

2025 10 23
발표 자료

광운대학교 로봇학과
FAIR Lab

김한서

이번 주 진행사항

- LTSF-Linear
 - 논문 리뷰
 - 실험 세팅 및 결과
 - 시각화
 - 결과 정리

Are Transformers Effective for Time Series Forecasting?

Ailing Zeng^{1*}, Muxi Chen^{1*}, Lei Zhang², Qiang Xu¹

¹The Chinese University of Hong Kong

²International Digital Economy Academy (IDEA)

{alzeng, mxchen21, qxu}@cse.cuhk.edu.hk

{leizhang}@idea.edu.cn

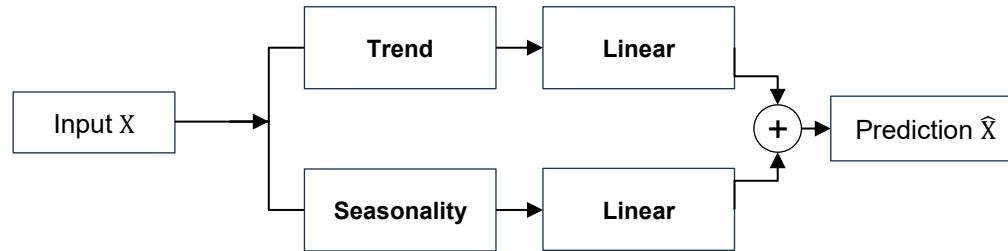
- 최근 들어 장기 시계열 예측에서 Transformer 기반 모델 급증
- 시계열 예측 문제의 경우 시간 순서가 중요
 - Transformer 기반 모델은 주로 Self-Attention을 사용
 - Self-Attention은 ‘Permutation-Invariant’ 특성을 가져 시간 정보가 손실될 수 있음
- 위 문제점 해결을 위해 LTSF-Linear를 사용
 - Transformer 기반 모델보다 성능이 20~50% 뛰어남
 - 모델이 단순해 학습 속도가 빠르고 메모리 사용량이 적음

LTSF-Linear

논문 리뷰

• DLinear

- 시계열 분해 방식 Seasonality, Trend를 선형 레이어와 결합한 모델
- 이동 평균으로 Seasonality 성분과 Trend 성분으로 분해
- 각 구성 요소에 선형 레이어 적용하고, 두 개를 합산하여 최종 예측을 계산



• NLinear

- Distribution shift 문제 방지를 위한 모델
- 모델 입력의 가장 마지막 값을 입력마다 뺀다. → 정규화
- 정규화된 입력을 선형 레이어로 예측한 뒤, 그 예측 결과에 뺀 값을 다시 더함. → 후처리



LTSF-Linear

주요 모델 성능 비교

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
<i>DLinear-S*</i>	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
<i>DLinear-I*</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		<i>DLinear-S*</i>		FEDformer		Autoformer		Informer		Pyraformer*		LogTrans		Reformer	
	Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
<i>ETTh1</i>	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
<i>ETTh2</i>	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
<i>ETTm1</i>	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
<i>ETTm2</i>	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].

- DLinear와 같은 선형 모델이 Transformer 기반 모델들보다 더 좋은 성능을 내는 것을 확인함
- Transformer가 시간 패턴을 찾으려다 노이즈에 과적합되는 경향이 있는 반면, 단순 선형 모델이 시계열의 핵심 특성을 더 효율적으로 파악함

실험 세팅

- 사용한 모델: DLinear
- 재현 실험 데이터셋: ETTh1
- 비교 실험 데이터셋: ETTm1, Weather, Exchange

Experiment	ETTh1
Learning rate	10^{-4}
Epoch	10
Batch size	32
Loss function	MSE Loss
Sequence Length	96
input_feature	7
Pred_len	96/192/336/720

