주의 모델에서 사용하는 알고리즘에 의해, 유닛의 개수가 증가하는 것은 모델의 학습에 걸리는 시간이 기하급수적으로 증가한다는 것을 의미한다. 비록 병렬처리기술이나 알고리즘의 개선 등을 통해 이전보다 많은 부분이 개선되었으나, 대규모의 자료를 사용할 경우 여전히 문제가 남게 된다.

둘째로, 최종적인 출력층을 계산 시에 제약이 발생한다. 지역표상에서 출력층을 계산 시에 흔히 이용되는 함수로는 소프트맥스와 시그모이드의 두 종류가 존재한다. 소프트맥스는 층의 전체 정보에서 가장 유력하다고 생각되는 정보를 확률적으로 나타내는 함수이다. 계속해서 알파벳을 예로 든다면, 입력된 정보가 26개의 알파벳 중 어느 알파벳인지 알아내는 단일 활성화 구조에선 최적이라고 할 수 있다. 하지만 만약 입력된 정보가 문자열에 대한 정보라면 소프트맥스는 속성 구조 기반에서 각 자리에 대한 알파벳에 대하여 각각의 확률을 모두 구하거나 혹은 모든 가능한 패턴에 대하여 단일 활성화 구조에서 가장 유의한 패턴을 선택하거나 의 둘 중 하나의 가능성 만이 남게 된다. 즉, 최대 10자의 문자열 정보를 입력 받는다면 연산횟수 또한 10배가 되는 것이다.

다른 한편으로 시그모이드 함수를 사용할 경우 소프트맥스에서의 문제는 발생하지 않는다. 하지만 지역표상에서의 시그모이드 이용은 목표(target) 정보의 유닛 활용도에 대한 비대칭성에 기인한 수렴문제가 발생하게 된다. 이는 유닛 활용도의 비대칭성에 의한 일부 정보가 학습되지 않는 현상으로 정의할 수 있다. ‘x’로 시작하는 영어 단어는 이에 대한 대표적인 예시이다. 알파벳 단어를 학습하기 위해 지역표상 구조로 구성한다면, 첫 26개의 유닛은 단어의 첫 번째 문자를 나타낸다. ‘x’로 시작하는 단어는 상대적으로 매우 한정적이기 때문에, 첫 번째 위치의 ‘x’를 나타내는 유닛은 극단적으로 0일 확률이 높게 된다. 이러한 데이터의 분포 상 비대칭성은 모델이 ‘해당 유닛은 언제나 0인 경우가 정답에 근접한다’라는 잘못된 방향성을 제시하게 되고, 결과적으로 모델은 첫 번째 위치의 ‘x’의 모든 정보에 대한 학습에 실패하게 된다. 이 문제의 핵심적인 부분은 학습하는 세대(epoch)의 증가나 은닉층(hidden layer) 증가, 학습률(learning rate)의 감소와 같은 학습능력의 향상으로 해결되지 않는다는 점에 있다. 유희조, 남기춘, 그리고 남호성(2015)의

경우 의미 정보 표상에서 이와 같은 문제를 해결하기 위해 활성과 억제에 대한 학습률을 나누는 방안을 적용하였으나, 이는 학습률 파라미터를 찾는데 추가적인 노력이 요구되며, 일반적인 모델에 적용하기에는 어려움이 있는 방법이다(유희조, 남기춘, 남호성, 2015).

분산표상은 이와 같은 지역표상의 제한점들을 해결하는데 적절하다. 분산표상이 갖고 있는 장점은 첫째로, 분산표상은 층 전체의 패턴을 통해 정보를 표상하므로, 표상 개수에 대한 유닛의 개수가 비교적 자유롭다는 점이다. 설사 패턴을 이진수(binary)로 표상한다고 할지라도 100개의 $2^{100}$개의 정보가 표상 가능하다. 물론 학습에는 유사성 정보가 요구되므로 실제 표상 가능 개수는 이보다는 훨씬 작다(가령 철자 정보를 바탕으로 발음 정보를 학습하는 모델을 가정할 때 철자 정보의 표상에서 단어 'have'와 'halt'가 일체의 관련성이 없다면, 모델은 'ha'가 어떻게 발음되는가에 대한 어떠한 규칙성도 학습할 수 없을 것이다). 하지만 이와 같은 규칙성 문제가 있음에도 불구하고, 분산 표상을 통해 표상을 구축할 경우 모델이 사용하는 전체 표상 개수와 이를 표현하기 위해 요구되는 유닛의 숫자는 지역표상과 달리 선형적으로 증가하지는 않는다.

둘째로, 상대적으로 비대칭성에 따른 수렴문제에서 자유롭다. 모든 유닛 전체가 표상에 이용되는 분산표상 구조는 같은 시그모이드를 사용하더라도 지역 표상에서 나타날 수 있는 비대칭성에 따른 수렴문제의 영향을 덜 받을 수 있다.

