

FEDformer: Frequency Enhanced Decomposed Transformer for Long-term Series Forecasting

<https://arxiv.org/pdf/2201.12740>

0. Introduction

- 기존 시계열 예측 모델은 파라미터 수가 많아 리소스 제한된 환경에서 쓰기 어려움
- 특히 Transformer 계열은 계산량 크고 구조 복잡해서 경량화에 불리함
- FEDformer는 시계열 분해와 주파수 기반 attention으로 이 문제를 해결하려고 함

1. Overview

- FEDformer는 시계열을 trend랑 seasonal로 나눠서 다르게 처리함
- trend는 복잡한 계산 없이 간단히 예측하고,
seasonal은 Fourier 변환으로 주파수 성분 뽑아서 중요한 것만 골라 sparse attention 적용함
- 이렇게 하면 계산량 줄이면서도, 중요한 정보는 놓치지 않게 예측 가능함
- Transformer 구조는 유지하되, 효율성과 정확도 모두 높인 게 특징임

2. Challenges

- 계산 부담 큼: Transformer는 입력 길이에 따라 계산량이 빠르게 늘어남 → 긴 시계열 예측에 비효율적임

- 추세 파악 어려움: 기존 모델은 전체적인 흐름(trend)보다 순간적인 패턴에 더 집중함
- 시계열 특성 반영 부족: 대부분 시계열은 특정 주파수에서만 정보가 집중되는데, 기존 모델은 이 구조를 활용하지 못함
- 자원 제약에 취약함: 복잡한 구조와 많은 파라미터 때문에 작은 디바이스에서 쓰기 어려움

3. Method

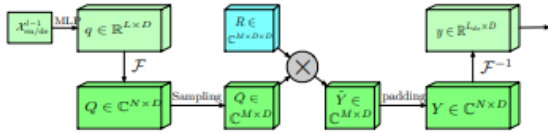
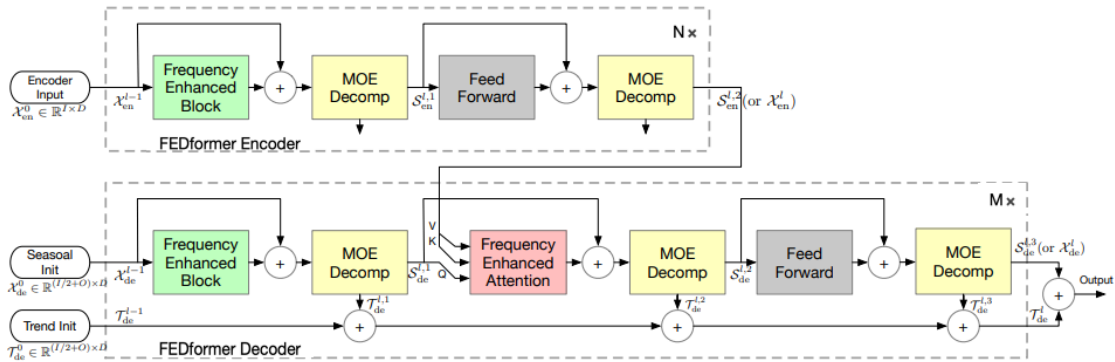


Figure 3. Frequency Enhanced Block with Fourier transform (FEB-f) structure.

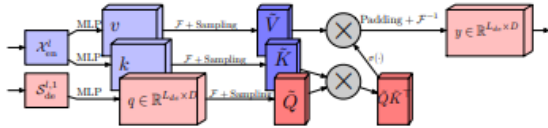


Figure 4. Frequency Enhanced Attention with Fourier transform (FEA-f) structure, $\sigma(\cdot)$ is the activation function.

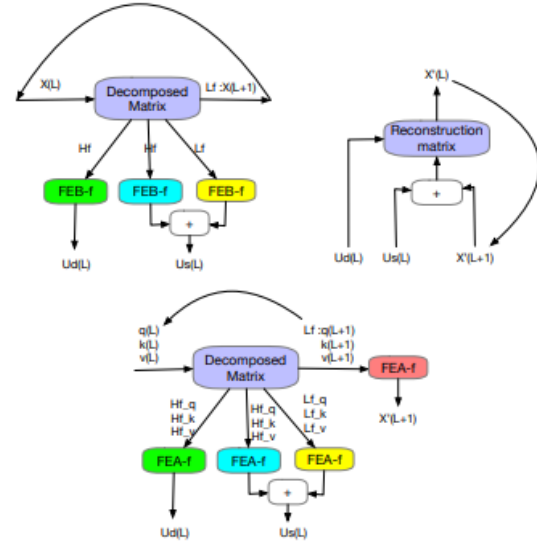


Figure 5. Top Left: Wavelet frequency enhanced block decomposition stage. Top Right: Wavelet block reconstruction stage shared by FEB-w and FEA-w. Bottom: Wavelet frequency enhanced cross attention decomposition stage.

