

시계열의 자기주도적 표현 학습을 위한 시간-주파수 영역에서의 계층적 손실을 활용하는 대조적 학습에 관한 연구

김시온, 김진모, 김상희, 최지웅
대구경북과학기술원

slon11@dgist.ac.kr, jmkim@dgist.ac.kr, sanghee123@dgist.ac.kr, jwchoi@dgist.ac.kr

Contrastive learning for self-supervised representation learning of time series using hierarchical loss in the time-frequency domain

Sion Kim, Jinmo Kim, Sanghee Kim, Ji-Woong Choi
Daegu Gyeongbuk Institute of Science and Technology (DGIST)

요약

본 논문은 다양한 시계열 데이터를 분석하기 위한 방법으로 hierarchical loss 를 time-frequency domain 에 적용하여 contrastive learning 을 수행하는 기술을 제안한다. 기존 연구에서 수행했던 time domain 에서의 hierarchical loss 최적화 방식이 시계열의 특성을 효과적으로 반영하는 representation 을 생성할 수 있다는 것에 착안하여, 이 hierarchical loss 를 time-frequency domain 으로 확장함으로써 시계열의 특성을 더 효과적으로 반영할 수 있다는 것을 제안한다. 이러한 방법을 통해 실제로 classification task 에서 보다 더 뛰어난 성능을 나타낼 수 있도록 표현 벡터를 학습시킬 수 있음을 제시한다.

I. 서론

자기주도학습 (self-supervised learning)은 label 이 없는 데이터에 대해서 데이터의 특성을 스스로 도출하여 각종 task 를 효과적으로 수행할 수 있도록 하는 기계 학습 방법이다. 자기주도학습은 스스로 label 을 생성하고 지도학습 (supervised learning)을 수행하는 pretext tasks 를 수행하며, 이러한 pretext tasks 방법 중 하나로써 많이 사용되는 방법은 대조적 학습 (contrastive learning)이다 [1]. 대조적 학습은 데이터에 대해 positive pair 와 negative pair 를 생성하고, 어떤 embedding 된 벡터 공간에서 positive pair 간의 거리는 가깝게, negative pair 간의 거리는 멀리 위치시키도록 함으로써 표현 벡터(representation vector)를 스스로 학습한다 [1]. 이러한 대조적 학습은 최근 컴퓨터 비전이나 자연어 처리 분야에서 각광받으며 뛰어난 성능을 나타내고 있다 [1]. 시계열 데이터의 또한 데이터의 복잡성으로 인해 대조적 학습을 비롯한 각종 방법으로 자기주도학습을 수행하는 연구가 시도되고 있다 [2,3]. 그러나 아직까지 시계열 데이터의 특성을 효과적으로 학습하는 표준적인 자기주도적 학습 모델이 정립되지 않은 상황이다.

본 논문은 기존 연구에서 제시하였던 시간 영역 (time domain)에서의 계층적 손실 (hierarchical loss)이 시계열 데이터의 패턴을 효과적으로 학습시킬 수 있다는 점[3]과 더불어 주파수 영역 (frequency domain)의 특성 또한 시계열 데이터에 대한 특성을 효과적으로 반영할 수 있다[2]는 기존 연구에 착안하여 time-frequency (TF) 영역에 대해 계층적 손실을 적용하는 방법을 제안한다. 본 연구에서는 최종적으로 TF 영역에서의 계층적 손실을 고려하여 학습한 모델과 기존 연구에서 제시한 모델들을 classification task 를 통해 비교함으로써, TF 영역에서의

계층적 손실을 고려할 경우 시계열 데이터의 특성을 효과적으로 반영할 수 있음을 제시한다.

II. 본론

2.1 계층적 손실을 TF 영역으로 확장

기존 연구에서 제안한 Ts2Vec (Time Series to Vector)에서는 시간 영역에서 계층적 대조 (hierarchical contrasting)을 활용하여 시계열의 특성을 효과적으로 추출 및 학습한다 [3]. 가장 최근에 제시된 방법인 TF-C (Time-Frequency Contrasting)는 시간 영역과 주파수 영역 각각에서 대조적 학습을 수행한 뒤, 하나의 영역으로 결합하여 표현 벡터를 학습한다 [2]. 즉, 시계열 데이터에 대해 표현 학습을 진행하기 위해서 시간 영역의 특성뿐만 아니라 주파수 영역까지 고려했다는 차별성이 존재한다. 그러나 TF-C 는 주파수 영역뿐만 아니라 시간 영역에서조차도 계층적 대조를 이용하지 않았기 때문에 아키텍처와 성능이 개선될 여지가 있다.