이러한 장점에도 불구하고, 많은 선행연구들은 지역표상을 이용한 모델들을 제안해왔다. 이는 규칙성이 확보된 분산표상의 구축 자체가 힘들기 때문이다. 사실상 대부분의 분산표상은 정보의 무작위적 분산 할당을 통해 이루어진다. 예를 들어, Seidenberg와 McClelland(1989)의 모델에서 이용된 분산표상 구조는, 단일한 정보(트라이그램 1개)에 대해 복수의 유닛이 활성화되고, 하나의 유닛이 복수의 정보(10개 트라이그램 할당)에 관련된다는 점에서 적절한 분산표상 구조라고 할 수 있다. 하지만 동시에 각 유닛에 트라이그램을 할당하는 방식이 무작위이기 때문에, 모델이 최종적으로 어떤 정보를 표상하였는지 확인이 어렵다는 문제가 있다. 즉, 트라이그램을 많이 공유한 단어 간에 비슷한 표상구조가 나타난다는 것까진 알 수 있으나, 출력층의 활성화 패턴만 보고 모델이 어떤 정보를 산출했는가에 대해서는 알지 못한다. 이는 직접적으로 실제 정보와 유닛이 직접적으로 대응되는 지역 표상 구조에 비해 크게 대비되는 부분이라고 할 수 있다

〈표 1〉 지역 표상과 분산 표상의 장단점

| | 지역 표상 | 분산 표상 |
|---|---|---|
| 구성 및 이해의 난이도 | 쉬움 | 어려움 |
| 학습 결과의 확인 | 쉬움 | 어려움 |
| 요구되는 유닛의 크기 | 정보에 비례하여 증가 | 적은 유닛을 사용 |
| 소프트맥스 출력층 계산 | 패턴에 따라 연산량 증가 | 사용되지 않음 |
| 시그모이드 출력층 계산 | 비대칭성에 의한 학습실패 가능성 | 비대칭성 문제에서 자유로움 |

(Seidenberg & McClelland, 1989).

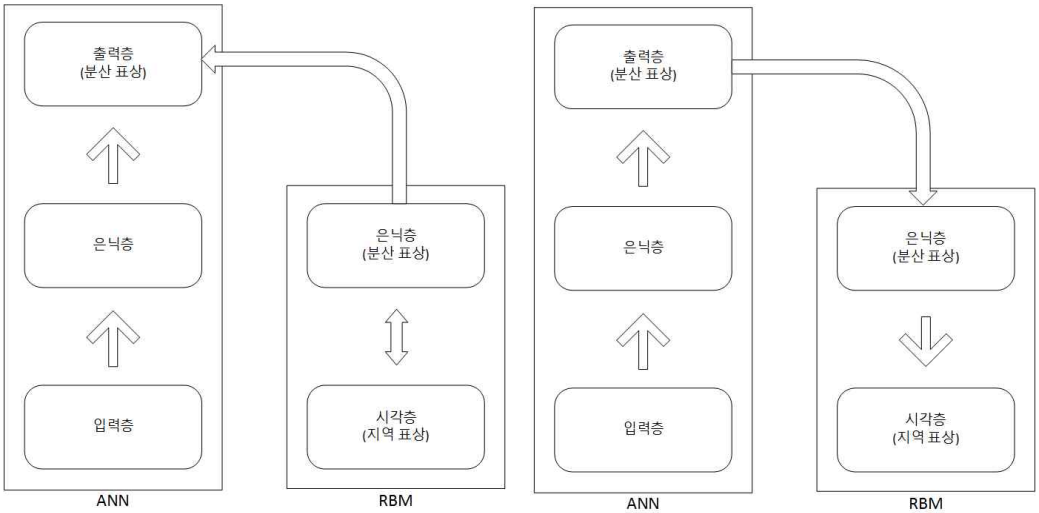
Kawamoto(1993), Laszlo와 Plaut(2012)의 의미 정보 표상에서 이용된 완전하게 무작위화 된 구조나, Plaut(1997) 등과 같은 원형과 본보기 구조 또한 문제는 크게 다르지 않다. 표상 구조를 형성할 때 무작위를 이용할 경우 어떻게 하더라도 규칙성의 손실이 나타나게 된다. 심지어 이 경우 속성이 공유되지 못하므로 표상 간 관계성까지 손상을 입게 된다. 예를 들어 현실세계에서 'bus'와 'taxi'는 교통수단이라는 점과 같이 일부 속성이 공유된다. 하지만 해당 연구의 분산표상은 이와 같은 속성의 공유가 일절 일어날 수 없다는 문제가 발생하게 된다(Kawamoto, 1993; Laszlo & Plaut, 2012; Plaut, 1997).

최근 들어 의미 정보에 대해서는 한 가지 큰 진전이 있었다. Word2Vec은 문장(sentence) 정보를 기반으로 학습하여, 각 단어들을 고정된 크기의 벡터(vector)로 표현해주는 알고리즘이다. Word2Vec으로 구현된 단어 정보들은 의미적인 유사성 및 관계성을 갖는 것으로 알려져 있으며, 이는 의미처리에 관해서는 위에서 언급한 종래 분산표상이 갖고 있던 문제점을 해결하는 중요한 발전으로 여겨지고 있다 (Mikolov, Corrado, Chen, & Dean, 2013). 하지만, 이는 의미 정보에 한정된 이야기로, 언어 처리에서 사용되는 다른 정보인 철자 정보나 발음 정보에 대해서는 더 많은 연구가 필요하다.

## 5. 제한된 볼츠만 머신(restricted Boltzmann machine)

RBM은 분산표상의 이러한 문제점에 대한 한 가지 해결 방안이 될 수 있다. RBM은 입력 값과 목표 값이 동일하며, 오토 인코더로서 입력된 정보를 온전하게 환원해 내도록 학습을 진행하는 알고리즘이다. 본 연구에서 초점을 맞춘 곳은 RBM 알고리즘의 은닉층이다. RBM에서 정보의 환원 과정에서 활성화되는 은닉층 각 유닛들의 값은 입력층의 정보에 대해 추상화된 정보가 표상된다. 그리고 만약 은닉층의 유닛 개수를 입력층보다 줄여서 학습을 진행할 경우, 더 적은 유닛으로 더 많은 정보를 표상해내야 하므로 정보의 압축이 일어나게 된다. 결과적으로 RBM을 통해 도출된 충분히 압축된 정보를 목표 패턴으로 사용할 경우 표상 구조의 크기를 줄일 수 있으며, 동시에 지역표상 구조에서 나타나는 목표 정보의 비대칭성을 해소할 수 있게 된다 (그림 2).

