

SWIFT: Mapping Sub-series with Wavelet Decomposition Improves Time Series Forecasting

<https://arxiv.org/abs/2501.16178>

0. Introduction

- 시계열 데이터는 다양한 주파수 성분을 포함해 효과적인 분해가 중요함
- 기존 Transformer 계열 모델은 전체 시계열을 한꺼번에 처리해 주파수 특성 반영에 한계 있음
- SWIFT는 Wavelet Decomposition으로 시계열을 여러 sub-series로 분해해 각각 처리함
- plug-and-play 구조로 기존 예측 모델과 쉽게 결합 가능함
- 다양한 벤치마크에서 기존 SOTA 대비 뛰어난 예측 성능을 보임

1. Overview

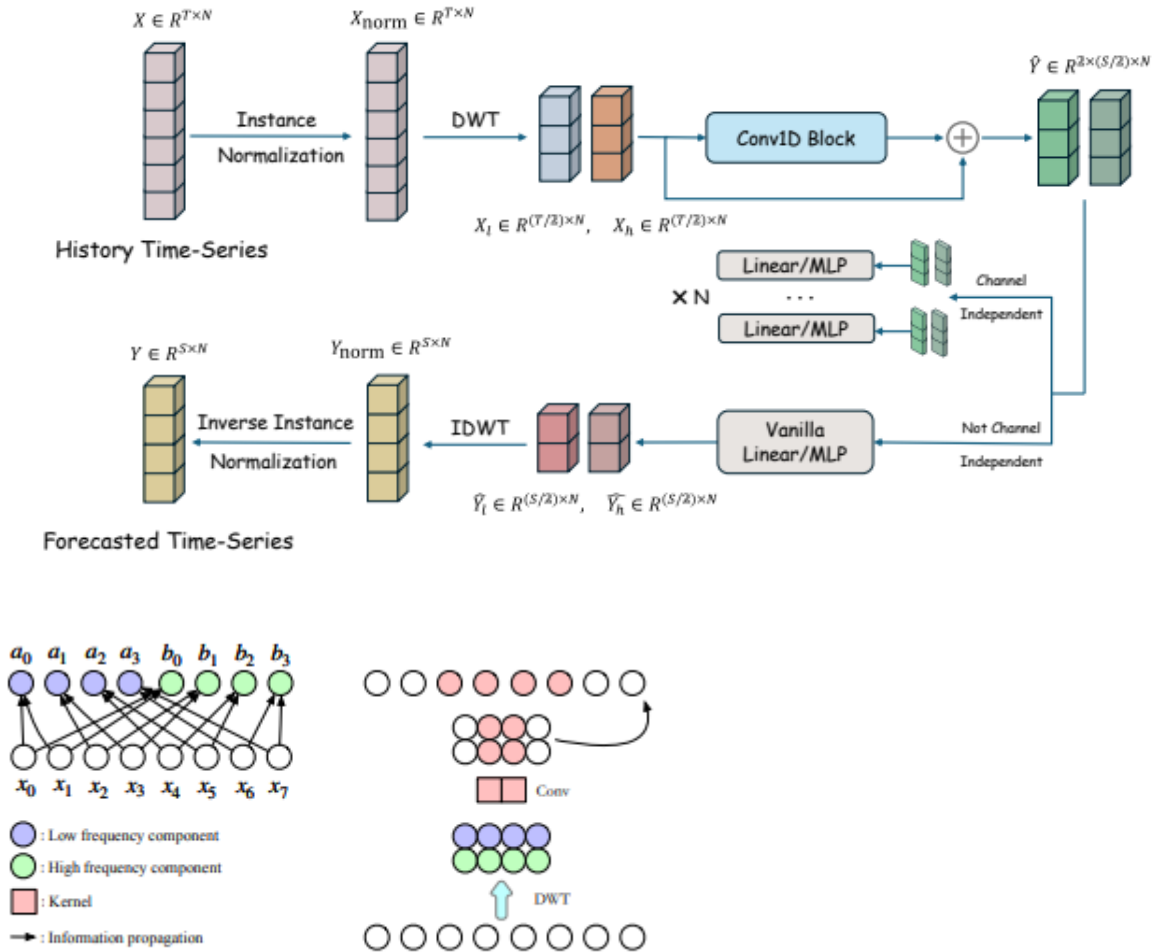
- SWIFT는 시계열을 wavelet을 통해 다중 해상도 sub-series로 분해함
- 각 sub-series는 개별 예측기로 독립 처리되고 결과를 다시 합침
- inverse wavelet transform으로 예측값 재조합
- 고주파 및 저주파 정보를 별도로 학습해 다양한 패턴을 포착함
- 기존 예측기(PatchTST, DLinear 등)에 쉽게 적용 가능함

