

AutoHFormer: Efficient Hierarchical Autoregressive Transformer for Time Series Prediction

<https://arxiv.org/html/2506.16001v1>

0. Introduction

- 시계열 예측은 금융, 교통, 에너지 등 다양한 분야에서 중요한 문제임.
- 기존 Transformer 계열 모델들은 자기회귀 기반으로 강력한 성능 보였지만, 긴 시계열 다룰 때 연산 복잡도가 커짐.
- 또한 다중 스케일 패턴을 동시에 포착하기 어려워 성능 제한 있음.
- AutoHFormer는 이러한 한계를 극복하기 위해 제안됨.
- 핵심은 효율적이고 계층적인 autoregressive 구조로 coarse-to-fine 방식으로 시계열을 점진적으로 예측하는 아이디어.
- 이를 통해 더 긴 시계열도 효율적으로 처리하면서도 일반화 성능 확보 가능.

1. Overview

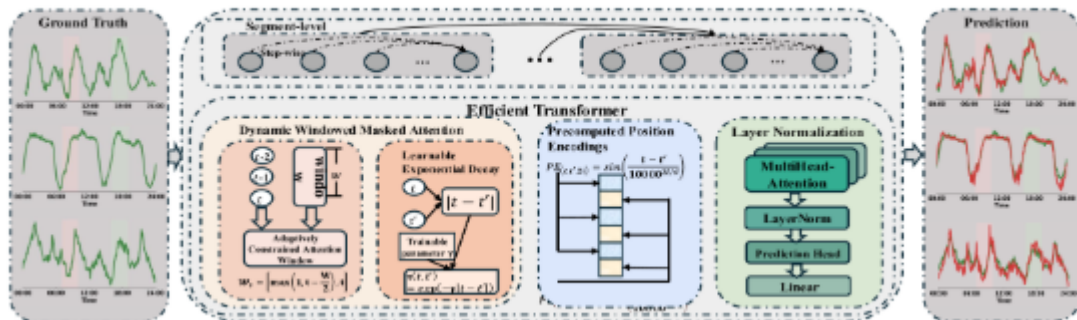
- AutoHFormer는 시계열 예측을 위한 계층적 autoregressive Transformer
- 기존 Transformer는 모든 시점 간의 의존성을 직접 계산해야 해서 긴 시계열에서는 비효율적
- AutoHFormer는 이를 개선하기 위해 coarse-to-fine 구조 도입.
- 먼저 큰 시간 구간의 거친 패턴(coarse pattern)을 예측하고, 이후 점점 작은 구간으로 내려가며 세밀한 fine-grained 패턴을 보완함.
- 이렇게 계층적으로 예측하기 때문에 연산량이 크게 줄어들고, 장기 의존성도 안정적으로 포착 가능.
- 추가로, 병렬적인 autoregressive 예측을 지원해 학습·추론 모두에서 효율적임.

- AutoHFormer는 긴 시계열 처리 시 효율성 + 정확성 두 가지를 동시에 잡는 모델임.

2. Challenges

- 긴 시계열 예측에서는 여러 가지 문제가 있음.
- Transformer 구조는 시퀀스 길이가 길어질수록 연산량이 크게 증가해 비효율적임.
- 시계열에는 단기 패턴과 장기 패턴이 동시에 존재하지만 기존 모델은 두 패턴을 균형 있게 학습하기 어려움.
- 또한 큰 주기의 흐름과 작은 변동 같은 다중 스케일 패턴을 함께 고려해야 하는데, 기존 방식은 이를 한 번에 처리하려다 보니 정보 손실이 발생함.
- 더불어 autoregressive 방식 특성상 순차적으로 예측을 진행해야 해서 긴 구간을 예측 할 때 속도가 느려지는 문제가 있음.
- AutoHFormer는 이런 한계를 해결하기 위해 계층적 구조와 병렬 예측 전략을 도입함.