동시에 이 방식은 과거 분산표상이 갖고 있던 문제 또한 해결 할 수 있게 된다. 은닉층 전체가 입력정보의 재구성에 사용되므로 은닉층에 표상된 정보는 분산표상이 되며, 이 정보는 입력정보를 기반으로 추상화되었으므로, 유사 표상 간에 연관성을 갖게 된다. 거기에 (그림 2)의 우측과 같이, 모델에서 출력된 값을 분산표상 제작에 사용된 RBM 모델에 집어넣음으로써, 지역표상과 같이 모델이 실제로 어떤 표상을 해냈는가에 대해서 직접 확인하는 것 또한 가능하게 된다(Hinton, 2010; Smolensky, 1986).




(그림 2) 분산 표상 구축과 활용을 위한 RBM 적용방식왼쪽) 제한된 볼츠만 머신(restricted Boltzmann machine, RBM)으로 분산 표상을 추출하는 흐름. 추출된 분산 표상은 인공신경망(artificial neural network, ANN) 모델의 학습을 위한 목표 패턴으로 이용된다.
오른쪽) 분산 표상으로 구성된 출력 패턴으로부터 지역 표상 패턴을 추출해내는 과정. RBM에서의 가중치는 양방향(bi-directional)이므로, 분산 표상 구조의 출력 값에 RBM의 가중치를 곱함으로써 지역 표상으로 변환될 수 있다.

## 6. 목 표

본 연구는 RBM을 이용한 지역 표상의 분산 표상화가 두 표상의 문제점을 해결할 수 있는지 검증하고자 하였다. 이를 위해 RBM 알고리즘으로 지역표상으로 이뤄진 단어들에 대한 분산표상 정보를 도출하고, 이 정보들을 이용하여 인공신경망(artificial neural network, 이하 ANN) 모델을 학습시켰다. 그리고 이하의 두 가지 분석을 실시하여 도출된 정보에 대해 종래 표상구조들이 갖고 있던 문제점이 해결될 수 있었는지 여부를 확인하고자 하였다. 첫째, RBM에 의해 도출된 분산 표상에 대해 유사 표상 간의 연관성을 보여주는 분석을 실시한다. 둘째, 지역표상을 이용하여 학습된 ANN에서 도출된 정보와, 분산표상을 이용하여 학습된 ANN에서 도출된 정보에 RBM 정보를 이용하여 수복된 정보(그림 2의 오른쪽)를 비교하여 목표정보의 비대칭성 문제를 해결될 수 있음을 보인다.

# 방 법

## 1. 분산 표상 구축을 위한 RBM 모델

서론에서 언급한 바와 같이 지역 표상을 분산 표상으로 변환하기 위한 RBM 모델을 구성하였다. 분산 표상을 구성하기 위한 RBM은 시각층(visual layer) 260개 유닛, 은닉층(hidden layer) 50개 유닛으로 구성되었다. 시각층의 유닛 개수는 $26 \times 10 = 260$으로 결정되었고, 이를 통해 RBM 모델은 최대 10자로 이뤄진 단어까지 학습이 가능하였다. 은닉층의 유닛 개수는 260개 보다 적을 것이라는 조건을 걸고 몇 번의 선행 테스트를 통해 결정되었다. 260개보다 적다는 조건의 이유는 첫째로 정보를 압축하기 위해서, 둘째로 이후 ANN의 부담을 줄이기 위해서였다. 서론에서 언급한 바와 같이 지역표상의 단점 중 하나는 유닛의 개수가 과도하게 커질 수 있다는 점이므로, RBM의 효용성 증명을 위해서는 260개보다 작은 것이 전제될 필요가 있었다. 이에 따라 선행 테스트는 200, 150, 100, 50개의 은닉층 유닛 개수를 설정하여 10,000 epoch의 학습을 진행하였다. 선행 테스트 결과 50개 유닛을 사용했을 때 모든 유닛을 충분히 활용한다고 판단하여 최종적으로 50개 유닛으로 결정되었다.

학습에 사용된 패턴은 총 18,006 개의 길이 1에서 10까지의 단어를 기반으로 구성되었으며, 본 연구에서 대소문자, 빈도 그리고 품사는 고려되지 않았다.

RBM의 학습은 총 100,000 세대(epoch) 동안 이루어졌고, 학습률은 0.0001이었다. 학습 방식은 한 번의 학습에 18,006개 전부를 학습하는 배치(batch) 방식으로 이루어졌다.

모델은 CPU i7-6700k, RAM 64GB, 그래픽카드 NVIDIA Titan X Pascal, OS Linux Mint 18 Sarah, CUDA 8.0에 Python 3.5 및 Tensorflow 0.11RC 환경에서 학습되었다.

## 2. 표상 간 차이를 확인하기 위한 ANN 모델

지역 표상과 분산 표상 간의 차이를 확인하기 위해 하나는 출력층이 지역표상인 모델, 다른 하나는 RBM에 의해 도출된 분산표상인 모델로 두 가지 형태의 ANN 모델이 구성되었다.

