

GPHT: Generative Pretrained Hierarchical Transformer for Time Series Forecasting

<https://arxiv.org/pdf/2402.16516>

0. Introduction

- 시계열 예측에서 기존 모델들은 일반화 성능과 예측 유연성에 한계가 있음.
- 단일 데이터셋에 의존하거나, 고정된 예측 범위를 갖는 모델들이 많음.
- 이러한 문제를 해결하기 위해, 다양한 데이터셋을 혼합하여 사전 학습하고, 자기회귀 방식으로 예측하는 새로운 모델인 Generative Pretrained Hierarchical Transformer (GPHT)를 제안함.
- GPHT는 하나의 모델로 다양한 예측 범위에 대응할 수 있으며, 사전 학습을 통해 데이터셋 간 공통 패턴을 학습하여 일반화 성능을 향상시킴.
- 8개의 벤치마크 데이터셋에서 실험한 결과, GPHT는 기존 모델들을 능가하는 성능을 보였음.

1. Overview

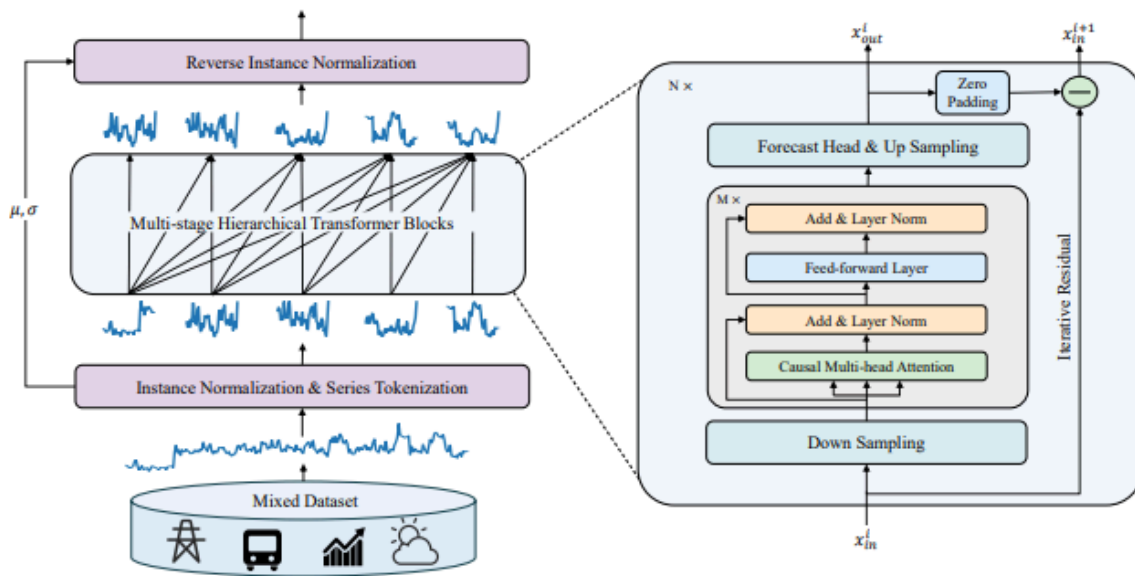
- GPHT는 다양한 시계열 데이터셋에서 공통 패턴을 학습하도록 사전 학습(pretraining)된 Transformer 모델
- 계층적 구조로 시계열 패턴을 장기, 단기 수준에서 동시에 모델링
- 자기회귀 방식으로 유연하게 다양한 예측 범위에 대응
- 단일 모델로 여러 데이터셋과 예측 범위에 일반화 가능
- 목표는 기존 모델 대비 장기 예측 안정성과 데이터셋 간 일반화 성능 향상

2. Challenges

- 기존 시계열 모델들은 단일 데이터셋이나 고정 예측 범위에 최적화되어 일반화 어려움

- 장기 예측에서 정확도와 안정성을 동시에 확보하기 어려움
- 다양한 도메인과 시계열 길이를 동시에 다루기 위해 모델 구조와 학습 전략 설계 필요
- 자기회귀 방식 사용 시 누적 오차 문제 발생 가능
- 사전 학습(pretraining) 시 데이터셋 간 패턴 차이로 인한 학습 불안정성 존재
- 효율적인 학습과 추론 속도 유지하면서 높은 예측 성능 확보가 과제

