



Are Transformers Effective for Time Series Forecasting?

AILING ZENG, MUXI CHEN, LEI ZHANG , QIANG XU(AAAI,2023)

2025.05.15
발표자: 박기웅

CONTENTS

- 1. 배경 지식
- 2. 연구 배경 및 문제 정의
- 3. 핵심 아이디어
- 4. 결과
- 5. 느낀점

1. 배경 지식 – Time Series Data



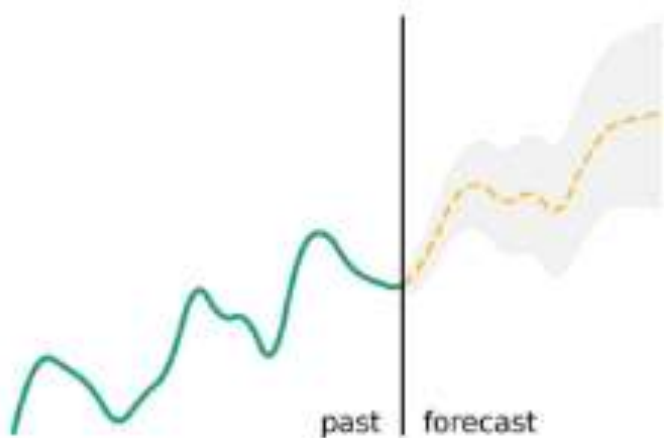
Time Series Data ?

1. 일정한 시간동안 수집된 순차적으로 정해진 데이터
2. 데이터들은 시간에 관하여 순서가 매겨져 있고, 연속한 관측치들은 서로 상관관계를 가진다.

특징

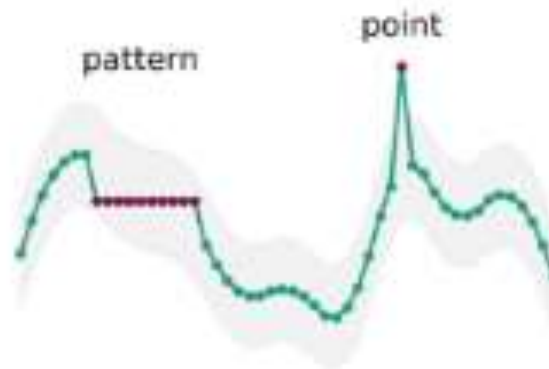
1. 계절성(Seasonality) : 규칙적인 시간(월, 년) 동안에 동일한 패턴이 나타남
2. 추세 (Trend) : 데이터가 장기적으로 보여주는 전반적인 방향성
3. 정상성(Stationary) : 시간이 지나도 시계열 데이터의 통계적 특성이 변하지 않는 것

