UN GEONOW 2024
首届联合国地信周

Data-Driven Decision Intelligence with Time Series and Spatio-Temporal Data

Prof. Bin Yang
School of Data Science and Engineering
East China Normal University

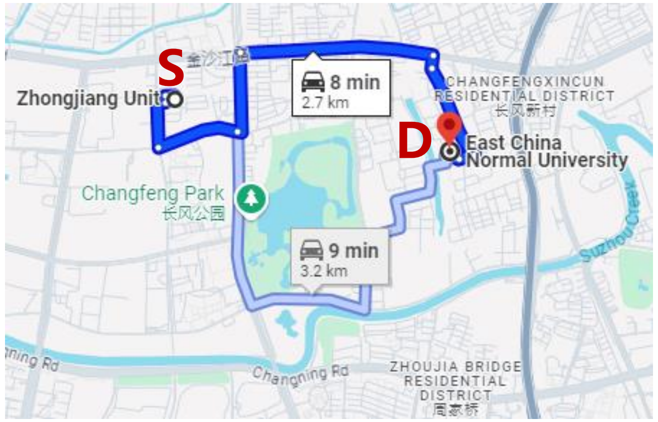
DaSE Data Science & Engineering

Outline

1. Motivating Examples
2. Data, Governance, Analytics, Decision
3. Automated Time Series Forecasting
4. Benchmarking Time Series Analytics
5. Conclusion

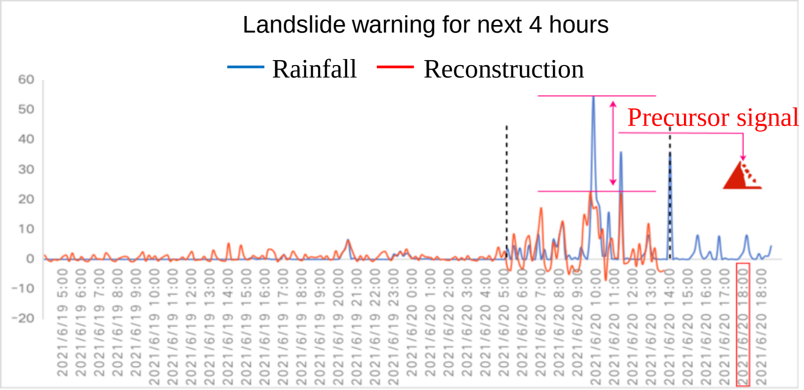
Data-driven Decision Intelligence

- Navigation from a source S to a destination D
  - Current and historical traffic flow — **Data**
  - Predicting traffic flows, and identifying congested and uncongested streets — **Analytics**
  - Selecting fastest/green paths; whether using highways, bridges vs. tunnels — **Decision**