3. Method



- 다양한 시계열 데이터셋을 혼합하여 사전 학습(pretraining) 진행
- 계층적 Transformer 구조 사용 장기 패턴과 단기 패턴을 동시에 모델링
- 입력 시계열을 패치 단위로 나누어 임베딩 후 Transformer에 투입
- 자기회귀 방식으로 시점별 예측 수행, 이전 예측값을 다음 입력에 반영
- 사전 학습 단계에서 데이터셋 간 공통 패턴 학습, 미세 조정(fine-tuning) 시 특정 도메인 적합
- 학습 과정에서 예측 범위와 도메인에 따라 유연하게 대응 가능
- 모델 구조 변경 최소화, 기존 Transformer 모듈 활용 가능

4. Experiments

- 평가에 사용된 데이터셋 8개 전력, 기상, 금융, 교통 등 다양한 도메인 포함
- 시계열 길이와 샘플 수가 다른 고차원 데이터셋 혼합
- 사전 학습(pretraining) 후 특정 도메인 데이터로 미세 조정(fine-tuning) 수행
- Ablation study 준비 사전 학습 제거, 계층적 구조 단순화, 자기회귀 제거 등 조건별 성능 확인
- 장기 및 단기 예측 모두 평가되도록 데이터셋 구성

5. Results

Methods	Metric	GPHT		FPT		PatchTST		DLinear	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Exchange	96	0.098	0.219	0.104	0.226	0.102	0.227	0.169	0.316
	192	0.183	0.305	0.218	0.333	0.205	0.325	0.230	0.374
	336	0.321	0.411	0.391	0.460	0.362	0.440	0.334	0.444
	720	0.824	0.682	0.978	0.734	0.991	0.745	0.560	0.591
Traffic	96	0.411	0.291	0.447	0.331	0.433	0.314	0.453	0.328
	192	0.435	0.302	0.461	0.335	0.447	0.319	0.464	0.330
	336	0.460	0.316	0.477	0.343	0.465	0.329	0.481	0.340
	720	0.521	0.353	0.503	0.356	0.504	0.354	0.506	0.351
Weather	96	0.202	0.244	0.216	0.264	0.207	0.259	0.239	0.297
	192	0.248	0.283	0.260	0.301	0.257	0.299	0.275	0.325
	336	0.306	0.324	0.328	0.351	0.340	0.350	0.323	0.360
	720	0.389	0.377	0.414	0.403	0.414	0.402	0.392	0.405

Dataset	Variables	Frequency	Length	Scope
ETTh1/ETTh2	7	1 Hour	17420	Energy
ETTm1/ETTm2	7	15 Minutes	69680	Energy
Electricity	321	1 Hour	26304	Energy
Exchange	8	1 Day	7588	Finance
Traffic	862	1 Hour	17544	Transportation
Weather	21	10 Minutes	52696	Weather