1. 배경 지식 – Time Series Data의 Main Task



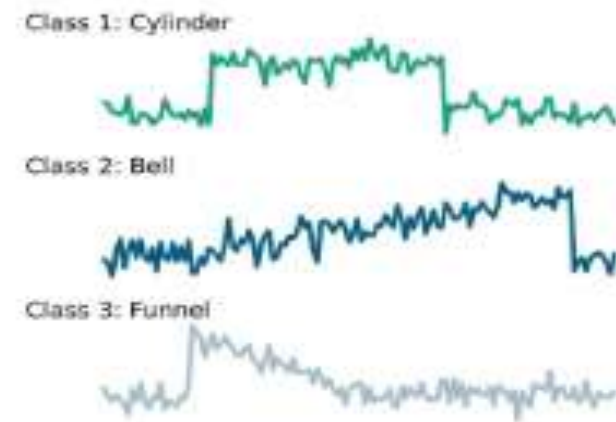
Time Series Forecasting

과거 및 현재 데이터를
학습하여, 미래 시점의 값들을
예측



Anomaly Detection

Time Series Data 내에서
정상분포와는 다른 양상을
보이는 이상치를 찾음



Classification

주어진 특정 Time Series
Data가 어떠한 패턴 or
레이블에 속하는지 분류

1. 배경 지식 – Time Series Forecasting의 방법론

Recurrent Neural Networks

- 딥러닝 모델 중 가장 고전적
- 현재 시점의 데이터가 이전 시점의 데이터의 영향을 받는 시계열 데이터의 특성상, RNN과 가장 부합
- 깊은 신경망을 학습시키는 과정에서, 기울기가 소실되는 문제인 vanishing gradient 문제 존재
- 시점 간의 간격이 커진다면, 현재 시점으로부터 멀리 떨어진 과거 시점 정보의 영향력이 약해지는 장기 의존성 문제(Long-Term Dependency Problem)
- LSTM, GRU
- 병렬 처리 어려움

Convolutional Neural Networks

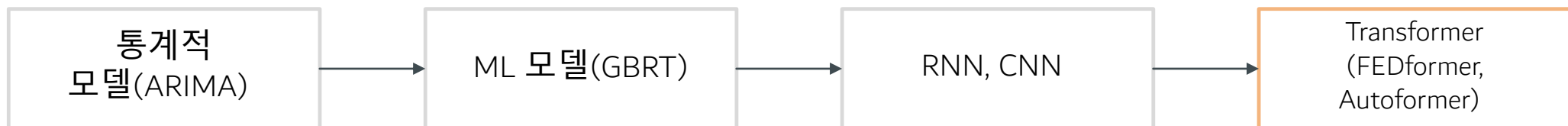
- CNN은 공간 차원에서 변하지 않는 로컬 관계를 추출
- Standard CNN에 대한 공간 불변성 가정에 따라, temporal CNN은 각 시간 단계에서 모든 시간에 걸쳐 동일한 필터 가중치 세트를 사용하여, 관계가 time-invariant라고 가정
- RNN과 결합하여 주로 사용(CNN-lstm 모델)
- CNN은 지역 패턴은 잘 잡지만, 장기 시점 간 의존성 포착은 어려움

Attention mechanisms

- Attention mechanism에 기반한 모델, 대표적으로 Transformer
- Transformer는 서로 다른 시점의 정보들 간의 관계를 바탕으로, (attention) score를 부여하여 활용하고자 함
- 현재 Transformer 기반의 모델들이 RNN기반의 모델 보다 성능이 우월함
- Self-attention을 바탕으로 각 시점 정보 간의 관계를 모델링 하기에 CNN 또는 RNN 대비 long range dependency를 학습하기에 용이하다는 장점
- 높은 복잡도 및 연산량, 메모리 부담

2. 연구 배경 및 문제 정의

TSF(Time Series Forecasting) 의 발전



TSF에서 Transformer 모델의 문제점

1. Multi-head attention mechanism
 - long sequence에서, 요소들 간에 의미적인 상관관계를 뽑아내는 능력은 우수
 - 그러나, self-attention은 어느 정도 길이의 sequence 까지는 Permutation invariant한 특성을 가져서, 순서가 바뀌어도 큰 상관이 없음

나는 어제 운동을 했다

나는 운동을 했다 어제

어제 운동을 했다 나는

- 이러한 특성은 NLP 처럼, Semantic이 풍부한 분야에서는 문제가 없다.
- NLP에서는 문장에서 일부 순서를 바꾸어도, Semantic meaning은 보존이 된다.