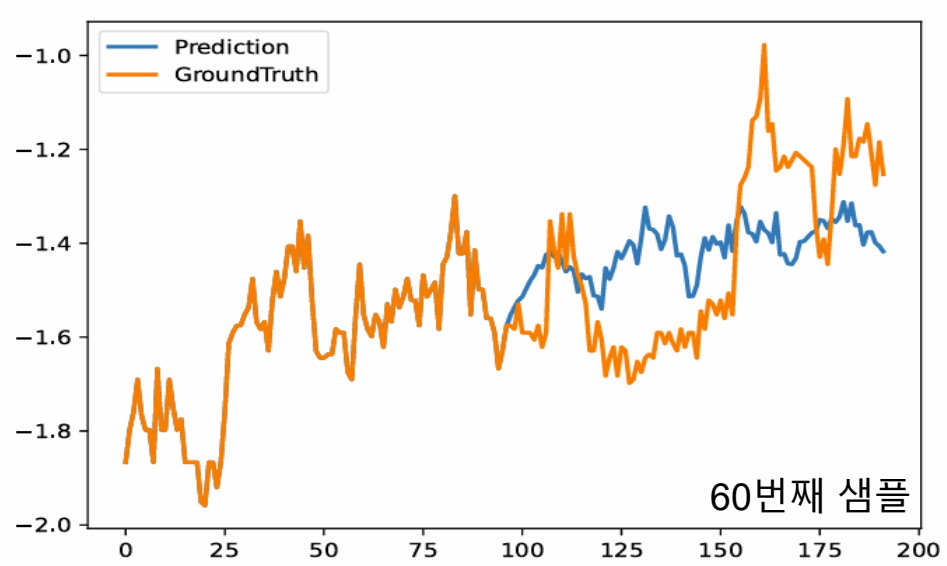
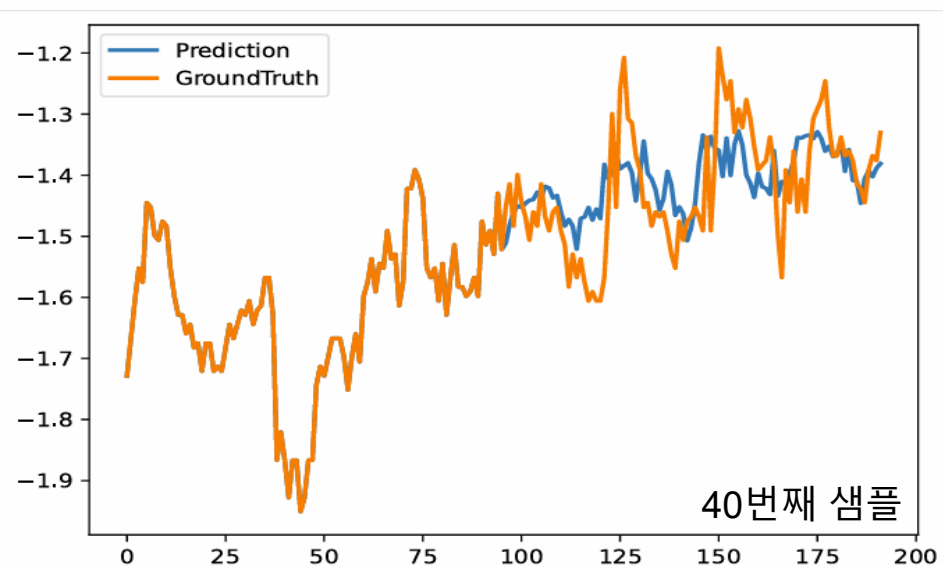
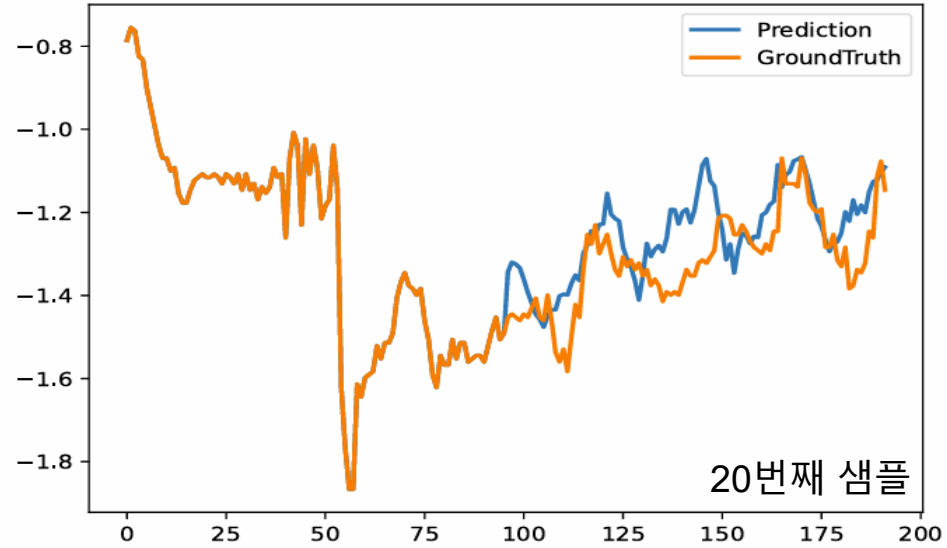
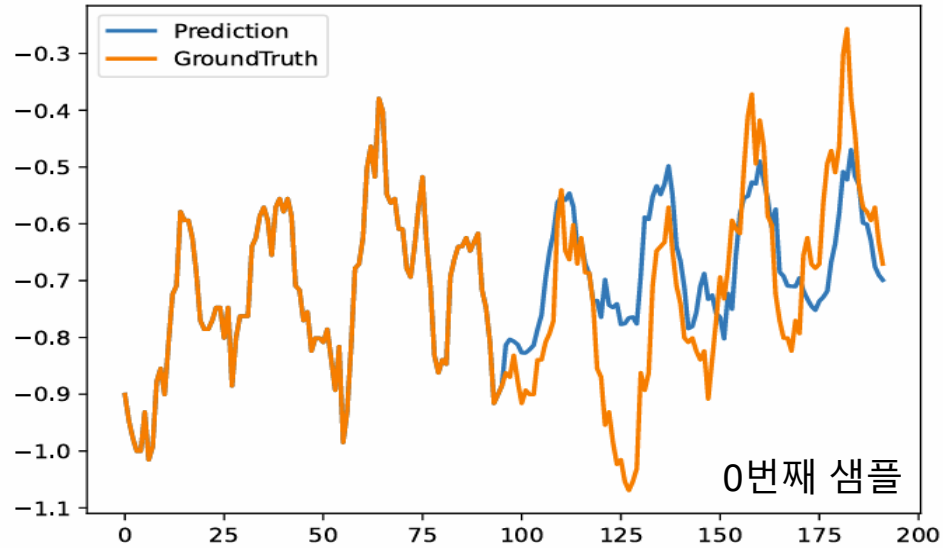
DLinear 재현 실험 결과 (ETTh1)

