

Deep Time Series Models: A Comprehensive Survey and Benchmarking

<https://arxiv.org/abs/2407.13278>

0. Introduction

- 시계열 데이터는 금융, 기상, 헬스케어, 산업 등 다양한 분야에서 핵심적
- 기존 통계 기반 모델은 장기 의존성, 비선형성, 다변량 관계를 잘 처리하지 못함
- 최근 심층 학습 기반 시계열 모델들이 등장하며 예측 정확도 향상, 이상치 탐지, 결측치 보간 등 다양한 응용에서 효과를 보임
- 이 논문은 심층 시계열 모델들을 체계적으로 정리하고, 공정한 벤치마크 실험을 통해 비교함

1. Overview

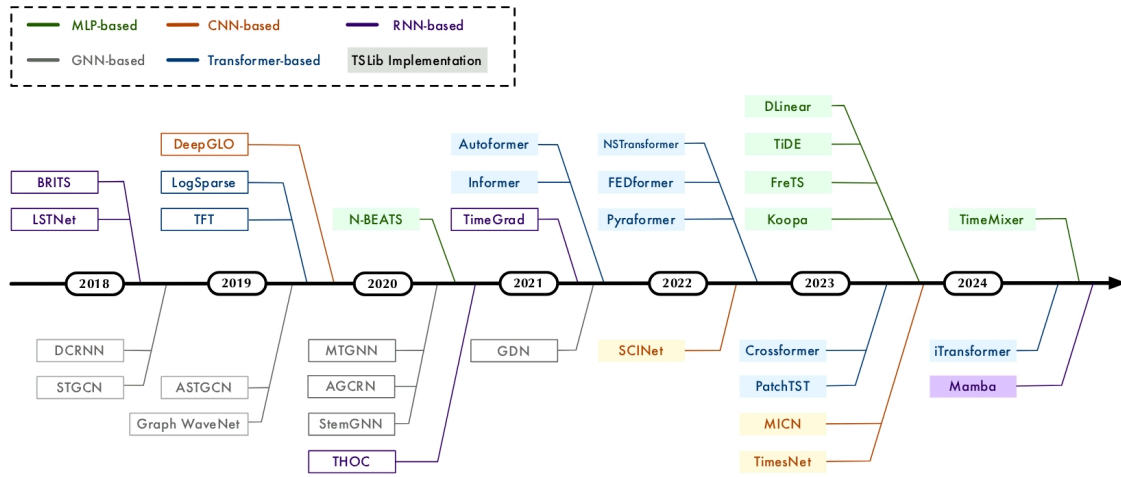
- 시계열 모델은 크게 RNN 계열, CNN 계열, Transformer 계열, 혼합 모델로 분류됨
- RNN 계열 : LSTM, GRU 등 순차적 데이터 학습에 강점 있으나 긴 시계열에서는 학습 어려움
- CNN 계열 : 지역적 패턴 포착에 유리, 연속적 특징 학습에 강점 있음
- Transformer 계열 : 장기 의존성 학습에 탁월, 병렬 연산 가능, 최근 시계열 예측에서 두각
- 혼합 모델 : CNN+RNN, Transformer+CNN 등 다양한 구조를 결합해 성능 개선 시도함

2. Challenges

- 시계열 데이터의 특성상 불규칙 샘플링, 결측치, 이상치가 많음

- 다변량 시계열에서는 변수 간 상관관계 학습 필요
- 장기 의존성 학습과 계산 비용 간 trade-off 문제 존재
- 다양한 도메인 데이터에 범용적으로 적용 가능한 모델 개발 어려.