- FEDformer는 시계열 데이터를 주파수 영역에서 효율적으로 다루기 위해 설계된 Transformer 모델임

- 시계열 전체 푸리에 성분을 다 쓰면 과적합 위험 있고 계산량 많아짐
- 그래서 푸리에 성분 중 일부를 무작위로 골라서 시계열을 간결하게 표현함
- 이렇게 하면 중요한 저주파와 고주파 정보를 모두 반영하면서 계산 효율도 높임
- 모델 구조는 인코더와 디코더로 나뉘고, 주요 블록은 Frequency Enhanced Block(FEB), Frequency Enhanced Attention(FEA), Mixture of Experts Decomposition(MOEDecomp)임
- FEB는 시계열을 푸리에 변환 또는 웨이블릿 변환해서 주파수 성분 기반으로 처리함
- FEA는 인코더-디코더 간 어텐션을 주파수 영역에서 계산함
- MOEDecomp는 여러 크기의 필터로 시계열을 계절성과 추세로 나눠 분해함
- 복잡도는 입력 길이 L 에 대해 선형($O(L)$)으로, 기존 Transformer보다 훨씬 가벼움
- 기본적으로 64개의 주파수 성분을 사용하고 3단계 분해를 적용함

4. Experiments

- 6개 공개 시계열 데이터셋 사용: ETTm1, ETTm2, ETTh1, ETTh2, Weather, Traffic, Electricity
- 평가 지표: MSE, MAE 기준으로 모델 성능 평가
- 다양한 실험 진행: 모델 변형(ablation), 주파수 샘플링 방식 비교 등

5. Results

Table 2. Multivariate long-term series forecasting results on six datasets with input length $I = 96$ and prediction length $O \in \{96, 192, 336, 720\}$ (For ILI dataset, we use input length $I = 36$ and prediction length $O \in \{24, 36, 48, 60\}$). A lower MSE indicates better performance, and the best results are highlighted in bold.

Methods	Metric	ETTh2				Electricity				Exchange				Traffic				Weather				ILI			
		96	192	336	720	96	192	336	720	96	192	336	720	96	192	336	720	96	192	336	720	24	36	48	60
FEDformer-f	MSE	0.203	0.269	0.325	0.421	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	0.287	0.328	0.366	0.415	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
FEDformer-w	MSE	0.204	0.316	0.359	0.433	0.183	0.195	0.212	0.231	0.139	0.256	0.426	1.090	0.562	0.562	0.570	0.596	0.227	0.295	0.381	0.424	2.203	2.272	2.209	2.545
	MAE	0.288	0.363	0.387	0.432	0.297	0.308	0.313	0.343	0.276	0.369	0.464	0.800	0.349	0.346	0.323	0.368	0.304	0.363	0.416	0.434	0.963	0.976	0.981	1.061
Autoformer	MSE	0.255	0.281	0.339	0.422	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	0.339	0.340	0.372	0.419	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	0.365	0.533	1.363	3.379	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	0.453	0.563	0.887	1.338	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
LogTrans	MSE	0.768	0.989	1.334	3.048	0.258	0.266	0.280	0.283	0.968	1.040	1.659	1.941	0.684	0.685	0.7337	0.717	0.458	0.658	0.797	0.869	4.480	4.799	4.800	5.278
	MAE	0.642	0.757	0.872	1.328	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	0.658	1.078	1.549	2.631	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	0.619	0.827	0.972	1.242	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

Table 3. Univariate long-term series forecasting results on six datasets with input length $I = 96$ and prediction length $O \in \{96, 192, 336, 720\}$ (For ILI dataset, we use input length $I = 36$ and prediction length $O \in \{24, 36, 48, 60\}$). A lower MSE indicates better performance, and the best results are highlighted in bold.