두 모델 모두 입력에 철자 표상을 사용하였다. 이는 입력되는 표상의 영향을 차단하기 위해서였다. 발음 표상이나 의미 표상과 같은 다른 정보를 사용할 경우, 입력되는 정보 자체로 인한 편향(bias)이 발생할 가능성이 있다. 하지만 입력되는 정보와 출력되는 정보를 동일할 경우 1:1로 직접적으로 대응하므로 둘 간의 관계성이 명확하게 되며, 학습이 제대로 진행되었는가를 명확하게 판단 할 수 있다고 보았다. 따라서 두 모델 모두 입력층은 최대 10개의 문자를 표현할 수 있는 260개 유닛의 지역 표상으로 구성되었다.

두 모델 모두 은닉층은 100개의 동일한 유닛 개수로 구성되었다. 일반적으로 ANN의 연산능

력은 각 유닛들 간의 연결(connection)의 개수에 의해서 결정된다. 그리고 모델을 설계함에 있어서 입력층과 출력층은 고정되므로 많은 경우 모델의 연산 능력은 은닉층의 크기로 조절된다. 따라서 두 모델의 명확한 비교를 위해 은닉층의 크기 역시 동일할 필요가 있었다.

출력층의 경우 지역 표상은 260개, 분산 표상인 경우 50개였다. 지역 표상의 경우 입력층과 동일하게 10개 문자가 표현될 필요가 있었으며, 분산 표상의 경우 RBM에서 도출된 적정한 표상 크기에 맞추었다. 일부 문제는 출력층의 크기의 차이에 의해서 실질적으로 두 ANN 모델의 연산능력에 차이가 발생한다는 점이었다. 다만 이는 상대적으로 많은 유닛을 갖고 있는 지역 표상 모델이 더 나은 연산능력을 갖고 있다는 의미이므로, 개선된 분산 표상의 성능에 초점을 맞춘 본 연구의 방향성에는 지장이 없다고 판단하여 그대로 진행되었다.

학습에 사용된 패턴은 RBM에서 사용된 것과 동일하였다. 즉 출력층이 지역표상인 모델은 각 입력 값과 목표 값이 동일한 패턴이었으며, 분산표상인 모델은 입력 값은 지역표상인 모델과 동일하나 목표 값은 분산표상인 형태였다.

ANN의 학습은 총 100,000 세대 동안 이루어졌고, 학습률은 0.0001이었다. 학습 방식은 한 번의 학습에 18,006개 전부를 학습하는 배치 방식으로 이루어졌다.

학습 환경은 RBM과 동일하였다.

## 결 과

### 1. RBM모델을 통한 분산표상 구성 결과

#### 1) 각 Unit의 최솟값 & 최댓값 범위

전체 18,006개의 패턴에 대해 각 50개 유닛의 최솟값과 최댓값, 그리고 평균값을 확인하였다. 이는 분산표상시의 각 유닛의 활용도를 확인하기 위해 중요한 관측값이라 할 수 있다. 만약 구축된 분산표상에서 어떤 유닛이 모든 패턴에 대해 비슷한 값을 갖는 경우, 해당 유닛은 패턴을 구성하는데 거의 기여하는 바가 없다고 판단할 수 있다. 가령 모든 패턴이 0.000001에서 0.000002 사이라면, 0.0에서 1.0을 표현할 수 있음에도 불구하고 매우 제한적인 범위의 값만이 사용되었으므로 이 유닛은 표상에 기여하는 바가 거의 없다는 것을 의미한다.

확인 결과 50개 최솟값 내에서의 최댓값은 1.08E-09였으며, 50개 최댓값 내에서의 최솟값은 1.0이었다. 각 유닛의 전체 패턴에 대한 활성화 값 평균은 0.442에서 0.674 이었다. 이와 같은 결과는 50개 유닛 전부가 정보 표상에 이용되었음을 의미한다.

