

Long-term Forecasting with TiDE : Time-series Dense Encoder

<https://arxiv.org/pdf/2304.08424>

0. Introduction

- 장기 시계열 예측은 에너지, 교통, 재고, 수요 등 다양한 산업에서 핵심 문제임
- Transformer 모델이 널리 쓰이지만 복잡하고 연산 비용 큼
- 단순 선형 모델이 특정 상황에서는 더 나은 성능을 보인다는 결과도 존재
- 하지만 선형 모델은 비선형 패턴, 계절성 변화, 공변량 반영에 한계 있음
- 논문은 단순 MLP 기반 구조로 비선형성·공변량 활용을 강화한 TiDE(Time-series Dense Encoder) 제안
- 이론적 분석으로 선형 근사 형태가 장기 예측에서 near-optimal 성능을 가질 수 있음 보여줌
- 실험 결과 여러 벤치마크에서 최신 Transformer보다 빠르고 경쟁력 있는 성능 보임

1. Overview

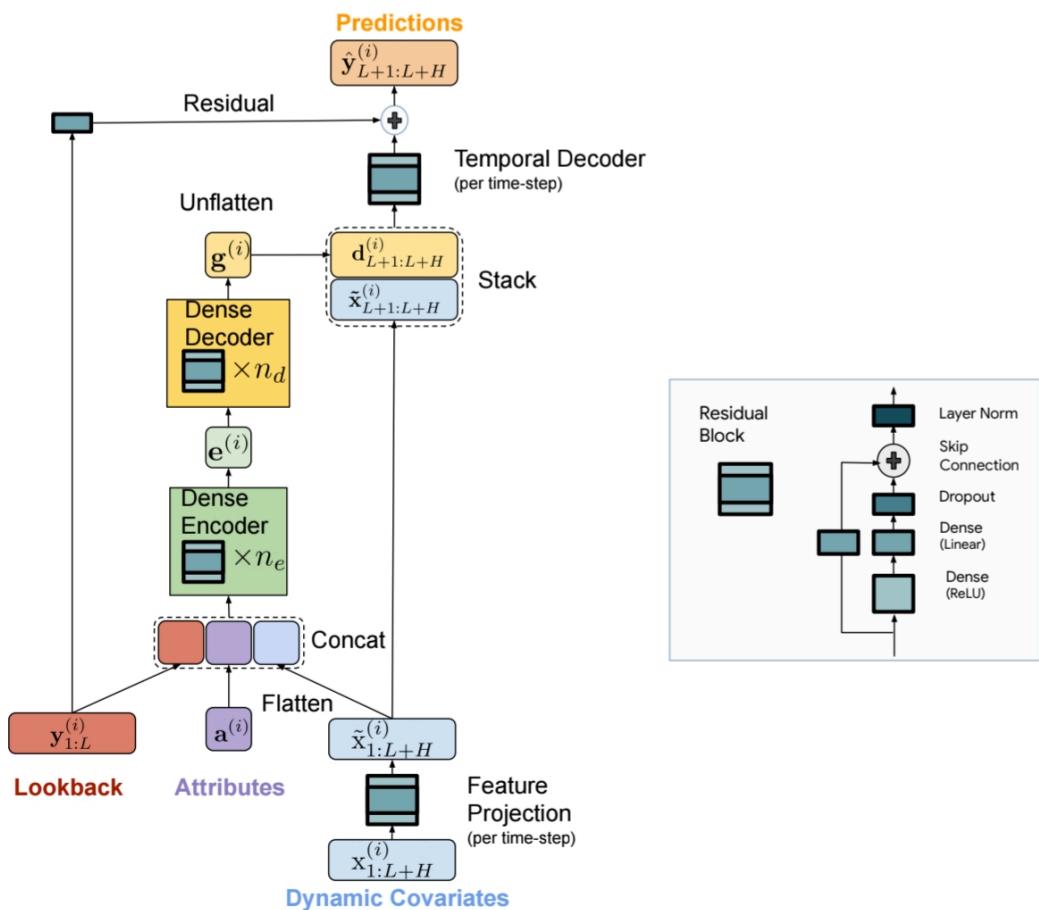
- 모델은 단순하고 효율적인 Dense MLP Encoder–Decoder 구조로 구성됨
- 구성요약
 - 인코더: 과거 시계열 + 공변량 + 정적 속성을 MLP residual 블록으로 인코딩
 - 디코더:
 - Dense Decoder: latent 예측 표현 생성
 - Temporal Decoder: 미래 공변량과 결합해 최종 출력 생성
 - Global Linear Residual: 선형 예측 경로 추가해 단순 선형 모델 포함 가능
- 학습은 MSE 손실 + Adam 기반