2. Challenges

- 기존 Transformer 모델은 전체 시계열을 고정된 범위로 처리해 주파수 정보 손실 발생
- 단일 해상도 방식은 장기 의존성이나 세부 패턴을 충분히 반영하지 못함

- 시계열 내 주파수별 특성이 다르므로 분리 학습이 필요함
- 복잡한 주파수 성분을 효과적으로 처리하는 모델 설계가 어려움

3. Method



- 시계열 입력에 대해 Discrete Wavelet Transform(DWT) 적용해 여러 해상도의 sub-series로 분해
- 각 sub-series는 개별 예측 모듈에 입력되어 병렬 처리
- 처리된 결과는 inverse wavelet transform을 통해 다시 합쳐져 최종 예측 생성
- 전체 과정은 end-to-end 학습 가능
- 기존 예측 모델(Linear, DLinear, PatchTST 등)을 SWIFT 프레임워크 내에서 그대로 사용할 수 있음

4. Experiments

- 총 8개의 시계열 벤치마크 데이터셋 사용함 (ETT, Weather, Electricity, Traffic, Exchange-rate 등)
- 기존 Transformer 계열 모델과 경량 모델을 비교 대상으로 설정함
- 다양한 입력 길이와 예측 구간에 대해 실험 진행함
- 동일한 데이터 분할과 평가 지표(MSE, MAE)로 공정한 성능 비교 수행함

5. Results

Dataset	ETTh1					ETTh2					ETTm1					ETTm2				
Horizon	96	192	336	720	Avg	96	192	336	720	Avg	96	192	336	720	Avg	96	192	336	720	Avg
FEDformer	0.375	0.427	0.459	0.484	0.436	0.340	0.433	0.508	0.480	0.440	0.362	0.393	0.442	0.483	0.420	0.189	0.256	0.326	0.437	0.302
TimesNet	0.384	0.436	0.491	0.521	0.458	0.340	0.402	0.452	0.462	0.414	0.338	0.374	0.410	0.478	0.400	0.187	0.249	0.321	0.408	0.291
Dlinear	0.384	0.443	0.446	0.504	0.444	0.282	0.350	0.414	0.588	0.409	<u>0.301</u>	<u>0.335</u>	0.371	0.426	0.358	0.171	0.237	0.294	0.426	0.282
PatchTST	0.385	0.413	0.440	0.456	0.424	0.274	0.338	0.367	0.391	0.343	0.292	0.330	0.365	0.419	0.352	0.163	0.219	0.276	0.368	0.257
iTransformer	0.386	0.441	0.487	0.503	0.454	0.297	0.380	0.428	0.427	0.383	0.334	0.377	0.426	0.491	0.407	0.180	0.250	0.311	0.412	0.288
FITS	<u>0.372</u>	<u>0.404</u>	<u>0.427</u>	0.424	<u>0.407</u>	<u>0.271</u>	<u>0.331</u>	<u>0.354</u>	0.377	<u>0.333</u>	0.303	0.337	<u>0.366</u>	<u>0.415</u>	<u>0.355</u>	0.162	<u>0.216</u>	<u>0.268</u>	0.348	<u>0.249</u>
CycleNet	0.379	0.416	0.447	0.477	0.430	<u>0.271</u>	0.332	0.362	0.415	0.345	0.307	0.337	0.364	0.410	<u>0.355</u>	0.159	0.214	<u>0.268</u>	<u>0.353</u>	<u>0.249</u>
SWIFT / MLP	0.383	0.439	0.469	0.476	0.442	0.305	0.349	0.372	0.416	0.361	0.305	0.330	0.368	0.444	0.362	0.170	0.233	0.278	0.355	0.259
SWIFT / Linear	0.367	0.395	0.420	<u>0.430</u>	0.403	0.268	0.329	0.351	<u>0.383</u>	0.333	0.307	0.336	0.364	0.413	<u>0.355</u>	<u>0.161</u>	0.214	0.267	0.348	0.248
STD	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
IMP	0.005	0.009	0.007	-0.006	0.004	0.003	0.002	0.003	-0.006	0.000	-0.015	-0.006	0.002	0.002	-0.003	-0.002	0.000	0.001	0.000	0.001