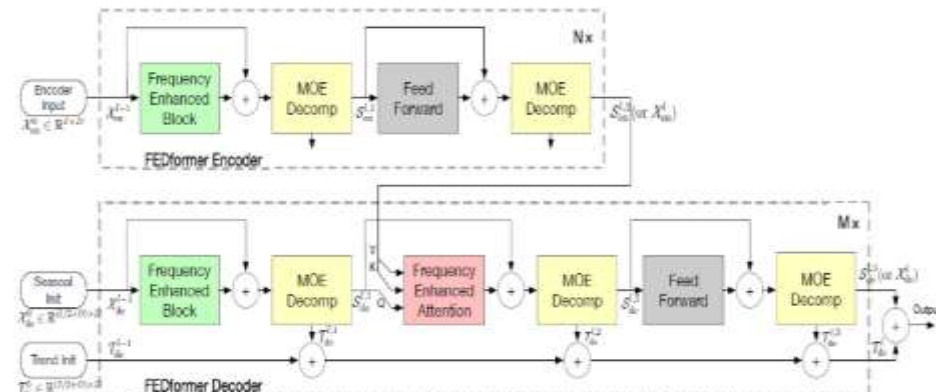
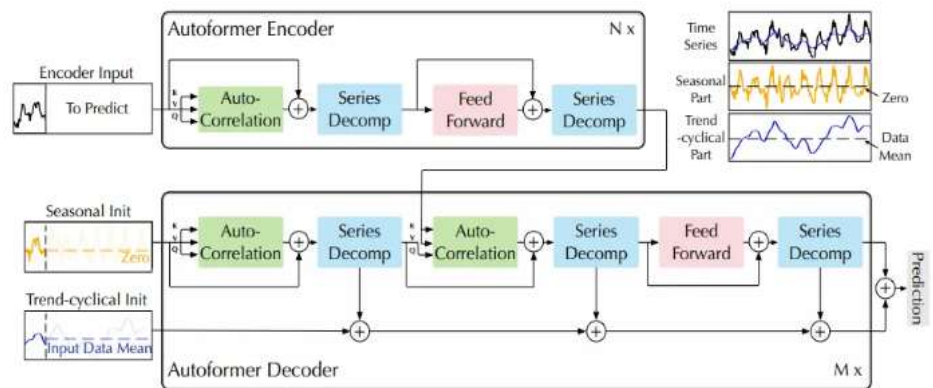
→ 그러나, Time Series data는 Numerical data로 인해 Semantic한 정보가 부족하다.(순서 그 자체가 가장 중요)

→ 즉, Long-Term sequence에서 Temporal 한 정보를 손실하는 단점이, Time Series data에는 크리티컬하게 작용

2. 연구 배경 및 문제 정의

TSF에서 Transformer 모델의 문제점

2. Efficiency



- 기존 Vanilla transformer의 복잡도: Time Complexity $O(L^2)$, Memory Complexity $O(L^2)$
 - AutoRegressive한 Decoder 구조로 인하여 Error Accumulation 문제 발생
 - 이러한 문제를 해결하기 위해 제시된 다양한 모델들은 그 자체로 매우 복잡한 모습을 보임
- 이론적으로 Time/Memory 복잡도는 줄어들었지만, 실제로도 그럴지는 의문

정말 Transformer 모델이 TSF에서 효과적일까?

