

## 계층적 시간-기억 모델 기반 기계학습을 위한 시계열 데이터 인코더

이건재, 조주필, 강규창

군산대학교

flrjswo2003@naver.com, stefano@kunsan.ac.kr, kc.kang@kunsan.ac.kr

## Time Series Data Encoder for Hierarchical Temporal Memory Model-Based Machine Learning

Gunjae Lee, Juphil Cho, Kyuchang Kang

Kunsan National University

## 요약

본 연구는 계층적 시간-기억 (HTM, Hierarchical Temporal Memory) 모델 기반 기계학습에서 데이터 입력 측 필수 요구사항인 데이터 인코더에 관한 것이다. 특히 입력 데이터가 시계열 정보를 나타내는 데이터일 경우 HTM의 기계학습 알고리즘 입력으로 사용 가능한 데이터 형식으로 변환하는 시계열 데이터 인코더를 제안하고 공개 데이터셋을 사용한 실험을 통해 데이터 인코더 요구사항 충족여부를 검증하였다.

## I. 서론

기계학습은 기계가 코드를 명시하지 않은 동작을 데이터로부터 학습하여 실행할 수 있도록 하는 알고리즘을 개발하는 연구 분야로 최근 인공지능 기반의 다양한 방법론이 등장하였고 여러 분야에서 실제 적용되고 있다. 인공지능망은 기계학습과 인지과학에서 생물학의 신경망에서 영감을 얻은 통계학적 학습 알고리즘으로 시냅스의 결합으로 네트워크를 형성한 인공 뉴런이 학습을 통해 시냅스의 결합 세기를 변화시켜, 문제 해결 능력을 가지는 모델 전반을 일컫는다. 그러나 실제 뇌에서의 신경망 구조는 인공신경망에서 사용한 모델과는 다소 차이가 있다.

생물학적 신경망 구조와 유사한, 특히 대뇌 신피질의 생물학적 구조를 모방한 HTM (Hierarchical Temporal Memory) 기계학습 모델이 호킨스에 의해 온인텔리전스[1] 문헌에서 소개되었다. HTM은 인간의 대뇌 신피질 구조와 동작 알고리즘 특성을 이용하기 위한 기술의 일부로 모델링되었고 참조모델 구현이 오픈 소스로 공개되었다.

일반적으로 신피질은 포유류의 뇌에서 지능이 자리하고 있는 곳으로 생각되는 곳이다. 고차원의 시각, 청각, 촉각, 운동, 언어 그리고 계획과 같은 기능이 모두 신피질에 의해 일어난다. 다양하고 유연한 인지 기능의 특성을 고려할 때, 신피질에도 여러 가지 특화된 신경 알고리즘이 있을 것으로 생각할 수도 있지만 실제로는 놀라울 정도로 같은 패턴의 신경구조로 되어 있다는 것이 밝혀졌다. 센서 즉 사람의 오감 기관으로부터 유입되는 입력 데이터는 신피질에서 희소하게 분포된 신피질의 활동으로 나타난다. 이처럼 희소하게 분포된 신피질의 활동은 희소분포표현 (SDR, Sparse Distributed Representation)이라는 데이터 구조로 모델링 된다. SDR은 소위 뇌의 데이터 구조를 활성화와 비활성의 이진 벡터 또는 행렬로 표현할 수 있다는 가정으로 대용량의 정보를 표현할 수 있고 잡음에 강건하며 효율적으로 저장 가능하다고 알려졌다[2].

HTM 모델은 분류, 추론, 이상 상황 탐지 등의 응용 분야에서 활용되고 있다. HTM 모델의 기본적인 운용 구조는 원시데이터가 데이터에 특화된 인코더에 의해 이진 벡터 형태의 SDR로 변환되고 이 변환된 데이터가 HTM의 학습 아키텍처에서 기계학습을 위한 입력 데이터로 사용된다. 따라서 데이터 인코더는 입력 원시데이터의 특징을 잘 반영할 수 있는 구조

로 구현되어야 한다.

본 논문에서는 입력이 연속적인 시간과 값을 포함하는 시계열 데이터일 경우 HTM의 기계학습 데이터로 사용하기 위한 시계열 데이터 인코더를 제안한다. 제안한 인코더 검증을 위한 실험에서는 공개 시계열 데이터셋을 사용하여 SDR로 변환하는 인코더를 설계 및 구현하고 테스트를 실시하였다. 검증에서는 인코더가 갖춰야 하는 주요 요구사항에 대해 점검하였고 구현한 인코더에 의해 변환된 SDR이 이러한 요구사항을 충족하고 있음을 확인하였다.