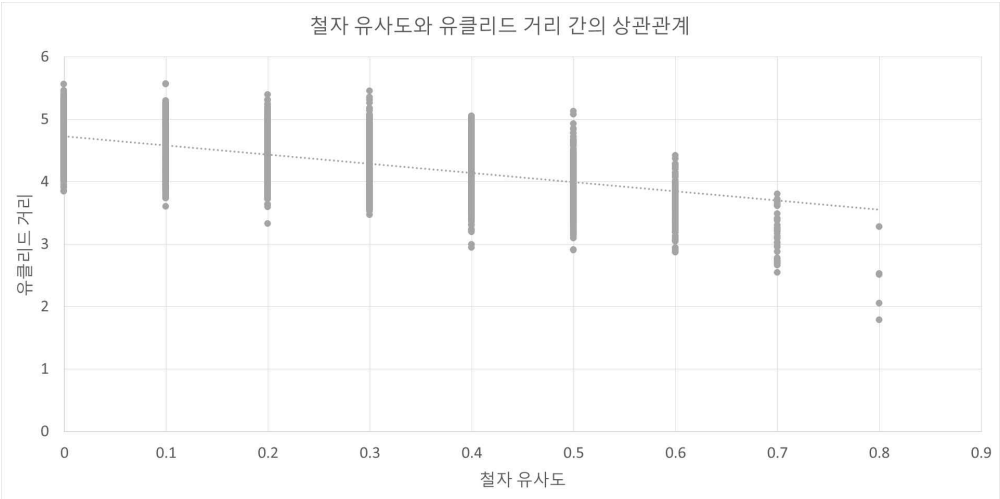
2) 구성된 분산 표상 간의 유사성 확인

두 단어 간의 철자유사성과 분산표상의 패턴 유사성 간의 상관관계를 확인하였다. 철자 유사성은 같은 위치에 같은 철자가 있을 때마다 0.1씩 더하였으며, 만약 같은 위치가 공백일 경우에도 0.1이 더해졌다. 가령 'line'과 'lane'은 세 철자가 같고 똑같이 6자가 공백이므로 $0.3+0.6=0.9$, 'although'와 'anchorite'는 세 철자가 같고 1자가 똑같이 공백이므로 0.4가 되었다. 이에 따라 완전하게 동일한 단어일 경우 1.0이 되었으므로, 적어도 1개의 자소가 다른 두 단어의 철자유사성 최댓값은 0.9였다.

다른 한 편으로 분산 표상 패턴 유사성은 이하의 공식으로 유클리드 거리를 통해서 도출되었다.

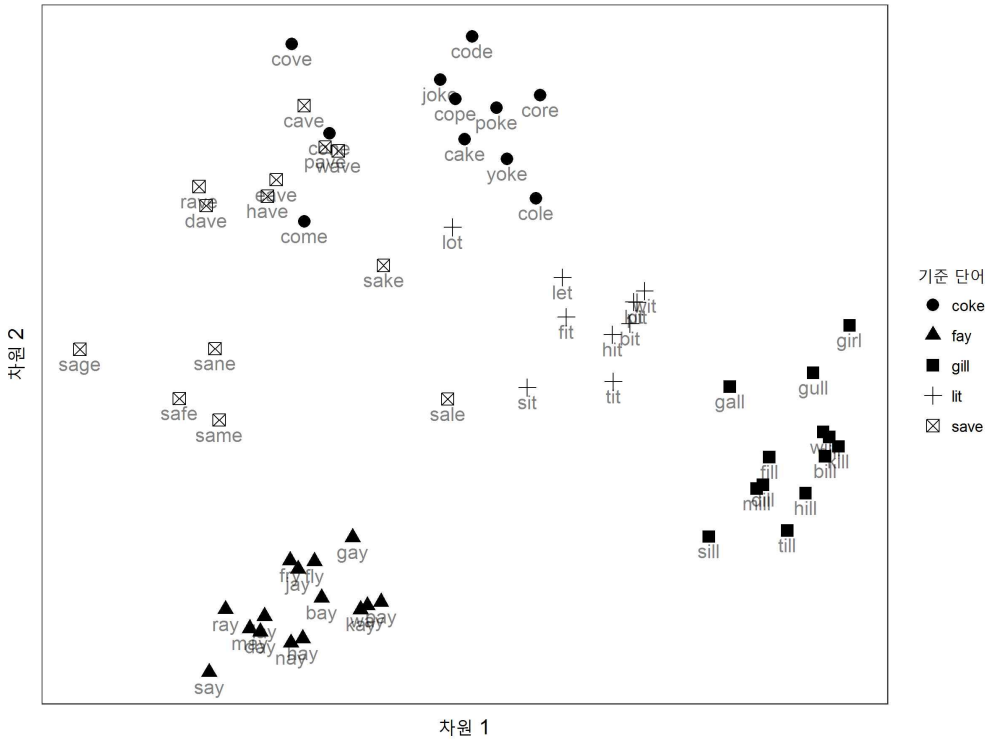
$$E.D = \sqrt{\sum(pattern_1 - pattern_2)^2}$$

다만 분석 상에서 전체 ${}_{18,006}C_2 = 162,099,015$개의 전체 단어 간의 거리를 구하는 것은 너무 많은 연산을 요구하였으므로 18,006개 단어 중 100개를 무작위로 선정하여 선정된 단어들 간의 거리를 도출, 총 4950개의 유클리드 거리와 철자유사성 간의 상관관계를 비교하였다. 상관분석 결과는 $r$=-0.608로 $p$<0.001수준에서 유의하였다 (그림 3).




(그림 3) 두 단어의 유클리드 거리와 철자 유사성 간의 상관관계. 두 단어가 유사할수록, 분산 표상 패턴의 유클리드 거리 또한 줄어들고 있다.

추가로 일부 10개 이상의 이웃을 가진 단어 중 무작위로 5 단어를 선정, 해당 단어들의 이웃 단어들에 대한 다차원척도(Multidimensional Scaling, MDS) 그래프를 유클리드 거리 기준으로 작성하였다(그림 4). (그림 4)에서 보여주는 것과 같이 한 단어의 이웃 단어들은 다른 단어의 이웃들




(그림 4) 임의로 선택된 5단어 coke, fay, gill, lit, save의 이웃 단어들에 대한 다차원 척도 그래프. 각 단어들의 이웃 단어들끼리 모여 있으며, 기준 단어와 다른 철자의 위치에 따라 이웃단어들 내의 밀집 역시 영향을 받고 있다.

