

빈도 효과에 기반한 DBN 알고리즘의 언어 획득 모델링 적용 가능성 탐구*

유 회 조	구 민 모	이 광 오	남 기 춘*
고려대학교 심리학과	고려대학교의료원 의생명연구센터	영남대학교 심리학과	고려대학교 심리학과

최근 컴퓨터 공학의 패턴인식 분야에서는 딥러닝 알고리즘이 도입 및 활용되고 있다. 하지만, 언어 처리에 관한 계산주의적 접근을 위해 패턴인식 알고리즘들을 적용해 왔던 연결주의 모델링 분야에서는 아직까지 딥러닝 알고리즘이 제대로 활용되지 못하고 있다. 본 연구에서는 딥러닝 알고리즘 중 하나인 deep belief network 알고리즘을 이용한 모델링을 구축하고, 단어와 의미 간의 관계를 학습시켰다. 학습 과정 이후 모델 용의 어휘 판단 과제를 수행하여, 빈도 효과(frequency effect)를 중심으로 행동 실험과 유사한 결과가 도출되는지 통계적으로 검증하였다. 모델링 수행 결과, 모델은 행동실험과 유사하게 빈도 효과를 도출해냄으로써, 모델이 정상적인 언어 처리가 가능함을 보였다. 본 연구의 결과는, 딥러닝 알고리즘으로 구축된 모델이 연결주의 모델링, 더 나아가 실제 사람의 언어 처리를 모사해 낼 수 있음을 제시한다. 아울러서 본 연구에서는, 어떻게 deep belief network 알고리즘이 연결주의 모델링에 적용이 가능한가에 대해서도 논의하였다.

주제어 : 연결주의 모델링, 딥러닝, 빈도 효과, 어휘 판단 과제

* 이 논문은 2013년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원(NRF-2013R1A2A2A03006323)으로 수행되었음.

이 논문은 2012년 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임.
(NRF-2012S1A5A2A03034555)

본 연구는 보건복지부의 재원으로 한국보건산업진흥원의 보건의료기술연구개발사업 지원에 의하여 이루어진 것임(H 113C 1 611).

† 교신저자 : 남기춘, 고려대학교 심리학과, (02841) 서울특별시 성북구 안암로 145 구법학관 405호,

Email : kichun@korea.ac.kr

딥러닝(deep learning)은 복수의 은닉층(hidden layer)을 상징하고 해당 은닉층들에 정보를 여러 번 거쳐 가도록 함으로써 연산의 개선을 추구하는 알고리즘들을 통칭한다. 최근 10여 년 동안 컴퓨터 공학 분야에서는 컴퓨터의 패턴인식 능력 개선을 위해 이와 같은 딥러닝 알고리즘을 개발 및 적용해왔다(Ciresan, Meier, & Gambardella, 2010; Deng & Yu, 2014; Kim, Nan, & Zhang, 2015). 하지만 과거 컴퓨터 공학의 패턴인식 알고리즘을 도입 및 적용하여 언어 처리에 관한 계산주의적 시뮬레이션을 시도해 온 연결주의(connectionist) 모델링 연구에서는, 현재까지 딥러닝 알고리즘의 적용이 이루어지지 못하고 있다.

컴퓨터 공학 분야에서 딥러닝 알고리즘이 주목받고, 더 나아가 적극적으로 도입되고 있는 이유에는 몇 가지가 있으나, 무엇보다 큰 이유는 패턴인식 능력의 직접적인 개선에 있다. 현재 컴퓨터 공학 분야에서 진행되고 있는 음성 인식 및 사진 태깅과 같은 다양한 패턴인식 연구들은, 딥러닝 알고리즘을 패턴인식에 적용, 이를 통한 성능 향상을 보고하고 있다(Ciresan, Meier, & Gambardella, 2010; Deng & Yu, 2014; Gwak, Bak, & Yi, 2015). 정확도 향상을 추구하는 패턴인식 분야에서 이와 같은 성능 향상은 딥러닝이 주목받는 가장 기본적인 이유라고 할 수 있다.

그 중에서도 특히 두드러지는 세부 분야는 자연언어처리 분야로서 대규모의 단어 및 문장과 같은 언어 정보 데이터에서 효율적으로 의미 정보를 산출 및 처리하기 위해 적용되고 있다. 가령 Bowman, Angeli, Potts, Manning (2015)은 딥러닝의 적용을 통해 의미상속과 모

순이 고려된 자연 언어 추론(natural language interface)이 가능함을 제시하였고, Bowman, Manning, Potts(2015), Chen, Manning(2014), Socher, Huval, Manning, Ng(2011)는 딥러닝을 적용하여 문장 내 각 세부 단어들의 품사를 분류하는 파싱 성능의 개선을 시도하였다. 단어 표상을 구축하는 알고리즘에서도 많은 개선이 있었다. Luong, Socher, Manning(2013) 및 Huang, Socher, Manning, Ng(2012)은 문맥정보를 활용하여 개선된 단어 표상이 가능한 모델을 구축하였고, Pennington, Socher, Manning(2014)은 제시된 대규모의 문장데이터에서 단어별 동시 등장(co-occur) 정도에 의해 각 단어별 벡터 표상을 구축해내는 비교사 학습모델을 제안하였다. 특히 Socher, Chen, Manning, Ng(2013)은 재귀적 텐서 신경망(recursive neural tensor network, 이하 rNTN)을 통한 대규모의 개념 간 관계들을 구축하고자 시도하였다. Bowman, Potts, Manning(2014) 및 Bowman(2013)은 재귀적 텐서 신경망(rNTN)을 적용하여 단어 간의 논리적 관련성 처리 능력을 향상시켰으며, Socher, Huang, Pennington, Manning, Ng(2011)은 재귀적 오토인코더(recursive autoencoder)를 통해 목표값이 따로 주어지지 않는 형태로 학습하는 방식인 비교사학습을 통해 문장 단위에서의 유사성 판단 과제를 수행할 수 있는 모델을 구축하였다. 이에 더해 문장 혹은 문단의 감정 정보를 분석하여 해당 발언자의 의견이나 태도를 측정해내는 감정 분석(sentiment analysis) 역시 딥러닝을 적용하면서 크게 발전하고 있다. Socher, Pennington, Huang, Ng, Manning (2011)은 문장 단위에서 각 하부 구조의 정서 분석을 가능하게 하는 재귀적 오토인코더를

구축하였고, Socher, Perelygin, Wu, Chuang, Manning, Ng, Potts(2013)는 한발 더 나아가 rNTN 모델에 Stanford sentiment treebank 정보를 활용하여 개선된 감성 분석(sentiment analysis) 모델 구축을 시도하였다 (Bowman, 2013; Bowman, Angeli, Potts, & Manning, 2015; Bowman, Manning, & Potts, 2015; Bowman, Potts, & Manning, 2014; Chen & Manning, 2014; Huang, Socher, Manning, & Ng, 2012; Luong, Socher, & Manning, 2013; Pennington, Socher, & Manning, 2014; Socher, Chen, Manning, & Ng, 2013; Socher, Huang, Pennington, Manning, & Ng, 2011; Socher, Huval, Manning, & Ng, 2011; Socher, Pennington, Huang, Ng, & Manning, 2011; Socher, Perelygin, Wu, Chuang, Manning, Ng, & Potts, 2013).

음성학 분야 또한 자연언어처리 분야만큼 딥러닝을 통한 기존 알고리즘의 개선을 시도하고 있다. 가장 대표적인 분야는 인간의 발화를 지각 및 분석하여 도출해내는 음성인식으로서, Dahl, Yu, Deng, Acero(2011), Dahl, Yu, Deng, Acero(2012) 및 Hinton 등(2012)은 deep belief network(이하 DBN)와 은닉 마코프 모델(hidden markov model)을 적용하여 음성인식 성능의 향상을 시도하였다. 또한 Ishii, Komiyama, Shinozaki, Horiuchi, Kuroiwa(2013)은 노이즈 제거를 통한 음질 개선으로 음성 인식 성능을 향상시키는 방식을 시도하였다. 그 외에도 Richmond, Hoole, King(2012)은 소리 정보를 통해 조음기관의 움직임을 역산하는 speech inversion에 DBN 알고리즘을 적용하여 조음기관에 대한 예측 성능을 향상시켰다(Dahl, Yu, Deng, & Acero, 2011; Dahl, Yu, Deng, & Acero,

2012; Hinton et al., 2012; Ishii, Komiyama, Shinozaki, Horiuchi, Kuroiwa, 2013; Richmond, Hoole, & King, 2012)

하지만 연결주의 모델링 입장에서 봤을 때 딥러닝에는 정확성 향상 이상의 장점이 있다. 바로 정보를 추상화 하는 방식으로 처리한다는 점이다. 과거 연결주의 모델링에서 주로 이용되었던 구조에서는, 모델의 연산 능력을 은닉층(hidden layer)의 크기에 의존하였다(Figure 1). 문제는 은닉층을 과도하게 크게 만들 경우, 모델이 패턴들이 갖고 있는 규칙성에 대한 학습이 이루어지지 못하고, 각 패턴을 개별적으로 학습하는 과적합(over-fitting)이 일어나게 된다는 점이다(Caruana, Lawrence, & Giles, 2001; Harm & Seidenberg, 1999). 즉 소규모의 패턴만을 가지고 이뤄지는 학습에서는 문제가 없지만, 대규모의 패턴에서 규칙을 학습하고자 시도할 경우 모델이 제대로 된 인간의 학습을 모사하기 어렵게 된다.

그러나 딥러닝의 경우 이와는 다르게, 비교적 작은 규모의 은닉층을 여러 단계 쌓고 이를 한 단계씩 거치면서, 점차적으로 입력된 패턴에 대한 추상화를 시도한다(Bengio, Courville, & Vincent, 2013; LeCun, Bottou,



Figure 1. The usual structure of connectionist modeling. The activation is transferred from left to right. The unit sizes of input and output layer are related with the purpose of model, and the computing power of model depend on the unit size of hidden layer.

Bengio, & Haffner, 1998). 이와 같은 추상화 방식은 종래 연결주의 모델링의 제한점을 해결할 뿐 만 아니라, 과거 연구들이 주장해 온 인간의 지식 형성 방식과 매우 유사하다고 알려져 있다(Tenenbaum, Kemp, Griffiths, & Goodman, 2011).

하지만 그럼에도 불구하고 처음에 언급한 바와 같이, 아직까지 딥러닝을 이용한 연결주의 모델링 연구는 이루어지지 못하고 있다. 이에 대한 가장 큰 원인은, 패턴인식에서 사용되는 딥러닝 알고리즘이 연결주의 모델링 연구에 적합한지에 대한 검증이 이루어지지 않았기 때문이다.

비록 연결주의 모델링은 컴퓨터공학에서 개발된 패턴인식의 알고리즘을 적용하여 진행하지만, 추구하는 방향성에 있어서 패턴인식과 언어 처리 연결주의 모델링은 전혀 다르다. 위에서 언급한 바와 같이, 패턴인식의 일차적인 목표는 최대한의 정확성이다. 따라서 일반적으로 패턴인식에 대한 연구에서는 학습 이후 가능한 한 많은 패턴에 대해 올바른 반응을 이끌어내는 것을 목표로 학습을 진행한다(Ciresan et al., 2010; Hinton, Osindero, & Teh, 2006). 하지만 연결주의 모델링은 정확성 이상으로 인간과의 유사성을 중요시한다. 따라서 연결주의 모델링에서는, 실제 참여자에 의해 도출된 행동실험 결과와 유사한 반응이 나타나는가를 확인하는 것을 목표로 학습을 진행하게 된다(Harm & Seidenberg, 1999; Hutzler, Ziegler, Perry, Wimmer, & Zorzi, 2004; Park & Lim, 2014; Seidenberg & McClelland, 1989; You, Nam, & Nam, 2015).

이와 같은 연결주의 모델링과 패턴 인식 분

야 간 방향성 차이는 연결주의 모델이 패턴인식 알고리즘을 사용할 때 몇 가지의 차이점을 만들게 된다. 그리고 이 차이가 가장 극명하게 나타나는 경우가 빈도(frequency)이다. 단어의 빈도는 특정 말뭉치 상에서 해당 단어가 얼마나 자주 등장하였는가에 따라 결정된다. 그리고 해당 단어가 자주 등장함으로써 언어 사용자는 해당 단어에 더 자주 노출되게 되고, 그 결과 언어 사용자는 해당 단어의 처리능력이 증가하게 된다. 결과적으로 언어 사용자는 자주 노출된 단어에 대한 처리능력을 선택적으로 향상시키게 되어, 자주 사용되는 단어를 그렇지 않은 단어에 비해서 더 빠르고 정확하게 인식하게 된다(Grainger, 1990). 빈도 효과(frequency effect)는 빈도에 따른 처리 속도 및 정확도의 차이를 의미하며, 여러 언어심리학적 현상 중에서도 참여자가 사용하는 언어에 관계없이 가장 일반적이고, 강하게 효과가 나타나는 현상으로 알려져 있다. 이에 따라 국내외를 가리지 않고 과거 연구들은 빈도를 실험에 있어서 반드시 고려해야 하는 변인으로 다뤄 왔다(Andrews, 1992; Grainger, 1990; Kwon & Lee, 2014; Kwon & Nam, 2011; Monsell, Doyle, & Haggard, 1989; Nam, Seo, Choi, Lee, Kim, & Lee, 1997; Yi, 1993; Yi, Park, Abe, Liu, & Zhang, 2010).

빈도에 관해서 연결주의 모델링 연구와 패턴 인식 간의 입장 차이는 명확하다. 엄밀히 말해서 최대한의 정확성을 중시하는 패턴인식에서 빈도는 고려 대상이 아니다. 패턴인식은 많은 패턴을 학습할수록 정확도가 향상된다. 따라서 미리 학습 가능한 패턴이 있다면, 그것을 가능한 한 전부 학습 하는 것이 패턴인

식의 정확성을 향상시키는 데 유리하다. 따라서 빈도 효과와 같은 인간의 수행 능력과의 유사성을 위하여 학습 가능한 패턴이 있을 때 이를 고의로 학습하지 않는 경우를 전제할 이 유가 없다.