3. Method



- AutoHFormer는 시계열을 coarse-to-fine 방식으로 단계적으로 예측하는 구조임.
- 먼저 긴 시계열을 큰 단위의 coarse 패턴으로 압축해 거친 예측을 수행함.
- 그 후 세밀한 fine 단위로 점차 분해해 추가적인 예측을 이어감.
- 이 계층적 autoregressive 구조 덕분에 큰 흐름과 세부 변동을 동시에 포착 가능함.
- 또한 각 단계의 예측은 병렬적으로 계산할 수 있어 기존 autoregressive 모델보다 훨씬 효율적임.

- 최종적으로 AutoHFormer는 연산량을 줄이면서도 장기 의존성과 다중 스케일 패턴을 잘 학습할 수 있는 구조를 제공함.

4. Experiments

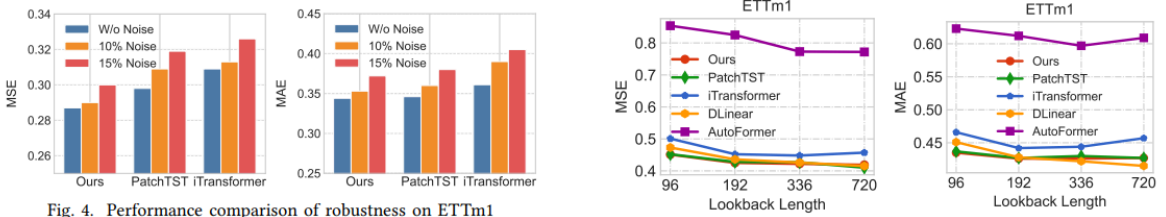
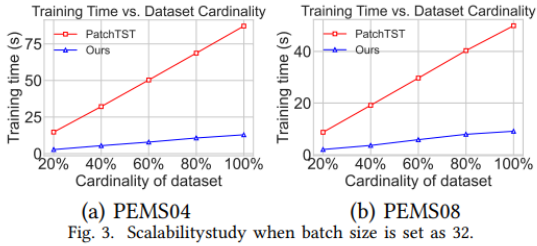
- 실험에는 다양한 시계열 데이터셋 사용함.
- 금융, 에너지, 기상 등 실제 환경에서 수집된 장기 시계열 데이터 포함.
- 데이터는 훈련·검증·테스트로 분리해 모델 성능 평가함.
- 비교 대상은 기존 Transformer 기반 시계열 모델과 autoregressive 모델들임.
- 실험에서는 AutoHFormer가 긴 시계열 처리 시 연산 효율과 예측 정확성을 동시에 확보하는지 검증함.
- 데이터 전처리와 스케일링, 패치 분할 등 계층적 학습에 필요한 세부 설정도 적용함.

5. Results

Model	Time Complexity	Space Complexity
Full Attention	$\mathcal{O}(L^2d)$	$\mathcal{O}(L^2d)$
AutoHFormer ($W \ll L$)	$\mathcal{O}(LWd)$	$\mathcal{O}(LWd)$
RNN-based	$\mathcal{O}(Ld^2)$	$\mathcal{O}(d)$

Datasets	Variates	Time steps	Granularity	Datasets	Variates	Time steps	Granularity
ETTh1	7	69,680	1 hour	ETTh2	7	69,680	1 hour
ETTm1	7	17,420	15 minutes	ETTm2	7	17,420	15 minutes
PEMS04	307	16,992	5 minutes	PEMS08	170	17,856	5 minutes
Weather	21	52,696	10 minutes	Electricity	321	26,304	1 hour

