

객체 인식을 위한 계층적 시간 기억 모델에 의한 이미지 표현 사례 연구

강규창, 조주필

군산대학교

kc.kang@kunsan.ac.kr, stefano@kunsan.ac.kr

A Case Study on Image Representation by Hierarchical Temporal Memory Model for Object Recognition

Kyuchang Kang, Juphil Cho

Kunsan National University

요 약

본 사례 연구는 계층적 시간 기억(HTM) 모델 기반으로 객체 인식을 하는 데 필요한 시각 입력에 대한 최소 분포 표현(SDR) 생성에 관한 사례 연구이다. 본 고에서는 HTM 모델의 소개와 SDR을 생성하기 위한 실험 구성, 그리고 이미지 데이터셋을 입력으로 하여 생성한 SDR을 직관적으로 살펴보기 위해 시각화를 통해 그 결과를 도시하고 의미를 살펴보았다.

I. 서론

인공지능이 내린 결정을 인공지능이 스스로 설명할 수 있는 ‘설명 가능한 인공지능(XAI, eXplainable Artificial Intelligence)[1]’은 인공지능의 불확실한 의사 결정 과정을 해소할 수 있어 인공지능에 대한 신뢰성을 높여 줄 수 있다. 설명 가능한 범위를 사물의 인식으로 제한하여 사람처럼 사물, 예를 들어 고양이를 설명한다고 하면 고양이는 수염과 꼬리가 있다는 근거와 같은 사람이 이해할 수 있는 판단 근거를 가지고 사물을 인식할 수 있다. 현재 설명 가능한 인공지능 연구의 주요 접근법은 역합성곱 신경망(Deconvolutional Network)과 같은 기존의 학습모델에 역산하는 과정을 추가하여 구현하고 있다.

본 사례 연구는 기존의 접근법과는 상이한 것으로 HTM(Hierarchical Temporal Memory)[2]이나 SPAUN(Semantic Pointer Architecture Unified Network)[3]과 같은 사람의 두뇌 구조나 기능을 모사한 모델을 활용하여 설명 가능한 인공지능을 구현하고자 하는 시도의 한 부분이다.

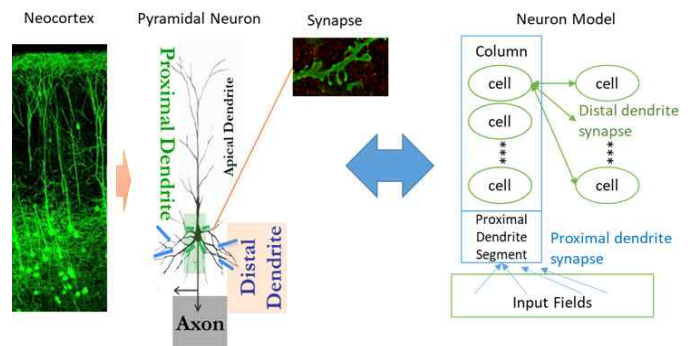
본 사례 연구에서는 HTM 모델을 기반으로 사람의 기억 과정을 모사하여 사물을 인식할 수 있도록 사물에 대한 시각적 입력에 대해 대뇌 신피질에서 뉴런의 발화패턴으로 나타나는 사물에 대한 표현을 소프트웨어적으로 모사하기 위한 실험을 실시하고 그 결과에 대한 검토 및 향후 연구방향에 관해 설명한다.

본 사례 연구에서는 사물에 대한 시각 정보가 입력되어 뉴런 모델을 거쳐 대뇌 신피질에서 뉴런 발화패턴으로 표현되는 전체 프로세스를 확인하는 것이 주요 목적이기 때문에 데이터 및 모듈의 구성을 가능한 단순한 구조로 구성하였다. 실험에 사용되는 사물에 대한 시각적 정보는 MNIST 데이터셋[4]을 사용하고 HTM 모델의 입력을 위한 이진화 과정은 MNIST 그레이 이미지를 평균값을 기준으로 흑색과 백색의 바이너리로 단순변환하여 HTM 모델의 입력으로 사용하고 출력으로는 사람이 사물을 바라볼 때 대뇌 신피질에서 나타나는 뉴런의 발화패턴과 유사하게 희박하게 분포하는 바이너리 표현(SDR, Sparse Distributed Representation)을 생성하였다.