우리는 기존 연구의 이러한 한계점에 착안하여 시간 영역뿐만 아니라 주파수 영역에서도 계층적 대조를 수행하는 TF-HC (Time-Frequency Hierarchical Contrasting) 모델을 제안한다. TF-HC 는 각 데이터 포인트에 대해 시간과 주파수 영역에서의 정보를 획득하기 위하여 주어진 데이터에 대해 wavelet transform (WT)을 수행하고, 이에 따라 계층적 손실을 계산하도록 그림 1A 와 같이 아키텍처를 구성한다.

TF-HC 모델이 임의의 시계열 데이터에 대해 표현 벡터를 학습하는 과정은 다음과 같다. 임의의 시계열 데이터는 자기주도 대조적 학습을 위해 복제되고, 복제된 데이터와 기존 데이터는 두 개의 과정을 거친다. 첫째로 시간 영역에서의 특성 추출을 위해 데이터는 time encoder 에

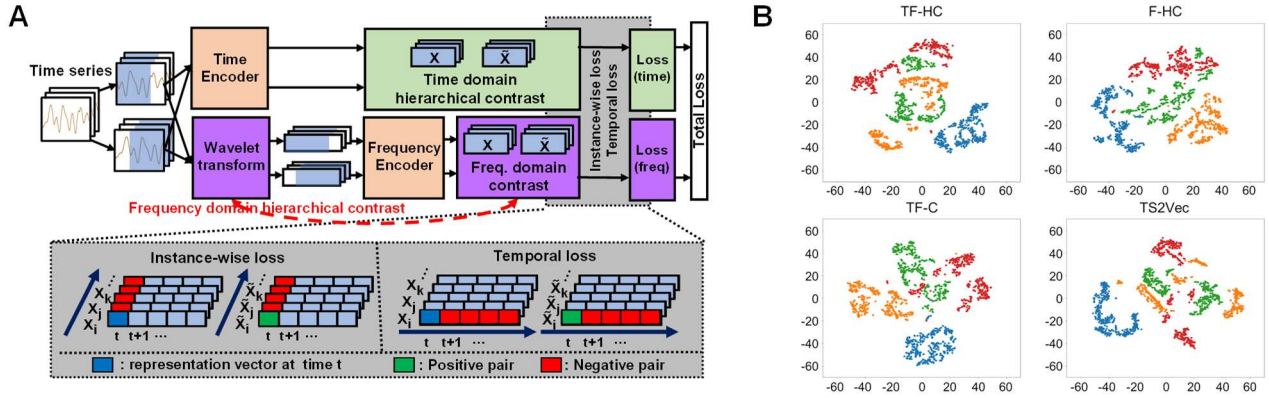


그림 1. TF-HC의 개략도 및 학습 결과 예시. (A) 제안된 TF-HC의 아키텍처, (B) CinCEGTorso 데이터에 대해 각 모델의 학습된 표현 벡터.

입력된다. Time encoder는 linear layer와 dilated convolution layer로 구성된다. 이때 데이터의 길이는 유지되도록 convolution을 수행함으로써, 해당 layer를 통해 증가된 채널이 각 데이터 포인트의 표현 벡터를 의미하게 한다. 이러한 표현 벡터들에 대해 각각 positive pair와 negative pair를 정의하여 대조적 학습을 시킴으로써 encoder가 보다 더 유의미한 표현 벡터를 생성하게끔 학습한다. 학습의 성능 향상을 위해 Ts2Vec에서 제안된 시간 영역의 계층적인 대조적 손실 계산 방법을 이용하였다 [3].

둘째로 주파수 영역에서 정보를 효과적으로 추출하기 위해 시간 대조뿐만 아니라 주파수 대조를 병행하여 학습시킨다. 먼저 모델의 입력 데이터에 대해 WT를 수행하여 scalogram을 획득한다. Wavelet은 Morlet을 사용했으며, scale factor를 10, translation factor를 1로 설정하여 데이터의 길이가 시간 영역의 측면에서 유지되도록 하였다. 획득한 scalogram은 frequency encoder에 입력되고, 각 데이터 포인트의 WT 결과가 표현 벡터로 변환된다. 해당 표현 벡터에 대한 대조적 학습은 시간 영역에서 시행한 것과 동일하게 학습하였다. 이러한 방법을 통해 시간과 주파수 영역의 정보를 다양한 범위로 고려한 표현 벡터를 학습할 수 있다.