Models	AutoHFormer(Ours)		PatchTST [4]		TimeMixer [19]		iTransformer [18]		Autoformer [17]		Informer [12]		Linear [16]		DLinear [16]		NLinear [16]		
Metric	MSE ↓	MAE ↓	MSE ↓	MAE ↓	MSE ↓	MAE ↓	MSE ↓	MAE ↓	MSE ↓	MAE ↓	MSE ↓	MAE ↓	MSE ↓	MAE ↓	MSE ↓	MAE ↓	MSE ↓	MAE ↓	
ETTm1	96	0.287	0.344	<u>0.298</u>	<u>0.346</u>	0.302	0.351	0.309	0.361	0.723	0.569	1.293	0.862	0.311	0.354	0.301	0.345	0.308	0.350
	192	0.329	0.371	0.339	0.374	0.347	0.377	0.347	0.385	0.692	0.549	1.328	0.919	0.345	0.374	<u>0.336</u>	0.376	0.345	<u>0.372</u>
	336	0.363	0.392	<u>0.369</u>	0.395	0.401	0.414	0.386	0.407	0.727	0.523	1.483	0.963	0.377	0.394	0.371	0.397	0.380	<u>0.393</u>
	720	0.422	0.426	<u>0.427</u>	0.430	0.459	0.445	0.448	0.444	0.773	0.579	1.667	1.014	0.432	<u>0.427</u>	<u>0.426</u>	0.429	0.434	0.428
	Avg	0.350	0.383	<u>0.358</u>	0.386	0.377	0.396	0.372	0.399	0.728	0.555	1.442	0.939	0.366	0.387	0.360	0.386	0.366	<u>0.385</u>
Improved	-	-	2.28%	0.78%	7.71%	3.39%	6.26%	4.17%	108%	44.90%	226.28%	145.16%	4.57%	1.04%	2.85%	0.78%	4.57%	0.52%	
ETTm2	96	0.172	0.259	0.170	<u>0.260</u>	0.172	0.262	0.182	0.276	0.277	0.349	0.726	0.648	0.190	0.278	<u>0.171</u>	0.267	0.173	0.261
	192	0.236	0.308	<u>0.238</u>	0.305	0.249	0.309	0.243	0.315	0.306	0.365	0.632	0.627	0.249	0.329	0.239	0.320	0.241	0.324
	336	0.289	0.342	0.294	<u>0.343</u>	0.360	0.365	<u>0.290</u>	0.344	0.343	0.385	0.683	0.680	0.323	0.381	0.312	0.372	0.291	0.344
	720	0.379	0.398	<u>0.380</u>	<u>0.399</u>	0.437	0.413	0.384	0.399	0.433	0.437	0.645	0.589	0.450	0.458	0.445	0.455	0.383	0.417
	Avg	0.269	0.326	<u>0.270</u>	<u>0.327</u>	0.304	0.336	0.274	0.333	0.339	0.384	0.671	0.645	0.303	0.361	0.291	0.354	0.271	0.336
Improved	-	-	0.37%	0.30%	13.01%	3.06%	1.85%	2.14%	26.02%	17.79%	149.44%	97.85%	12.63%	10.74%	8.17%	8.59%	0.74%	3.07%	
ETTth1	96	<u>0.382</u>	<u>0.407</u>	0.374	0.400	0.384	0.410	0.398	0.418	0.608	0.529	1.178	0.826	0.431	0.441	0.389	0.412	0.426	0.438
	192	<u>0.427</u>	0.436	0.417	0.420	0.432	0.441	0.448	0.453	0.519	0.492	1.163	0.830	0.462	0.459	0.428	<u>0.434</u>	0.458	0.454
	336	0.431	0.442	0.457	<u>0.451</u>	0.450	0.453	0.465	0.468	0.643	0.541	1.161	0.819	0.472	0.471	<u>0.447</u>	<u>0.452</u>	0.470	0.464
	720	0.464	0.480	0.587	0.529	0.608	0.546	0.547	0.533	1.044	0.710	1.428	0.928	0.506	0.516	0.484	0.499	<u>0.473</u>	<u>0.481</u>
	Avg	0.426	0.441	0.458	0.450	0.468	0.462	0.464	0.468	0.703	0.568	1.232	0.850	0.467	0.471	<u>0.436</u>	<u>0.449</u>	0.459	0.459
Improved	-	-	7.51%	2.04%	9.85%	4.76%	8.92%	6.12%	65.02%	28.79%	189.20%	92.74%	9.62%	6.80%	2.34%	5.39%	7.74%	4.08%	
ETTth2	96	0.287	0.350	<u>0.294</u>	<u>0.353</u>	0.297	0.358	0.309	0.363	0.371	0.417	0.901	0.759	0.344	0.397	0.324	0.381	0.294	0.354
	192	0.353	0.390	0.374	0.404	<u>0.354</u>	<u>0.394</u>	0.389	0.412	0.426	0.453	0.631	0.629	0.435	0.454	0.416	0.440	0.357	0.395
	336	0.347	0.396	0.357	0.399	<u>0.348</u>	0.404	0.372	0.410	0.363	0.465	0.579	0.608	0.491	0.492	0.471	0.480	0.349	<u>0.401</u>
	720	0.399	0.436	<u>0.384</u>	<u>0.425</u>	0.411	0.441	0.446	0.462	0.491	0.506	0.734	0.662	0.781	0.629	0.747	0.614	0.403	0.441
	Avg	0.346	0.393	<u>0.353</u>	<u>0.395</u>	0.352	0.399	0.379	0.411	0.412	0.460	0.711	0.664	0.512	0.493	0.489	0.478	<u>0.350</u>	0.397