Dataset	Weather					Electricity					Traffic				
Horizon	96	192	336	720	Avg	96	192	336	720	Avg	96	192	336	720	Avg
FEDformer	0.246	0.292	0.378	0.447	0.341	0.188	0.197	0.212	0.244	0.210	0.573	0.611	0.621	0.630	0.609
TimesNet	0.172	0.219	0.280	0.365	0.259	0.168	0.184	0.198	0.220	0.193	0.593	0.617	0.629	0.640	0.620
Dlinear	0.174	0.217	0.262	0.332	0.246	0.140	0.153	0.169	0.204	0.167	0.413	0.423	0.437	0.466	0.435
PatchTST	0.151	0.195	0.249	0.321	0.229	<u>0.129</u>	0.149	0.166	0.210	0.164	0.366	<u>0.388</u>	<u>0.398</u>	0.457	0.402
iTransformer	0.174	0.221	0.278	0.358	0.258	0.148	0.162	0.178	0.225	0.178	0.395	0.417	0.433	0.467	0.428
FITS	<u>0.143</u>	<u>0.186</u>	<u>0.236</u>	0.307	<u>0.218</u>	0.134	0.149	0.165	0.203	0.163	0.385	0.397	0.410	<u>0.448</u>	0.410
CycleNet	0.149	0.192	0.242	<u>0.312</u>	0.224	0.127	0.144	0.159	0.196	0.157	0.374	0.390	0.405	0.441	<u>0.403</u>
SWIFT / Linear	0.159	0.201	0.243	0.325	0.232	0.133	<u>0.148</u>	0.164	0.203	<u>0.162</u>	0.385	0.396	0.410	0.448	0.410
SWIFT / MLP	0.140	0.183	0.235	0.307	0.216	0.127	0.144	<u>0.160</u>	<u>0.197</u>	0.157	<u>0.368</u>	0.382	0.396	0.430	0.394
STD	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
IMP	0.003	0.003	0.001	0.000	0.002	0.000	0.000	-0.001	-0.001	0.000	-0.002	0.006	0.002	0.018	0.009

Model	Parameters	MACs	Train./epoch (GPU)
DLinear	138.4k	44.61 M	19.062s
FITS	116.2k	1189.91 M	25.070s
CycleNet / Linear	123.7k	22.42M	28.268s
CycleNet / MLP	472.9k	134.84M	30.200s
SWIFT / MLP	53.1k	33.53 M	19.717s
SWIFT / Linear	18.1k	11.09 M	18.571s

Dataset	ETTh1				ETTm1				traffic			
Horizon	96	192	336	720	96	192	336	720	96	192	336	720
SWIFT	0.367	0.395	0.420	0.430	0.307	0.336	0.364	0.413	0.368	0.382	0.396	0.430
<i>w/o Conv</i>	0.376	0.413	0.463	0.443	0.310	0.337	0.377	0.439	0.371	0.384	0.427	0.470
<i>w/o DWT</i>	0.365	0.399	0.427	0.446	0.321	0.338	0.365	0.414	0.521	0.674	0.533	0.717

Dataset	ETTh1				ETTh2				ETTm1				ETTm2			
Horizon	96	192	336	720	96	192	336	720	96	192	336	720	96	192	336	720
<i>Share</i>	0.367	0.395	0.420	0.430	0.268	0.329	0.351	0.383	0.307	0.336	0.364	0.413	0.161	0.214	0.267	0.348
<i>Split</i>	0.368	0.393	0.421	0.433	0.266	0.329	0.353	0.380	0.305	0.334	0.366	0.414	0.162	0.215	0.266	0.348
IMP	0.001	-0.002	0.001	0.003	-0.002	0.000	0.002	-0.003	-0.002	-0.002	0.002	0.001	0.001	0.001	-0.001	0.000

- 대부분 데이터셋에서 기존 SOTA 대비 MSE, MAE 수치 개선
- 특히 PatchTST에 SWIFT 적용 시 성능 향상폭 큼
- 장기 예측 시에도 성능 저하 적음
- 예를 들어 ETTm2 데이터셋에서 PatchTST 대비 10~15% MSE 개선
- 다양한 다변량 시계열에서도 안정적인 성능 유지

6. Insight

- Wavelet Decomposition을 통해 시계열을 주파수별로 분해해 정보 손실을 줄이고 다양한 패턴을 효과적으로 포착함
- sub-series별 개별 학습으로 모델이 세부 주파수 성분에 집중할 수 있음
- plug-and-play 구조로 기존 예측 모델에 쉽게 적용 가능해 확장성이 높음
- 데이터 특성에 집중한 접근법이 강점이나, 분해 및 재조합 과정에서 계산 비용과 복잡도가 증가할 수 있음
- 특정 도메인이나 데이터에 따라 분해 방법의 최적성은 불확실하며, 소규모 또는 비정형 데이터에 대한 일반화가 추가 검증 필요함
- 모든 Transformer 기반 모델에 동일한 성능 향상을 보장하지 않을 수 있음