Methods	Metric	ETTM2				Electricity				Exchange				Traffic				Weather				ILI			
		96	192	336	720	96	192	336	720	96	192	336	720	96	192	336	720	96	192	336	720	24	36	48	60
FEDformer-f	MSE	0.072	0.102	0.130	0.178	0.253	0.282	0.346	0.422	0.154	0.286	0.511	1.301	0.207	0.205	0.219	0.244	0.0062	0.0060	0.0041	0.0055	0.708	0.584	0.717	0.855
	MAE	0.206	0.245	0.279	0.325	0.370	0.386	0.431	0.484	0.304	0.420	0.555	0.879	0.312	0.312	0.323	0.344	0.062	0.062	0.050	0.059	0.627	0.617	0.697	0.774
FEDformer-w	MSE	0.063	0.110	0.147	0.219	0.262	0.316	0.361	0.448	0.431	0.277	0.426	1.162	0.170	0.173	0.178	0.187	0.0035	0.0054	0.008	0.015	0.693	0.554	0.699	0.828
	MAE	0.189	0.252	0.301	0.368	0.378	0.410	0.445	0.501	0.284	0.420	0.511	0.832	0.263	0.265	0.266	0.286	0.046	0.059	0.072	0.091	0.629	0.604	0.696	0.770
Autoformer	MSE	0.065	0.118	0.154	0.182	0.341	0.345	0.406	0.565	0.241	0.300	0.509	1.260	0.246	0.266	0.263	0.269	0.011	0.0075	0.0063	0.0085	0.948	0.634	0.791	0.874
	MAE	0.189	0.256	0.305	0.335	0.438	0.428	0.470	0.581	0.387	0.369	0.524	0.867	0.346	0.370	0.371	0.372	0.081	0.067	0.062	0.070	0.732	0.650	0.752	0.797
Informer	MSE	0.080	0.112	0.166	0.228	0.258	0.285	0.336	0.607	1.327	1.258	2.179	1.280	0.257	0.299	0.312	0.366	0.004	0.002	0.004	0.003	5.282	4.554	4.273	5.214
	MAE	0.217	0.259	0.314	0.380	0.367	0.388	0.423	0.599	0.944	0.924	1.296	0.953	0.353	0.376	0.387	0.436	0.044	0.040	0.049	0.042	2.050	1.916	1.846	2.057
LogTrans	MSE	0.075	0.129	0.154	0.160	0.288	0.432	0.430	0.491	0.237	0.738	2.018	2.405	0.226	0.314	0.387	0.437	0.0046	0.0060	0.0060	0.007	3.607	2.407	3.106	3.698
	MAE	0.208	0.275	0.302	0.322	0.393	0.483	0.483	0.531	0.377	0.619	1.070	1.175	0.317	0.408	0.453	0.491	0.052	0.060	0.054	0.059	1.662	1.363	1.575	1.733
Reformer	MSE	0.077	0.138	0.160	0.168	0.275	0.304	0.370	0.460	0.298	0.777	1.833	1.203	0.313	0.386	0.423	0.378	0.012	0.0098	0.013	0.011	3.838	2.934	3.755	4.162
	MAE	0.214	0.290	0.313	0.334	0.379	0.402	0.448	0.511	0.444	0.719	1.128	0.956	0.383	0.453	0.468	0.433	0.087	0.044	0.100	0.083	1.720	1.520	1.749	1.847

Table 4. Ablation studies: multivariate long-term series forecasting results on ETTm1 and ETTm2 with input length $I = 96$ and prediction length $O \in \{96, 192, 336, 720\}$. Three variants of FEDformer-f are compared with baselines. The best results are highlighted in bold.

Methods	Metric	Transformer		Informer		Autoformer		FEDformer V1		FEDformer V2		FEDformer V3		FEDformer-f	
		Self-att	Cross-att	FullAtt	ProbAtt	AutoCorr	AutoCorr	FEB-f(Eq. 4)	AutoCorr	FEA-f(Eq. 7)	FEA-f(Eq. 7)	FEA-f(Eq. 7)	FEA-f(Eq. 7)	FEB-f(Eq. 4)	FEA-f(Eq. 7)
ETTM1	96	0.525	0.486	0.458	0.465	0.481	0.463	0.378	0.419	0.539	0.490	0.534	0.482	0.379	0.419
	192	0.526	0.502	0.564	0.521	0.628	0.526	0.417	0.442	0.556	0.499	0.552	0.493	0.426	0.441
	336	0.514	0.502	0.672	0.559	0.728	0.567	0.480	0.477	0.541	0.498	0.565	0.503	0.445	0.459
	720	0.564	0.529	0.714	0.596	0.658	0.548	0.543	0.517	0.558	0.507	0.585	0.515	0.543	0.490
	Count	0	0	0	0	0	0	5	5	6	6	7	7	8	8
ETTM2	96	0.268	0.346	0.227	0.305	0.255	0.339	0.259	0.337	0.216	0.297	0.211	0.292	0.203	0.287
	192	0.304	0.355	0.300	0.360	0.281	0.340	0.285	0.344	0.274	0.331	0.272	0.329	0.269	0.328
	336	0.365	0.400	0.382	0.410	0.339	0.372	0.320	0.373	0.334	0.369	0.327	0.363	0.325	0.366
	720	0.475	0.466	1.637	0.794	0.422	0.419	0.761	0.628	0.427	0.420	0.418	0.415	0.421	0.415
	Count	0	0	0	0	0	0	5	5	6	6	7	7	8	8