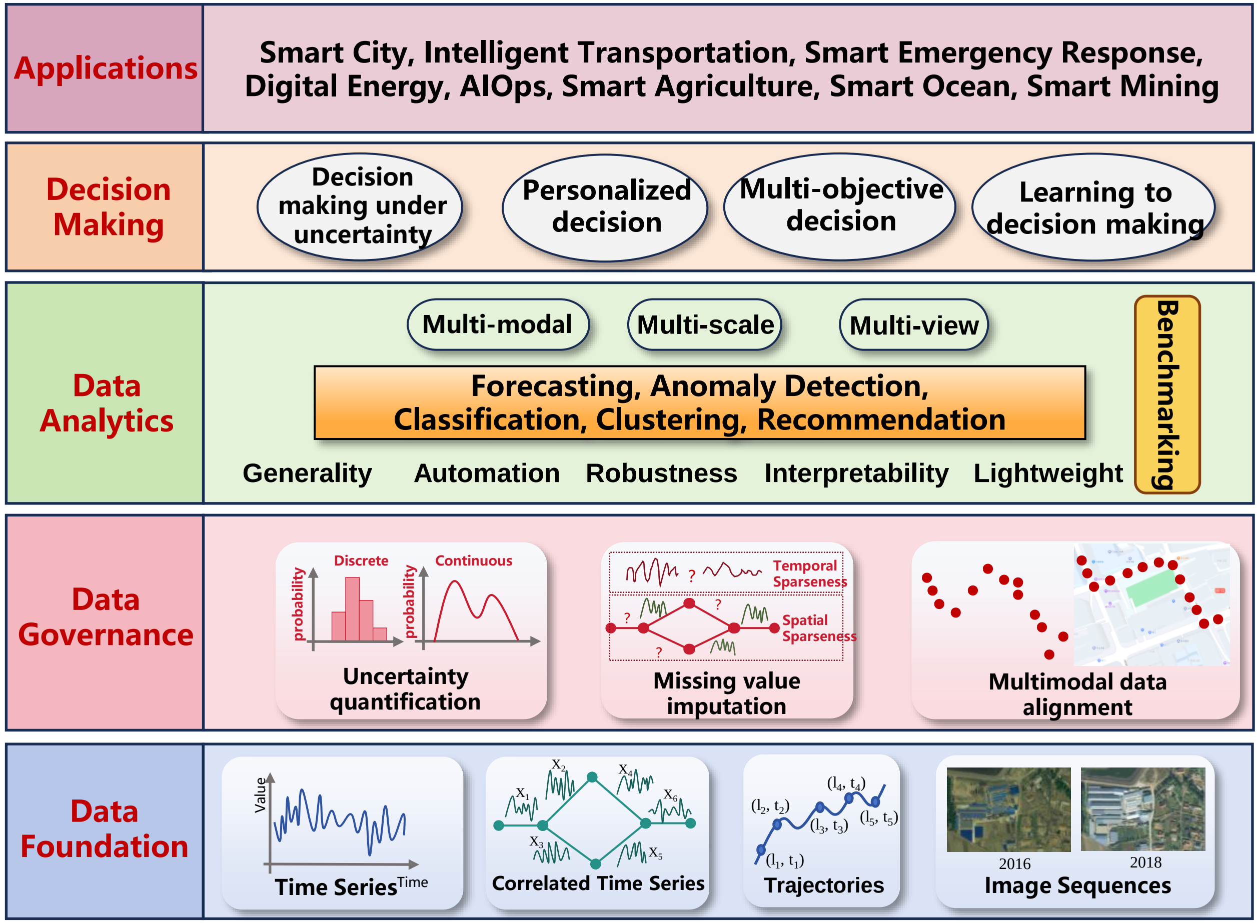
Spatio-temporal Data: Trajectories, Traffic Flows

- Geological disaster responses
  - Current and historical rainfall — **Data**
  - Predicting future rainfall, and whether geological disaster will happen in the future — **Analytics**
  - When to take what precautionary measures — **Decision**



Time series: Rainfall

Data - Governance - Analytics - Decision



Support various "Smart+" applications

Diverse data-driven decision-making strategies

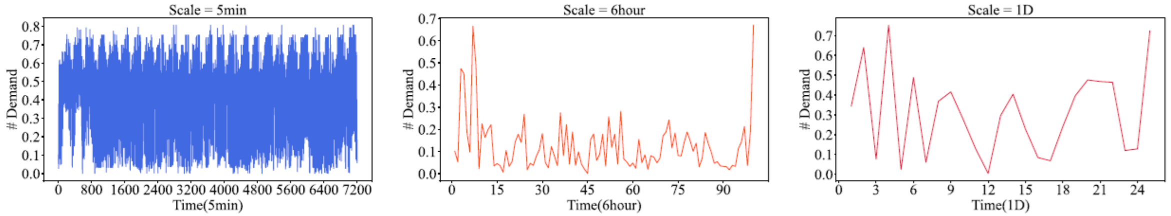
Spatio-temporal Foundational Model
Time Series Foundation Model

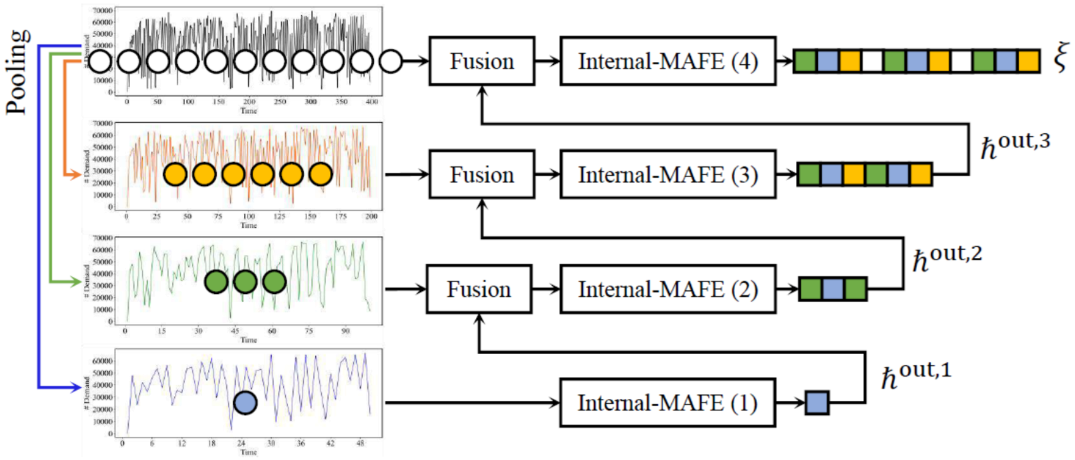
Improve data quality
Provide efficient data accesses