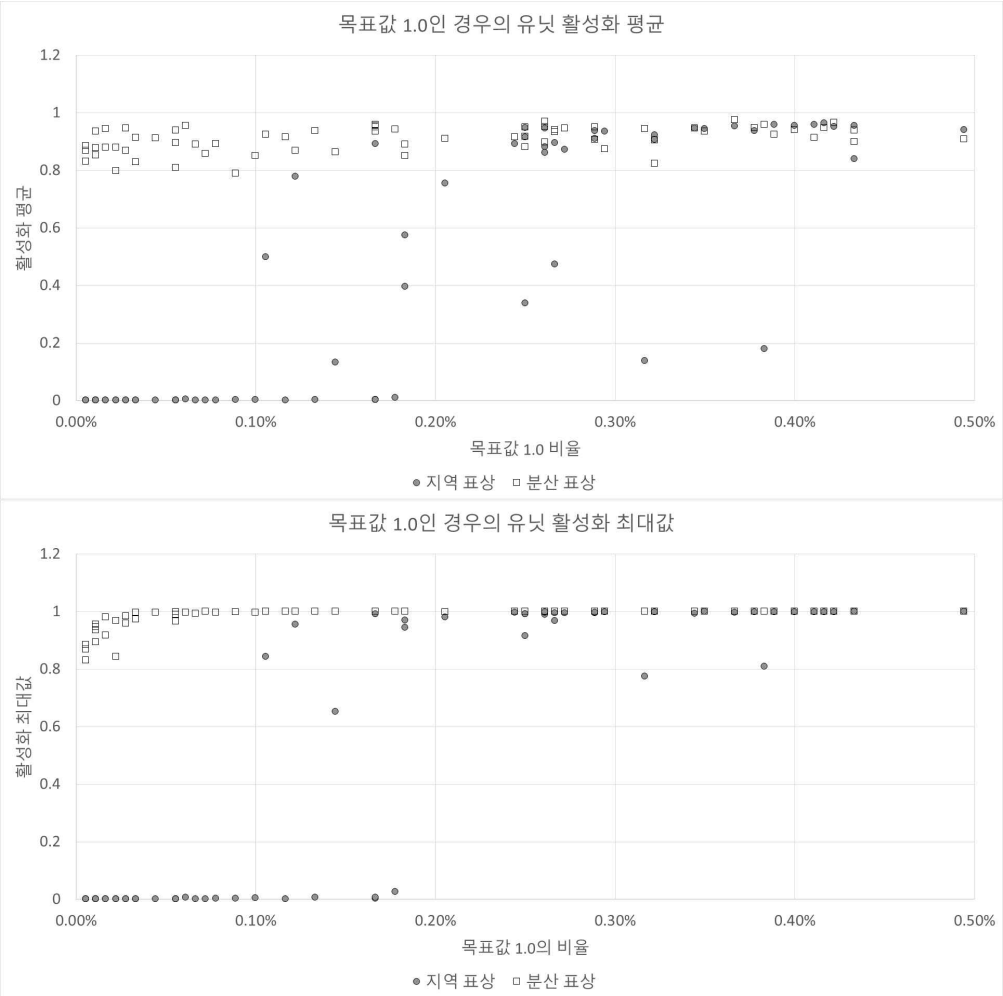
보다 밀집되어 있다. 또한 save의 이웃 단어들에서 v가 다른 집단과 s가 다른 집단의 밀집 위치가 약간씩 차이가 나는 것처럼, 철자가 다른 위치에 따라서 이웃 단어들 간의 거리 또한 다르게 나타남을 확인 할 수 있다.

## 2. 지역 표상과 분산 표상 간의 활성화 패턴 비교

목표정보의 비대칭성 문제를 확인하기 위해 두 ANN 모델의 활성화 패턴을 비교하였다. 우선 지역표상을 사용한 모델의 목표값 패턴들 중 활성화 비율이 0.5%이하인 극단적 비대칭성을 보유하고 있는 유닛들에 대해 그 출력 분포를 분석하였다. 비율이 0%인 7개의 유닛(즉, 어떤 패턴에서도 사용되지 않은 유닛)를 제외하고 총 67개의 유닛을 선출하였다. 구체적으로는 18,006개 단어 패턴 중 1개 이상 90개 미만의 패턴만이 해당 유닛의 목표값을 1로 할당한 경우가 이에 해당하였다. 이후 각 유닛 별로 목표값이 1인 패턴들을 입력하고, 해당 패턴들에 대한 유닛의

출력값을 확인, 그 평균 및 최대 활성화값을 확인하였다. 즉, 확인하고자 한 출력값의 이상적인 값은 평균과 최대값 모두 1이며, 둘 다 1에 근접할수록 올바른 산출이 가능하였다는 것을 의미한다.

분석 결과 지역표상의 경우 전체 목표 패턴의 0.1% 이하에 대해서만 활성화 비율을 가지는




(그림 5) 목표 값이 1.0인 경우에서 유닛들의 실제 출력값. x축은 전체 18,006개 패턴 중 해당 유닛이 1.0의 목표값을 가지는 비율을 의미한다.
위) 목표 값이 1.0인 경우에서 각 유닛들의 평균 출력값. 지역 표상을 사용한 ANN에서는 매우 낮은 비율을 가진 유닛들이 정상적으로 활성화되지 않고 있다. 반면에 분산 표상을 사용한 뒤에 해당 출력값을 RBM을 통해 지역 표상으로 변환한 경우, 낮은 비율의 유닛들도 일정 수준 이상의 활성화를 보이고 있다.
아래) 목표 값이 1.0인 경우에서 각 유닛들의 최대 출력값. 평균 출력값과 비슷한 결과를 보이고 있다.

모든 유닛이 0.01 미만의 평균 출력값을 가짐으로써 제대로 활성화되지 못하였고, 0.18% 까지도 활성화되지 못하는 유닛이 발생하였다. 반면 분산표상을 통해 학습하고 결과물을 RBM 모델로 복구한 경우는 가장 낮은 확률인 0.006% 미만의 활성화 비율(18,006개 중 단 1개만이 활성화 패턴)을 가진 유닛 3개도 0.8이상의 활성화 값을 도출하였다.