II. 실험 및 결과 검토

가. 실험 구성 및 절차

본 사례 연구는 HTM 모델을 기반으로 하고 있으므로 먼저 이 모델의 기본적인 뉴런 모델 및 구조에 대해 (그림 1)에서 도시하였다.



(그림 1). HTM 뉴런 모델

HTM 뉴런 모델은 기존 신경망과는 다른 형태의 구조를 가지며 대뇌 신피질(neocortex) 구조를 모사하여 모델링 된다. HTM 뉴런 모델은 고차원의 시각, 청각, 촉각, 운동, 언어 및 추측과 같은 기능이 모두 대뇌 신피질 영역에서 같은 메커니즘으로 처리된다는 생물학적인 근거를 기준으로 모델링 되었다. (그림 2)는 본 사례 연구에서 실험을 위해 구성한 모듈, 연결 관계, 입출력을 나타내고 있다.



(그림 2). 실험 시스템 구성도

구성 모듈 및 입출력에 대한 간략 설명은 아래와 같다.

- 입력 데이터(Raw data): MNIST 이미지 데이터셋을 사용함
- Binary Encoder: HTM 모델의 Spatial Pooler 입력 조건을 맞추기 위해 입력을 바이너리 벡터 형식으로 변환함. 본 사례 연구에서는 가장

단순한 구조로 입력 이미지를 흑백 이미지로 변경함

- Spatial Pooler: HTM 모델의 구성요소로 바이너리 벡터를 입력으로 SDR 형식의 출력을 생성함
- Sparse Distributed Representation: Spatial Pooler의 출력 형식으로 대부분의 비트가 '0'이고 희소하게 분포하는 '1'의 비트를 가지는 비트 스트림

HTM 모델의 Spatial Pooler (SP)의 입출력 및 동작 구조에 대해 부가적으로 설명하면 아래와 같다.

- SP의 입력은 바이너리 비트인데 이 입력은 센서를 통해 직접 획득될 수도 있고 하위 영역(region)의 출력이 입력으로 사용될 수도 있는데 본 사례 연구에서는 단층으로 모델을 구축하였으므로 입력은 이미지가 바이너리로 변경된 데이터임
- SP는 각 입력에 대해 한정된 숫자의 칼럼(column)을 할당하는데 각 칼럼의 수상돌기 부분(dendrite segment)은 입력 비트들에 대한 하위 집합을 나타내는 잠재적 시냅스(potential synapse)를 가지고 있음. 각 잠재적 시냅스는 영속치(permanence value)를 가지는데 이 영속치에 따라 잠재적 시냅스는 연결된 시냅스로 상태가 변경될 수 있음
- SP는 각 칼럼에 연결된 활성화(active)된 시냅스 수를 세고 상황에 따라 부스트(boost) 만큼 곱을 하기도 함. 부스트는 해당 칼럼이 주변 칼럼에 비해서 얼마나 자주 활성화되었는지에 따라 결정됨. 즉 주변 칼럼보다 덜 자주 활성화되었다면 부스트를 더 크게 조정함
- SP는 부스트 과정 후에 가장 크게 활성화된 칼럼을 기준으로 억제 반경(inhibition radius) 내에 있는 일정량의 다른 칼럼을 비활성화(disable) 시키는데 이러한 억제 반경은 입력 바이너리 벡터들의 퍼진 정도에 따라서 동적으로 설정됨. 이러한 과정을 통해 활성화 칼럼의 희소 집합(sparse set)이 결정됨
- SP는 각 활성화 칼럼에 있어서 칼럼에 연결된 시냅스의 영속치를 조정할 수 있는데, 활성화 입력과 연결된 시냅스의 영속치는 증가시키고 비활성 입력과 연결된 시냅스의 영속치는 감소시킴으로서 시냅스의 연결이 유효/비유효 상태로 변경될 수 있음