하지만 패턴인식에서 빈도를 고려하지 않는다는 점은 딥러닝 알고리즘을 적용하는데 있어서 커다란 장애가 된다. 왜냐하면 위에서 언급한 바와 같이 연결주의 모델링 연구는 인간과의 유사성을 중요시하기 때문이다. 이에 대해 좀 더 정확하게 말하자면, 인간과 유사한 결과를 내는 것 자체가 연결주의 모델링의 목표라고 할 수 있다. 인간과 유사한 결과를 낸 모델을 통해 간접적으로 인간의 인지적 처리과정을 알아보려고 하는 것이기 때문이다. 그리고 상기한대로 빈도 효과는 언어심리학적 현상에 있어서 가장 일반적이고 강하게 나타나는 현상이다. 즉, 연결주의 모델링의 목적 달성을 위해서 인간의 행동에서 일반적인 현상인 빈도 효과를 모사해내는 것은 해당 모델의 유효성을 검증하는데 가장 기본적이며 중요한 요소라 할 수 있다. 실제로 과거 진행되었던 다수의 연결주의 모델링 연구들은, 빈도 효과를 보일 수 있도록 모델을 구성 및 제시하였다. 연결주의 모델링 연구들은 빈도가 높은 학습 패턴이 낮은 패턴보다 더 빈번하게 학습하는 형태로 인간 빈도별 단어 학습과 유사한 학습환경을 구축해 왔으며, 그 결과 대부분의 연결주의 모델링 연구들은 제시한 모델이 빈도 효과를 정상적으로 모사함을 보여왔다(Park & Lim, 2014; Plaut, 1996; Seidenberg & McClelland, 1989; Yim, Lim, Park, & Nam, 2005; You, Nam, & Nam, 2015).

당초 언급한 바와 같이 딥러닝의 연결주의 모델링에의 적용은 과적합 문제의 해결 및 좀 더 인간의 인지처리 과정 근접한 모델 구축을 위한 중요한 자원이 될 수 있다. 하지만 아직까지 ‘연결주의 모델에 딥러닝이 적용될 경우 빈도 효과와 같은 기본적인 언어심리학적 현상이 설명 가능한가의 여부조차 검증되지 않았다는 것도 사실이다. 자연언어처리 내지는 음성학 분야와 같이 적극적인 딥러닝 알고리즘 적용을 위해서는 이와 같은 검증작업이 필요하며, 딥러닝에 기반을 둔 모델에서 빈도에 따른 수행의 차이가 나타나는가를 알아보는 것은 딥러닝의 연결주의 모델 적용 가능성을 확인하는 검증작업의 일환이 될 수 있으리라 판단된다.

이에 따라 본 연구에서는 딥러닝 알고리즘에서 빈도 효과가 나타나는가를 확인함으로써, 딥러닝의 차후 연결주의 모델링에의 적용 가능성을 확인해보고자 하였다. 이를 위하여 본 연구에서는 참여자를 대상으로 한 행동 실험과 딥러닝 알고리즘에 기반을 둔 모델링 시뮬레이션을 진행하였다.

행동 실험에서는 참여자를 대상으로 한 어휘 판단 과제(lexical decision task)를 진행함으로써 빈도 효과를 재확인하고, 시뮬레이션 결과와 비교할 행동 실험 데이터를 확보하고자 하였다.

시뮬레이션에서는 딥러닝에서 주로 이용되는 알고리즘 중 하나인 DBN 알고리즘을 적용하여 어휘 판단 과제를 모사한 시뮬레이션을 진행하였고, 행동 실험의 결과와 유사한 빈도 효과가 나타나는가를 검증하고자 하였다.

본 연구는 시뮬레이션이 두 가지 가능성이

있다고 판단한다. 만약 딥러닝 알고리즘이 연결주의 모델링에 적합하지 않다면 학습 단계에서 빈도를 고려한다 하더라도 학습 빈도에 관계없이 과도한 정확성에 의해 천장 효과(ceiling effect)가 나타날 것이다. 반대로 딥러닝 알고리즘이 연결주의 모델링에 충분히 적용 가능하다면, 빈도에 따른 수행의 차이를 보임으로써 빈도 효과가 정상적으로 나타날 것이다.

실 험

앞에서 언급한 바와 같이, 모델링을 진행하기에 앞서 빈도 효과를 재확인하고 차후 시뮬레이션 결과와의 비교를 사용될 데이터를 수집하기 위해 행동 실험¹⁾을 실시하였다.

방 법

참여자 총 16명의 고려대학교 학생이 실험에 참가하였다. 참여자의 성별은 남성 7명, 여성 9명이었으며, 참여자의 평균 연령은 20.25세였다. 16명의 참여자 전부 실험에 자원하여 참가하였으며, 각 참여자에게 감사의 표시로 일정금액을 사례금으로 지급하였다.

자극 30943개의 단어와 30917개의 비단어를 포함하여 총 61860개의 1, 2, 3, 4음절 자극이

실험에 사용되었다. 빈도 효과를 보기 위한 본 연구에서는 몇 가지 과정을 거쳐 분석에 사용할 자극을 선정하였다. 먼저 2음절 명사 단어에 한정하여 1차적으로 500개의 단어를 무작위로 선출하였다. 그 다음 500개의 단어를 빈도 크기 별로 정렬한 후 각각 100개씩으로 5등분하였고, 최상위의 빈도 상위 100개의 단어를 고빈도로, 하위 100개의 단어를 저빈도로 정의하여 분석용 자극으로 선출하고 나머지는 제외하였다. 이와 같은 과정은 고빈도와 저빈도 범주 간의 빈도 차이를 명확히 하기 위해 진행되었다. 선출 결과 고빈도 범주 단어들의 평균 빈도는 1786.41이었고, 저빈도 범주 단어들의 평균 빈도는 22.92였다.

절차 참여자들은 어휘 판단 과제(lexical decision task)를 통해 실험을 진행하였다. 각 자극마다 참여자에게는 1000ms의 빈 화면이 제시된 이후, 단어 혹은 비단어 자극이 화면에 제시되었다. 제시된 자극이 단어일 경우 버튼 박스의 오른쪽 버튼을, 비단어일 경우 버튼 박스의 왼쪽 버튼을 누르도록 하였다. 실험 프로그램은 자극이 제시된 시점부터 참여자가 반응을 한 시점까지의 시간을 반응시간으로 기록하였으며, 반응시간이 2000ms가 넘었을 경우 무응답으로 오반응 처리하였다(Figure 2). 실험은 총 126블록이 존재하였고, 마지막 블록을 제외하고 각 블록마다 506개의 자극이 배치되었다. 참여자는 8주에서 14주에 걸쳐서 블록 별로 실험을 진행하였다. 각 블록에 대한 실험 종료 후에 오류율이 15%이상인 블록에 대해서는, 참여자의 컨디션이 실험을 진행하기 어려운 상황이었다고 판단하였고, 해당

1) 본 연구에서 사용된 행동실험 데이터는 “The Korean Lexicon Project” 연구사업의 일환으로 고려대학교에서 수집한 자료이다. 고려대학교 외에도 영남대학교, 이화여자대학교, 전남대학교에서도 자료를 수집하였으며 현재 통합 분석을 진행하고 있다.

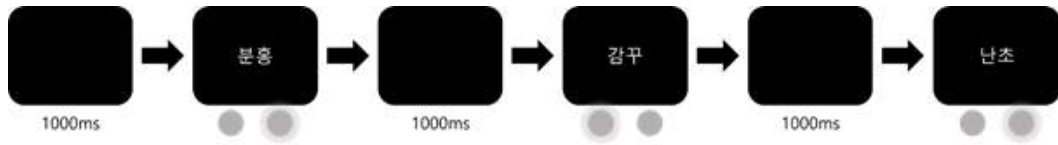


Figure 2. The process of human experiment. At first, the blank screen was displayed while 1000ms. After that, a word or nonword were displayed on the monitor. Participant inserted their response by the button box which there are two buttons. The participants clicked the right button by the right hand when the displayed string was a word, and clicked the left button by the left hand when the string was a nonword. After the response, the screen returned to the blank, and process were repeated.

블록에 대해서는 모든 블록에 대한 실험이 종료된 후에 재실험을 진행하였다. 실험은 DMDX 5.0을 사용하여 제작 및 진행되었다.

결 과

위에서 언급한 바와 같이 500개의 단어 중, 빈도 상위 100개의 단어를 고빈도로, 하위 100개의 단어를 저빈도로 정의하여 빈도효과를 확인하고자 하였다.

빈도 효과 빈도 효과를 확인하기 위하여 고 빈도 단어와 저빈도 단어 간의 반응시간과 정확도에 대한 대응표본 t-검정을 실시하였다.

각 고빈도, 저빈도 자극목록에 대한 반응시간, 정확도의 평균값은 Table 1과 같았으며, 두 변인 모두 두 집단 간의 차이는 유의하였다

(반응시간: $t(15)=10.468, p<0.001$, 정확도: $t(15)=15.571, p<0.001$).

또한 자극의 타당성 확인을 위해, 고빈도와 저빈도 단어 범주 간의 독립표본 t-검정을 실시하였다. 분석결과, 두 자극 범주 간의 차이 또한 유의하였다(반응시간: $t(198)=9.722, p<0.001$, 정확도: $t(198)=8.296, p<0.001$)

상관분석 결과 극단값 추출로 인한 분산분석에서의 1종 오류의 가능성을 차단하기 위해, 500개 자극 전체에 대한 빈도와 반응시간 및 정확도 간의 상관분석을 실시하였다. 분석결과 반응 시간과 빈도는 $r=-0.176$ 로 $p<0.01$ 수준, 정확도와 빈도는 $r=0.142$ 로 $p<0.01$ 수준에서 유의하게 나타남으로써, 두 변인 모두 빈도효과가 유의함을 확인하였다(Figure 3).

Table 1. The reaction time and accuracy of high, low, and total stimuli.

	High frequency	Low frequency	Difference
Reaction time(ms)	588.198	688.691	100.494***
Accuracy(%)	96.22	71.27	24.95***

※ *** $p<0.001$

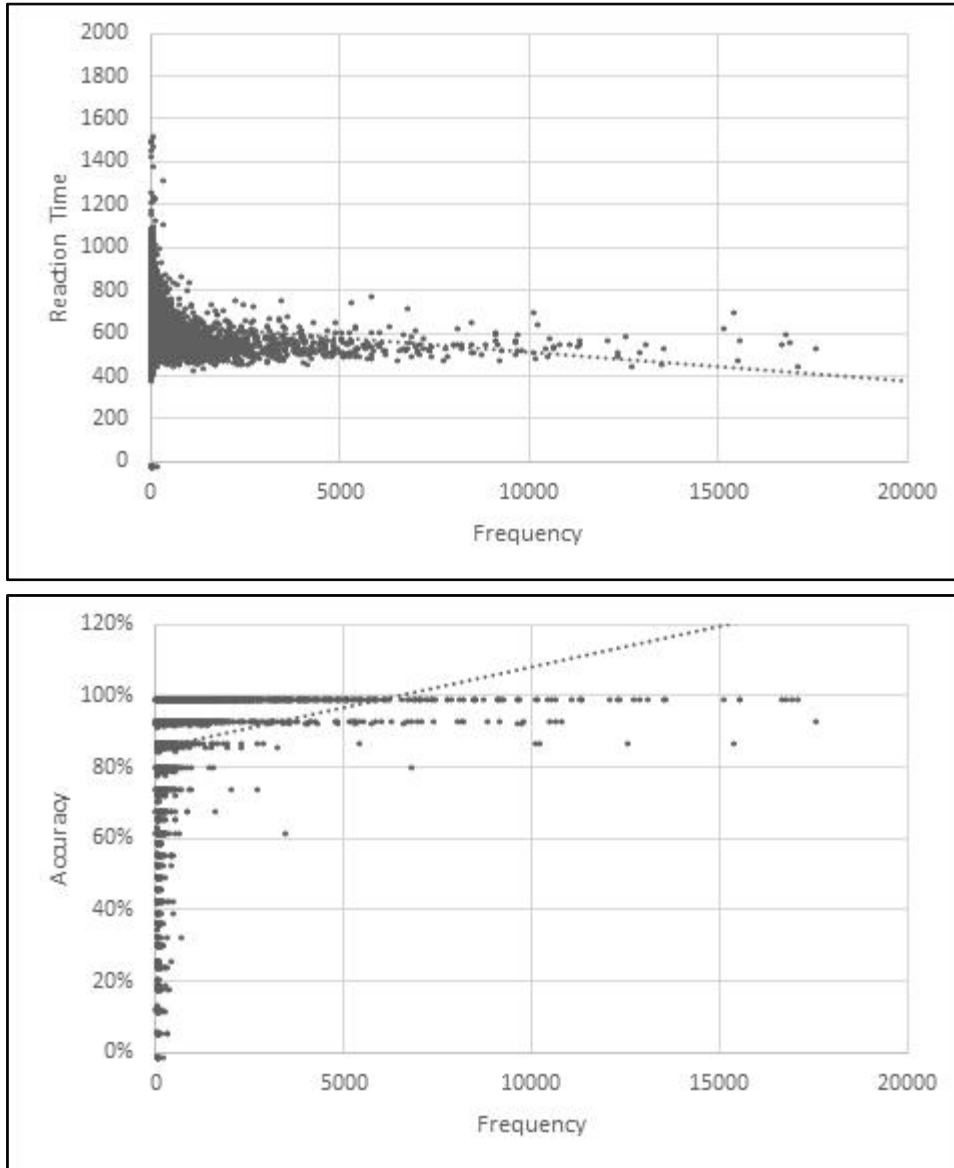


Figure 3. The correlation between the frequency of each stimulus and the reaction time and the accuracy. There are some significant correlation of $p < 0.01$ level in the both of them.