## 논 의

본 연구는 RBM을 이용하여 지역 표상을 분산 표상화하는 것의 가능성을 검증하고, 도출된 분산표상이 지역표상과 어떤 차이를 갖는지를 확인하고자 하였다. 이를 위해 지역표상으로 이뤄진 철자 정보를 RBM 알고리즘을 이용하여 분산표상 정보로 변환하고, 이 정보들을 이용하여 인공신경망(artificial neural network, 이하 ANN) 모델을 학습시켰다. 결과적으로 RBM에 의해 도출된 분산표상은 유사 표상 간의 연관성 표현해 낼 수 있었으며, 지역 표상이 갖고 있던 문제점인 목표 정보의 비대칭성 문제를 해결할 수 있었다. 또한 분산표상으로 산출된 출력값을 지역표상으로 재변환하여 과거 분산표상이 갖고 있던 제한점을 개선해냈다.

### 1. 유사성 표현의 원리

무작위의 할당과는 다르게 RBM을 통해 분산패턴을 도출하는 경우, 동일한 입력패턴은 동일한 Weight에 의해 Hidden activation이 계산되게 된다. 즉 부분적으로 동일한 단어의 경우 해당 부분에 대해서는 동일한 Weight를, 그렇지 않은 부분에 대해서는 다른 Weight를 통해 패턴이 도출되므로, 철자유사성의 정도와 RBM을 통해 도출된 패턴들의 유클리드 거리는 필연적으로 상관관계를 갖게 된다.

### 2. 목표 값의 비대칭성 문제 극복

분산표상이 비대칭성 문제가 제거 될 수 있었던 가장 근본적인 요인은 표상정보가 압축되어 있었다는 점이다. 지역표상과 다르게 모든 유닛들은 고르게 일정수치의 활성화 정보를 갖게 되고 이는 ANN에서 특정 정보가 소실되는 것을 막을 수 있도록 하였다.

또한 분산표상을 도출해내는 알고리즘으로 RBM을 사용한 것 역시 비대칭성 문제 해결에 중요한 요인 중 하나라고 할 수 있다. 오토 인코더(auto-encoder)는 비단 RBM만이 아니라 ANN과 같이 다양한 알고리즘을 통해 구현될 수 있다. 하지만 RBM 알고리즘 수준의 강한 복원력을 갖지 못할 경우 최초로 형성되는 분산표상에서 비대칭성 문제가 발생하였을 것이다. 이는 3.2의

결과에 의해 추론할 수 있다.

2.2에서 언급한 바와 같이 3.2에서 지역표상을 사용한 ANN 모델은 입력과 출력이 동일한 지역표상 구조였으므로, 일종의 오토 인코더라고 할 수 있다. 그리고 3.2의 결과에서 지역표상을 사용한 ANN은 을 통해 재산출된 지역표상 정보는 비대칭성에 매우 취약하다는 것을 확인할 수 있었다. 목표값이 1임에도 불구하고 최대값과 평균이 0에 가깝다는 것은 학습된 모델이 해당 정보를 표현해내지 못했다는 것이며, 이는 다시 말해서 모델이 그 정보를 복원해낼 수 없다는 것을 의미한다.

물론 3.2에서 사용된 지역표상의 오토인코더 모델이 구성에 실패한 영역이 입력층과 은닉층 사이인지, 아니면 은닉층과 출력층 사이인지, 그것도 아니면 둘 다인지는 알 수 없다. 하지만 입력층과 은닉층 사이가 정상적으로 구성되지 못했다면 분산 표상으로 이용되어야 할 은닉층의 활성화값을 신뢰할 수 없으며, 은닉층과 출력층 사이가 정상적으로 구성되지 못했다면 차후 지역 표상으로의 수복이 불가능하다는 것을 의미한다. 즉 어느 쪽이든 모델의 비대칭성 문제가 발생한 시점에서 ANN의 복원력으로는 한계가 있다고 할 수 있다.

### 3. 초극단적 비대칭성 문제

그림 5를 보면 분산 표상에서 유닛 활성화 값이 1에 근접하지 못하는 구간이 있음을 확인 할 수 있다. 비록 결론에서 말한 바와 같이 가장 극단적인 경우조차 0.8이상의 활성화 값을 보였지만, 이와 같이 활성화 값이 감소되는 구간이 있다는 것 자체가 분산표상 구조 또한 목표값 비대칭성 문제에서 자유롭지 못하다는 것을 의미하는 것은 아닌가? 물론 가능성은 있다. RBM의 복원력은 강력하지만, 오차에 기반을 두어 활성화 값을 수정하는 기계학습 알고리즘들은 이와 같은 문제에서 완전하게 자유로울 수는 없다. 다만 그럼에도 불구하고, 분산표상 구조가 충분히 안정적이라고 말 할 수 있는 것은 분산표상에 의해 도출된 활성화 최대값이 선형적이라는 점에 있다. 그림 5에서 보는 바와 같이 활성화 비율에 따라 극단적인 값차이를 나타낸 지역표상과는 다르게, 분산표상은 선형적으로 값이 하락하였다. 분산표상은 정보를 전체 층에 분산하여 표상함으로써, 해당 정보가 일부 소실된다 하더라도 나머지 층에서 정보를 복원할 수 있다. 이에 따라 여전히 RBM을 이용해 구성된 분산표상은 비대칭성 문제에 대한 해결책이 될 수 있다.