3. Method



Category	Method	Architecture	Embedding	Attention Mechanism
Vanilla	Transformer	Enc-Dec	Standard	FullAttention($\mathbf{Q}, \mathbf{K}, \mathbf{V}$) = $\text{Softmax}(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d}})$
Point-wise	LogSparse [152]	Dec-only	Standard	$\hat{\mathbf{Q}}, \hat{\mathbf{K}} = \text{CausalCov}(\mathbf{H})$ FullAttention($\hat{\mathbf{Q}}, \hat{\mathbf{K}}, \mathbf{V}$) = $\text{Softmax}(\frac{\hat{\mathbf{Q}}\hat{\mathbf{K}}^T}{\sqrt{d}})$
	Informer [21]	Enc-Dec	Standard	$\bar{\mathbf{Q}} = \text{Top}_u(\overline{\mathbf{M}}(\mathbf{q}_i, \mathbf{K}))$ ProbSparse-Attention($\mathbf{Q}, \mathbf{K}, \mathbf{V}$) = $\text{Softmax}(\frac{\bar{\mathbf{Q}}\mathbf{K}^T}{\sqrt{d}})\mathbf{V}$
	Pyraformer [153]	Enc-Dec	Standard	Pyramid-Attention($\mathbf{Q}, \mathbf{K}, \mathbf{V}$) = $\text{Masked}(\text{Softmax}(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d}})\mathbf{V})$
Patch-wise	Autoformer [22]	Enc-Dec	Standard	Auto-Correlation($\mathbf{Q}, \mathbf{K}, \mathbf{V}$) = $\sum_{i=1}^k \text{Roll}(\mathbf{V}, \tau_i) \hat{\mathbf{R}}_{\mathbf{Q}, \mathbf{K}}(\tau_i)$
	Crossformer [29]	Enc-Dec	Patch-Wise	FullAttention($\mathbf{Q}, \mathbf{K}, \mathbf{V}$) = $\text{Softmax}(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d}})$
	PatchTST [23]	Enc-only	Patch-Wise	FullAttention($\mathbf{Q}, \mathbf{K}, \mathbf{V}$) = $\text{Softmax}(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d}})$
Variate-wise	iTransformer [30]	Enc-only	Variate-Wise	FullAttention($\mathbf{Q}, \mathbf{K}, \mathbf{V}$) = $\text{Softmax}(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d}})$

