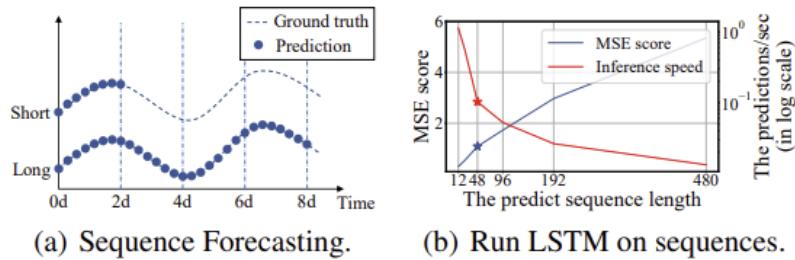


# Informer : Beyond Efficient Transformer for Long Sequence Time-Series Forecasting

<https://arxiv.org/pdf/2012.07436>

## 0. Introduction



- 장기 시계열(Long Sequence Time-Series Forecasting, LSTF) 문제는 전력 예측, 센서 데이터, 기후 예측 등 다양한 실제 응용에서 중요함.
- 기존 모델(LSTM, RNN, Transformer)은 장기 의존성 학습과 계산 효율에서 한계가 있음.
- Transformer는 장거리 의존성을 캡처 가능하지만 계산량과 메모리 사용이 급격히 증가함.
- 핵심 기여:
  1. ProbSparse Self-Attention을 통한 연산 효율 개선
  2. Self-Attention Distilling으로 feature map 압축
  3. Generative Style Decoder로 전체 시퀀스 예측 속도 향상

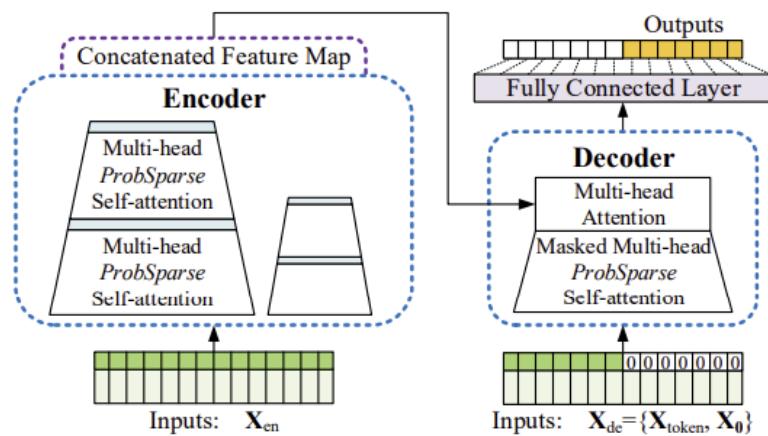
## 1. Overview

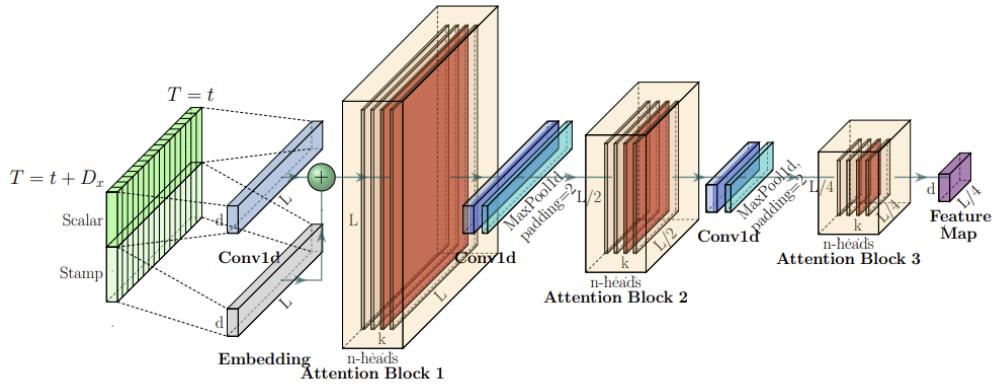
- Encoder-Decoder 기반 Transformer 구조
- ProbSparse Self-Attention: 입력 시퀀스  $L$  길이 대비  $O(L \log L)$  계산량
- Distilling 구조: 여러 layer를 거치며 dominant feature 추출, 입력 길이 감소
- Generative Decoder: 전체 시퀀스를 한 번에 예측, 누적 오류 최소화
- 적용 분야: 전력, 센서, 금융, 기후 등 장기 시계열 예측

## 2. Challenges

- Self-Attention 연산 복잡도:  $O(L^2)$ , 장기 시퀀스에서 비효율적
- 긴 시퀀스 의존성 학습 어려움
- 기존 RNN 기반 접근 방식은 장기 시퀀스 처리 한계 존재
- 대규모 시계열 처리 시 계산량, 메모리 문제