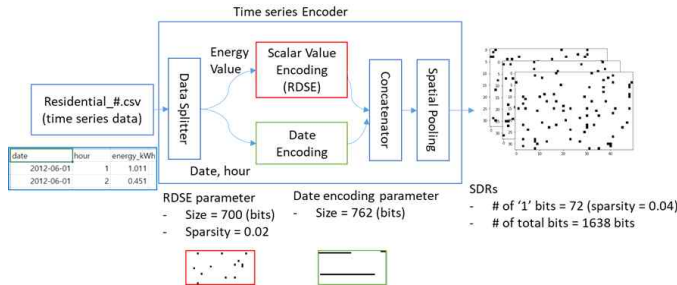
## II. 실험 및 검증

## 가. 실험의 구성 및 절차

본 연구는 HTM 모델을 기반으로 하는 기계학습 아키텍처의 입력단에서 요구되는 데이터 인코더에 관한 것으로 특히 입력 데이터가 시계열 정보를 나타내는 데이터일 경우 HTM 모델기반 기계학습 입력으로 데이터를 변환하기 위한 시계열 데이터 인코더를 구현하고 데이터 인코더 요구사항을 충족하는지 여부를 실험을 통해 확인하였다. HTM 모델에서 사용되는 데이터 인코더의 요구사항[2]은 세 가지로 다음과 같다.

- 인코더는 유사한 입력 데이터에 대해 중복된 활성 비트를 가지는 SDR 출력을 생성해야 한다.
- 인코더의 SDR 출력은 인코더 모든 입력에 대해 같은 차원을 가져야 한다.
- 인코더의 SDR 출력은 인코더 모든 입력에 대해 유사한 희소도 (sparsity)를 가져야 한다.

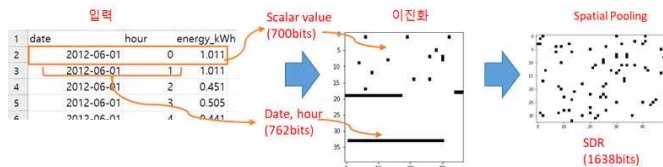
본 연구의 시계열 데이터 인코더 검증을 위한 실험의 구성은 (그림 1)과 같다. 먼저 인코더 입력으로는 주거용 건물의 시간별 에너지 소비량을 수집한 시계열 데이터셋 HUE[3]를 사용하였다. 데이터셋은 (date, hour, consumption energy) 값을 포함하고 있는 CSV(Comma Separated Value) 파일이다. 출력은 시계열 데이터 인코더의 내부 모듈을 통해 총 1,638개 비트로 구성되는 벡터로 72개의 활성비트를 가지는 희소도 4% SDR이다.



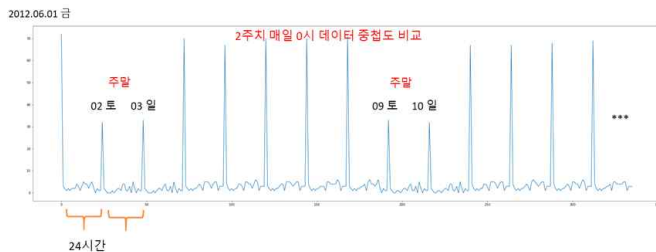
(그림 1) 데이터 인코더 내부 구성 및 데이터 입출력

시계열 데이터 인코더의 내부 구성은 (그림 1)에서와 같이 시계열 입력 데이터를 시간 정보와 에너지 소비량 값으로 분리하는 분리부 (Data Splitter), 에너지 소비량 값을 임의의 분포 스칼라 인코딩 (RDSE, Random Distributed Scalar Encoding)을 통해 이진화하는 모듈과 시간 정보를 데이터 인코딩을 통해 이진화하는 모듈, 그리고 이진화 결과를 물리적으로 연결하여 하나의 이진 데이터로 만드는 연결부 (Concatenator), 그리고 사용자가 설정한 희소도(sparsity)의 이진화 벡터 (SDR)로 변환하는 공간 풀링(Spatial Pooling)으로 구성된다.