- 전체 목표는 빠르고 강건하며 확장 가능한 장기 예측 모델 개발

2. Challenges

- Transformer 구조는 길이가 길어질수록 비효율적임
- 선형 모델은 효율적이지만 복잡한 비선형 패턴·공변량 영향 포착 어려움
- 공변량이 많은 실제 예측 문제에서 구조적으로 잘 통합하기 어려움
- 다양한 벤치마크에서 안정적으로 동작하는 구조 설계가 필요함
- 이론적 보장은 비선형 모델 전체에 적용하기 어려움

3. Method



- Residual MLP Block

- 1층 MLP + ReLU + Dropout + LayerNorm + skip connection
- Encoder
 - 시계열·공변량·정적 속성을 저차원 embedding으로 투영
 - Dense residual MLP 통해 compact representation 만들기
- Dense Decoder
 - 인코더 latent를 MLP로 변환해 예측용 중간 표현 생성
- Temporal Decoder
 - 각 미래 시점 공변량과 latent 결합
 - highway 구조로 미래 공변량이 직접 예측에 영향 주도록 설계
- Global Linear Residual
 - 선형 모델 결과를 그대로 더해 선형성과 비선형성 조화
- Training
 - MSE 사용
 - Adam optimizer
 - rolling window 기반 평가
- Theoretical View
 - 선형 근사 모델이 특정 조건에서 near-optimal long-term forecasting 오차율 달성함

4. Experiments

Dataset	#Time-Series	#Time-Points	Frequency
Electricity	321	26304	1 Hour
Traffic	862	17544	1 Hour
Weather	21	52696	10 Minutes
ETTh1	7	17420	1 Hour
ETTh2	7	17420	1 Hour
ETTm1	7	69680	15 Minutes
ETTm2	7	69680	15 Minutes

- 데이터셋

- Weather, Electricity, Traffic, ETT(ETTh1/ETTh2/ETTm1/ETTm2)
- M5 forecasting dataset (정적·공변량 포함)

- 비교 모델

- Transformer 계열 (Fedformer, Autoformer, Informer, PatchTST)
- 선형 계열 (DLinear)
- 기타 모델 (N-HiTS 등)

- 평가 지표

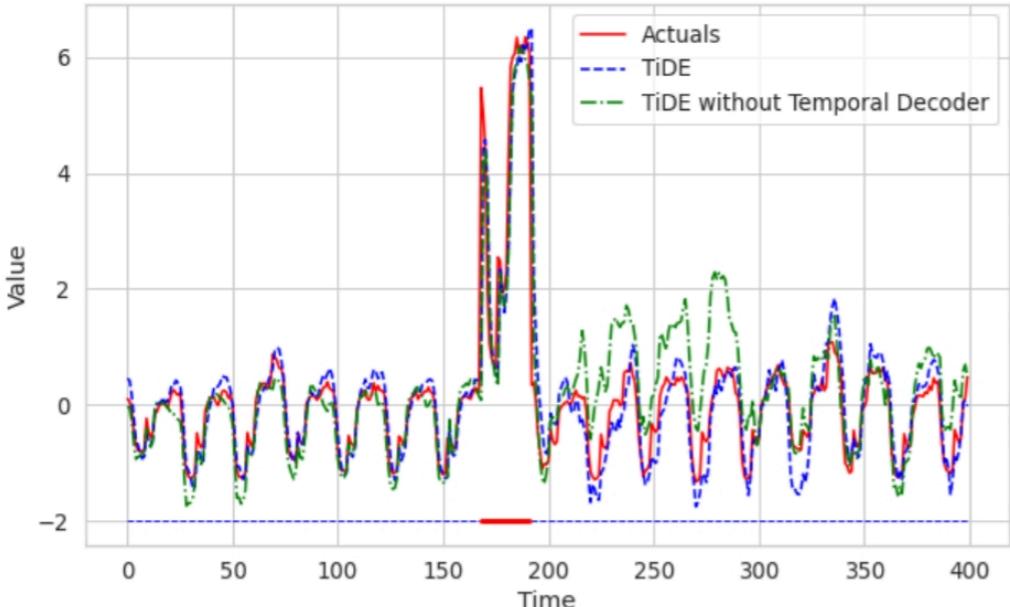
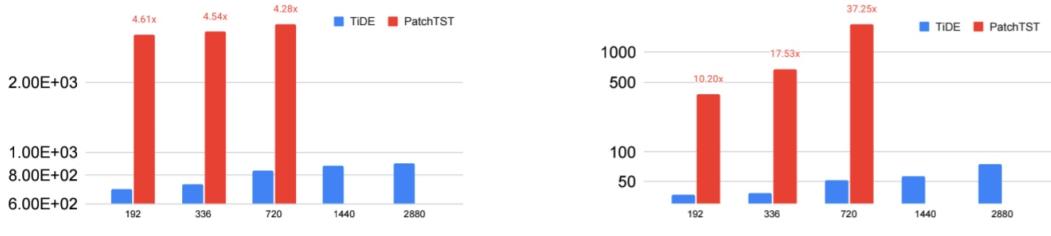
- MSE, MAE