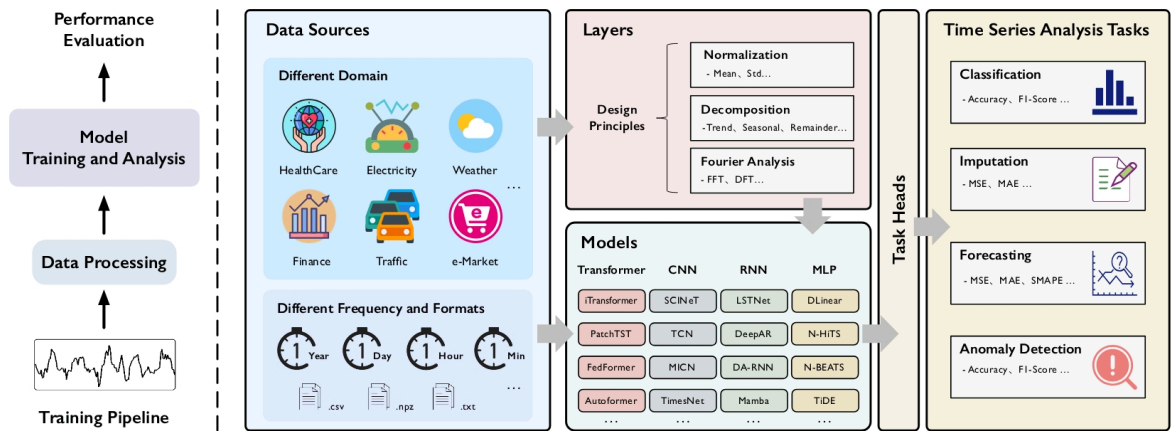
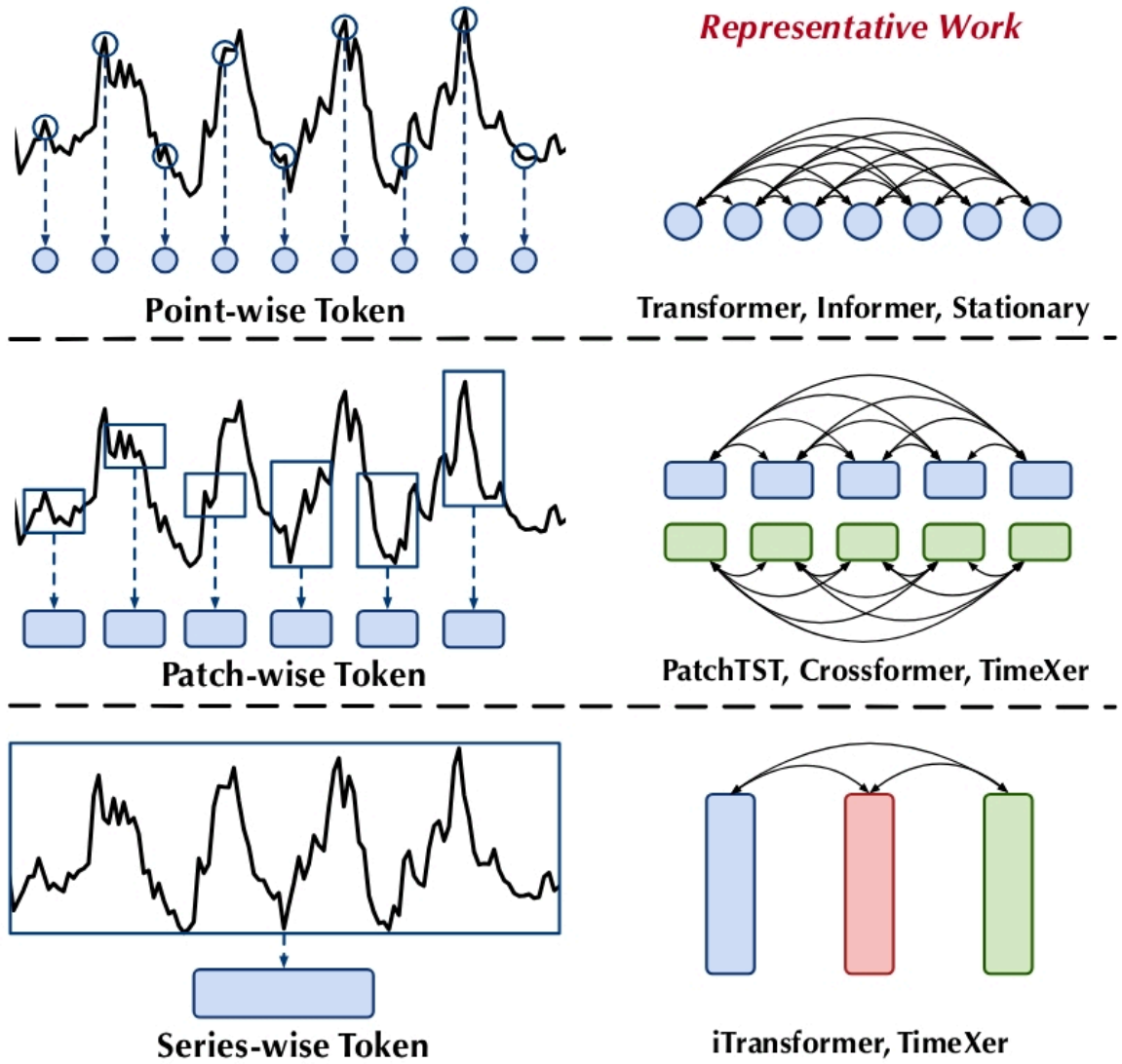


Fig. 9. Architecture and experiment pipeline of Time Series Library. Left: Unified training and evaluation process. Right: Overall Architecture.