2.2 TF-HC의 성능 실험

TF-HC의 성능을 검증하기 위해 오픈소스 데이터에 대해 표현 벡터를 획득한 뒤, radial basis function과 함께 support vector machine (SVM)을 이용해 classification task를 수행하였다. SVM에 대한 hyperparameter는 grid search method를 사용하여 최적화되었다. 주파수 영역에서 계층적 대조의 효과를 조사하기 위해 TF-HC에서 시간 영역의 계층적 대조를 제외한 모델로서 F-HC (Frequency Hierarchical Contrasting)를 구현하여 추가 비교하였다.

Dataset	TS2Vec	TF-C	F-HC	TF-HC
ACSF1	0.900	0.667	0.800	0.890
CinCEGTorso	0.820	0.779	0.852	0.912
EOGHorizontal-Signal	0.517	0.494	0.235	0.544
Ethanol-Level	0.370	0.514	0.262	0.316
Mallat	0.927	0.902	0.743	0.939
MixedShapes-RegularTrain	0.918	0.935	0.663	0.941
PLAID	0.555	0.163	0.488	0.527
SemgHand-GenderCh2	0.967	0.894	0.702	0.832
UWaveGesture-LibraryAll	0.912	0.927	0.490	0.912
Worms	0.636	0.672	0.636	0.792

표 1. 오픈소스 데이터에 대한 각 모델의 분류 성능 비교

기존 모델을 비롯한 각 모델의 classification 성능 결과는 표 1에 표기하였다. 모델의 성능을 비교했을 때, F-HC 자체만으로는 뚜렷한 성능 향상을 보이지 않았으며, TF-HC는 특정 데이터셋에서 다른 모델들보다 뛰어난 성능을 나타내어 최대 12%의 정확도 향상을 보였다. 각 모델을 통해 학습한 표현 벡터를 직관적으로 비교하기 위하여 CinCEGTorso 데이터셋에 대해 획득한 표현 벡터를 t-SNE로 그림 1B에 도시하였다.

III. 결론

본 논문에서는 시계열 데이터의 특성을 자기주도적으로 학습하기 위한 방법으로 시간 영역의 계층적 손실과 더불어 주파수 영역의 계층적 손실을 고려하여 학습하는 TF-HC 모델을 제안하였다. 제안된 기술을 classification 성능으로 비교한 결과, 많은 데이터셋에 대해 기존에 제시된 모델 대비 비슷한 성능을 내거나 더 뛰어난 성능을 나타냈다. F-HC 모델이 뚜렷한 성능 향상을 나타내지 않은 것을 미루어 보았을 때, 단순히 주파수 영역의 정보만을 고려하여 학습하는 것은 성능이 좋지 않다는 것을 시사한다. 또한 학습된 표현 벡터가 embedding된 벡터 공간에서 각각의 실제 label 별로 구분되어 분포한 것을 미루어 보았을 때, TF-HC는 자기주도적인 방식으로 시계열의 특성을 효과적으로 학습할 수 있음을 시사한다. 따라서 이후 TF-HC를 활용할 경우 시계열 데이터의 특성을 효과적으로 분석 가능할 것이라 기대한다.

ACKNOWLEDGMENT

이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단 바이오의료기술개발사업의 지원(No. 2017M3A9G8084463)과 정부(과학기술정보통신부)가 지원하는 한국뇌연구원(KBRI)을 통한 KBRI 기초연구 프로그램의 지원(22-BR-04-04)을 받아 수행된 연구임.

참고 문헌

- [1] Le-Khac, P. H. "Contrastive representation learning: A framework and review," IEEE Access, 8, 193907–193934, 2020.
- [2] Zhang, X. "Self-supervised contrastive pre-training for time series via time-frequency consistency," arXiv preprint arXiv:2206.08496, 2022.
- [3] YUE, Z. "Ts2vec: Towards universal representation of time series," Proceedings of the AAAI Conference on Artificial Intelligence, pp. 8980–8987, Jun. 2022.