## 결 론

서론에서 말한 바와 같이 연결주의 모델을 구성함에 있어서 연구자는 모델의 구조 뿐 만 아니라 표상의 구조 역시 충분한 고려가 필요하다. 그리고 이에 대해 본 연구의 결과는 표상의 구

조를 결정함에 있어서 매우 강력한 선택지를 제시하였다고 판단된다. 이를 통해 차후 연결주의 모델의 연구에 있어서 다양한 개선이 이뤄질 수 있을 것으로 기대한다.

## 참고문헌

유희조, 남기춘, 남호성. (2015). 속성적 관점에 기반한 단어 재인시의 의미 처리 가능성 연구: 연결주의 모델링, **한국심리학회지: 인지 및 생물, 27**, 613-638.

Cree, G. S., McNorgan, C., & McRae, K. (2006). Distinctive features hold a privileged status in the computation of word meaning: Implications for theories of semantic memory. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *32*, 643-658. https://doi.org/10.1037/0278-7393.32.4.643

Dilkina, K., McClelland, J. L., & Plaut, D. C. (2008). A single-system account of semantic and lexical deficits in five semantic dementia patients. *Cognitive Neuropsychology*, *25*, 136-164. https://doi.org/10.1080/02643290701723948

Dilkina, K., McClelland, J. L., & Plaut, D. C. (2010). Are There Mental Lexicons? The Role of Semantics in Lexical Decision. *Brain Research*, *1365*, 66-81. https://doi.org/10.1016/j.biotechadv.2011.08.021.Secreted

Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: insights from connectionist models. *Psychological Review*, *106*, 491-528. https://doi.org/10.1037/0033-295X.106.3.491

Hinton, G. (2010). A Practical Guide to Training Restricted Boltzmann Machines A Practical Guide to Training Restricted Boltzmann Machines. *Computer*, *9*, 1. https://doi.org/10.1007/978-3-642-35289-8_32

Hinton, G., & Shallice, T. (1991). Lesioning an attractor network: investigations of acquired dyslexia. *Psychological Review*, *98*, 74-95. https://doi.org/10.1037/0033-295X.98.1.74

Kawamoto, A. (1993). Nonlinear dynamics in the resolution of lexical ambiguity: A parallel distributed processing account. *Journal of Memory and Language*, *32*, 474-516. Retrieved from http://psycnet.apa.org/psycinfo/1994-07885-001

Laszlo, S., & Plaut, D. C. (2012). A neurally plausible parallel distributed processing model of event-related potential word reading data. *Brain and Language*, *120*, 271-281. https://doi.org/10.1016/j.bandl.2011.09.001.A

McClelland, J., & Rumelhart, D. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review*, (September). Retrieved from http://psycnet.apa.org/journals/rev/88/5/375/

Mikolov, T., Corrado, G., Chen, K., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. In *Proceedings of the International Conference on Learning Representations* (pp. 1 – 12).

https://doi.org/10.1162/153244303322533223

Plaut, D. C. (1997). Structure and Function in the Lexical System: Insights from Distributed Models of Word Reading and Lexical Decision. *Language and Cognitive Processes*, *12*, 765-806. https://doi.org/10.1080/016909697386682

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: computational principles in quasi-regular domains. *Psychological Review*, *103*, 56-115. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8650300

Plaut, D. C., & Shallice, T. (1993). Deep dyslexia: A case study of connectionist neuropsychology. *Cognitive Neuropsychology*, *10*, 377-500.

Rogers, T. T., Lambon Ralph, M. a, Garrard, P., Bozeat, S., McClelland, J. L., Hodges, J. R., & Patterson, K. (2004). Structure and deterioration of semantic memory: a neuropsychological and computational investigation. *Psychological Review*, *111*, 205-235. https://doi.org/10.1037/0033-295X.111.1.205

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, *96*, 523-568. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2798649

Smolensky, P. (1986). Information processing in dynamical systems: Foundations of harmony theory. In *Parallel Distributed Processing Explorations in the Microstructure of Cognition* (Vol. 1, pp. 194-281). Retrieved from http://portal.acm.org/citation.cfm?id=104279.104290

1차 원고 접수: 2017. 02. 02
1차 심사 완료: 2017. 05. 12
2차 원고 접수: 2017. 06. 05
최종 게재 확정: 2017. 06. 21

(*Abstract*)

# RBM-based distributed representation of language

Heejo You[1)] Kichun Nam[1)] Hosung Nam[2)]

[1)]Department of Psychology, Korea University

[2)]Department of English language and literature, Korea University

The connectionist model is one approach to studying language processing from a computational perspective. And building a representation in the connectionist model study is just as important as making the structure of the model in that it determines the level of learning and performance of the model. The connectionist model has been constructed in two different ways: localist representation and distributed representation. However, the localist representation used in the previous studies had limitations in that the unit of the output layer having a rare target activation value is inactivated, and the past distributed representation has the limitation of difficulty in confirming the result by the opacity of the displayed information. This has been a limitation of the overall connection model study. In this paper, we present a new method to induce distributed representation with local representation using abstraction of information, which is a feature of restricted Boltzmann machine, with respect to the limitation of such representation of the past. As a result, our proposed method effectively solves the problem of conventional representation by using the method of information compression and inverse transformation of distributed representation into local representation.

*Key words : connectionist modeling, localist representation, distributed representation, restricted Boltzmann machine*