Two main dimensions: time and space
Multi-modal, multi-source heterogeneous data

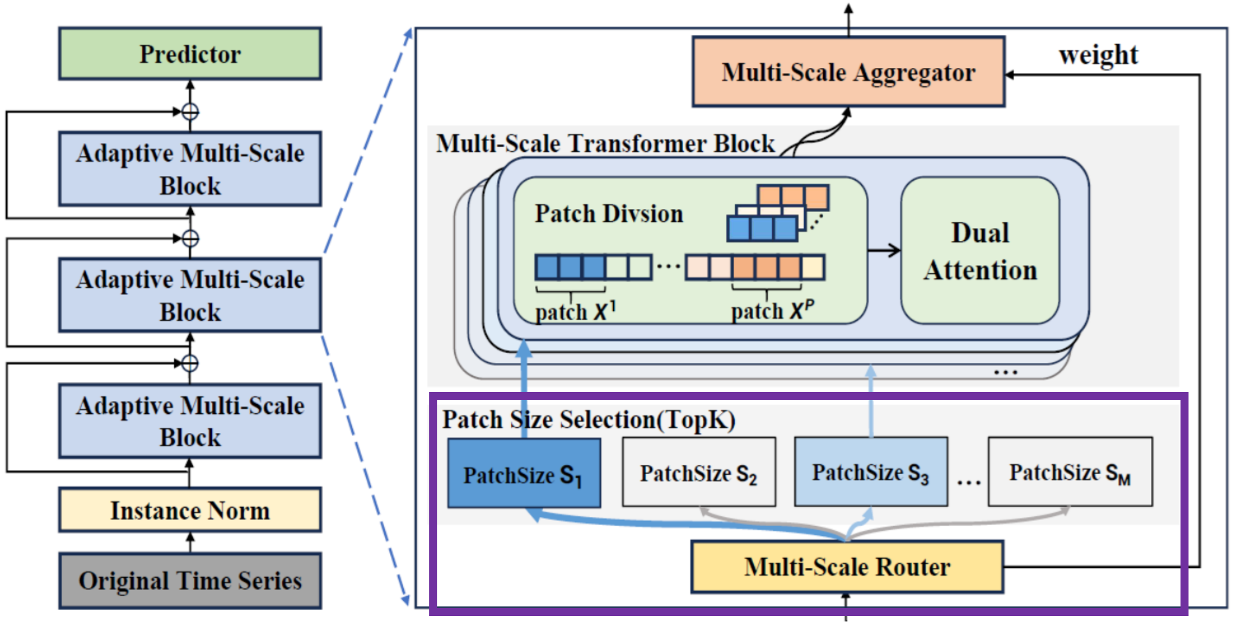
Data Foundation

- Multi-scale modeling
  - Temporal scales, spatial scales
  - Multi-scale modeling strategies
    - Multiple predefined scales
    - Adaptive selection of the most appropriate scales from multiple scales





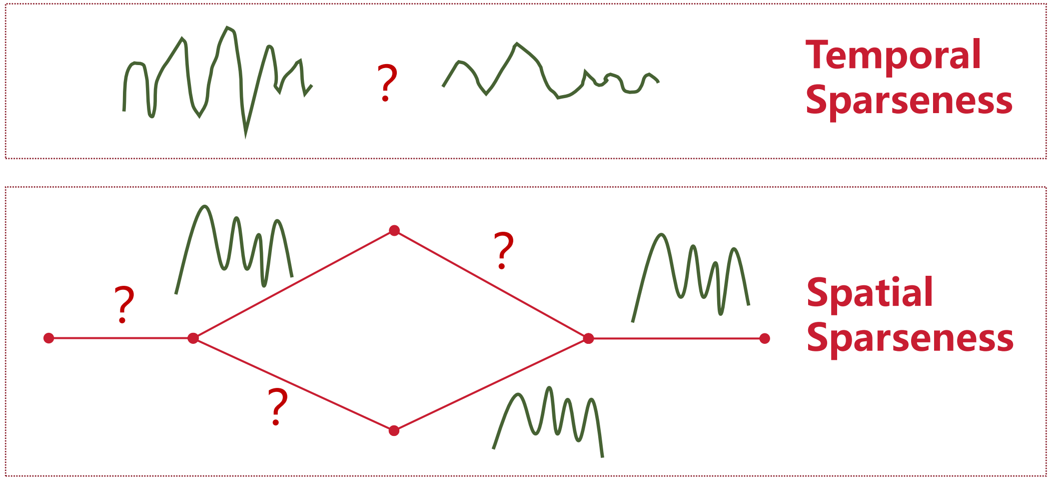
MagicScaler: Uncertainty-aware, Predictive Autoscaling. PVLDB 2023.