- 심층 시계열 모델을 RNN, CNN, Transformer 기반으로 나눠 구조, 학습 방법, 특징을 정리함
- 주요 방법: LSTM, GRU, TCN, Informer, Autoformer, TimesNet 등
- 각 모델의 장단점과 연산 효율성을 비교함

4. Experiments

Task	Dataset	Dimension	Length	Domain	Size
Classification	EthanolConcentration	3	1,751	Alcohol Industry	20.3 MB
	FaceDetection	144	62	Face (250 Hz)	789.1 MB
	Handwriting	3	152	Motion	3.9 MB
	Heartbeat	61	405	Health (0.061 secs)	87.8 MB
	JapaneseVowels	12	29	Voice	1.1 MB
	PEMS-SF	963	144	Transportation (1 day)	420.1 MB
	SelfRegulationSCP1	6	896	Health (256 Hz)	17.8 MB
	SelfRegulationSCP2	7	1,152	Health (256 Hz)	17.7 MB
	SpokenArabicDigits	13	93	Voice (11025 Hz)	37.6 MB
	UWaveGestureLibrary	3	315	Gesture	3.4 MB
Imputation	ETTh1, ETTh2	7	17,420	Electricity (1 hour)	10.4 MB
	ETTh1, ETTh2	7	69,680	Electricity (15 mins)	2.6 MB
	Electricity	321	26,304	Electricity (1 hour)	95.6 MB
	Weather	21	52,696	Environment (10 mins)	7.2 MB
Long-term Forecasting	ETTh1, ETTh2	7	17,420	Electricity (1 hour)	10.4 MB
	ETTh1, ETTh2	7	69,680	Electricity (15 mins)	2.6 MB
	Electricity	321	26,304	Electricity (1 hour)	95.6 MB
	Weather	21	52,696	Environment (10 mins)	7.2 MB
	Traffic	862	17,544	Transportation (1 hour)	136.5 MB
	Exchange	8	7,588	Economic (1 day)	623 KB
	ILI	7	966	Health (1 week)	66 KB
Short-term Forecasting	M4-Yearly	1	6	Demographic	589.5 MB
	M4-Quarterly	1	8	Finance	
	M4-Monthly	1	18	Industry	
	M4-Weakly	1	13	Macro	
	M4-Daily	1	14	Micro	
	M4-Hourly	1	48	Other	
Anomaly Detection	SMD	38	100	Industry (1 min)	436.4 MB
	MSL	55	100	Industry (1 min)	58.2 MB
	SMAP	25	100	Industry (1 min)	113.0 MB
	SwaT	51	100	Industry (1 min)	903.2 MB
	PSM	25	100	Industry (1 min)	107.1 MB

- 벤치마크 데이터 :
 - 금융 : 주가, 거래량, 환율
 - 기상 : 온도, 강수량, 풍속
 - 헬스케어 : 심박수, 혈압, 환자 모니터링
 - 산업 : 센서 데이터, 기계 상태
 - 전력/에너지 : 전력 소비량, 에너지 생산량
- 작업 종류 : 시계열 예측, 분류, 결측치 보간, 이상 탐지
- 모든 모델을 동일한 환경과 지표로 평가

5. Results

- Transformer 계열 모델이 대부분의 데이터셋에서 높은 예측 정확도 기록
- TimesNet이 다중 작업에서 안정적 성능을 보여 범용 모델로 평가됨
- CNN 기반 모델은 지역적 패턴 학습에서 강점
- RNN 기반 모델은 단기 시계열 예측에서 경쟁력 있음
- 결측치 보간 및 이상 탐지에서도 Transformer와 혼합 모델이 우수
- 모델별 계산 효율성과 성능 간 trade-off 분석 제공

6. Insight

- 장기 의존성 학습과 다변량 시계열 처리 능력이 핵심 경쟁력
- 프리트레이닝과 대형 모델 활용 가능성 높음
- 외생 변수 통합, 이질적 데이터 처리 등 연구 필요
- 범용 시계열 모델 개발이 향후 연구 방향임