논 의

행동실험 결과 과거 연구들과 마찬가지로 어휘 판단 과제에서 빈도효과가 나타남을 확

인하였다. 본 연구는 이와 같은 행동실험 결과가 딥러닝 알고리즘이 적용된 모델링에서도 나타날 수 있는가를 확인하고자 하였다.

시물레이션 1

위에서 언급한 바와 같이 딥러닝 알고리즘에서의 빈도 효과를 확인하기 위하여 딥러닝 알고리즘에 기반을 둔 모델링을 구축, 학습 후 어휘 판단 과제를 모사한 시물레이션을 진행하였다.

모델링 구조 모델의 구조는 Figure 4와 같다. 이와 같은 구조는, 과거의 철자-발음 간에 대해 진행된 연결주의 모델을 참고하여 구성되었다(Park & Lim, 2014; You, Nam, & Nam, 2015). 다만, 본 연구의 목적에 따라 입력층과 출력층 사이의 은닉층을 5개 할당하여, 딥러닝을 위한 구조로 변경하였다.

입력층의 세부구조는 You, Nam, Nam(2015)을 참조하여 구성하였다. 이 구조는 다수의 선행 연구들과 마찬가지로 시간적인 이유와 서체에 따른 인식 자체가 또 다른 학습을 요한다는 이유로 시지각 단계가 생략되었고, 문자 정보가 추상화된 형태로 철자정보를 표현하였다(Hutzler et al., 2004; Kwon, Park, Lim, Jung, & Nam, 2006; Park & Lim, 2014; Plaut, 1999; Plaut, McClelland, Seidenberg, & Patterson, 1996; You, Nam, & Nam, 2015).

입력층의 각 단위는 하나의 자소를 표상하

며 67개(초성 19개, 중성 21개, 종성 27개)의 단위가 하나의 음절을 구성하였다. 또한, 음절 위치에 상관없이 어느 음절에서 표상되더라도 활성화되는 67개의 단위가 추가로 포함되었다. Figure 5의 0행은 이에 대한 예시를 보여준다. 1음절의 초성 ‘ㅎ’ 중성 ‘ㅏ’ 종성 ‘ㄱ’에 대한 단위만이 아니라 2음절의 초성 ‘ㄱ’, 중성 ‘ㅇ’에 대한 단위 또한 활성화되어 있음을 볼 수 있다. 행동실험에서 언급한 바와 같이 본 연구에서는 빈도 효과를 확인하기 위해 나머지 요인을 통제하고자 2음절 단어만을 사용하였으므로, $67 \times (2 + 1) = 201$ 개의 단위로 구성되었다. 식의 2는 2음절을 나타내며 더해진 1은 음절 위치에 상관없이 활성화되는 추가 단위를 의미한다(You, Nam, & Nam, 2015).

출력층은 단어를 의미하는 단위와 비단어를 의미하는 단위의 총 두 개의 단위로 구성하였다.

본 연구에서는 You, Nam, Nam(2015)과 같은 방식으로 모델의 통계적인 분석을 시도하고자 하였다. 구체적으로는 우선 동일 구조의 모델을 총 8개 구성하고, 각 모델은 이후 언급될 학습과 과제를 수행하였다. 이들 모델들은 초기 가중치가 일정 범위 안에서 무작위로 할당되며, 학습 시의 자극 노출 순서가 무작위로 이루어짐으로써 개인차를 구성, 독립된 참여

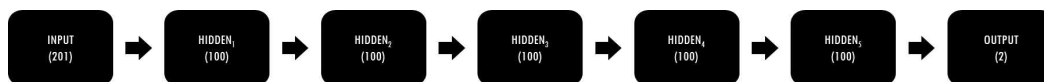


Figure 4. The structure of model. The activation was transferred from left to right. Input layer was constituted by 201 units, and the letter information was inserted to the layer. Each hidden layer had 100 units. The hidden layers received the activation from the previous layer and sent their activation to next layer. There are two units in the output layer. Output layer finally judged whether the inserted information is a word.

	ㄱ	ㄴ	...	ㅅ	...	ㅇ	ㅋ	ㆁ	ㅇ	ㅈ	ㅊ	...	ㅌ	ㅍ	...	ㄱ	ㄴ	...	ㅇ	...	ㅇ		
0	1	0	...	0	...	0	0	0	1	1	0	...	0	1	0	...	1	0	0	0	...	0	0
1	0	0	...	0	...	0	0	0	1	1	0	...	0	0	0	...	1	0	0	0	...	0	0
2	1	0	...	0	...	0	0	0	0	0	0	...	0	1	0	...	0	0	0	0	...	0	0

Figure 5. The example of input activation about Korean word '학교'. The figure of You, Nam, Nam(2015) was modified and reused. The units of first row(label '0') have the value 1 when the assigned letter was represented at any location. The units of second and third row(label '1' and '2') have the value 1 when the assigned letter was represented at assigned location. Except the units, every unit has the value 0.

자료서의 데이터로 이용될 수 있도록 하였다 (You, Nam, & Nam, 2015).

본 연구의 모델링은 Microsoft Visual Studio 2015 Community 프로그램을 이용하여 C# 언어를 통해 이루어졌고, Intel Xeon(R) E3-1230 v3 프로세서와 24,576MiB 램, Windows 10 64-bit 운영체제인 컴퓨터에서 모델의 학습 및 과제가 수행되었다.

자극 행동실험에서 사용된 자극 중 2음절 명사 단어 12,125개와 2음절 비단어 12,881개가 모델링에서 사용되었다. 또한 과거연결주의 모델링 연구들에서 시행된 것과 같이 학습 진행시에 학습 확률의 극단화를 방지하기 위해 이하의 공식으로 변환된 압축 빈도를 사용하였다(Park & Lim, 2014; Seidenberg & McClelland, 1989; You, Nam, & Nam, 2015).

$$P_i = 0.15 \times \log_{10}(F_i + 2)$$

P는 학습 확률을, F는 단어의 빈도를 나타낸다. 이 공식에 의해 도출된 학습 확률은 최저 7.16%에서 최고 72.80%까지 다양하였다.

동시에 학습하는 비단어의 확률은 각각 0.25로 할당하였다. 이는 단어 자극의 전체 학습가능성의 합을 비단어 자극의 총 개수로 나눈 값 0.288에 근접한 값으로, 모델이 단어와 비단어를 균형 있게 학습할 수 있도록 하기 위해 설정되었다. 이와 같이 비단어의 학습을 설정하는 이유는 모델이 모든 철자 패턴을 '단어'라고 잘못 판단하는 현상을 막기 위해서이다. 모델에게 '비단어가 존재함'이라는 것을 학습시키기 위해서는 단어와 비슷한 수준의 비단어 학습이 요구된다.

모델이 학습하는 12,125개의 단어 중, 행동실험에서 선정된 고빈도 범주 단어 100개와 저빈도 범주 단어 100개가 실제 빈도 효과의 분석에 이용되었다.

도입에 언급한 바와 같이 연구는 본 연구는 DBN 알고리즘을 딥러닝 알고리즘으로 이용하였다. 이는 DBN 알고리즘이 연결주의 모델링과 적합성이 높았기 때문이었다. DBN의 학습 과정은 선행학습(pre-training)과 미세 조정(fine-tuning)의 두 단계로 구성되며, 이 중 미세조정 단계는 오류 역전파 및 classification RBM과 같은 몇 가지 알고리즘 선택적으로 이용할 수

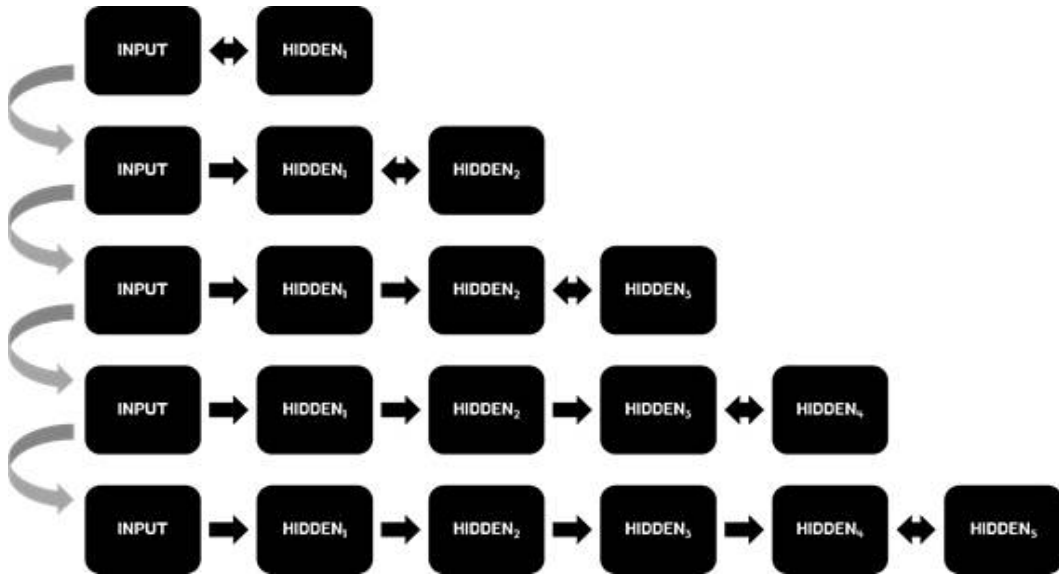


Figure 6. The pre-training method of deep belief network(DBN). In first stage, the input patterns of fine-tuning are used as the training pattern. Curved arrows mean the training flow. One way arrows mean the pattern making for the pre-training of the stage. The bidirectional arrows mean the adjusting part at the stage.

있다(Dahl et al., 2011; Gwak, Bak, & Yi, 2015). 이 중 오류 역전파 알고리즘은 연결주의 모델링의 선행연구들에서 주로 사용해온 알고리즘으로, 차후 선행연구들을 딥러닝으로 개선하고자 할 때 쉽게 적용이 가능하다는 장점이 있다.

선행 학습 선행학습은 복수의 은닉층이 정상적으로 학습될 수 있도록 하는 과정으로서, 목표 값을 비교하여 오류를 수정하는 교사 학습(supervised learning) 없이, 오직 입력값만을 사용하여 진행된다.

선행학습은 일반적으로 greedy layer-wise training 로 불리는 방법을 이용하여 진행된다. 먼저 제한된 볼츠만 머신(restricted Boltzmann machine, RBM) 알고리즘을 이용하여, 최초의

두 층(입력층과 은닉층1) 간의 연결 강도를 조정한다. 이 때 입력층에는 학습될 패턴들을 사용한다. 그리고 두 층의 연결 강도 조정이 완료되면, 같은 방식으로 다음 두 층(은닉층1과 은닉층2) 간의 연결 강도를 조정한다. 이 때, 은닉층1에는 입력층에서 학습될 패턴들을 입력한 뒤 은닉층1에서 도출된 패턴들을 연결 강도 조정에 사용한다. 그리고 ‘입력층-은닉층1’, ‘은닉층1-은닉층2’의 연결 강도 조정과 같은 과정을 반복하여 은닉층4와 은닉층5 사이의 연결강도까지 조정한다(Hinton, 2010; Hinton et al., 2006).

본 연구에선 선행 학습 중의 학습률을 0.1로 설정하였으며, 총 25,006개의 단어, 비단어 패턴 전체를 한 번에 학습하는 배치방식(batch mode)으로 학습이 진행되었다. 적절한 연결강

도 조절 정도를 확인하기 위해 10, 100, 1000 세대로 1회씩 예비 시뮬레이션을 한 결과 10 세대의 선행학습으로도 충분하다고 판단하여, 모델들의 각 층간 연결 강도 조절은 10세대 (epoch)씩 진행하였다.

미세 조정 위에서 언급한 바와 같이 DBN에서 선행 학습 이후에 진행되는 미세 조정 단계에는 몇 가지 방법이 있다(Dahl et al., 2011; Gwak, Bak, & Yi, 2015). 본 연구에서는 이들 중, 일반적인 연결주의 모델에서 사용하는 오류 역전파(back-propagation) 알고리즘을 사용한 학습을 진행하였다. 미세 조정 단계의 한 세대 내에서 각 자극은 압축 빈도에 의해 도출된 학습 확률에 의해 학습 여부가 결정되었으며, 매 세대 별로 배치 학습이 진행되었다. 각 단위의 활성화값은 은닉층의 경우 0에서 1사이로 변경해주는 시그모이드(sigmoid) 함수를, 출력층의 경우 전체 단위의 활성화값의 합이 1이 되도록 확률분포의 형태로 변경해주는 소프트맥스(softmax) 함수를 사용하였다. 미세 조정은 한 번에 하나의 패턴만을 학습하는 온라인(on-line) 방식으로 진행되었으며, 매 세대마다 학습되는 패턴의 순서는 무작위로 재조정되었다. 또한 각 모델들 간의 수행 비교를 위해, 안정화 상태를 기준으로 학습을 멈추는 가변 세대는 적용되지 않고 모든 모델들은 고정적으로 100세대 동안 학습이 진행되었다. 시그모이드(sigmoid) 함수에서 이용되는 모멘텀(momentum) 변수는 0.9, 학습률(learning rate)은 0.001로 설정하였으며, 학습률은 종료 시까지 변동되지 않았다.

과제 시뮬레이션 세 패턴 집단에 대한 어휘 판단 과제(lexical decision task)를 모사한 시뮬레이션 과제를 수행하였다. 첫 번째 패턴 집단은, 25,006개의 학습한 자극 전체였으며, 두 번째와 세 번째 패턴 집단은 자극에서 언급한 고빈도와 저빈도 단어들이었다.