- Ablation

- temporal decoder 유무
- global residual 제거 실험
- 효율성 비교 (training/inference 속도)

5. Results

Models	TiDE		PatchTST/64		N-HiTS		DLinear		FEDformer		Autoformer		Informer		Pyraformer		LogTrans		
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	
Weather	96	0.166	0.222	0.149	0.198	0.158	0.195	0.176	0.237	0.238	0.314	0.249	0.329	0.354	0.405	0.896	0.556	0.458	0.490
	192	0.209	0.263	0.194	0.241	0.211	0.247	0.220	0.282	0.275	0.329	0.325	0.370	0.419	0.434	0.622	0.624	0.658	0.589
	336	0.254	0.301	0.245	0.282	0.274	0.300	0.265	0.319	0.339	0.377	0.351	0.391	0.583	0.543	0.739	0.753	0.797	0.652
	720	0.313	0.340	0.314	0.334	0.401	0.413	0.323	0.362	0.389	0.409	0.415	0.426	0.916	0.705	1.004	0.934	0.869	0.675
Traffic	96	0.336	0.253	0.360	0.249	0.402	0.282	0.410	0.282	0.576	0.359	0.597	0.371	0.733	0.410	2.085	0.468	0.684	0.384
	192	0.346	0.257	0.379	0.256	0.420	0.297	0.423	0.287	0.610	0.380	0.607	0.382	0.777	0.435	0.867	0.467	0.685	0.390
	336	0.355	0.260	0.392	0.264	0.448	0.313	0.436	0.296	0.608	0.375	0.623	0.387	0.776	0.434	0.869	0.469	0.734	0.408
	720	0.386	0.273	0.432	0.286	0.539	0.353	0.466	0.315	0.621	0.375	0.639	0.395	0.827	0.466	0.881	0.473	0.717	0.396
Electricity	96	0.132	0.229	0.129	0.222	0.147	0.249	0.140	0.237	0.186	0.302	0.196	0.313	0.304	0.393	0.386	0.449	0.258	0.357
	192	0.147	0.243	0.147	0.240	0.167	0.269	0.153	0.249	0.197	0.311	0.211	0.324	0.327	0.417	0.386	0.443	0.266	0.368
	336	0.161	0.261	0.163	0.259	0.186	0.290	0.169	0.267	0.213	0.328	0.214	0.327	0.333	0.422	0.378	0.443	0.280	0.380
	720	0.196	0.294	0.197	0.290	0.243	0.340	0.203	0.301	0.233	0.344	0.236	0.342	0.351	0.427	0.376	0.445	0.283	0.376
ETTh1	96	0.375	0.398	0.379	0.401	0.378	0.393	0.375	0.399	0.376	0.415	0.435	0.446	0.941	0.769	0.664	0.612	0.878	0.740
	192	0.412	0.422	0.413	0.429	0.427	0.436	0.412	0.420	0.423	0.446	0.456	0.457	1.007	0.786	0.790	0.681	1.037	0.824
	336	0.435	0.433	0.435	0.436	0.458	0.484	0.439	0.443	0.444	0.462	0.486	0.487	1.038	0.784	0.891	0.738	1.238	0.932
	720	0.454	0.465	0.464	0.464	0.472	0.561	0.501	0.490	0.469	0.492	0.515	0.517	1.144	0.857	0.963	0.782	1.135	0.852
ETTh2	96	0.270	0.336	0.274	0.337	0.274	0.345	0.289	0.353	0.332	0.374	0.332	0.368	1.549	0.952	0.645	0.597	2.116	1.197
	192	0.332	0.380	0.338	0.376	0.353	0.401	0.383	0.418	0.407	0.446	0.426	0.434	3.792	1.542	0.788	0.683	4.315	1.635
	336	0.360	0.407	0.363	0.397	0.382	0.425	0.448	0.465	0.400	0.447	0.477	0.479	4.215	1.642	0.907	0.747	1.124	1.604
	720	0.419	0.451	0.393	0.430	0.625	0.557	0.605	0.551	0.412	0.469	0.453	0.490	3.656	1.619	0.963	0.783	3.188	1.540
ETTm1	96	0.306	0.349	0.293	0.346	0.302	0.350	0.299	0.343	0.326	0.390	0.510	0.492	0.626	0.560	0.543	0.510	0.600	0.546
	192	0.335	0.366	0.333	0.370	0.347	0.383	0.335	0.365	0.365	0.415	0.514	0.495	0.725	0.619	0.557	0.537	0.837	0.700
	336	0.364	0.384	0.369	0.392	0.369	0.402	0.369	0.386	0.392	0.425	0.510	0.492	1.005	0.741	0.754	0.655	1.124	0.832
	720	0.413	0.413	0.416	0.420	0.431	0.441	0.425	0.421	0.446	0.458	0.527	0.493	1.133	0.845	0.908	0.724	1.153	0.820
ETTm2	96	0.161	0.251	0.166	0.256	0.176	0.255	0.167	0.260	0.180	0.271	0.205	0.293	0.355	0.462	0.435	0.507	0.768	0.642
	192	0.215	0.289	0.223	0.296	0.245	0.305	0.224	0.303	0.252	0.318	0.278	0.336	0.595	0.586	0.730	0.673	0.989	0.757
	336	0.267	0.326	0.274	0.329	0.295	0.346	0.281	0.342	0.324	0.364	0.343	0.379	1.270	0.871	1.201	0.845	1.334	0.872
	720	0.352	0.383	0.362	0.385	0.401	0.413	0.397	0.421	0.410	0.420	0.414	0.419	3.001	1.267	3.625	1.451	3.048	1.328