## 3. Method





- ProbSparse Self-Attention: 전체 attention 중 상위 중요 쌍만 계산,  $O(L \log L)$  복잡도
- Self-Attention Distilling: feature map down-sampling, dominant features 유지
- Generative Style Decoder: 한 번의 forward pass로 전체 시퀀스 예측, 추론 속도 향상
- 입력 처리: 시퀀스 embedding, positional encoding
- 학습: Cross-entropy loss, optimizer 및 하이퍼파라미터 조정

## 4. Experiments

- 데이터셋: 다양한 공개 장기 시계열 데이터
- 비교 대상 모델: 기존 Transformer, RNN 계열 모델
- 평가 지표: MSE, MAE, 메모리 사용량, 추론 속도
- 실험 구성: Train/Validation/Test 분리, 하이퍼파라미터 조정

## 5. Results

Methods	Informer		Informer <sup>†</sup>		LogTrans		Reformer		LSTM		DeepAR		ARIMA		Prophet		
Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	
ETTh <sub>1</sub>	24	0.098 0.247	<b>0.092 0.246</b>		0.103 0.259	0.222 0.389	0.114 0.272	0.107 0.280	0.108 0.284	0.115 0.275							
	48	<b>0.158 0.319</b>	0.161 0.322	0.167 0.328	0.284 0.445	0.193 0.358	0.162 0.327	0.175 0.424	0.168 0.330								
	168	<b>0.183 0.346</b>	0.187 0.355	0.207 0.375	1.522 1.191	0.236 0.392	0.239 0.422	0.396 0.504	1.224 0.763								
	336	0.222 0.387	<b>0.215 0.369</b>	0.230 0.398	1.860 1.124	0.590 0.698	0.445 0.552	0.468 0.593	1.549 1.820								
	720	0.269 0.435	<b>0.257 0.421</b>	0.273 0.463	2.112 1.436	0.683 0.768	0.658 0.707	0.659 0.766	2.735 3.253								
ETTh <sub>2</sub>	24	<b>0.093 0.240</b>	0.099 0.241	0.102 0.255	0.263 0.437	0.155 0.307	0.098 0.263	3.554 0.445	0.199 0.381								
	48	<b>0.155 0.314</b>	0.159 0.317	0.169 0.348	0.458 0.545	0.190 0.348	0.163 0.341	3.190 0.474	0.304 0.462								
	168	<b>0.232 0.389</b>	0.235 0.390	0.246 0.422	1.029 0.879	0.385 0.514	0.255 0.414	2.800 0.595	2.145 1.068								
	336	0.263 0.417	<b>0.258 0.423</b>	0.267 0.437	1.668 1.228	0.558 0.606	0.604 0.607	2.753 0.738	2.096 2.543								
	720	<b>0.277 0.431</b>	0.285 0.442	0.303 0.493	2.030 1.721	0.640 0.681	0.429 0.580	2.878 1.044	3.355 4.664								
ETTm <sub>1</sub>	24	<b>0.030 0.137</b>	0.034 0.160	0.065 0.202	0.095 0.228	0.121 0.233	0.091 0.243	0.090 0.206	0.120 0.290								
	48	0.069 0.203	<b>0.066 0.194</b>	0.078 0.220	0.249 0.390	0.305 0.411	0.219 0.362	0.179 0.306	0.133 0.305								
	96	0.194 0.372	<b>0.187 0.384</b>	0.199 0.386	0.920 0.767	0.287 0.420	0.364 0.496	0.272 0.399	0.194 0.396								
	288	<b>0.401 0.554</b>	0.409 0.548	0.411 0.572	1.108 1.245	0.524 0.584	0.948 0.795	0.462 0.558	0.452 0.574								
	672	<b>0.512 0.644</b>	0.519 0.665	0.598 0.702	1.793 1.528	1.064 0.873	2.437 1.352	0.639 0.697	2.747 1.174								
Weather	24	<b>0.117 0.251</b>	0.119 0.256	0.136 0.279	0.231 0.401	0.131 0.254	0.128 0.274	0.219 0.355	0.302 0.433								
	48	<b>0.178 0.318</b>	0.185 0.316	0.206 0.356	0.328 0.423	0.190 0.334	0.203 0.353	0.273 0.409	0.445 0.536								
	168	<b>0.266 0.398</b>	0.269 0.404	0.309 0.439	0.654 0.634	0.341 0.448	0.293 0.451	0.503 0.599	2.441 1.142								
	336	<b>0.297 0.416</b>	0.310 0.422	0.359 0.484	1.792 1.093	0.456 0.554	0.585 0.644	0.728 0.730	1.987 2.468								
	720	<b>0.359 0.466</b>	0.361 0.471	0.388 0.499	2.087 1.534	0.866 0.809	0.499 0.596	1.062 0.943	3.859 1.144								
ECL	48	0.239 0.359	0.238 0.368	0.280 0.429	0.971 0.884	0.493 0.539	<b>0.204 0.357</b>	0.879 0.764	0.524 0.595								
	168	0.447 0.503	0.442 0.514	0.454 0.529	1.671 1.587	0.723 0.655	<b>0.315 0.436</b>	1.032 0.833	2.725 1.273								
	336	0.489 0.528	0.501 0.552	0.514 0.563	3.528 2.196	1.212 0.898	<b>0.414 0.519</b>	1.136 0.876	2.246 3.077								
	720	<b>0.540 0.571</b>	0.543 0.578	0.558 0.609	4.891 4.047	1.511 0.966	0.563 0.595	1.251 0.933	4.243 1.415								
	960	<b>0.582 0.608</b>	0.594 0.638	0.624 0.645	7.019 5.105	1.545 1.006	0.657 0.683	1.370 0.982	6.901 4.264								
Count	32		12		0		0		0		6		0		0		