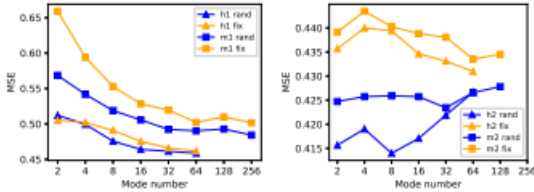


Figure 6. Comparison of two base-modes selection method (Fix&Rand). Rand policy means randomly selecting a subset of modes, Fix policy means selecting the lowest frequency modes. Two policies are compared on a variety of base-modes number $M \in \{2, 4, 8, \dots, 256\}$ on ETT full-benchmark (h1, m1, h2, m2).

Table 5. P-values of Kolmogrov-Smirnov test of different transformer models for long-term forecasting output on ETTm1 and ETTm2 dataset. Larger value indicates the hypothesis (the input sequence and forecasting output come from the same distribution) is less likely to be rejected. The best results are highlighted.

Methods		Transformer	Informer	Autoformer	FEDformer	True
ETTM1	96	0.0090	0.0055	0.020	0.048	0.023
	192	0.0052	0.0029	0.015	0.028	0.013
	336	0.0022	0.0019	0.012	0.015	0.010
	720	0.0023	0.0016	0.008	0.014	0.004
ETTM2	96	0.0012	0.0008	0.079	0.071	0.087
	192	0.0011	0.0006	0.047	0.045	0.060
	336	0.0005	0.00009	0.027	0.028	0.042
	720	0.0008	0.0002	0.023	0.021	0.023
Count		0	0	3	5	NA

- FEDformer는 기존 Transformer 모델들(Informer, Autoformer 등) 대비
 - 다변량 장기 예측에서 평균 14.8% MSE 감소
 - 단변량 예측에서 평균 22.6% MSE 감소
- 긴 예측 구간(336, 720 타임스텝)에서 성능 향상 두드러짐
- ablation 실험에서 시계열 분해와 frequency-enhanced attention이 성능 향상에 중요함 확인
- 랜덤 주파수 샘플링 방식이 고정 주파수 선택보다 더 효과적임
- 메모리 사용량과 연산 속도 측면에서 기존 Transformer 대비 크게 개선됨

- 긴 시계열에 대해 선형에 가까운 시간 복잡도 달성
- 주파수 기반 sparse attention이 핵심 주파수 정보 효과적으로 활용하며, 분포 변화나 잡음에도 강인함 입증됨

6. Insight

- 시계열을 trend와 seasonal로 분해해 각 성분에 맞게 처리하는 방식이 효과적임
- 주파수 도메인에서 sparse attention을 적용해 계산 효율과 예측 정확도를 동시에 개선함
- trend는 단순 선형 예측으로도 충분히 복잡한 계산 없이도 성능 유지 가능
- random sampling 방식의 주파수 선택이 예상외로 좋은 일반화 효과를 냄
- Transformer의 계산 복잡도 문제를 frequency 기반 방법으로 해결할 가능성을 보여줌
- trend 성분을 단순 선형으로만 예측하는 방식이 복잡한 패턴에선 한계일 수 있음
- 주파수 샘플링이 랜덤에 의존해 불안정할 가능성 존재
- 외부 요인이나 멀티모달 데이터를 반영하지 않아 실제 환경 적용 시 제한적일 수 있음
- 일부 데이터셋에서는 기존 모델과 큰 차이가 없거나 개선 폭이 크지 않은 경우도 있음
- 구현 복잡도와 실제 적용 시 계산 자원 문제에 대한 구체적 분석은 부족함