Models		TiDE		TiDE (no res.)	
Electricity	96	0.132 ± 0.003	0.229 ± 0.001	0.136 ± 0.001	0.235 ± 0.002
	192	0.147 ± 0.003	0.243 ± 0.001	0.153 ± 0.001	0.253 ± 0.001
	336	0.161 ± 0.001	0.261 ± 0.002	0.172 ± 0.003	0.274 ± 0.002
	720	0.196 ± 0.002	0.294 ± 0.001	0.196 ± 0.003	0.295 ± 0.002

- 대부분 벤치마크에서 Transformer 계열 모델과 동등 혹은 우수한 성능
- 장기 horizon(특히 720)에서 큰 성능 향상
- M5 예측에서는 기존 모델 대비 큰 오차 감소
- 효율성
 - 추론·학습 속도 모두 Transformer보다 빠름
 - 긴 look-back에서도 메모리 안정성 높음
- Ablation
 - temporal decoder가 이벤트·공변량 영향 큰 데이터에서 빠른 적응
 - global residual 제거 시 성능 크게 악화 → 구조적 중요성 확인됨

6. Insight

- attention 없이도 강력한 장기 예측 가능하다는 점을 실증
- MLP 기반 구조의 재평가를 촉발하는 연구임
- Transformer 대비 표현력 제한 가능
- 이론 분석 범위가 비선형 전체 모델까지 확장되진 않음
- 복잡한 계절·비선형 패턴을 세밀하게 모델링하는 데 한계 있을 수 있음
- 실무형 noisy·missing 데이터에서의 강건성 검증 필요
- 온라인·실시간 예측 시나리오 적용 연구 필요