매 10세대의 학습이 진행될 때 마다 세 집단에 대한 교차 엔트로피(cross entropy)를 구하였다. 교차 엔트로피는 다음과 같은 공식으로 계산되었다.

$$\text{Cross Entropy} = -\frac{1}{n} \sum ((T_i \times \log_e A_i) + (1 - T_i) \times \log_e (1 - A_i))$$

교차 엔트로피는 모델의 수행에 있어서 ‘숙련도’와 유사한 의미를 지니며, 교차 엔트로피가 낮을수록 모델의 수행이 좋은 것으로 판단된다. 이와 같은 지표를 사용하는 것은 모델이 행동실험의 ‘반응 시간’ 지표를 그대로 사용하는 것이 불가능하기 때문이다. 모델의 연산 속도는 시뮬레이션이 진행되는 컴퓨터의 성능에 따라 달라지므로, 정량적인 지표로 이용될 수 없다. 이에 따라 과거 연결주의 모델링 연구들에서는 반응시간에 대응하는 지표를 고려하였고, 그 중 하나로 이용되었던 것이 교차 엔트로피이다(Hutzler et al., 2004; Kwon et al., 2006; Park & Lim, 2014; Plaut, 1999; Plaut, McClelland, Seidenberg, & Patterson, 1996; You, Nam, & Nam, 2015).

또한 매 10세대마다 교차 엔트로피 외에 각 패턴 집단에 대해 정확하게 단어/비단어를 판단했는가 여부를 확인하였다. 위에서 언급한 바와 같이 소프트맥스 함수는 전체 단위들의

활성화값을 확률분포의 형태로 변경해주는 함수이다. 이를 통해 본 연구에서는 단어와 비단어의 두 단위 각각에 더 높은 활성화값이 나온 단위를 모델의 반응으로 결정하였다. 따라서 단어 패턴이 입력되었을 때 출력층의 단어에 해당하는 단위가 0.5 이상의 활성화값을 가졌을 경우 해당 판단이 정확했다고 판정하였다.

결 과

모델링 구조 절에서 언급한 바와 같이, 본 연구에서는 You, Nam, Nam(2015)과 같은 방식으로 모델의 통계적인 분석을 시도하고자 하였다. 구체적으로, 본 연구에서는 학습 및 과제를 수행한 모델 하나가 하나의 참여자로 가정되어 총 8명의 참여자가 실험을 수행한 것과 같은 형태로 분석이 진행되었다. 동일한 모델의 구조는 실험의 참여자 각 개인은 동일한 집단이라는 통계학적 가정을 충족시키고, 동시에 무작위로 할당된 초기 연결 강도와 학습 중 각 세대의 학습 순서의 무작위성 및 학습 확률에 따른 학습 불규칙성은 참여자 개인의 개인 차이를 반영함으로써 현실에 존재하는 개인차를 충족시킨다(You, Nam, & Nam, 2015).

학습 진행 간 교차 엔트로피 변화 Figure 7의 위는 모델들의 교차 엔트로피의 흐름을 보여준다. 그래프에서 나타난 바와 같이, 모델의 교차 엔트로피는 10세대에서 0.619로 감소한 이후, 해당 수준을 유지하였으며, 이는 모델의 발음 학습 과정이 10세대 시점에 이미 안정화

상태에 들어가 어트랙터(attractor)가 되어있음을 보여준다.

학습 진행간 정확도 변화 Figure 8의 아래는 모델들의 정확도의 흐름을 보여준다. 교차 엔트로피의 흐름과 비슷하게, 모델의 정확도는 10세대에서 약 70%정도의 정확도를 달성한 이후 큰 변동 없이 유지되고 있다.

빈도 효과 모델의 빈도 효과가 나타났는가를 확인하기 위해 대응표본 t-검정을 실시하였다. 검정 결과, $p < 0.001$ 수준에서 유의하게 고빈도 범주의 자극 목록이 저빈도 범주의 자극 목록보다 교차 엔트로피가 높았다($t(7) = 11.067$, $p < 0.001$). Figure 8은 고빈도와 저빈도 자극목록에 대한 교차 엔트로피의 평균값을 보여준다.

교차 엔트로피의 반응시간과의 직접적 비교는 제한적이므로, 정확도를 중점적으로 살펴본다면, 모델의 수행 정확도는 행동실험과 비교해서 낮다고 할 수 있다(행동실험 결과: 고빈도 96.22%, 저빈도 71.27%). 이에 따르면 모델의 연산능력이 실제 인간의 수행에 비해 떨어진다고 할 수 있다.

자극별 반응시간과 교차 엔트로피의 상관

12,125개의 단어 자극 중 행동실험에서 참여자 전원이 오반응하여 반응시간을 측정할 수 없었던 자극 28개를 제외한 12,097개의 행동실험 자극이 분석에 사용되었다. 행동실험 자극의 반응 시간 평균과 시뮬레이션의 교차 엔트로피 간의 상관 계수는 $r = 0.048$ 로 나타났으며, 상관은 $p < 0.01$ 수준에서 유의하였다. 다만 이

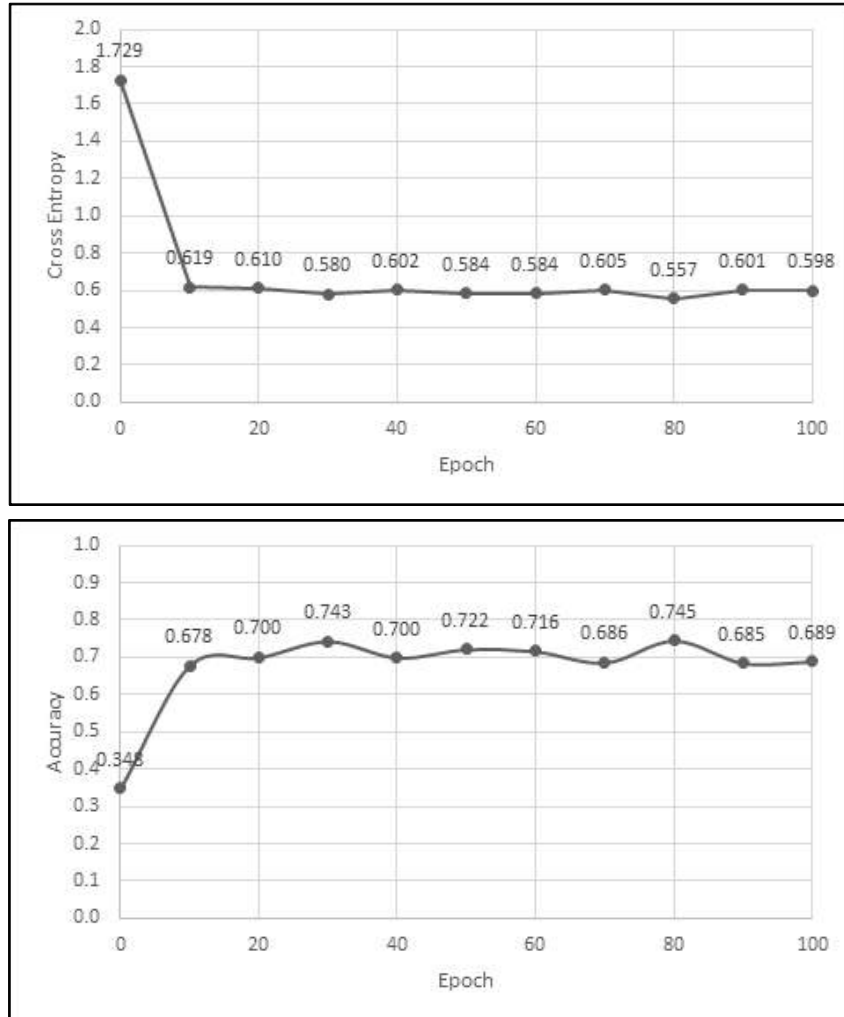


Figure 7. The cross entropy and accuracy by the progress of epoch. Upper and below graphs display about the cross entropy and accuracy, respectively. The graphs show that there is no big change in both of them after 10 epoch. This result means that training was already finished at the 10 epoch because the error does not decrease after 10 epoch.

와 같은 상관계수는 비록 유의하긴 하였으나 교차 엔트로피와 반응 시간이 직접적으로 관련되어 있다고 말하기엔 다소 제한이 있는 수치라고 할 수 있다(Table 2).

자극별 반응시간과 정확도의 상관 교차 엔트로피와 마찬가지로의 12,097개의 행동실험 자극의 정확도 평균과 시뮬레이션의 정확도 평균 간의 상관 계수는 $r=0.000$ 으로 유의하지 않으므로써, 완전히 관련 없는 것으로 나타났다(Table 2).

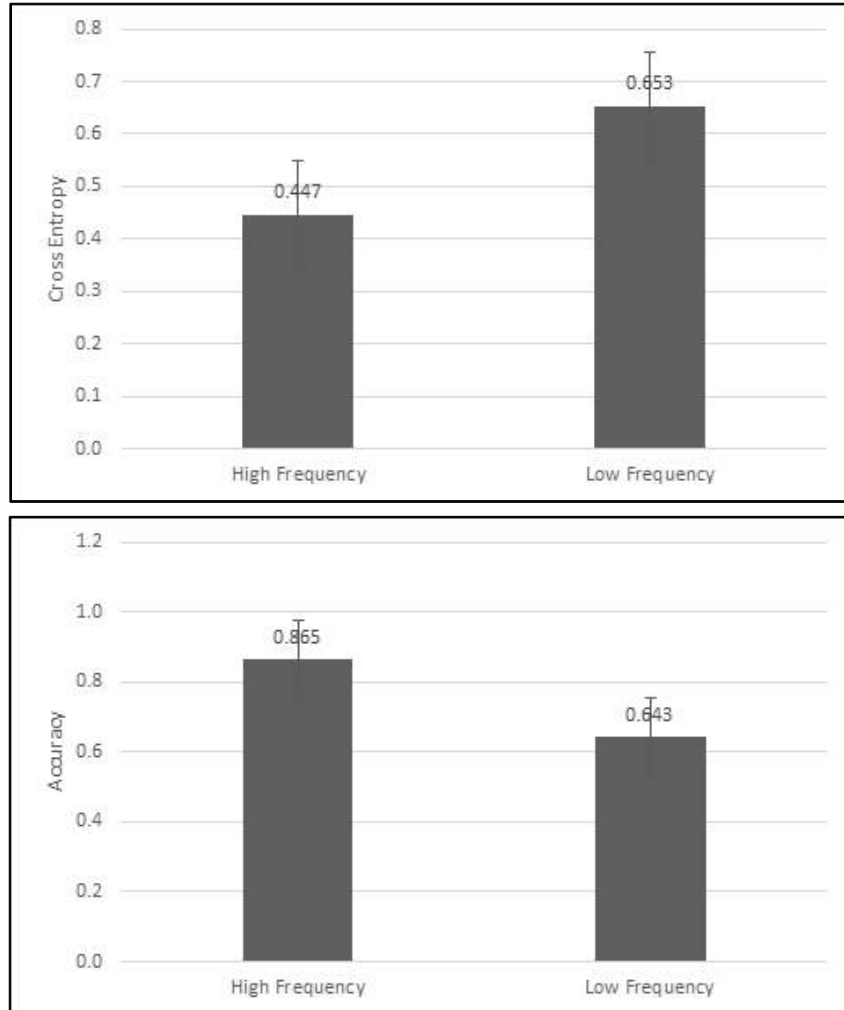


Figure 8. The difference between high and low frequencies. Model calculated lower cross entropy and high accuracy at the high frequency. Lower frequency means the short reaction time, so this result shows that model could simulate the frequency effect.

Table 2. The relations between human and model data. There was a positive relation between human reaction time and cross entropy, but the size of correlation is very small. On the other hand, in accuracy, there was no correlation.

	Correlation	p-value
Human RT to Cross Entropy	0.048	0.000**
Human ACC to Model ACC	0.000	0.970

※ RT, ACC = Reaction time, Accuracy

※ **: $p < 0.01$

논 의

시뮬레이션 1에서 구성된 모델의 자체적인 결과는 딥러닝 알고리즘이 빈도 효과를 모사해 낼 수 있음을 보였다. 하지만 행동실험 데이터와의 비교 결과는, 제시된 모델이 인간의 언어처리를 적절하게 반영해내지 못하고 있음을 보임으로써, 모델이 모사해 낸 빈도 효과를 신뢰하기 어렵다는 제한점 또한 보여주었다.

이와 같은 결과는 본 연구의 목적과 정반대 방향의 흐름이라고 할 수 있다. 본 연구의 시작의 목적은 딥러닝의 과도한 정확성에도 불구하고 빈도 효과와 같은 인간의 행동 패턴을 보일 수 있는가에 대한 의문이었으나, 시뮬레이션 1의 결과는 오히려 인간의 수행능력에 한참 못 미치는 결과가 나타남으로써, 정확성이 떨어짐을 보였다.

이 결과에 대한 가장 가능성이 높은 가설은, 구성된 모델의 연산능력이 전체 패턴을 처리하기엔 모자라다는 것이다. 딥러닝은 은닉층의 개수를 증가시킴으로써 그 연산능력을 향상시키며, 시뮬레이션 1에서 사용한 100개 단위로 구성된 은닉층 5개는 25,006개의 전체 패턴을 처리하기에 연산능력이 모자랐을 가능성이 있다.

이 제한점 해소를 위해, 본 연구는 시뮬레

이션 2를 진행하여 시뮬레이션 1의 제한된 결과가 연산능력의 부족으로 발생하였음을 보이고, 나아가 연산능력을 지속적으로 향상시키더라도 빈도 효과가 계속적으로 나타날 수 있음을 보임으로써, 연구의 주목적인 과도한 정확성에 의한 제한이 없음을 보이고자 하였다.

시뮬레이션 2

시뮬레이션 1의 논의에서 언급한 바와 같이, 시뮬레이션 1의 제한된 결과가 연산능력의 부족에서 기인하였음을 보이고, 연산능력을 향상시키더라도 빈도 효과가 나타날 수 있음을 보이고자 시뮬레이션 2를 진행하였다.

모델링 구조 201개 단위를 보유한 입력층에서 은닉층을 거쳐 2개의 출력층까지 일렬로 연결되도록 함으로써, 기본 구조는 시뮬레이션 1과 동일하게 구성하였다. 다만 시뮬레이션 1과의 차이점으로 은닉층의 개수를 1개에서 10개까지 다양화하였다. 이와 같은 은닉층 개수의 다양화는 시뮬레이션에서 연산능력의 스펙트럼을 형성하는 것이 목적이었다.

