

다중 스케일 특징 융합을 통한 트랜스포머 기반 장기 시계열 예측 정확도 향상 기법

민희수, 채동규
한양대학교 컴퓨터소프트웨어학과
hsmin@hanyang.ac.kr, dongkyu@hanyang.ac.kr

Fusion of Multi-Scale Features towards Improving Accuracy of Long-Term Time Series Forecasting

Heesu Min, Dong-Kyu Chae
Dept. of Computer and Software, Hanyang University, South Korea

요 약

본 논문에서는 정확한 장기 시계열 예측을 위해 시계열 데이터의 다양한 스케일 (시간 규모)에서 표현을 학습하는 트랜스포머 모델을 제안한다. 제안하는 모델은 시계열의 다중 스케일 특징을 추출하고, 이를 트랜스포머에 반영하여 예측 시계열을 생성하는 구조로 되어 있다. 스케일 정규화 과정을 통해 시계열의 전역적 및 지역적인 시간 정보를 효율적으로 융합하여 종속성을 학습한다. 3 가지의 다변량 시계열 데이터를 이용한 실험을 통해 제안하는 방법의 우수성을 보인다.

1. 서론

장기 시계열 예측이란 관측된 과거 시계열 데이터로부터 장기간의 미래 시계열 값을 예측하는 것으로 장비 수명 주기 예측, 날씨 예측 및 교통 흐름 예측과 같은 많은 응용 분야에 적용될 수 있다. 최근 트랜스포머[1] 기반의 모델이 순차적인 데이터를 모델링하는데 있어서 높은 성능을 나타내며 시계열 예측에도 많은 연구가 진행되고 있다. 그러나 기존의 트랜스포머 기반 모델 방법은 예측 시점 간격이 길어질수록 데이터의 주기성이나 경향으로 복잡성이 증가하여 예측 정확도가 떨어지는 한계점이 존재한다.

본 논문에서는 이러한 문제를 해결하기 위하여 다양한 시간 패턴을 고려하여 학습하고자 트랜스포머 모델에 다중 스케일 정보를 융합하는 방법을 제안한다. 제안하는 방법은 시계열 데이터가 다중 스케일 (서로 다른 시간 스케일)에서 혼합된 변화를 나타낸다는 아이디어에서 착안했다. 이는 기존의 트랜스포머 모델과 다르게 여러 시간 스케일에서 종속성을 집계함으로써 동적인 시간 패턴을 가지는 시계열에 대해 더욱 강인한 특성을 가진다. 다중 스케일 구조는 입력된 각 시계열의 스케일과 창 사이의 분포 이동을 최소화하여 시간 종속성을 학습한다. 또한, 모델의 효율성을 향상시키기 위해 트랜스포머 모델에 합성곱과 풀링 연산을 결합한다. 결과적으로 이러한 방식을

통해 모델에 세분화된 시계열의 스케일 정보를 융합하여 종속성을 더 잘 학습시키고자 한다.

2. 시계열 예측을 위한 트랜스포머 기반 모델

최근 트랜스포머 모델은 시계열의 장기 의존성 문제를 해결하며 많은 관심을 받고 있다. 특히 트랜스포머의 attention 메커니즘을 효율적으로 사용한 트랜스포머 기반 시계열 예측 모델 연구가 진행되고 있다. 이는 Self-attention 으로 시계열의 각 시점 정보 간에 관련이 높은 시점에만 집중할 수 있도록 유도하여 데이터의 복잡한 패턴을 학습한다.

Informer[2]는 ProbSparse self-attention 으로 연산과 메모리 효율을 향상시키며 장기 예측을 수행한다. 이는 유의미한 내적 쌍을 구분해내기 위해서 쿼리와 키 분포 간 유사도를 그 지표로써 활용하고, 키 샘플링과 분포 차이의 극대값을 기준으로 적용한다. 이러한 작업은 단순히 인접, 또는 무작위로 일부의 입력 값에만 attention 을 적용하는 Sparse attention 대신 확률 분포에 입각한 ProbSparse self-attention 를 얻는다.

Autofomer[3]는 기간 기반의 종속성을 찾기 위해 자기상관 (auto-correlation)을 활용하여 장기 예측의 성능을 높였다. Self-attention 을 대체하는 자기상관과 시간지연 집계를 활용하여 시리즈 별 상관관

계를 찾고, 이를 집계하여 사용한다. 마지막으로 시계열 분해를 통해 계절성과 추세를 분리해 복잡한 시계열의 패턴을 학습에 반영한다.

3. 제안하는 방법

본 연구는 다변량 시계열로부터 다중 스케일 정보를 학습하는 장기 시계열 예측 모델을 제안한다. 제안하는 방법론은 다중 스케일 추출과 인코더-디코더 기반의 트랜스포머 절차로 구성된다. 전체적인 구조는 그림 1 과 같다.