- 재현 실험 진행한 결과, 모든 예측 구간에서 논문 성과와 거의 동일한 수치가 나오는 것을 확인
- Prediction length가 길어질수록 MSE, MAE 값이 증가하지만 모든 예측 구간에서 훨씬 낮은 오차를 기록함

	Paper		Reproduction	
Pred_len	MSE	MAE	MSE	MAE
96	0.386	0.400	0.396	0.410
192	0.437	0.432	0.445	0.440
336	0.481	0.459	0.487	0.465
720	0.519	0.516	0.512	0.510

LTSF-Linear

DLinear 재현 실험 시각화 (ETTh1)

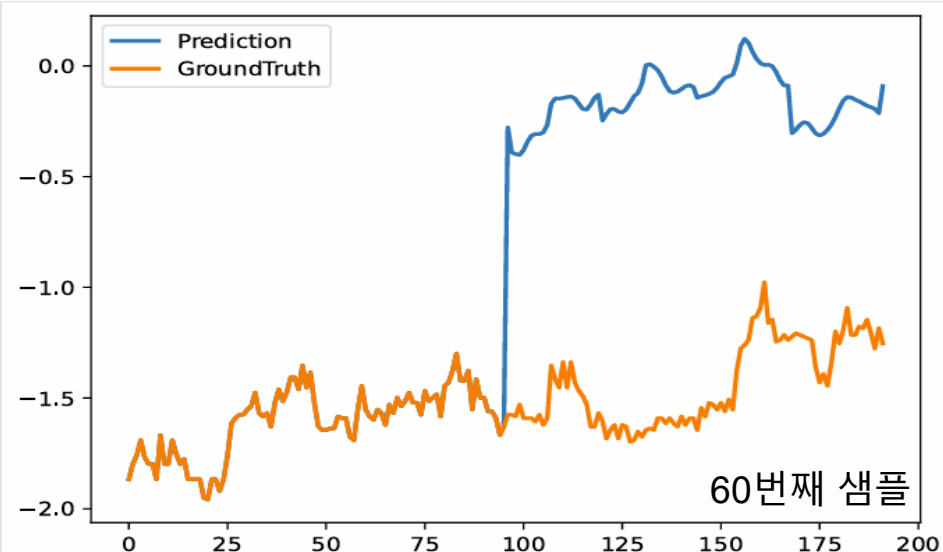
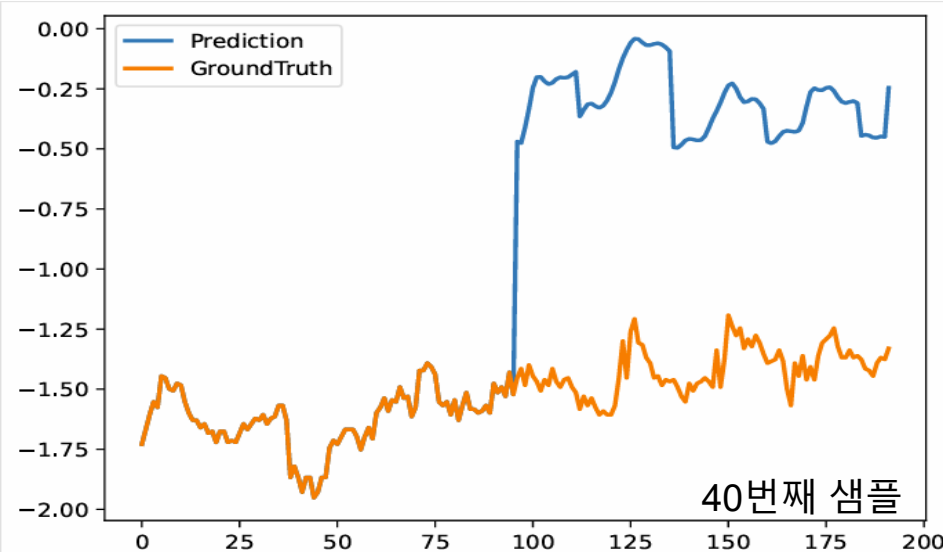
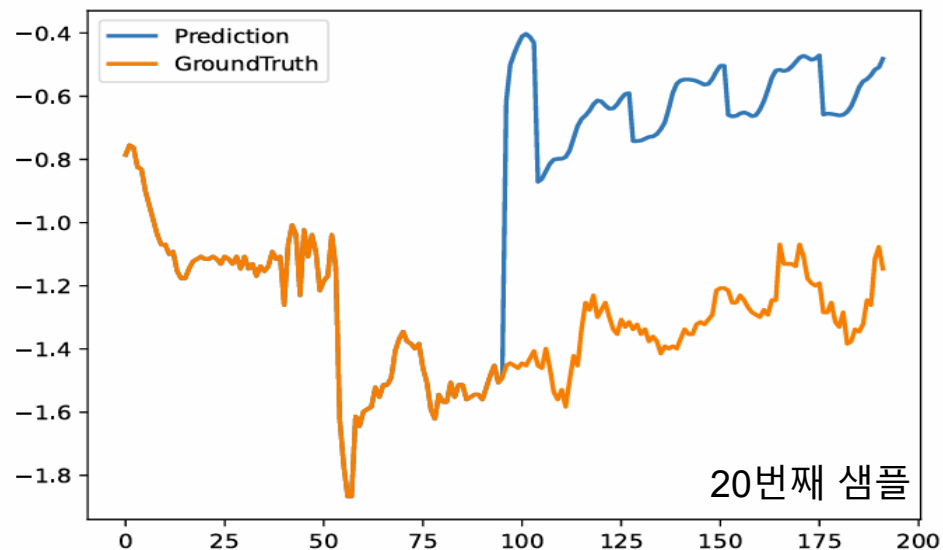
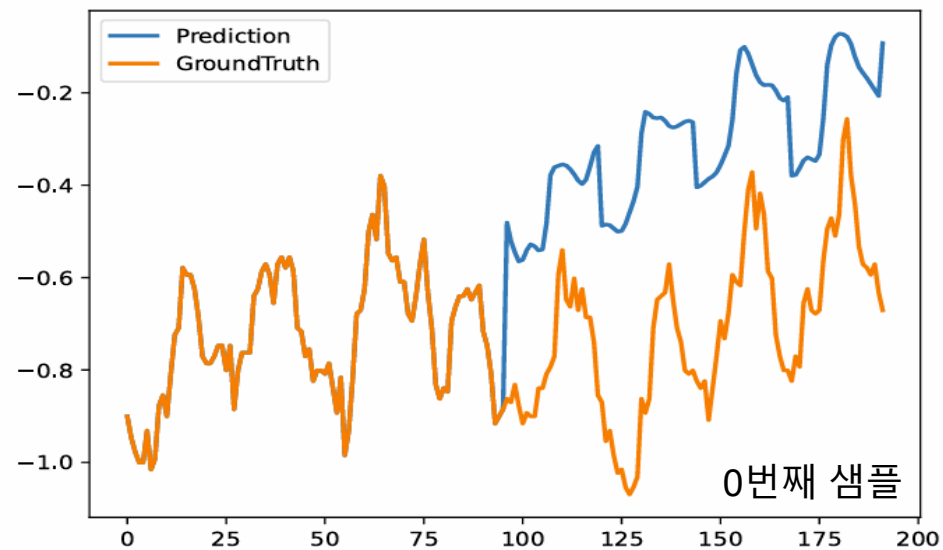