자극 시뮬레이션 2는 시뮬레이션 1과 동일하게 단어 12125, 비단어 12886개의 패턴을 사용하였다.



Figure 9. The structure of model in simulation 2. The condition which model has 5 hidden layer is same to the model of simulation 1(see Figure 4). Other models are also same to the model of simulation 1, except the amount of hidden layer.

선행 학습 은닉층의 개수 변화에 따른 단계의 증감을 제외하곤 모델링 2의 각 모델들의 선행 학습 과정은 모델링 1과 동일하였다. 차이점인 은닉층의 개수에 따라, 해당 은닉층에 도달할 때까지 반복적으로 선행학습이 진행되었다. 즉, 은닉층의 개수가 3개인 모델의 경우 은닉층3까지, 은닉층의 개수가 10개인 모델의 경우 은닉층10까지 선행 학습이 진행되었다.

미세 조정 미세 조정 과정 또한 모델링 1과 동일하게 오류 역전파 알고리즘을 사용하여 학습이 진행되었다.

과제 시뮬레이션 시뮬레이션 1과 동일한 세 자극 집단에 대한 어휘 판단 과제(lexical decision task)를 모사한 시뮬레이션 과제를 수행하였다.

결 과

학습 진행간 교차 엔트로피 변화 Figure 10의 위는 모델들의 교차 엔트로피의 흐름을 보여준다. 시뮬레이션 1과 비슷한 형태로 모델의 교차 엔트로피는 10세대에서 일정 지점에 도달하고 큰 변동 없이 유지되었다. 다만 은닉층의 개수에 따라서 도달 지점이 약간씩 차이가 나타나고 있으며, 은닉층이 많을수록 그 도달 지점이 낮아졌다.

학습 진행간 정확도 변화 Figure 10의 아래는 모델들의 정확도의 흐름을 보여준다. 교차 엔트로피와 마찬가지로 모델의 정확도는 10세대에서 일정 지점에 도달하고 큰 변동 없이 유

지되었다. 또한, 교차 엔트로피와 마찬가지로 은닉층의 개수가 클수록 도달 지점이 높아졌다.

빈도 효과 은닉층의 개수에 관계없이 각 고빈도와 저빈도 자극목록에 대한 교차 엔트로피의 평균값은 유의한 차이를 보였다(Figure 11 위). 빈도의 주효과는 $F(1,70)=1184.480$ 으로 $p<0.001$ 수준에서 유의하게 차이가 나타났다. 또한 은닉층의 개수와 빈도의 상호작용의 경우, $F(9, 70)=3.082$ 로 $p<0.01$ 수준에서 유의하였다.

각 은닉층 개수 변인 내에서의 빈도 효과는 $p<0.001$ 수준에서 유의한 차이가 나타났다(Table 3).

교차 엔트로피와 비슷하게, 고빈도와 저빈도 자극목록에 대한 정확도의 평균 값 차이 역시 은닉층의 개수에 관계없이 유의한 차이를 보였다(Figure 11 아래). 빈도의 주효과는 $F(1,70)=836.769$ 로 $p<0.001$ 수준에서 유의하게 차이가 나타났다. 하지만 은닉층의 개수와 빈도의 상호작용은 경우, $F(9, 70)=1.068$ 로 유의하지 않았다($p=0.397$).

각 은닉층 개수 변인 내에서의 빈도 효과는 $p<0.001$ 수준에서 유의한 차이가 나타났다(Table 4).

은닉층 개수와 교차 엔트로피 차이의 상관

Figure 12의 위는 은닉층 개수와 저빈도와 고빈도의 교차 엔트로피의 차(저빈도 - 고빈도)를 상관 분석한 결과이다. 상관분석 결과 $r=0.487$ 로 $p<0.01$ 수준에서 유의하였다.

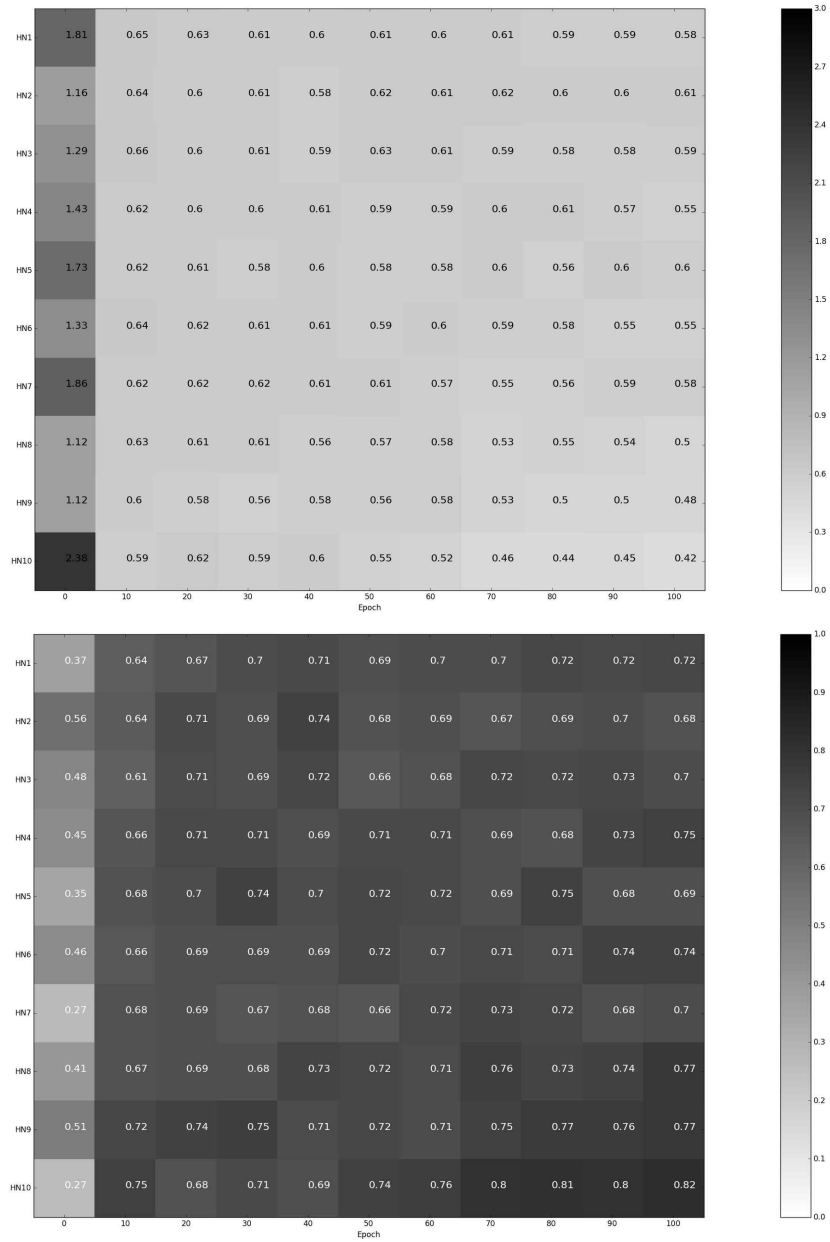


Figure 10. The cross entropy and accuracy by the progress of epoch. Upper and below graphs display about the cross entropy and accuracy, respectively. HN means the amount of hidden layer. The graphs show that there is no big change in both of them after 10 epoch like simulation 1. However, each model's the reaching point is different, and more hidden layer related with the lower and higher reaching point of cross entropy and accuracy, respectively.

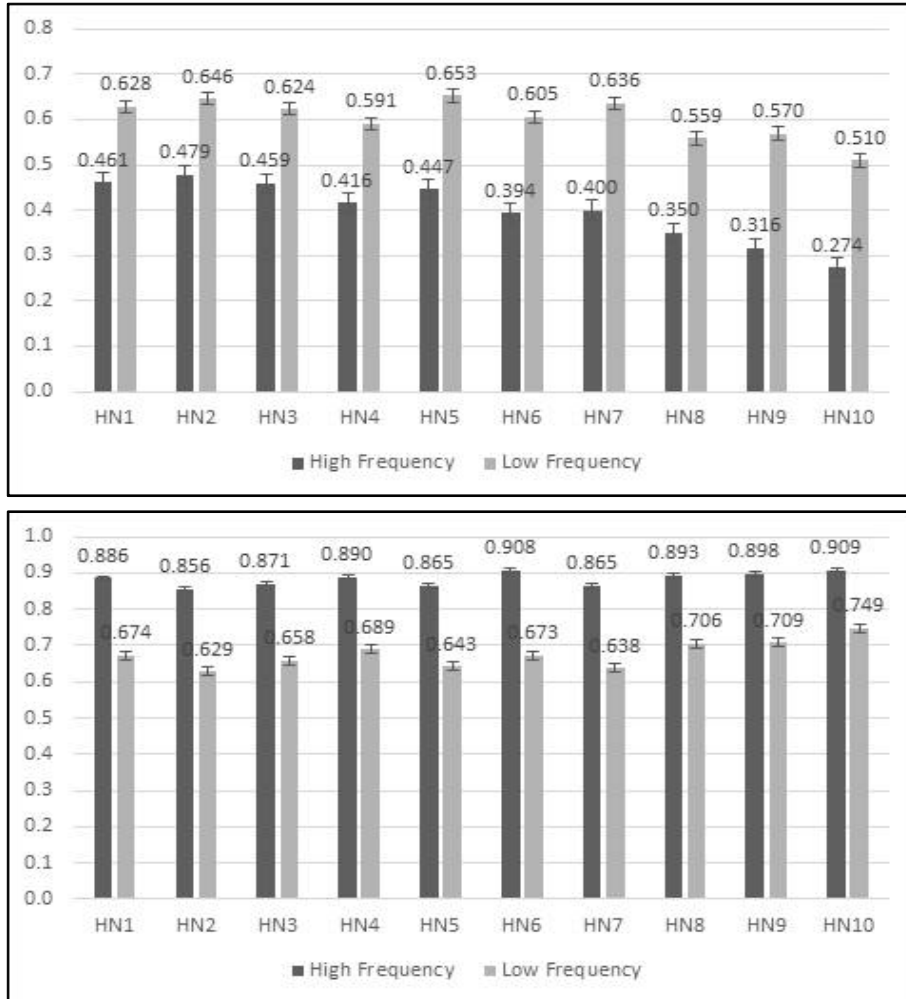


Figure 11. The difference between high and low frequency. Upper and below graphs display about the cross entropy and accuracy, respectively. HN means the amount of hidden layer. The graphs shows that both of cross entropy and accuracy are better when the amount of hidden layer increase.

은닉층 개수와 정확도 차이의 상관 Figure 12의 아래는 은닉층 개수와 저빈도와 고빈도의 정확도 차이(고빈도 - 저빈도)를 상관 분석한 결과이다. 상관분석 결과는 $r=-0.215$ 에 $p=0.056$ 으로 유의하지는 않았으나 부적의 경향을 보였다.

자극별 반응시간과 은닉층 10개 조건의 교차 엔트로피의 상관 시뮬레이션 1의 분석과 같은 방식으로, 12,097개의 행동실험자극의 반응 시간 평균과 은닉층이 10개인 시뮬레이션의 교차 엔트로피 간의 상관관계를 분석하였다. 둘 간의 상관 계수는 $r=0.160$ 로 나타났으며,

Table 3. The statistical analysis of the cross entropy difference between high and low frequency

	High frequency	Low frequency	Difference	t-distribution
HN1	0.461	0.628	0.167 ^{***}	18.115
HN2	0.479	0.646	0.168 ^{***}	23.210
HN3	0.459	0.624	0.165 ^{***}	16.285
HN4	0.416	0.591	0.175 ^{***}	14.648
HN5	0.447	0.653	0.206 ^{***}	11.067
HN6	0.394	0.605	0.212 ^{***}	18.831
HN7	0.400	0.636	0.236 ^{***}	22.764
HN8	0.350	0.559	0.209 ^{***}	11.990
HN9	0.316	0.570	0.254 ^{***}	7.513
HN10	0.274	0.510	0.236 ^{***}	7.241

- ※ HN = The amount of hidden layer
- ※ Difference = Low frequency - High frequency
- ※ *** $p < 0.001$, ** $p < 0.01$

Table 4. The statistical analysis of the accuracy difference between high and low frequency

	High frequency	Low frequency	Difference	t-distribution
HN1	0.886	0.674	0.213 ^{***}	8.580
HN2	0.856	0.629	0.228 ^{***}	13.831
HN3	0.871	0.658	0.214 ^{***}	13.722
HN4	0.89	0.689	0.201 ^{***}	7.358
HN5	0.865	0.643	0.223 ^{***}	7.942
HN6	0.908	0.673	0.235 ^{***}	16.767
HN7	0.865	0.638	0.228 ^{***}	18.200
HN8	0.893	0.706	0.186 ^{***}	8.747
HN9	0.898	0.709	0.189 ^{***}	5.597
HN10	0.909	0.749	0.16 ^{***}	6.867

- ※ HN = The amount of hidden layer
- ※ Difference = High frequency - Low frequency
- ※ *** $p < 0.001$, ** $p < 0.01$

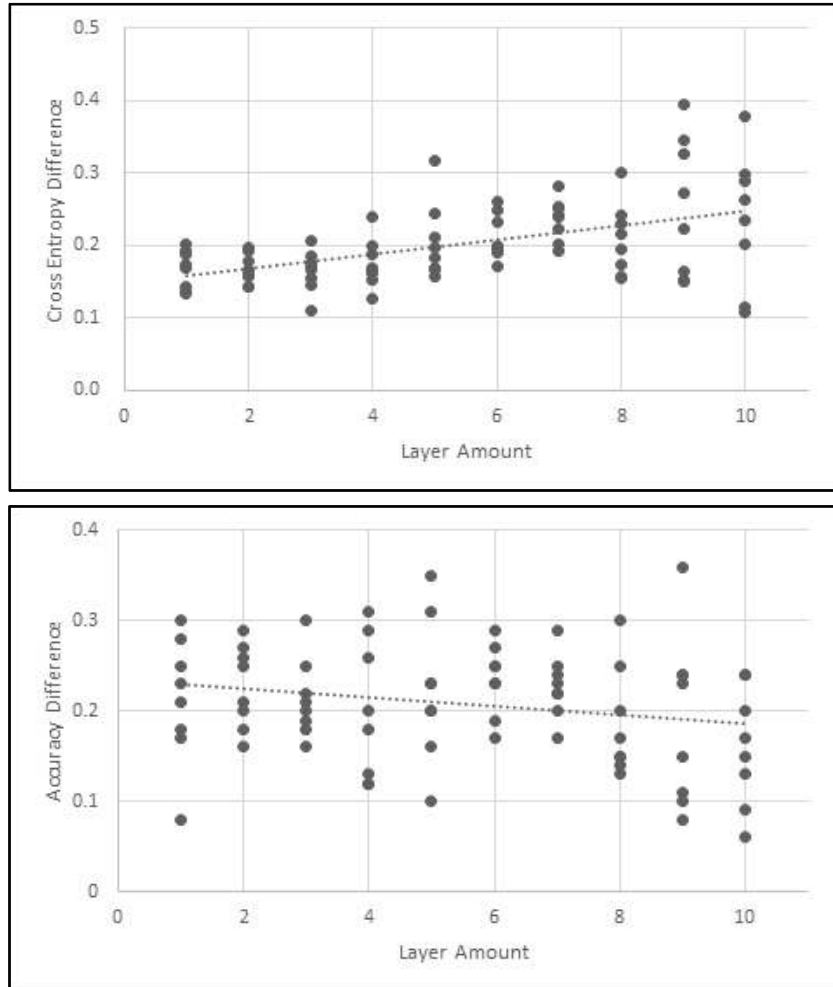


Figure 12. The dispersion between the amount of hidden layer and the difference between categories. Upper graph shows the relation between the amount of hidden layer and the difference of cross entropy (low frequency - high frequency), and there was a significant positive correlation. Below graph shows the relation between the amount of hidden layer and the difference of accuracy (high frequency - low frequency), and there was a negative correlation which is marginally significant.