Methods	Informer		Informer <sup>†</sup>		LogTrans		Reformer		LSTM		LSTnet	
Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh <sub>1</sub>	24	<b>0.577 0.549</b>	0.620	0.577	0.686	0.604	0.991	0.754	0.650	0.624	1.293	0.901
	48	<b>0.685 0.625</b>	0.692	0.671	0.766	0.757	1.313	0.906	0.702	0.675	1.456	0.960
	168	<b>0.931 0.752</b>	0.947	0.797	1.002	0.846	1.824	1.138	1.212	0.867	1.997	1.214
	336	1.128 0.873	<b>1.094 0.813</b>	1.362	0.952	2.117	1.280	1.424	0.994	2.655	1.369	
	720	<b>1.215 0.896</b>	1.241	0.917	1.397	1.291	2.415	1.520	1.960	1.322	2.143	1.380
ETTh <sub>2</sub>	24	<b>0.720 0.665</b>	0.753	0.727	0.828	0.750	1.531	1.613	1.143	0.813	2.742	1.457
	48	<b>1.457 1.001</b>	1.461	1.077	1.806	1.034	1.871	1.735	1.671	1.221	3.567	1.687
	168	3.489 1.515	3.485	1.612	4.070	1.681	4.660	1.846	4.117	1.674	<b>3.242</b>	2.513
	336	2.723 1.340	2.626	<b>1.285</b>	3.875	1.763	4.028	1.688	3.434	1.549	<b>2.544</b>	2.591
	720	<b>3.467 1.473</b>	3.548	1.495	3.913	1.552	5.381	2.015	3.963	1.788	4.625	3.709
ETTm <sub>1</sub>	24	0.323 0.369	<b>0.306 0.306</b>	0.371	0.419	0.412	0.724	0.607	0.621	0.629	1.968	1.170
	48	0.494 0.503	<b>0.465 0.470</b>	0.507	0.583	1.098	0.777	1.392	0.939	1.999	1.215	
	96	<b>0.678 0.614</b>	0.681	<b>0.612</b>	0.768	0.792	1.433	0.945	1.339	0.913	2.762	1.542
	288	<b>1.056 0.786</b>	1.162	0.879	1.462	1.320	1.820	1.094	1.740	1.124	1.257	2.076
	672	<b>1.192 0.926</b>	1.231	1.103	1.669	1.461	2.187	1.232	2.736	1.555	1.917	2.941
Weather	24	<b>0.335 0.381</b>	0.349	0.397	0.435	0.477	0.655	0.583	0.546	0.570	0.615	0.545
	48	0.395 0.459	<b>0.386 0.433</b>	0.426	0.495	0.729	0.666	0.829	0.677	0.660	0.589	
	168	<b>0.608 0.567</b>	0.613	0.582	0.727	0.671	1.318	0.855	1.038	0.835	0.748	0.647
	336	<b>0.702 0.620</b>	0.707	0.634	0.754	0.670	1.930	1.167	1.657	1.059	0.782	0.683
	720	<b>0.831 0.731</b>	0.834	0.741	0.885	0.773	2.726	1.575	1.536	1.109	0.851	0.757
ECL	48	0.344 0.393	<b>0.334 0.399</b>	0.355	0.418	1.404	0.999	0.486	0.572	0.369	0.445	
	168	0.368 0.424	<b>0.353 0.420</b>	0.368	0.432	1.515	1.069	0.574	0.602	0.394	0.476	
	336	0.381 0.431	0.381	0.439	<b>0.373</b>	0.439	1.601	1.104	0.886	0.795	0.419	0.477
	720	0.406 0.443	<b>0.391 0.438</b>	0.409	0.454	2.009	1.170	1.676	1.095	0.556	0.565	
	960	<b>0.460 0.548</b>	0.492	0.550	0.477	0.						