Transformer 재현 실험 결과 (ETTh1)

- Prediction length가 길어질수록 MSE, MAE 값이 증가하여 예측 성능이 점차 감소하는 모습을 확인

Pred_len	MSE	MAE
96	0.888	0.746
192	0.920	0.759
336	0.997	0.813
720	1.028	0.805

Transformer 재현 실험 시각화 (ETTh1)



LTSF-Linear

DLinear, Transformer 비교 실험

- 비교 실험을 통해 DLinear 기반 모델이 Transformer 기반 모델보다 성능이 20~50% 향상된 것을 확인
- 실험 결과를 통해 Prediction length가 길어질수록 DLinear의 성능 감소폭이 Transformer에 비해 훨씬 적다는 것을 확인, 이를 통해 장기 시계열 예측에서 DLinear가 훨씬 안정적이라는 것을 알 수 있음

	Transformer ETTm1		Transformer Weather		Transformer Exchange		DLinear ETTm1		DLinear Weather		DLinear Exchange	
Pred len	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
96	0.700	0.609	0.322	0.384	0.527	0.566	0.345	0.373	0.195	0.255	0.098	0.232
192	0.851	0.687	0.560	0.532	0.941	0.736	0.381	0.391	0.238	0.299	0.186	0.325
336	1.019	0.776	0.657	0.595	1.452	0.936	0.415	0.415	0.281	0.330	0.342	0.448
720	1.198	0.851	0.835	0.679	2.566	1.329	0.472	0.450	0.345	0.381	0.749	0.663

LTSF-Linear

입력 길이에 따른 성능 변화

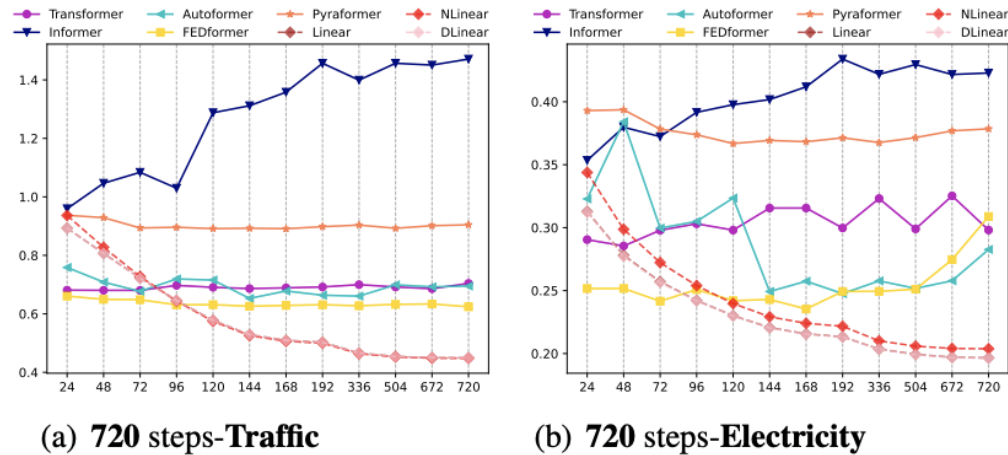


Figure 4. The MSE results (Y-axis) of models with different look-back window sizes (X-axis) of long-term forecasting (T=720) on the Traffic and Electricity datasets.

x: look-back window y: MSE 결과
Prediction: 720

- look-back window가 커질수록 더 많은 패턴을 학습해 성능이 향상되어야 하지만 Transformer 모델은 입력 길이가 길어지면 성능이 나빠지고, DLinear 모델은 반대로 성능이 좋아지는 모습을 보임
- Transformer 모델이 시간적 정보를 추출하지 못하고 많은 입력 정보를 노이즈로 처리해 과적합됨

실험 결과 정리

- DLinear 기반 모델이 Transformer 모델 대비 20~50%의 성능 향상을 보이며, 장기 시계열 예측에서 더 효과적임을 확인
- 비교 실험을 통해 Self-Attention을 사용하는 Transformer 모델보다 시계열의 Trend, Seasonality를 분리하는 단순한 선형 모델이 더 효율적인 것을 확인