Improved	-	-	2.02%	0.50%	1.73%	1.52%	9.53%	4.58%	19.07%	17.04%	105.49%	68.95%	47.97%	25.44%	41.32%	21.62%	1.15%	1.02%	
Electricity	96	0.127	0.222	<u>0.132</u>	<u>0.226</u>	0.161	0.272	0.151	0.253	0.251	0.364	1.652	1.021	0.165	0.271	0.159	0.264	0.175	0.280
	192	0.138	0.246	<u>0.148</u>	<u>0.261</u>	0.183	0.293	0.168	0.267	0.267	0.373	1.518	0.975	0.174	0.281	0.168	0.274	0.185	0.290
	336	0.160	0.253	<u>0.165</u>	<u>0.258</u>	0.200	0.312	0.191	0.292	1.054	0.783	1.013	0.827	0.191	0.297	0.186	0.219	0.203	0.305
	720	0.200	0.287	<u>0.202</u>	<u>0.290</u>	0.226	0.330	0.232	0.326	1.737	1.068	0.986	0.818	0.225	0.327	0.220	0.322	0.242	0.334
	Avg	0.156	0.252	<u>0.161</u>	<u>0.258</u>	0.192	0.301	0.185	0.284	0.827	0.647	1.292	0.910	0.188	0.294	0.183	0.269	0.201	0.302
Improved	-	-	3.21%	2.38%	23.07%	19.44%	18.58%	12.69%	430.12%	156.74%	728.20%	261.11%	20.51%	16.67%	17.30%	6.74%	28.84%	19.84%	
Weather	96	<u>0.153</u>	<u>0.201</u>	0.150	0.198	0.181	0.230	0.159	0.210	0.261	0.325	0.830	0.665	0.177	0.237	0.175	0.234	0.182	0.233
	192	<u>0.199</u>	<u>0.246</u>	0.194	0.241	0.217	0.264	0.204	0.250	0.289	0.340	0.687	0.612	0.218	0.275	0.216	0.273	0.225	0.268
	336	0.247	0.282	<u>0.248</u>	<u>0.282</u>	0.272	0.302	0.255	0.288	0.340	0.391	0.843	0.677	0.263	0.312	0.261	0.310	0.272	0.302
	720	0.317	0.332	<u>0.318</u>	<u>0.334</u>	0.359	0.360	0.321	0.335	0.377	0.401	0.934	0.893	0.322	0.362	0.321	0.361	0.338	0.349
	Avg	<u>0.229</u>	<u>0.265</u>	0.227	0.263	0.257	0.289	0.234	0.270	0.316	0.339	0.825	0.711	0.245	0.296	0.243	0.296	0.254	0.288
Improved	-	-	-0.87%	-0.75%	12.22%	9.05%	2.18%	1.88%	37.99%	27.92%	260.26%	168.30%	6.99%	11.70%	6.11%	0.699%	10.92%	8.68%	
PEMS04	12	0.072	0.173	<u>0.080</u>	0.185	0.073	<u>0.178</u>	0.081	0.186	0.594	0.613	0.252	0.376	0.125	0.244	0.096	0.204	0.104	0.209
	24	0.083	0.187	0.103	0.204	<u>0.086</u>	<u>0.197</u>	0.098	0.207	0.527	0.580	0.670	0.591	0.152	0.268	0.126	0.236	0.136	0.250
	48	0.101	0.205	0.131	<u>0.229</u>	<u>0.125</u>	0.239	0.126	0.234	0.845	0.763	1.353	0.871	0.191	0.299	0.169	0.273	0.184	0.287
	96	0.124	0.225	<u>0.150</u>	<u>0.262</u>	0.196	0.302	0.154	0.263	1.149	0.875	0.949	0.738	0.224	0.322	0.204	0.300	0.227	0.316
	Avg	0.095	0.198	0.116	<u>0.220</u>	0.120	0.229	<u>0.114</u>	0.222	0.778	0.707	0.806	0.644	0.173	0.283	0.148	0.253	0.171	0.265
Improved	-	-	22.11%	11.11%	26.31%	15.65%	20.00%	12.12%	718.95%	257.07%	748.42%	225.25%	82.11%	42.93%	55.79%	27.78%	80.00%	33.84%	
PEMS08	12	0.066	0.161	<u>0.074</u>	0.177	0.080	<u>0.175</u>	0.081	0.178	0.769	0.686	0.405	0.426	0.139	0.246	0.103	0.207	0.105	0.213
	24	0.082	0.175	<u>0.097</u>	0.199	0.104	<u>0.193</u>	0.108	0.197	0.885	0.751	0.786	0.629	0.183	0.275	0.151	0.243	0.154	0.250
	48	0.111	0.192	<u>0.140</u>	0.227	0.149	<u>0.212</u>	0.149	0.225	0.959	0.759	1.293	0.842	0.271	0.319	0.242	0.294	0.254	0.301
	96	0.157	0.211	<u>0.201</u>	0.254	0.229	<u>0.246</u>	0.209	0.253	1.226	0.910	1.023	0.750	0.352	0.353	0.328	0.333	0.367	0.341
	Avg	0.104	0.185	<u>0.128</u>	0.214	0.120	<u>0.206</u>	0.136	0.213	0.959	0.776	0.646	0.661	0.223	0.298	0.206	0.269	0.220	0.276
Improved	-	-	23.08%	15.68%	15.38%	11.35%	30.77%	15.14%	822.12%	319.46%	521.15%	257.30%	114.42%	61.08%	98.08%	45.41%	111.54%	49.19%	
1 st /2 nd	34/5	34/5	6/22	6/19	0/4	0/8	0/2	0/0	0/0	0/0	0/0	0/0	0/0	0/1	0/5	0/2	0/2	0/5	