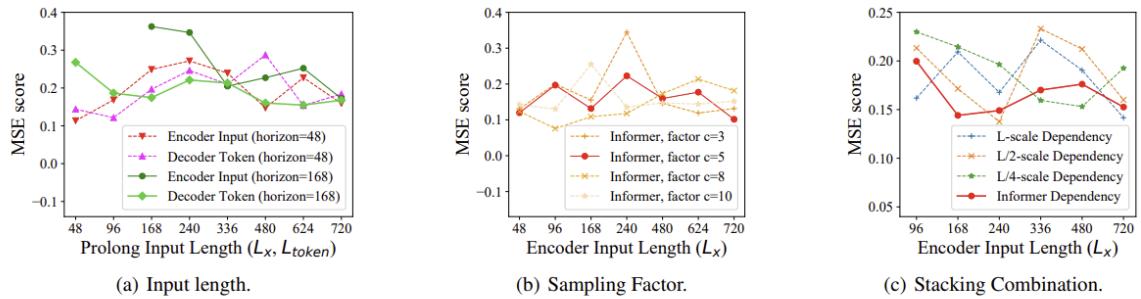


Figure 4: The parameter sensitivity of three components in Informer.

Prediction length	336			720		
Encoder's input	336	720	1440	720	1440	2880
Informer	MSE	0.249	0.225	0.216	0.271	0.261
	MAE	0.393	0.384	0.376	0.435	0.422
Informer <sup>†</sup>	MSE	0.241	0.214	-	0.259	-
	MAE	0.383	0.371	-	0.423	-
LogTrans	MSE	0.263	0.231	-	0.273	-
	MAE	0.418	0.398	-	0.463	-
Reformer	MSE	1.875	1.865	1.861	2.243	2.174
	MAE	1.144	1.129	1.125	1.536	1.497
						2.113

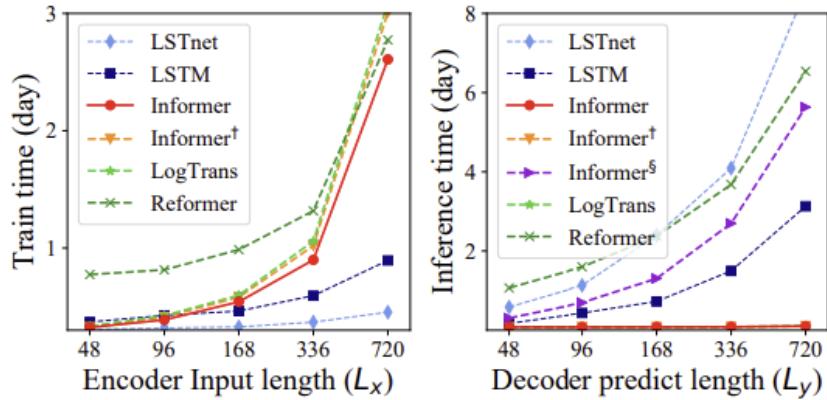
<sup>1</sup> Informer<sup>†</sup> uses the canonical self-attention mechanism.

<sup>2</sup> The '-' indicates failure for the out-of-memory.

Methods	Training		Testing
	Time	Memory	
Informer	$\mathcal{O}(L \log L)$	$\mathcal{O}(L \log L)$	1
Transformer	$\mathcal{O}(L^2)$	$\mathcal{O}(L^2)$	$L$
LogTrans	$\mathcal{O}(L \log L)$	$\mathcal{O}(L^2)$	$1^*$
Reformer	$\mathcal{O}(L \log L)$	$\mathcal{O}(L \log L)$	$L$
LSTM	$\mathcal{O}(L)$	$\mathcal{O}(L)$	$L$

<sup>1</sup> The LSTnet is hard to present in a closed form.

<sup>2</sup> The '\*' denotes applying our proposed decoder.



- Informer는 기존 모델 대비 정확도 향상, 계산량/메모리 사용 감소, 추론 속도 향상
- 핵심 성능 지표:
  - 계산량:  $O(L \log L)$
  - 예측 정확도: 기존 Transformer 대비 향상
  - 추론 속도: 전체 시퀀스 한 번에 예측 가능

## 6. Insight

- 장기 시계열 예측에서 Transformer의 성능을 유지하면서 연산 효율 극대화 가능
- 장점: 긴 시퀀스 처리 성능, 모델 확장성, 빠른 추론
- 단점: 구조 복잡성, 하이퍼파라미터 민감도
- 후속 연구:
  - Exogenous 변수 반영 확장
  - Graph-Augmented Informer
  - Transformer 계열 효율적 구조 연구