#### 나. 실험 결과 및 고찰



(a) 데이터 변환 예시



(b) 데이터 유사도 비교 (매일 0시 데이터의 SDR 기반 비교)

(그림 2) 실험 결과 예시

(그림 2)에서 실험 결과를 도시하였다. (a)는 실험에 사용된 29,232개 데이터 샘플의 하나가 스칼라 인코딩과 데이터 인코딩의 이진화 과정 및 공간 풀링 과정을 거쳐 SDR로 변환되는 데이터 형상의 예를 보였다. (b)는 입력으로 사용한 데이터셋의 첫 번째 레코드에 해당하는 ‘2012년 6월 1일 (금요일) 0시 데이터’에 대한 변환된 SDR이 다른 레코드의 변환된 SDR과의 유사도 비교를 한 것이다. 유사도 비교는 SDR간 중첩된 활성비트의 개수로 측정하였다. (b)의 그림에서 가로축은 0시부터 23시까지 24시간마다 반복되는 연속된 시간 정보이고 세로축은 SDR간 활성비트 중첩의 개수를 나타낸다. (b)의 그림에서 확인할 수 있는 내용은 매일 같은 시간(현재 그림에서는 매일 0시)의 데이터 패턴은 비슷하게 나타난다는 것이고 주말의 경우에는 주중과는 다소 다르게 나타남을 확인할 수 있다. 이는 주말의 경우 에너지 소비 패턴이 달라져서 그렇다고 예상해 볼 수 있다.

시계열 데이터 인코더 설계의 관점에서 고려했던 특징들이 반영되었는지 정리해 보면 <표 1>과 같다.

&lt;표 1&gt; 데이터 인코더 요구사항에 대한 실험 결과

|         | 인코더 요구사항   | 구현된 인코더의 실험결과   |
|---------|--|---|
| 요구 사항 1 | 인코더는 유사한 입력 데이터에 대해 중복된 활성비트를 가지는 SDR 출력을 생성해야 함     | 요구사항 만족함<br><ul style="list-style-type: none"> <li>유사한 값은 유사도가 높게 나타남 (매일 같은 시간 데이터가 변환된 SDR의 활성비트 중첩도가 높음)</li> </ul>              |
| 요구 사항 2 | 인코더의 SDR 출력은 인코더 모든 입력에 대해 같은 차원을 가져야 함              | 요구사항 만족함<br><ul style="list-style-type: none"> <li>최종 출력 SDR의 차원은 시계열 인코더 내부 설정으로 가변할 수 있음 (현재 구현에서는 1,638 비트 출력으로 설정했음)</li> </ul> |
| 요구 사항 3 | 인코더의 SDR 출력은 인코더 모든 입력에 대해 유사한 희소도 (sparsity)를 가져야 함 | 요구사항 만족함<br><ul style="list-style-type: none"> <li>SDR의 희소도는 사용자 설정에 의해 가변할 수 있음 (현재 구현에서는 4%의 희소도로 설정했음)</li> </ul>                |

상기와 같이 인코더 요구사항을 모두 만족하는 시계열 데이터 인코더가 구현되고 검증되었다. 현재의 구현 결과는 HTM 기반 기계학습을 지원하기 위한 데이터 전처리 단계이다. 향후 추가로 HMT 기반 비정상 검출, 예측 등의 학습을 위한 후속 연구가 필요하다.

### III. 결론

본 연구에서는 시계열 정보를 포함하는 데이터를 이용하여 HTM 모델 기반 학습 아키텍처를 지원하기 위한 전처리용 데이터 인코더를 제안하고 구현 및 검증하였다. 실험에서는 구현한 데이터 인코더가 HTM 전처리 인코더의 요구사항을 충족하는지 여부를 공개 데이터셋을 활용하여 검증하였고 요구사항을 만족함을 확인하였다.

향후 할 일은 본 연구의 인코더가 HTM 기반 기계학습에서 효율적으로 사용될 수 있을지 확인하는 작업으로, 비정상 검출이나 추론 등의 분야에서 HTM 기반 기계학습 알고리즘을 병행 개발하고 개발한 인코더를 적용하는 연구를 계속 진행할 예정이다.

### ACKNOWLEDGMENT

본 논문은 (재)전북테크노파크의 재원으로 지원을 받아 수행된 지역특성화산업 전문인력양성사업의 연구 결과입니다

### 참 고 문 헌

- [1] J. Hawkins, S. Blakeslee, "On Intelligence: How a New Understanding of the Brain Will Lead to the Creation of Truly Intelligent Machines," Macmillan, 2007
- [2] S. Purdy, "Encoding Data for HTM Systems," arXiv:1602.05925
- [3] Stephen Makonin, "HUE: The hourly usage of energy dataset for buildings in British Columbia," Journal of Data in Brief, Volume 23, April 2019