Type Methods	Metric	Ours				Self-supervised								Supervised							
		GPHT*		GPHT		PatchTST		FPT		SimMTM		TimeMAE		PatchTST		iTransformer		TimesNet		DLinear	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Electricity	96	0.128	0.219	0.128	0.219	<u>0.132</u>	0.225	0.139	0.238	0.133	<u>0.223</u>	0.133	0.230	0.138	0.233	<u>0.132</u>	0.228	0.177	0.281	0.141	0.238
	192	<u>0.147</u>	0.236	0.146	0.236	0.148	0.241	0.155	0.252	<u>0.147</u>	<u>0.237</u>	0.150	0.246	0.153	0.247	0.154	0.249	0.193	0.295	0.154	0.251
	336	0.165	0.255	0.165	0.255	0.167	<u>0.260</u>	0.170	0.267	<u>0.166</u>	0.265	<u>0.166</u>	0.265	0.170	0.263	0.172	0.267	0.206	0.306	0.170	0.269
	720	0.206	0.292	0.207	0.292	0.205	0.292	0.208	0.299	<u>0.203</u>	0.297	0.199	0.296	0.206	<u>0.295</u>	0.204	0.296	0.223	0.320	0.205	0.302
Exchange	96	0.096	<u>0.216</u>	0.087	0.207	<u>0.088</u>	0.207	0.098	0.222	0.100	0.226	0.229	0.352	0.094	<u>0.216</u>	0.099	0.225	0.166	0.305	0.087	0.217
	192	0.183	0.304	<u>0.172</u>	0.296	0.186	0.308	0.209	0.327	0.210	0.332	0.653	0.581	0.191	0.311	0.206	0.329	0.303	0.413	0.164	<u>0.298</u>
	336	<u>0.322</u>	<u>0.410</u>	0.309	0.400	0.374	0.446	0.398	0.463	0.389	0.460	1.524	0.887	0.343	0.427	0.370	0.448	0.445	0.511	0.333	0.437
	720	<u>0.833</u>	<u>0.685</u>	0.808	0.669	0.857	0.692	1.010	0.747	1.104	0.800	2.525	1.193	0.888	0.706	0.963	0.746	1.389	0.899	0.988	0.749
Traffic	96	<u>0.348</u>	<u>0.236</u>	0.346	0.234	0.382	0.262	0.388	0.279	0.368	0.262	0.365	0.252	0.395	0.272	0.361	0.266	0.600	0.323	0.411	0.284
	192	0.374	<u>0.248</u>	0.371	0.246	0.385	0.261	0.411	0.287	<u>0.373</u>	0.251	0.383	0.260	0.411	0.278	0.378	0.271	0.612	0.327	0.423	0.289
	336	0.392	0.259	0.388	<u>0.256</u>	0.409	0.275	0.423	0.293	0.395	0.254	0.399	0.269	0.424	0.284	<u>0.390</u>	0.274	0.628	0.344	0.437	0.297
	720	0.428	<u>0.284</u>	0.423	0.279	0.438	0.291	0.449	0.307	0.432	0.290	0.438	0.291	0.453	0.300	<u>0.424</u>	0.291	0.657	0.349	0.467	0.316
Weather	96	0.155	0.196	0.154	0.196	<u>0.148</u>	0.196	0.152	0.201	0.152	0.201	0.151	0.208	0.147	<u>0.197</u>	0.162	0.212	0.168	0.225	0.176	0.236
	192	0.203	0.240	0.201	0.240	<u>0.193</u>	0.240	0.197	<u>0.244</u>	0.198	0.245	0.198	0.256	0.191	0.240	0.205	0.251	0.218	0.268	0.217	0.275
	336	0.259	0.283	0.257	0.283	0.244	0.279	0.252	0.287	0.249	0.285	<u>0.246</u>	0.294	0.244	<u>0.282</u>	0.257	0.291	0.269	0.301	0.264	0.315
	720	0.338	0.337	0.335	0.337	0.321	0.334	0.329	0.340	0.324	<u>0.335</u>	0.316	0.351	<u>0.320</u>	0.334	0.325	0.337	0.340	0.350	0.325	0.364
ETTh1	96	0.378	<u>0.388</u>	0.363	0.382	0.384	0.401	0.388	0.405	0.383	0.411	0.431	0.450	0.382	0.403	0.405	0.419	0.421	0.438	<u>0.375</u>	0.396
	192	0.425	<u>0.416</u>	0.405	0.408	0.427	0.431	0.422	0.423	0.417	0.432	0.484	0.486	<u>0.416</u>	0.423	0.448	0.447	0.482	0.479	0.428	0.437
	336	0.456	<u>0.432</u>	<u>0.430</u>	0.423	0.461	0.450	0.442	0.435	0.425	0.439	0.515	0.507	0.441	0.440	0.482	0.470	0.528	0.505	0.448	0.449
	720	0.454	<u>0.449</u>	0.414	0.435	0.460	0.465	0.469	0.473	<u>0.437</u>	0.456	0.595	0.577	0.470	0.475	0.560	0.537	0.527	0.510	0.505	0.514
ETTh2	96	0.307	0.347	0.296	0.340	0.297	0.354	<u>0.291</u>	0.349	0.298	0.350	0.294	0.358	0.286	<u>0.342</u>	0.305	0.361	0.355	0.408	0.296	0.360
	192	0.373	0.389	0.363	0.384	0.388	0.406	<u>0.356</u>	0.390	0.360	<u>0.388</u>	0.352	0.397	0.357	0.389	0.391	0.412	0.403	0.434	0.391	0.423
	336	0.399	0.414	0.392	<u>0.410</u>	0.392	0.413	<u>0.387</u>	0.418	0.388	<u>0.410</u>	0.394	0.427	0.377	0.409	0.418	0.433	0.398	0.434	0.445	0.460
	720	0.412	<u>0.429</u>	<u>0.407</u>	<u>0.427</u>	0.413	0.442	0.415	0.448	0.412	0.435	0.539	0.510	0.406	0.440	0.437	0.455	0.443	0.465	0.700	0.592
ETTm1	96	0.301	0.345	0.291	0.339	0.281	<u>0.341</u>	<u>0.290</u>	0.346	0.296	0.349	0.301	0.348	0.298	0.345	0.306	0.360	0.331	0.372	0.303	0.346
	192	0.347	0.374	0.337	0.368	0.326	0.372	<u>0.330</u>	<u>0.371</u>	0.334	0.373	0.351	0.383	0.339	0.374	0.345	0.382	0.435	0.421	0.338	0.368
	336	0.388	0.401	0.377	<u>0.393</u>	0.348	0.384	<u>0.366</u>	<u>0.393</u>	0.371	0.398	0.390	0.408	0.381	0.401	0.378	0.402	0.457	0.445	0.373	<u>0.393</u>
	720	0.465	0.441	0.452	0.433	0.399	0.418	<u>0.416</u>	<u>0.421</u>	0.418	0.425	0.457	0.446	0.428	0.431	0.443	0.439	0.526	0.481	0.428	0.423
ETTm2	96	0.179	<u>0.257</u>	0.170	0.250	<u>0.171</u>	<u>0.257</u>	<u>0.171</u>	0.261	0.173	0.264	0.180	0.267	0.174	0.261	0.174	0.266	0.190	0.276	0.170	0.264
	192	0.242	<u>0.298</u>	0.230	0.291	0.236	0.304	<u>0.231</u>	0.302	0.230	0.299	0.243	0.312	0.238	0.307	0.247	0.315	0.244	0.311	0.233	0.311
	336	0.300	0.334	<u>0.285</u>	0.327	0.291	0.344	0.288	0.343	0.282	<u>0.332</u>	0.308	0.355	0.293	0.346	0.292	0.343	0.302	0.349	0.298	0.358
	720	0.400	0.393	0.380	0.386	0.388	0.404	0.389	0.406	<u>0.374</u>	<u>0.390</u>	0.395	0.407	0.373	0.401	0.375	0.395	0.406	0.406	0.423	0.437
#1 Counts		8		41		13		0		4		3		10		0		0		4	