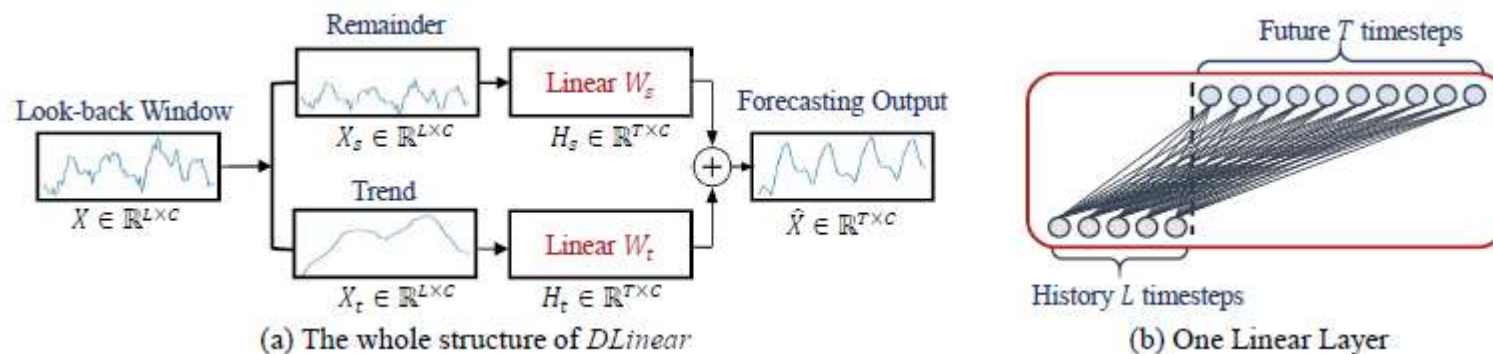
3. 핵심 아이디어

DMS + Decomposition + Linear Regression

IMS vs DMS

- IMS(Iterative Multi-Step) Forecasting: 한 시점만 예측하는 단일(1-step) 모델을 학습한 뒤, 예측 결과를 다시 입력으로 넣어 다음 시점을 순차적으로 예측하는 방식
- DMS(Direct Multi-Step) Forecasting: 목표 예측 step 마다 별도의 모델을 학습하여 시점 t 에서 한번에 전체 예측을 수행, 모델이 각 변수의 중요도를 동일하게 고려하고, 변수 간에 공간적인 상관관계를 반영 x
- 기존 Transformer 기반 TSF 모델들은 IMS 사용으로 인해, Error Accumulation이 심했다. 따라서 DMS를 사용한다.

DLinear



- Autoformer, FEDformer의 Decomposition scheme + linear layer의 결합
- Raw data -> 이동평균으로 Trend 와 Remainder(seasonal 이 포함됨) part로 분해 -> 각각의 part를 linear layer에 적용 -> 2개의 part를 합침
- 데이터 안에 Trend가 명확할때 성능이 좋음

NLinear



- Decomposition 사용 x , Linear layer만 사용
- 시계열은 다음 값이 이전 값의 영향을 받으므로, 이전 값 중 가장 최근 값을 기준으로 Normalization
- Linear layer에 입력 후, 예측 끝난 시점에 역정규화

4. 결과

- 데이터셋 - ETT(Electricity Transformer Temperature) : [ETTh1, ETTh2, ETTm1, ETTm2], Traffic, Electricity, Weather, ILI ,Exchange Rate
- 평가 지표 - MSE, MAE
- 비교 모델 - Transformer 기반(5개) : FEDformer, Autoformer, Informer, Pyraformer, LogTrans
- 단순한 DMS method(1개) : Closest Repeat (단순히 마지막 값 반복)