- AutoHFormer는 대부분 데이터셋에서 기존 Transformer 기반 모델 대비 예측 정확도 향상함.
- 특히 긴 시계열에서는 RMSE와 MAE 지표에서 유의미한 성능 개선 확인됨.
- 연산 효율도 높아져, 기존 모델 대비 추론 속도가 빠름.
- coarse-to-fine 계층 구조 덕분에 장기 패턴과 단기 변동을 동시에 잘 포착함.
- 다중 스케일 패턴 학습 능력 덕분에 금융, 에너지, 기상 데이터 등 다양한 도메인에서 안정적인 성능 보여줌.

6. Insight

- AutoHFormer는 긴 시계열 예측에서 효율성과 정확성을 동시에 달성할 수 있음을 보여줌.
- coarse-to-fine 구조는 장기 패턴을 먼저 잡고 세밀한 변동을 보완한다는 점에서 직관적이고 확장성 있는 접근임.
- 병렬적인 autoregressive 예측은 기존 모델이 가진 속도 문제를 완화시킨 점에서 의미가 큼.
- 다만 구조가 계층적으로 복잡해 실사용 환경에서 튜닝과 최적화가 쉽지 않을 수 있음.
- coarse-to-fine 방식이 데이터 특성에 따라 성능 편차를 보일 가능성 있음.
- 효율적인 구조적 아이디어는 인상적이지만 실제 응용에서는 데이터 도메인별 일반화 성능을 더 면밀히 검증할 필요 있음.