나. 실험 결과 및 고찰

(그림 3)은 MNIST 데이터셋을 사용하여 각 라벨 이미지를 입력으로 하여 생성한 출력 SDR을 표시하였다. 시각적 비교의 편의를 위하여 각 라벨에서 서로 다른 특징을 가지는 이미지 각 3장씩을 표시하여 입력 이미지에 따른 출력 SDR이 상이해짐을 가시적으로 나타내었다. 대표적으로 나타나는 특징은 입력 바이너리 이미지의 위치, 방향, 굵기에 따라 SDR로 변환된 표현에서 패턴의 차이를 보인다. 현재 구현에서는 각각의 패턴의 차이를 수학적인 척도(measure)를 이용하여 정량화하지는 못했고 직관적인 차이를 확인하기 위해 가시화에 초점을 맞춰 실험이 진행되었다.

본 사례 연구 실험에서 확인할 수 있었던 결과는 서로 다른 특징을 가지는 입력에 대해 SDR 표현상에서도 입력 정보의 특징에 따라 유의미한 결과로 나타난다는 점이다. 이러한 결과는 사람에게 실제 영상을 보여주고 기능적 자기 공명 영상 (fMRI, functional magnetic resonance imaging) 장치를 통해 획득한 뇌 활동과 물리적으로는 일치하지는 않지만, 논리적으로는 같은 양태로 나타나는 것으로 간주할 수 있다.

다음 단계로 진행되어야 할 일은 SDR로 변환된 표현에서 활성비트(SDR 그림에서는 흰색으로 표현된 부분으로 수치적으로 '1'의 값을 가짐)로 나타나는 각 지점이 HTM 모델의 이론에 따르면 입력 정보의 특징 시맨틱을 포함하고 있는 지점이므로 활성화 비트 지점의 시맨틱을 어떻게

정의할 것인가에 관한 후속연구가 필요하다.

입력 이미지						
출력 SDR						
입력 이미지						
출력 SDR						
입력 이미지						
출력 SDR						
입력 이미지						
출력 SDR						
입력 이미지						
출력 SDR						

(그림 3). MNIST 라벨별 입력 이미지 및 출력 SDR 비교 예

III. 결론

본 사례 연구에서는 HTM 모델을 기반으로 사람의 기억 과정을 모사하여 사물을 인식할 수 있도록 사물에 대한 시각적 입력에 대해 대뇌 신경망에서 뉴런의 발화패턴으로 표현되는 것을 소프트웨어적으로 모사하기 위한 실험을 시행하고 그 결과에 대해 살펴보았다.

향후 추가적으로 실험이 진행되어야 할 부분은 현 실험에서 고정된 하이퍼파라미터 값을 사용한 부분을 입력에 따라 동적으로 적용하고 출력에 대한 명세에 따라 SDR을 도출할 수 있도록 하는 기능 보완이 필요하다.

ACKNOWLEDGMENT

본 연구는 한국연구재단의 이공학개인지초연구 지원사업 재원에 의해 수행되었음 (과제번호: 2018R1D1A1B07049231)

참 고 문 헌

- [1] Defense Advanced Research Projects Agency, DARPA “Explainable Artificial Intelligence (XAI),” DARPA presentation, DARPA. Retrieved 29 July 2020.
- [2] Jeff Hawkins, Sandra Blakeslee, “On Intelligence: How a New Understanding of the Brain Will Lead to the Creation of Truly Intelligent Machines,” Macmillan, 2007
- [3] Chris Eliasmith, “How to Build a Brain: A Neural Architecture for Biological Cognition,” Oxford Series on Cognitive Models and Architectures
- [4] Li Deng, “The MNIST Database of Handwritten Digit Images for Machine Learning Research”, IEEE Signal Processing Magazine, Volume 29, Issue 6, pp.141-142, 2012