Portion		5%										10%									
Methods	Metric	GPHT		FPT		SimMTM		PatchTST		iTransformer		GPHT		FPT		SimMTM		PatchTST		iTransformer	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Electricity	96	0.143	0.237	0.148	0.246	0.152	0.255	0.188	0.292	0.155	0.256	0.140	0.233	0.149	0.248	0.146	0.246	0.147	0.245	0.148	0.247
	192	0.162	0.254	0.163	0.259	0.167	0.268	0.202	0.304	0.172	0.272	0.159	0.250	0.164	0.261	0.163	0.262	0.162	0.258	0.167	0.266
	336	0.184	0.275	0.181	0.277	0.187	0.287	0.219	0.318	0.197	0.295	0.180	0.271	0.183	0.280	0.184	0.280	0.181	0.276	0.192	0.290
	720	0.238	0.321	0.231	0.315	0.240	0.326	0.264	0.351	0.261	0.344	0.231	0.313	0.234	0.318	0.242	0.325	0.230	0.315	0.244	0.329
ETTh1	96	0.383	0.390	0.478	0.474	0.537	0.502	0.505	0.481	0.580	0.520	0.382	0.391	0.453	0.454	0.482	0.467	0.450	0.448	0.557	0.514
	192	0.426	0.416	0.705	0.577	0.580	0.525	0.576	0.514	0.670	0.557	0.424	0.418	0.522	0.494	0.532	0.498	0.523	0.489	0.668	0.562
	336	0.453	0.430	0.736	0.571	0.603	0.543	0.672	0.554	0.726	0.577	0.450	0.443	0.571	0.522	0.561	0.523	0.523	0.494	0.684	0.559
	720	0.433	0.440	0.718	0.579	0.708	0.597	0.759	0.625	0.802	0.626	0.427	0.442	0.574	0.535	0.734	0.617	0.508	0.502	0.709	0.587
ETTh2	96	0.298	0.343	0.476	0.457	0.381	0.401	0.502	0.475	0.395	0.420	0.298	0.343	0.330	0.371	0.332	0.373	0.320	0.366	0.365	0.398
	192	0.368	0.386	0.714	0.573	0.435	0.435	0.569	0.511	0.448	0.453	0.367	0.387	0.419	0.422	0.391	0.411	0.400	0.416	0.432	0.439
	336	0.402	0.412	0.683	0.573	0.431	0.441	0.540	0.506	0.453	0.462	0.396	0.413	0.419	0.434	0.410	0.429	0.405	0.425	0.437	0.450
	720	0.417	0.429	0.648	0.557	0.450	0.459	0.506	0.494	0.483	0.484	0.409	0.428	0.506	0.485	0.448	0.460	0.483	0.474	0.463	0.471
ETTm1	96	0.513	0.438	0.395	0.409	0.446	0.434	0.376	0.395	0.434	0.436	0.506	0.427	0.403	0.411	0.442	0.430	0.386	0.401	0.420	0.427
	192	0.552	0.464	0.410	0.417	0.461	0.436	0.391	0.402	0.472	0.456	0.563	0.458	0.430	0.426	0.454	0.431	0.406	0.413	0.472	0.456
	336	0.609	0.491	0.453	0.440	0.500	0.455	0.438	0.431	0.531	0.486	0.634	0.492	0.472	0.443	0.552	0.469	0.438	0.429	0.530	0.486
	720	0.685	0.581	0.742	0.566	0.590	0.503	0.549	0.495	0.615	0.527	0.721	0.534	0.665	0.526	0.715	0.539	0.499	0.464	0.629	0.533
ETTm2	96	0.186	0.271	0.196	0.278	0.216	0.293	0.196	0.276	0.211	0.295	0.173	0.256	0.198	0.275	0.203	0.283	0.191	0.270	0.198	0.284
	192	0.248	0.311	0.263	0.316	0.267	0.324	0.258	0.315	0.269	0.322	0.234	0.297	0.263	0.315	0.256	0.315	0.252	0.308	0.254	0.318
	336	0.307	0.347	0.336	0.363	0.315	0.356	0.318	0.353	0.325	0.370	0.293	0.335	0.320	0.350	0.305	0.345	0.310	0.345	0.305	0.352
	720	0.412	0.409	0.453	0.430	0.406	0.406	0.447	0.427	0.441	0.434	0.398	0.395	0.426	0.412	0.397	0.398	0.398	0.397	0.405	0.408