Best Results , Second Best Results

Methods	Metric	Electricity				Exchange-Rate				Traffic				Weather				ILI			
		96	192	336	720	96	192	336	720	96	192	336	720	96	192	336	720	24	36	48	60
DLinear-S*	MSE	0.194	0.193	0.206	0.242	0.078	0.159	0.274	0.558	0.650	0.598	0.605	0.645	0.196	0.237	0.283	0.345	2.398	2.646	2.614	2.804
	MAE	0.276	0.280	0.296	0.329	0.197	0.292	0.391	0.574	0.396	0.370	0.373	0.394	0.255	0.296	0.335	0.381	1.040	1.088	1.086	1.146
DLinear-I*	MSE	0.184	0.184	0.197	0.234	0.084	0.157	0.236	0.626	0.647	0.602	0.607	0.646	0.164	0.209	0.263	0.338	3.015	2.737	2.577	2.821
	MAE	0.270	0.273	0.289	0.323	0.216	0.298	0.379	0.634	0.403	0.375	0.377	0.398	0.237	0.282	0.327	0.380	1.192	1.036	1.043	1.091
FEDformer	MSE	0.193	0.201	0.214	0.246	0.148	0.271	0.460	1.195	0.587	0.604	0.621	0.626	0.217	0.276	0.339	0.403	3.228	2.679	2.622	2.857
	MAE	0.308	0.315	0.329	0.355	0.278	0.380	0.500	0.841	0.366	0.373	0.383	0.382	0.296	0.336	0.380	0.428	1.260	1.080	1.078	1.157
Autoformer	MSE	0.201	0.222	0.231	0.254	0.197	0.300	0.509	1.447	0.613	0.616	0.622	0.660	0.266	0.307	0.359	0.419	3.483	3.103	2.669	2.770
	MAE	0.317	0.334	0.338	0.361	0.323	0.369	0.524	0.941	0.388	0.382	0.337	0.408	0.336	0.367	0.395	0.428	1.287	1.148	1.085	1.125
Informer	MSE	0.274	0.296	0.300	0.373	0.847	1.204	1.672	2.478	0.719	0.696	0.777	0.864	0.300	0.598	0.578	1.059	5.764	4.755	4.763	5.264
	MAE	0.368	0.386	0.394	0.439	0.752	0.895	1.036	1.310	0.391	0.379	0.420	0.472	0.384	0.544	0.523	0.741	1.677	1.467	1.469	1.564
Pyraformer*	MSE	0.386	0.378	0.376	0.376	1.748	1.874	1.943	2.085	0.867	0.869	0.881	0.896	0.622	0.739	1.004	1.420	7.394	7.551	7.662	7.931
	MAE	0.449	0.443	0.443	0.445	1.105	1.151	1.172	1.206	0.468	0.467	0.469	0.473	0.556	0.624	0.753	0.934	2.012	2.031	2.057	2.100
LogTrans	MSE	0.258	0.266	0.280	0.283	0.968	1.040	1.659	1.941	0.684	0.685	0.734	0.717	0.458	0.658	0.797	0.869	4.480	4.799	4.800	5.278
	MAE	0.357	0.368	0.380	0.376	0.812	0.851	1.081	1.127	0.384	0.390	0.408	0.396	0.490	0.589	0.652	0.675	1.444	1.467	1.468	1.560
Reformer	MSE	0.312	0.348	0.350	0.340	1.065	1.188	1.357	1.510	0.732	0.733	0.742	0.755	0.689	0.752	0.639	1.130	4.400	4.783	4.832	4.882
	MAE	0.402	0.433	0.433	0.420	0.829	0.906	0.976	1.016	0.423	0.420	0.420	0.423	0.596	0.638	0.596	0.792	1.382	1.448	1.465	1.483
Repeat-C*	MSE	1.588	1.595	1.617	1.647	0.081	0.167	0.305	0.823	2.723	2.756	2.791	2.811	0.259	0.309	0.377	0.465	6.587	7.130	6.575	5.893
	MAE	0.946	0.950	0.961	0.975	0.196	0.289	0.396	0.681	1.079	1.087	1.095	1.097	0.254	0.292	0.338	0.394	1.701	1.884	1.798	1.677

- Methods* are implemented by us; Other results are from FEDformer [29].