Table 5. The relations between human and 10 hidden layer model data. There was a positive relation between human reaction time and cross entropy. And there was also a positive correlation between human accuracy and model accuracy. This is perfectly different from the result of simulation 1.

	Correlation	p-value
Human RT to Cross Entropy	0.160	0.000**
Human ACC to Model ACC	0.113	0.000**

※ RT, ACC = Reaction time, Accuracy

※ **: $p < 0.01$

상관은 $p < 0.01$ 수준에서 유의하였다(Table 5). 이와 같은 수치는 시뮬레이션 1의 결과와는 다르게 명확한 관련성을 보이고 있는 수치라고 할 수 있다.

자극별 반응시간과 은닉층 10개 조건의 정확도의 상관 마찬가지로 12,097개의 행동실험 자극의 정확도 평균과 은닉층 10개인 시뮬레이션의 정확도 평균 간의 상관관계를 분석하였다. 시뮬레이션 1의 결과와는 다르게 둘 간의 상관계수는 $r = 0.113$ 로 나타났으며, 상관은 $p < 0.01$ 수준에서 유의하였다(Table 5).

논 의

시뮬레이션 2의 목적은 시뮬레이션 1에서 나타난 제한점을 해소함과 동시에, 본 연구의 주목적인 과도한 정확성에 의한 제한이 없음을 보이는데 있다. 시뮬레이션 결과는 이 두 가지를 충족하고 있다고 판단된다.

첫째로 시뮬레이션 2는 은닉층의 개수를 다양화함으로써, 더 높은 연산능력을 가진 모델들이 시뮬레이션 1과는 다르게 행동실험과 유의미한 상관관계를 가지는 것을 보였다. 이는 시뮬레이션 1의 논의에서 세운 가설대로, 시뮬레이션 1의 제한점이 연산능력의 부족에서 왔음을 증명하는 것이다.

둘째로 시뮬레이션 2는 은닉층의 개수와 모델의 빈도 효과와의 상호작용을 확인함으로써, 연산능력이 증가한다 하더라도 천장효과에 의한 빈도 효과의 소실과 같은 문제가 발생하지 않음을 보였다. 이는 다시 말하자면 본 연구의 주목적인 과도한 정확성에 의한 연결주의

모델링에서의 활용제한을 심각하게 고려할 필요가 없다는 점을 의미한다. 이에 대해서는 종합논의에서 추가로 다루도록 하겠다.

종합논의

본 연구에서는 행동 실험과 딥러닝 알고리즘에 기반을 둔 모델링 시뮬레이션을 이용하여, 딥러닝 알고리즘에서 빈도 효과(frequency effect)가 나타나는가를 확인함으로써, 딥러닝의 차후 연결주의 모델링에의 적용 가능성을 알아보고자 하였다.

어휘 판단 과제를 이용한 본 연구의 행동 실험은 고빈도와 저빈도 간의 유의한 차이가 나타남으로써, 과거에 이뤄진 다수의 선행연구들과 일치하는 결과를 보였다.

DBN알고리즘을 적용하여 어휘 판단 과제를 모사한 시뮬레이션 또한, 행동 실험 결과와 유사하게 빈도 효과가 나타났다. 이는 딥러닝이 갖는 과도한 정확성에 의한 천장 효과에 대한 가능성을 부정하는 결과로서, 연결주의 모델링 연구에의 딥러닝 적용이 가능할 수 있음을 보였다고 할 수 있다.

DBN의 학습과 빈도 효과

그렇다면, 왜 DBN에서 천장 효과가 나타나지 않고 정상적인 빈도 효과가 나타났는가? 우리는 이에 대해서 크게 두 가지 이유가 있다고 판단하였다. 첫째로, 본 연구에서 사용한 미세 조정 알고리즘이 오류 역전파 알고리즘이었다는 점을 들 수 있다. 도입에서 언급한 바와 같이, 오류 역전파 알고리즘은 많은 연

결주의 모델에서 사용되어 왔으며, 해당 모델링 연구들의 결과는 일반적으로 빈도 효과를 보여 왔다(Park & Lim, 2014; Plaut, 1997; Seidenberg & McClelland, 1989; You, Nam, & Nam, 2015). 그런 점에서 봤을 때, 본 연구에서 미세 조정 알고리즘으로 사용한 오류 역전파 알고리즘이 빈도 효과를 나타내는데 긍정적인 영향을 미쳤을 것이라 생각할 수 있다.

둘째로, DBN의 선행 학습은 최종적으로 모델의 수행 능력을 개선하지만, 선행 학습 자체만으로는 패턴에 대한 수행이 상승하지 않는다는 점을 들 수 있다. 선행 학습은, 이후 학습할 모든 패턴들을 RBM 알고리즘을 이용해 1차적으로 학습함으로써, 각 층 간의 연결에 일정한 방향성을 부여하도록 조정한다. 이를 다시 말하면, 선행 학습의 목표는 차후 어떤 학습을 할 것인지 미리 모델에게 제시함으로써, 이후의 학습이 이상한 방향으로 가지 않고, 정상적으로 이뤄질 수 있도록 방향성을 지정하는 것이다(Deng & Yu, 2014; Gwak, Bak, & Yi, 2015; Hinton et al., 2006). 그렇기에 이 과정 하에서 모델은 목표를 제시 받지 않으며, 결과적으로 '정확성'에 대해서는 학습하지 않는다. RBM 알고리즘은 입력된 패턴을 그대로 되돌리도록 각 연결을 수정하는 알고리즘이므로, 전체적인 정확성이 과거의 연결주의 모델링 보다 개선되었다 하더라도, 빈도는 여전히 수행의 차이에 영향을 끼칠 수 있다.

은닉층의 개수에 따른 빈도 효과의 변화

Figure 12는 은닉층의 개수와 빈도 효과의

상관관계를 보여준다. 해당 분석에서 교차 엔트로피는 은닉층의 개수가 증가할수록 그 크기가 커지고 있으며, 그 결과가 유의하게 나타났다($p < 0.01$). 반면 은닉층의 개수와 정확도에 대한 빈도효과 간의 상관관계는 비록 유의하지는 않았지만 부적 상관관계에 대한 경향성을 보였다($p = 0.056$). 이 결과가 의미하는 바는 무엇인가?

먼저 본 연구의 목적과 상관관계의 결과를 그대로 인용한다면, 교차 엔트로피는 증가하는 형태로, 그리고 정확도의 경우 변화가 없는 형태로 모델의 은닉층이 증가한다 할지라도 빈도 효과가 여전히 나타날 수 있다고 추측할 수 있다. 하지만 정확도에서 보인 부적 상관관계의 경향성은 이를 의심하도록 만든다. 어째서 부적 상관관계가 보일 수 있는가에 대한 해석이 필요하다고 생각된다.

여러 결과에 의해 도출되는 원인으로는 이는 모델의 연산 능력의 증가가 고빈도 패턴에 비해 저빈도 패턴에 대해서 더 현저한 성능 향상을 보였기 때문이라고 생각된다. Figure 11의 아래쪽을 보면, 각 은닉층의 개수에 따른 고빈도와 저빈도의 정확도 평균이 제시되어 있다. 이에 따르면 고빈도의 정확도 평균은 은닉층 개수에 따른 증가가 거의 나타나지 않았으나, 저빈도의 정확도 평균은 상당한 차이가 존재한다. 즉, 현재의 모델은 고빈도에서는 그 성능의 개선이 미미하나, 저빈도의 경우 모델의 연산 능력에 따라 증가의 여지가 있었다고 볼 수 있다.

그렇다면 은닉층을 매우 크게 증가시킨다면, 본 연구가 당초 언급했던 상한 효과가 나타나는 문제점이 있을 수 있지 않은가 하는

의문이 생길 수 있다. 하지만 이 의문은 기본적으로 본 연구의 목표인 딥러닝 알고리즘을 연결주의 모델에 적용할 수 있는가와는 다른 문제이다. 시뮬레이션 2의 결과는 충분한 연산능력을 갖춘 시점에도 딥러닝 알고리즘에서 빈도 효과가 모사가 가능함을 보이고 있다. 이와 같은 결과는 오히려, 딥러닝을 적용함으로써 더 많은 형태의 사람들 간의 개인차까지 반영될 수 있음을 나타낸다고 보는 게 타당하다. 도입에서 언급한 바와 같이 종래의 연결주의 모델에서 갖는 최대 연산능력의 한계를 딥러닝이 극복가능하며, 이를 통해 더 넓은 범위의 개인 능력 간 차이를 반영할 수 있게 되었다고 볼 수 있다.

입력층의 형태가 갖는 제한점과 장점

시뮬레이션 1에서 제시한 바와 같이 본 연구에서 사용된 모델의 입력층 구조는 시지각 단계를 생략하고, 추상화된 형태의 문자정보를 표상한다. 이는 같은 단어 자극은 항상 동일한 입력 패턴을 가짐으로써 학습 시의 동일 자극일 경우의 노이즈가 전혀 없도록 만들기 때문에, 같은 단어 자극이 제시되었을 때 항상 동일한 형태의 처리를 시도하게 된다. 기계학습의 관점에서 보았을 때 이와 같은 구조는 빈도 효과를 당연하게 만든다는 문제점이 있다. 만약 이 관점을 따른다면 입력층의 구조를 비트맵 이미지로 처리하는 것이 더 타당할 수 있다. 그렇다면 본 모델에서 사용된 입력층 구조는 개선되어야 하는가? 본 연구는 그럼에도 불구하고 이 입력층 구조가 크게 세 가지의 장점을 갖고 있다고 판단한다.

첫째로 이미 언급한 바와 같이 해당 구조를 사용함으로써 모델의 간략화가 가능해진다. 비트맵 이미지와 같은 구조는 입력층의 단위의 개수를 극단적으로 증가시킴으로써 모델의 연산 시간을 증가시킨다. 반면 현재 사용된 모델은 이보다 훨씬 적은 양의 단위를 사용하며, 이는 모델의 학습 및 연산 속도를 향상시켜, 실제 시뮬레이션 시간 및 연구 진행을 용이하게 하는 효과가 있다.

둘째로 입력에 사용될 정보를 구성하는 부담이 감소한다. 추상화되지 않은 비트맵 이미지의 경우 노이즈를 구성할 수 있다는 점은 사실이나, 막상 해당 정보를 구성하기 위한 데이터를 확보하는데는 매우 커다란 노력이 필요하다. 노이즈를 구성하면서 동시에 모델이 습득할 수 있을 만한 특성을 갖기 위해서는 둘 내지는 셋의 소수의 서로 다른 자극으로는 해결되지 않기 때문이다. 자극에 대한 너무 적은 예시는 추상화된 구조와 사실상 동일한 문제점을 갖게 되기 때문이다. 동시에 현재 구성되어 있는 한국어의 각 단어에 대한 디지털화된 필기 정보도 확보하기 어려운 상황으로, 본 연구에서 사용된 단어량에 대응될 만한 비트맵 이미지가 확보가능할지는 완전히 미지수이다. 그런 점에서 봤을 때 입력에 사용될 정보 구성이 용이한 현재의 추상화된 구조는 충분한 메리트를 갖고 있다고 할 수 있다.

마지막으로 가장 중요한 장점은 모델의 세분화가 가능해진다는 점이다. 본 연구에서는 철자 정보가 추상화된 이후의 과정만을 나타내었다. 이는 다르게 말하면 '시지각 → 추상화 → 의미 표상'을 각각 다른 과정으로 전제

하고, 이들 과정 중 '시지각 → 추상화'의 과정을 분리하였다는 것을 의미한다. 이에 따라 모델은 철자에 의한 의미 표상 과정을 좀 더 세분화하여 이해할 수 있게 되며, 만약 이후 연구들에서 시지각 과정에서의 논의가 요구된다면 본 구조의 이전 과정으로 시지각 과정을 추가하는 형태 또한 가능해진다.