- 장기 예측에서도 안정적인 성능 유지
- 사전 학습(pretraining) 단계가 모델 일반화 성능 향상에 기여
- Ablation study에서 계층적 구조 제거 시 성능 하락, 자기회귀 제거 시 장기 예측 정확도 감소
- 도메인별 분석 전력과 금융 데이터에서 성능 향상폭이 가장 큼, 교통과 기상 데이터에서는 개선폭이 상대적으로 작음
- 전반적으로 GPHT 구조가 다양한 도메인과 예측 범위에 일반화 성능 개선에 효과적임

6. Insight

- GPHT는 사전 학습과 계층적 Transformer 구조로 장기 및 단기 예측 성능과 일반화 능력 향상
- 자기회귀 방식과 패치 단위 임베딩이 다양한 데이터셋에서 유연하게 대응 가능
- 사전 학습 단계가 데이터셋 간 공통 패턴 학습에 기여, 미세 조정으로 특정 도메인 성능 강화
- 장점 장기 예측 안정성, 다양한 도메인 일반화, 구조적 효율성
- 비판적 시야 데이터셋 간 차이가 큰 경우 사전 학습이 오히려 불안정성 유발 가능
- 계산 비용과 메모리 사용량이 계층적 구조와 자기회귀 결합으로 증가할 수 있음
- 향후 개선 가능성 데이터셋 차이를 반영한 사전 학습 전략, 계산 효율 개선, 실시간 예측 적용 연구 필요