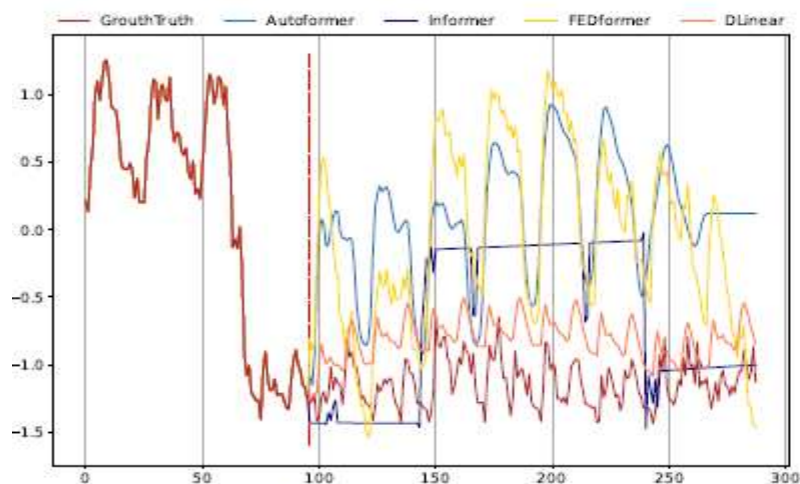
Methods		DLinear-S*		FEDformer		Autoformer		Informer		Pyraformer*		LogTrans		Reformer	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	0.386	0.400	0.376	0.419	0.449	0.459	0.865	0.713	0.664	0.612	0.878	0.740	0.837	0.728
	192	0.437	0.432	0.420	0.448	0.500	0.482	1.008	0.792	0.790	0.681	1.037	0.824	0.923	0.766
	336	0.481	0.459	0.459	0.465	0.521	0.496	1.107	0.809	0.891	0.738	1.238	0.932	1.097	0.835
	720	0.519	0.516	0.506	0.507	0.514	0.512	1.181	0.865	0.963	0.782	1.135	0.852	1.257	0.889
ETTh2	96	0.295	0.352	0.346	0.388	0.358	0.397	3.755	1.525	0.645	0.597	2.116	1.197	2.626	1.317
	192	0.452	0.462	0.429	0.439	0.456	0.452	5.602	1.931	0.788	0.683	4.315	1.635	11.12	2.979
	336	0.504	0.490	0.496	0.487	0.482	0.486	4.721	1.835	0.907	0.747	1.124	1.604	9.323	2.769
	720	0.577	0.538	0.463	0.474	0.515	0.511	3.647	1.625	0.963	0.783	3.188	1.540	3.874	1.697
ETTm1	96	0.345	0.372	0.379	0.419	0.505	0.475	0.672	0.571	0.543	0.510	0.600	0.546	0.538	0.528
	192	0.380	0.389	0.426	0.441	0.553	0.496	0.795	0.669	0.557	0.537	0.837	0.700	0.658	0.592
	336	0.413	0.413	0.445	0.459	0.621	0.537	1.212	0.871	0.754	0.655	1.124	0.832	0.898	0.721
	720	0.474	0.453	0.543	0.490	0.671	0.561	1.166	0.823	0.908	0.724	1.153	0.820	1.102	0.841
ETTm2	96	0.183	0.273	0.203	0.287	0.255	0.339	0.365	0.453	0.435	0.507	0.768	0.642	0.658	0.619
	192	0.260	0.325	0.269	0.328	0.281	0.340	0.533	0.563	0.730	0.673	0.989	0.757	1.078	0.827
	336	0.336	0.367	0.325	0.366	0.339	0.372	1.363	0.887	1.201	0.845	1.334	0.872	1.549	0.972
	720	0.415	0.423	0.421	0.415	0.433	0.432	3.379	1.338	3.625	1.451	3.048	1.328	2.631	1.242

- Methods* are implemented by us; Other results are from FEDformer [29].