이런 모든 장점에 더해 가장 중요한 점은 모델이 정상적으로 빈도 효과를 나타낼 수 있었다는 점이다. 비록 기계학습적 관점에서 일부 제한점이 있다 하더라도, 본 연구는 빈도 효과를 보임으로써 딥러닝의 연결주의 모델 적용 가능성을 보이고자 한 연구의 목적에서 벗어나진 않았다고 판단된다.

단순화된 출력층의 구조와 언어 정보 처리에서의 제약

많은 연결주의 모델을 이용하여 어휘 판단 과제의 시뮬레이션을 진행한 선행연구들은, 의미 정보들을 표상 가능한 구조로 출력층을 구성하였다(Park & Lim, 2014; Plaut, 1996, 1997; Plaut, McClelland, Seidenberg, & Patterson, 1996; Yim, Lim, Park, & Nam, 2005; You, Nam, & Nam, 2015). 이에 반해 본 연구에서 사용된 모델은, 시뮬레이션 1에서 언급한 바와 같이 출력층이 단어 그리고 비단어의 두 단위로 구성됨으로써 매우 단순화되어 있다. 이와 같은 구조는 의미 정보가 표현되지 않는다는 점에서는 실제 언어 처리 모델로서는 제한적일 수 있다. 그렇다면 의미 정보가 온전하게 반영되지 않은 본 연구의 결과는 언어 처리를 모사하였다고는 말하기 어려운 것이 아닌가? 이에

관해서 고려해야 할 점은, 과제를 진행하는 참여자들이 의미정보만을 사용하거나, 혹은 단어의 의미 정보를 완전하게 활용하는 것은 아니라는 점이다.

Ziegler, Besson, Jacobs, Nazir(1997)는 ERP를 이용하여 단어, 유사단어(pseudo-word), 비단어(nonword)를 처리하는 동안의 참여자의 뇌파의 차이를 확인하였다. 그리고 연구 결과는 과제 그리고 부위에 따라 유사단어가 단어와 비슷한 뇌파 형태를 보이기도 하고, 비단어와 유사한 형태를 보이기도 하였으며, 이를 통해 연구자들은 사람이 다양한 정보원을 바탕으로 언어 정보 처리를 진행함을 주장하였다.

Plaut(1997) 또한 의미 정보를 완전하게 사용하였는가를 판단 기준으로 삼는 것은 적당하지 않음을 지적하였다. 그에 따라 해당 연구에서는 어휘 판단 과제를 모사한 시뮬레이션을 진행하면서 목표값과의 차이가 아니라 출력층의 각 단위가 얼마나 명확하게 정보를 표현하였는가를 나타내는 의미 강조값(semantic stress)을 사용하였다.

이와 같은 선행 연구들은 여러 가지 정보가 과제 수행에서 활용되고 있으며, 의미 정보가 부분적으로 이용되고 있음을 제시하고 있다. 그럼 제시된 모델은 그와 같은 정보가 활용되고 있는가? 이에 관해서는 확실히 긍정이 가능하다. 비록 '단어/비단어'의 이분법적 구조는 극단적으로 제한된 의미 정보이나, '입력된 문자 패턴은 단어다'라는 의미 정보만은 확실히 학습이 가능하도록 한다. 그리고 자소 단위의 조합으로 구성된 입력 패턴 또한 어휘 판단 과제에서 사용할 수 있는 중요한 정보로서 기능한다. '뿔'과 같은 음절 단위에서의 패턴 조

합은 모델이 단어의 일부로서 학습한 적이 없으며, 이는 해당 패턴이 제시되었을 때 단어로써의 판단되지 않도록 유도할 수 있다. 특히 후자는 인간이 유사단어의 비단어 인식을 어려워하는 것과 마찬가지로, 각 문자열의 자소 조합이 단어와 비슷할 때 단어로써 잘못 판단하는 형태의 정보 처리도 가능하게 된다. 물론 그럼에도 불구하고 이와 같은 간략화된 구조가 제한이 있다는 것은 사실이다. 본 연구의 모델들의 고빈도 단어에 대한 정확도가 실험에서 제시된 실제 참여자들의 정확도에 비해 낮은 것은 의미 정보의 간략화에 따른 영향일 가능성도 있다. 하지만 제시된 구조로도 인간과 유사한 형태로 언어 처리를 모사하는 것이 가능하다는 것만은 분명하다.

선행 모델링 연구들의 딥러닝 적용 가능성

연결주의 모델링에서의 DBN 알고리즘 적용이 갖는 장점 중 하나는, 종래의 연결주의 모델링 구조를 거의 대부분 차용 가능하다는 점에 있다. 시뮬레이션 1에서 언급한 바와 같이, 미세 조정 과정은 다양한 방법이 존재한다(Bengio et al., 2013; Dahl et al., 2011; Gwak, Bak, & Yi, 2015; Hinton et al., 2006). 그리고 본 연구에 적용한 바와 같이 연결주의 모델링에서 주로 사용하는 알고리즘인 오류역전파 알고리즘 또한 미세 조정 과정으로 이용 가능하다.

이는 다시 말하자면 과거 연결주의 모델로 설명을 시도하였던 다양한 현상에도 딥러닝 알고리즘을 적용 가능함을 의미한다. 자소나

음절의 길이에 따라 반응시간이 증가하는 길이 효과(length effect), 비슷한 문자열 혹은 음소 구성의 단어가 많을수록 반응시간이 빨라지는 이웃 단어 효과(neighborhood effect), 보유한 의미 요소(semantic feature)가 많을수록 반응시간이 빨라지는 의미 충실도 효과(semantic richness effect)와 같은 현상들에 대해서는 선행 연구들에서 연결주의 모델을 통해 재현되어 왔으며, 이 또한 딥러닝을 통해 재현 가능하다 판단한다(Cree, McNorgan, & McRae, 2006; Park & Lim, 2014; Plaut, 1996, 1997; Plaut, McClelland, Seidenberg, & Patterson, 1996; Yim, Lim, Park, & Nam, 2005; You, Nam, & Nam, 2015). 특히 이웃단어효과나 의미 충실도 효과와 같은 현상들은 딥러닝의 적용이 중요하게 작용할 수 있다. 해당 현상에 대한 선행 연구들은 대규모 단어 사용에 어려움을 겪었기 때문에 패턴 개수를 제한하고, 모델이 학습한 패턴 내에서 재계산 하는 등의 형태로 시뮬레이션을 진행하는 등의 제한이 있었다. 하지만 딥러닝의 적용은 모델의 연산능력을 증가, 실제 언어에서 사용하는 전체 패턴을 이용함으로써 문제를 해결하여, 실제 인간의 행동실험 결과와 직접 비교하는 방식으로 제한점이 해소될 여지가 있다.

본 연구는 행동 실험과 딥러닝 알고리즘에 기반을 둔 모델링을 진행하여 빈도 효과를 성공적으로 시뮬레이션 함으로써, 딥러닝의 연결주의 모델링에의 적용 가능성을 보였다. 이와 같은 결과는 차후 연결주의 모델링 연구에 딥러닝 알고리즘이 이용될 수 있도록 함으로써, 종래의 모델링 연구가 갖고 있던 한계를 극복하는 기반이 될 것으로 기대한다.

하지만 동시에 종합논의에서 언급한 바와 같이 본 연구는 제한적인 모델 구조를 갖고 있으며, 2음절 명사라는 한정된 단어 패턴만을 이용하였다는 점, 그리고 빈도효과라는 단일 현상만을 설명하였다는 점에서 아직 개선의 여지가 있다. 본 모델의 제한점은 차후에 추가적인 연구를 통하여 해결해 나가야 할 것이다.

참고문헌

- Andrews, S. (1992). Frequency and neighborhood effects on lexical access: Lexical similarity or orthographic redundancy?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 234-254.
- Bengio, Y., Courville, A., & Vincent, P. (2013). Representation Learning: A Review and New Perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35, 1798-1828.
- Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning Long Term Dependencies with Gradient Descent is Difficult. *IEEE Transactions on Neural Networks*, 5, 157-166.
- Bowman, S. R. (2013). Can recursive neural tensor networks learn logical reasoning?. *arXiv preprint arXiv:1312.6192*.
- Bowman, S. R., Angeli, G., Potts, C., & Manning, C. D. (2015). A large annotated corpus for learning natural language inference. *arXiv preprint arXiv:1508.05326*.
- Bowman, S. R., Manning, C. D., & Potts, C. (2015). Tree-structured composition in neural networks without tree-structured architectures. *arXiv preprint arXiv:1506.04834*.
- Bowman, S. R., Potts, C., & Manning, C. D. (2014). Learning distributed word representations for natural logic reasoning. *arXiv preprint arXiv:1410.4176*.
- Caruana, R., Lawrence, S., & Giles, L. (2001). Overfitting in neural nets: Backpropagation, conjugate gradient, and early stopping. In Leen, T. K., Dietterich, T. G. & Tresp, V. (Eds.), *Advances in Neural Information Processing Systems* (pp. 402-408). Cambridge, MA: The MIT Press.
- Chen, D., & Manning, C. D. (2014). A Fast and Accurate Dependency Parser using Neural Networks. In *Conference on Empirical Methods in Natural Language Processing* (pp. 740-750). Doha: Association for Computational Linguistics.
- Ciresan, D. C., Meier, U., & Gambardella, L. (2010). Deep Big Simple Neural Nets Excel on Handwritten Digit Recognition. *Neural Computation*, 22, 1-14.
- 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.
- Dahl, G. E., Yu, D., Deng, L., & Acero, A. (2012). Context-dependent pre-trained deep neural networks for large-vocabulary speech

- recognition. *IEEE Transactions on Audio, Speech, and Language Processing*, 20, 30-42.
- Dahl, G. E., Yu, D., Li, D., & Alex, A. (2011). Large vocabulary continuous speech recognition with context-dependent DBN-HMMs. In *IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 4688-4691). Piscataway, NJ: IEEE.
- Deng, L., & Yu, D. (2014). *Deep Learning: Methods and Applications. Found. Trends Signal Process.* Hanover, MA: Now Publishers Inc.
- Grainger, J. (1990). Word frequency and neighborhood frequency effects in lexical decision and naming. *Journal of Memory and Language*, 29, 228-244.
- Graves, A., Mohamed, A. R., & Hinton, G. (2013). Speech recognition with deep recurrent neural networks. In *IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 6645-6649). Piscataway, NJ: IEEE.
- Gwak, D., Bak, S., & Yi, H. (2015). *Machine Learning to Deep Learning*. Seoul: Purple.
- Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: insights from connectionist models. *Psychological Review*, 106, 491-528.
- Hinton, G. (2010). A Practical Guide to Training Restricted Boltzmann Machines A Practical Guide to Training Restricted Boltzmann Machines. *Computer*, 9, 1.
- Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A. R., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Kingsbury, B., & Sainath, T. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, 29, 82-97.
- Hinton, G. E., Osindero, S., & Teh, Y. -W. (2006). A Fast Learning Algorithm for Deep Belief Nets. *Neural Computation*, 18, 1527-1554.
- Huang, E. H., Socher, R., Manning, C. D., & Ng, A. Y. (2012). Improving word representations via global context and multiple word prototypes. In *the 50th Annual Meeting of the Association for Computational Linguistics* (pp. 873-882). Jeju: Association for Computational Linguistics.
- Hutzler, F., Ziegler, J. C., Perry, C., Wimmer, H., & Zorzi, M. (2004). Do current connectionist learning models account for reading development in different languages?. *Cognition*, 91, 273-296.
- Ishii, T., Komiyama, H., Shinozaki, T., Horiuchi, Y., & Kuroiwa, S. (2013). Reverberant speech recognition based on denoising autoencoder. In *the 14th Annual Conference of the International Speech Communication Association* (pp. 3512-3516). Lyon: International Speech Communication Association.
- Kim, J., Nan, C., & Zhang, B. (2015). Deep Learning-based Video Analysis Techniques. *Communications of the Korea Information Science Society*, 33, 21-31.
- Kim, S. K., McAfee, L. C., McMahon, P. L., &