3.1 다중 스케일 추출

인코더의 입력 시퀀스는 풀링하고, 디코더 입력 시퀀스는 업 샘플링 하는 과정을 통해 윈도우의 차원을 맞춰준다. 이후 입력 윈도우의 시간 차원의 평균을 구하여 스케일 정규화 과정을 수행하고, 각 입력에 대한 스케일을 추출한다.

3.2 트랜스포머 인코더

인코더는 3 개의 독립된 인코더로 구성된다. 이때, 각 인코더는 N 개의 인코더 셀이 쌓여있는 구조이다. 입력은 이전 attention 블록의 출력을 절반으로 다운 샘플링 하여 다음 레이어의 입력으로 들어간다. 또한, ProbSparse attention 블록 사이에 합성곱과 풀링 연산을 통해 결합한다. 이를 통해, attention 출력에서 주요한 특징 표현을 추출하여 다음 레이어에 전달할 은닉 상태의 크기를 점진적으로 감소시킨다. 최종적으로 인코더를 거친 출력 값들은 연결되어 특징 맵을 구성한다.

3.3 트랜스포머 디코더

디코더는 인코더와 마찬가지로 마스킹 (masking) 된 Multi-head ProbSparse attention 을 적용했다. Attention 을 통해 나온 출력 값은 이전 정보들을 활용하여 최종 시계열을 예측한다.

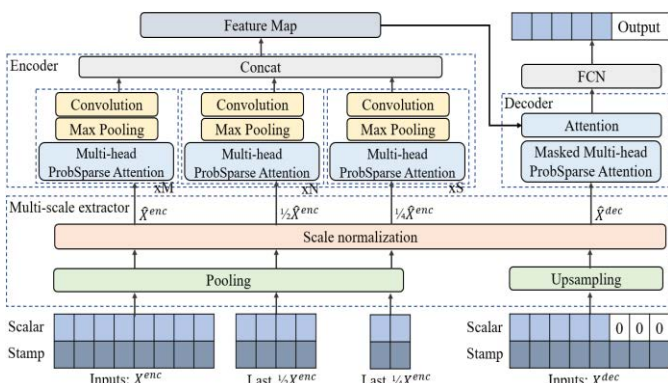


그림 1. 제안하는 모델

4. 실험 결과 및 결론

제안하는 방법의 정확도를 검증하기 위해 Weather, Electricity, Ocean 3 가지의 공개 시계열 데이터셋을

이용하였다. 기존 모델 중 Informer 모델과 제안하는 모델을 각각의 학습 데이터를 통해 학습시키고, 평가 데이터를 이용하여 성능을 비교하였다. 장기 시계열 예측을 위해 예측 길이를 96, 192, 336, 720 으로 설정하였다. 성능 비교 지표로는 MSE 와 MAE 를 사용하였다.

<표 1>은 전체 모형들에 대한 실험 결과를 나타내고 있다. 각 실험에서 제일 좋은 성능을 낸 지표는 굵은 글씨로 표기되어 있다. 실험 결과에서 제안한 예측 모델이 Informer 보다 대부분 지표에서 성능이 더 우수한 것을 볼 수 있다. 또한, 스케일 정규화 과정의 사용 여부를 비교해봤을 때 스케일 정보를 같이 학습하는 것이 예측 성능을 높이는데 기여하고 있음을 확인할 수 있다. 이는 제안하는 방법이 장기 예측에도 시계열의 다중 스케일 특징을 잘 학습했기 때문으로 보인다.

Method	제안하는 방법 (multi-scale)		제안하는 방법 (non multi-scale)		Informer		
	MSE	MAE	MSE	MAE	MSE	MAE	
Weather	96	0.259	0.302	0.350	0.403	0.383	0.446
	192	0.473	0.433	0.579	0.546	0.510	0.505
	336	0.397	0.435	0.650	0.572	0.812	0.649
	720	0.482	0.464	0.980	0.717	1.078	0.760
Electricity	96	0.188	0.304	0.263	0.359	0.318	0.406
	192	0.210	0.327	0.279	0.374	0.346	0.430
	336	0.227	0.339	0.261	0.358	0.346	0.430
	720	0.245	0.355	0.311	0.393	0.399	0.459
Ocean	96	0.174	0.179	0.183	0.196	0.189	0.204
	192	0.310	0.246	0.278	0.256	0.275	0.255
	336	0.345	0.277	0.356	0.301	0.360	0.314
	720	0.414	0.331	0.434	0.355	0.446	0.362

<표 1> 실험 결과

감사의 글

이 논문은 2022 년도 정부(과학기술정보통신부)의 재원으로 (1) 한국연구재단 바이오·의료기술개발사업의 지원 (No. NRF-2021M3E5D2A01021156)과 (2) 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2020-0-01373, 인공지능대학원지원(한양대학교))

참고문헌

- [1] VASWANI, Ashish, et al. Attention is all you need. *NIPS* 2017.
- [2] ZHOU, Haoyi, et al. Informer: Beyond efficient transformer for long sequence time-series forecasting. *AAAI* 2021.
- [3] WU, Haixu, et al. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *NIPS* 2021.