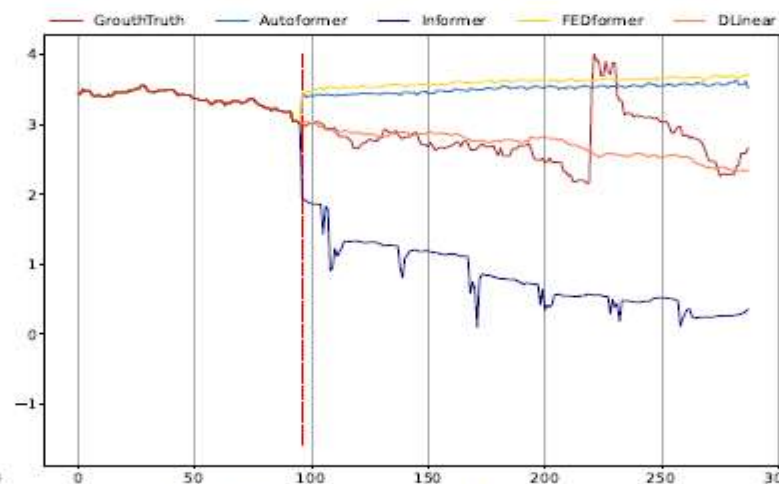
결과 해석

- 변수간 상관관계가 없더라도, 대부분 LTSF-Linear 적용 시 Transformer 보다 예측 성능이 좋음
- Nlinear 및 Dlinear 가 분포 이동, trend-seasonality 특징을 처리하는데 좋음
- Transformer based 방법들은 일부 잡음에 대해 train data에 대해 과적합 되어 정확도가 감소
- Transformer는 비주기적이고, 다른 시간적 패턴을 가진 데이터에 대해서는 추세 및 결과를 잘 예측하지 못함

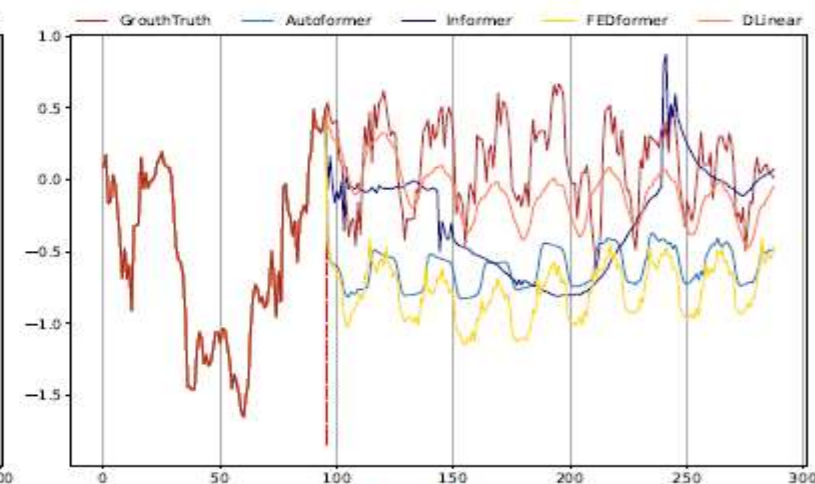
4. 결과



(a) Electricity



(b) Exchange-Rate



(c) ETTh2

효율성 비교

Method	MACs	Parameter	Time	Memory	Time	Memory	Test Step
<i>DLinear</i>	0.04G	139.7K	0.4ms	687MiB	$O(L)$	$O(L)$	1
Transformer \times	4.03G	13.61M	26.8ms	6091MiB	$O(L^2)$	$O(L^2)$	1
Informer	3.93G	14.39M	49.3ms	3869MiB	$O(L \log L)$	$O(L \log L)$	1
Autoformer	4.41G	14.91M	164.1ms	7607MiB	$O(L \log L)$	$O(L \log L)$	1
Pyraformer	0.80G	241.4M*	3.4ms	7017MiB	$O(L)$	$O(L)$	1
FEDformer	4.41G	20.68M	40.5ms	4143MiB	$O(L)$	$O(L)$	1

장점

- Temporal Relation을 낮은 연산량으로 포착 가능
- High-efficiency : 사용하는 메모리와 Parameter가 압도적으로 작으므로, 추론시에도 좋은 성능
- Input을 Trend와 Remainder로 분해한 후, Weight를 곱하는 선형 모델이기 때문에, Weight를 시각화 하여 예측 값에 대해 직관적인 해석 가능
- Hyper Parameter를 조정하지 않고도 쉽게 사용 가능

5. 느낀점

- 현업에서 실제로 중요한 것. (모델 성능 vs 사용 가능)
- 단순한 것이 가장 효과적일수도
- AI 시장은 Idea 싸움, Idea를 도출하기 위해서는 기본을 더욱 탄탄히 해야함