- Olukotun, K. (2009). A Highly Scalable Restricted Boltzmann Machine Fpga Implementation. In *2009 International Conference on Field Programmable Logic and Applications* (pp. 367-372). Prague: IEEE.
- Kwon, Y., & Lee, Y. (2014). Time course of Word Frequency and Word Length Effect in Visual Word Recognition: Evidence from Event-Related Brain Potential Study. *The Journal of Linguistic Science*, *69*, 43-62.
- Kwon, Y., & Nam, K. (2011). The Relationship Between Morphological Family Size and Syllabic Neighborhoods Density in Korean Visual Word Recognition. *The Korean Journal of Cognitive and Biological Psychology*, *23*, 301-319.
- Kwon, Y., Park, K., Lim, H., Jung, S., & Nam, K. (2006). Word frequency effect and word similarity effect in korean lexical decision task and their computational model. *Neural Information Processing*, *4234*, 331-340.
- LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*, 2278-2323.
- Luong, T., Socher, R., & Manning, C. D. (2013). Better Word Representations with Recursive Neural Networks for Morphology. In *17th Conference on Computational Natural Language Learning* (pp 104-113). Sofia: Special Interest Group on Natural Language Learning of the Association for Computational Linguistics.
- Monsell, S., Doyle, M. C., & Haggard, P. N. (1989). Effects of frequency on visual word recognition tasks: where are they?. *Journal of Experimental Psychology. General*, *118*, 43-71.
- Nam, K., Seo, K., Choi, K. -S., Lee, K., Kim, T., & Lee, M. (1997). The Word Length Effect on Hangul Word Recognition. *Korean Journal of Experimental and Cognitive Psychology*, *9*, 1-18.
- Park, K., & Lim, H. (2014). A computational model explaining language phenomena on Korean visual word recognition. *Cognitive Systems Research*, *27*, 11-24.
- Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global Vectors for Word Representation. In *Proceedings of Conference on Empirical Methods in Natural Language Processing* (pp. 1532-1543). Doha: Association for Computational Linguistics.
- Plaut, D. C. (1996). Relearning after damage in connectionist networks: toward a theory of rehabilitation. *Brain and Language*, *52*, 25-82.
- 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.
- Plaut, D. C. (1999). A Connectionist Approach to Word Reading and Acquired Dyslexia: Extension to Sequential Processing. *Cognitive Science*, *23*, 543-568.
- 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.
- Richmond, K., Hoole, P., & King, S. (2011). Announcing the Electromagnetic Articulatory (Day 1) Subset of the mngu0 Articulatory Corpus. In *12th Annual Conference of the International Speech Communication Association* (pp. 1505-1508). Portland: International Speech Communication Association.
- Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96, 523-568.
- Socher, R., Chen, D., Manning, C. D., & Ng, A. (2013). Reasoning with neural tensor networks for knowledge base completion. In *Advances in Neural Information Processing Systems* (pp. 926-934). Lake Tahoe, NV: Neural Information Processing Systems.
- Socher, R., Huang, E. H., Pennington, J., Manning, C. D., & Ng, A. Y. (2011). Dynamic pooling and unfolding recursive autoencoders for paraphrase detection. In *Proceeding of Advances in Neural Information Processing Systems* (pp. 801-809). Granada: Neural Information Processing Systems.
- Socher, R., Huval, B., Manning, C. D., & Ng, A. Y. (2012). Semantic compositionality through recursive matrix-vector spaces. In *the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning* (pp. 1201-1211). Jeju: Association for Computational Linguistics.
- Socher, R., Pennington, J., Huang, E. H., Ng, A. Y., & Manning, C. D. (2011). Semi-supervised recursive autoencoders for predicting sentiment distributions. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (pp. 151-161). Edinburgh: Association for Computational Linguistics.
- Socher, R., Perelygin, A., Wu, J. Y., Chuang, J., Manning, C. D., Ng, A. Y., & Potts, C. (2013). Recursive deep models for semantic compositionality over a sentiment treebank. In *the Conference on Empirical Methods in Natural Language Processing* (pp. 1642). Seattle: Association for Computational Linguistics.
- Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to Grow a Mind: Statistics, Structure, and Abstraction. *Science*, 331, 1279-1285.
- Yi, K. (1993). On the Role of Frequency and Internal Structure in the Processing of Kulca. *Korean Journal of Experimental and Cognitive Psychology*, 5, 26-39.
- Yi, K., Park, K., Abe, J., Liu, Y., & Zhang, Y. (2010). A Cross-Linguistic Study on Representation and Processing of Hanja Words: Naming and Lexical Decision. *The Korean Journal of Cognitive and Biological Psychology*, 22, 277-291.
- Yim, H., Lim, H., Park, K., & Nam, K. (2005). A Computation Model of Korean Lexical. *Advances in Natural Computation*, 3610, 844-849
- You, H., Nam, K., & Nam, H. (2015). Semantic

- Process Possibility Research in Featural View
Connectionist Modeling. *The Korean Journal of
Cognitive and Biological Psychology*, 27, 613-638.
- Ziegler, J. C., Besson, M., Jacobs, A. M., &
Nazir, T. A. (1997). Word, Pseudoword, and
Nonword Processing: A Multitask Comparison
Using Event-Related Brain Potentials. *Journal
of Cognitive Neuroscience*, 9, 758-775.
- 1 차원고접수 : 2016. 05. 04
수정원고접수 : 2016. 09. 19
최종게재결정 : 2016. 09. 22

The Frequency based Study of the Applicability of DBN Algorithm on Language Acquisition Modeling

Heejo You¹⁾ Min-Mo Koo²⁾ Kwangoh Yi³⁾ Kichun Nam¹⁾

¹⁾Department of Psychology, Korea University

²⁾Biomedical Research Center, Korea University Anam Hospital

³⁾Department of Psychology, Yeungnam University

Nowadays, computer science has recently been introduced and used deep learning algorithms in the field of pattern recognition. However, those deep learning algorithm has not been utilized in the field of connectionist modeling of language process which has used pattern recognition algorithms for computational perspective yet. In this study, we made a modeling which use the deep belief network which is a type of deep learning algorithm, and train the relation between words and semantic. After training, the model conducted the lexical decision task for model, and the results were statistically verified that it is similar with result of behavioral experiment with frequency effect as center. As the results of this study, the model showed that model was able to conduct language process through drawing the frequency effect. This result suggested that the model which used deep learning algorithm is able to be used as connectionist modeling, and to simulate the language process of human. In addition, in this study, we discussed how deep belief network can be applied to the connectionist modeling.

Key words : connectionist modeling, deep learning, frequency effect, lexical decision task

부 록

ANN의 제약과 DBN의 목적 Artificial neural network(이하 ANN)은 입력층(input layer)과 은닉층(hidden layer), 출력층(output layer)으로 구성되고, 각 층들이 순차적으로 가중치(weight)에 의해 연결된 네트워크이다. ANN에서 은닉층의 개수를 증가시키는 이유는 크게 두 가지가 있다. 하나는 연산능력을 증가시키기 위한 경우, 다른 하나는 연구자가 상정한 모델이 근본적으로 다층 구조에 의한 단계적 연산을 요구할 경우이다.

연산능력을 증가시키는 방법은 은닉층의 개수 외에도 은닉층이 보유한 단위(unit)를 증가시키는 방법이 있다. 하지만, 은닉층 단위를 증가시킴으로서 얻는 연산능력의 증가는 크게 두 가지 문제를 불러일으킨다. 첫째로 과적합(over-fitting) 문제가 발생할 소지가 있다. ANN은 일종의 분류 알고리즘이기 때문에, 최종적인 목표는 학습된 패턴을 바탕으로 아직 보지 못한 데이터를 올바르게 분류해 내는 것이다. 하지만 과적합 문제가 발생하면 학습 데이터들을 패턴으로 분류해 내는 게 아니라, 각각을 독립적으로 학습, 이후에 새로 제시되는 패턴을 적절하게 분류해 낼 수 없게 된다(Caruana, Lawrence, & Giles, 2001). 한국어로 이를 간단하게 예시를 들자면, 일반적으로 단어 ‘학교’와 ‘학원’, ‘학생’ 등에서 ‘학’의 공통의미를 도출해내고 차후에 ‘학습’이라는 단어를 처음 본다 할지라도, ‘배우다’의 의미가 간접적으로 유추되는 경험을 들 수 있다. 정상적인 학습구조라면 이와 같은 유추가 일어날 수 있겠지만 과적합 상황에서는 이를 보장할 수 없게 된다. 또 다른 문제는 연산시간이 기하급수적으로 늘어난다는 점이다. 많은 경우 ANN은 행렬 단위의 연산을 진행하게 되는데(이는 연산의 편의성 외에도 최적화된 행렬알고리즘이 반복연산보다 빠르다는 점에서도 기인한다) 특정한 알고리즘을 사용하지 않는 행렬 곱의 연산횟수는 $O(n^3)$ 로 알려져 있다. 이는, 100×100 의 행렬이 10^6 번, 1000×1000 의 행렬이 10^9 번의 연산을 요구한다는 의미로서, 은닉층의 크기가 증가할수록 엄청난 연산 횟수를 요구하게 된다. 은닉층의 개수를 증가시키는 것은 위와 같은 문제가 발생하지는 않는다. 소규모의 층을 반복해서 통과함으로써 과적합 문제가 발생할 여지가 줄어들고, 각 층의 크기가 증가하지 않기 때문에 연산횟수도 선형으로 증가하게 된다.

문제는 은닉층의 개수를 과도하게 증가시켰을 경우 학습 자체가 일어나지 않는 문제가 발생한다. 학습되기 전의 ANN가 보유하는 각 가중치들은 일반적으로 무작위값을 갖는다. ANN에서의 학습은 학습되거나 되지 않은 다양한 패턴들을 올바르게 인식하도록 이 가중치들을 제시되는 패턴들에 맞게 조정하는 것이다. 문제는 은닉층의 개수가 과도하게 증가하면 학습 자체가 일어나지 못한다는 점에 있다. ANN에서 이용하는 학습 알고리즘인 오차역전파(back-propagation) 알고리즘은, 출력층에 도달하는 시점에 목표값과 실제 산출된 활성화값을 비교하여 그 오차를 줄여나가는 방식이다. 하지만 은닉층이 많아짐에 따라 입력층에서부터 전달되었던 값이 출력층에 도달하기 전에 사라지게 된다. 무작위하게 할당되어있는 가중치들을 곱하면서 값이 정상적으로 남아있지 못하는 것이다(Bengio, Simard, & Frasconi, 1994).

Deep belief network(이하 DBN) 알고리즘은 이와 같은 다층 구조의 ANN 네트워크가 갖는 제한점

을 해결하기 위한 수단으로 제안되었다. 선행학습을 통해 각 무작위값에 일정한 방향성을 부여함으로써, 여러 단계의 은닉층을 거치더라도 온전한 값을 도출해 낼 수 있게 하고, 그 결과로 학습이 정상적으로 이루어질 수 있도록 한다(Hinton, Osindero, & Teh, 2006).

DBN의 선행학습 본문에서 언급한 바와 같이 DBN은 선행학습(pre-training)과 미세 조정(fine-tuning)의 두 단계로 나뉘어 진행된다. 이 중 선행학습은 실제 DBN이라 부를 수 있는 곳으로서, 무작위로 값이 할당되어있는 모델의 각 가중치를 일정한 방향성을 갖도록 조정한다. 개략적인 방법은 본문에 제시된 바와 같이 각 층 별로 restrict Boltzmann machine(이하 RBM)을 적용하여 가중치를 변경하는 방식이다. RBM의 흐름은 다음과 같다(Gwak, Bak, & Yi, 2015; Kim, McAfee, McMahon, & Olukotun, 2009).

1. 시각층(visual layer)에 실제 학습에서 사용될 입력 패턴을 넣는다.

$$visual_0 = pattern_i$$

2. 시각층의 활성화값에 가중치를 곱하고 은닉층(hidden layer)의 편향치(bias)를 더한 후 시그모이드 함수(sigmoid function)를 적용하여 활성화값을 구한다(hidden₀).

$$hidden_0 = sigmoid(visual_0 \times weight + bias_h)$$

3. 은닉층의 활성화값과 가중치를 역방향으로 곱하고 시각층의 편향치를 더한 후 시그모이드 함수를 적용하여 활성화값을 구한다(visual₁).

$$visual_1 = sigmoid(hidden_0 \times weight^T + bias_v)$$

4. 2를 다시 시행하여 hidden₁을 구한다.

$$hidden_1 = sigmoid(visual_1 \times weight + bias_h)$$

5. 필요에 따라서 3-4를 반복하여 계속적으로 값을 구한다. 보통은 1회로 종료한다.

6. 가중치를 갱신한다.

$$weight += \alpha (visual_0^T \times hidden_0 - visual_1^T \times hidden_1)$$

$$\alpha = \text{학습률}$$

7. 시각층의 편향치를 갱신한다.

$$b_v = \alpha (visual_0 - visual_1)$$

8. 은닉층의 편향치를 갱신한다.

$$b_h = \alpha (hidden_0 - hidden_1)$$

9. 지정된 세대(epoch)까지 1-8을 반복한다.

본문에서 언급한 바와 같이 각 층 간의 학습은 greedy layer-wise training이 이용된다. 이는 각 층 내에서 확실하게 안정화된 상태에 도달하면 다음 층으로 넘어가는 방식이다. 구체적인 방식은 다음

과 같다(Gwak, Bak, & Yi, 2015; Kim, McAfee, McMahon, & Olukotun, 2009).

1. 입력층과 은닉층1에 대하여 RBM을 수행한다.
2. 학습이 종료된 시점에 RBM의 1-3을 수행하여 다음 학습을 위한 입력 패턴을 도출한다.
3. 2로 도출된 입력패턴을 가지고 은닉층1과 은닉층2에 대하여 RBM을 수행한다.
4. 2-3을 반복해서 시행하면서 은닉층n 까지 진행한다.

DBN의 미세조정 선형학습이 끝나면 미세조정을 실시한다. 본문에 언급된 바와 같이, 미세조정 단계는 다양한 알고리즘이 적용가능하며 본 연구에서는 일반적인 ANN에서 사용하는 오류역전파 알고리즘이 다음과 같은 흐름으로 사용되었다.

1. 입력층(input layer)에 무작위로 선택된 입력 패턴을 넣는다(input).
 $input = input\ pattern$
2. 입력층의 활성화값에 두 층 간 연결에 해당하는 가중치를 곱하고 은닉층1(hidden layer1)의 편향치를 더한 후, 시그모이드 함수를 적용하여 활성화값을 구한다(hidden₁).
 $hidden_1 = sigmoid(input \times weight_{ih_1} + bias_{h_1})$
3. 마찬가지로 방식으로 은닉층n까지 활성화값을 구한다(hidden_n).
 for(i=2 to n) $hidden_i = sigmoid(hidden_{i-1} \times weight_{h_{i-1}h_i} + bias_{h_i})$
4. 출력층의 활성화값을 구한다(output).
 $output = softmax(hidden_n \times weight_{h_no} + bias_o)$
5. 목표값(target)과 출력층의 활성화값을 비교하여 오류를 구한다.
 $output\ error = target - output$
6. 먼저 계산된 상위 층의 오류를 이용하여 각 은닉층의 오류를 구한다.
 $hidden_n\ error = output\ error \times weight_{h_no}^T \times hidden_n \times (1 - hidden_n)$
 for(i = n-1 to 1) $hidden_i\ error = hidden_{i+1}\ error \times weight_{h_ih_{i+1}}^T \times hidden_i \times (1 - hidden_i)$
7. 각 가중치를 갱신한다.
 $weight_{ih_1} += \alpha(input^T \times hidden_0\ error)$
 for(i=1 to n-1) $weight_{h_ih_{i+1}} += \alpha(hidden_i^T \times hidden_{i+1}\ error)$
 $weight_{h_no} += \alpha(hidden_n^T \times output\ error)$
8. 각 편향치를 갱신한다.
 for(i=1 to n) $bias_{h_i} += \alpha(hidden_i\ error)$
 $bias_o += \alpha(output\ error)$