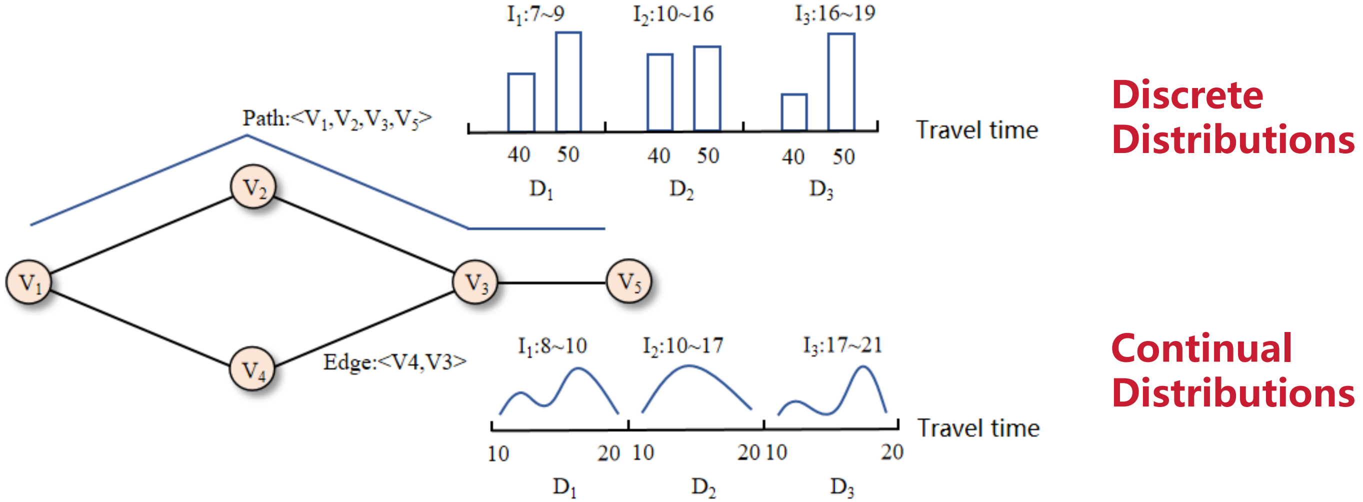


Pathformer: Multi-scale Transformers with Adaptive Pathways for Time Series Forecasting. ICLR 2024.

Data Governance

- Data governance prepares high-quality data and offers efficient data access methods for the subsequent data analytics
- Efficient data accesses: storage, query, indexing
- Improve data quality
  - Missing value imputation
  - Uncertainty quantification
  - Fusion of multimodal data





Stochastic Weight Completion for Road Networks Using Graph Convolutional Networks. ICDE 2019: 1274-1285
Stochastic Origin-Destination Matrix Forecasting Using Dual-Stage Graph Convolutional, Recurrent Neural Networks. ICDE 2020: 1417-1428
Enabling time-dependent uncertain eco-weights for road networks. GeoInformatica 21(1): 57-88 (2017)

Data Analytics

- Five types of analytics
  - Prediction, classification, clustering, anomaly detection, recommendation
- Five desired characteristics

Automation

- Traditional analytics models are designed by human experts, which is resource-intensive.
- Automated design of neural architectures and hyperparameters.

Generality

- Traditional analytics models are task-specifics, requiring large amounts of task-specific labels.
- General analytics models are pre-trained on abundant unlabeled data, which can be quickly adapt to different downstream tasks using limited task-specific labelled data.

Interpretability

- Deep learning's "black-box" nature hampers interpretability.
- Improve model interpretability, by incorporating prior knowledge into data-driven models, and offering post-hoc explainability metrics.

Robustness

- Spatio-temporal data often have noises and face distribution shifts.
- Strong robustness better deals with noisy data and streaming scenarios, so that data-driven decision-making schemes can be used in more real and complex scenarios.

Lightweight

- When decisions need to be made in resource–limited environments, it calls for lightweight models.
- Lightweight models (space/computation/power consumption), and real-time analytics and decision making on resource-constrained edge devices.

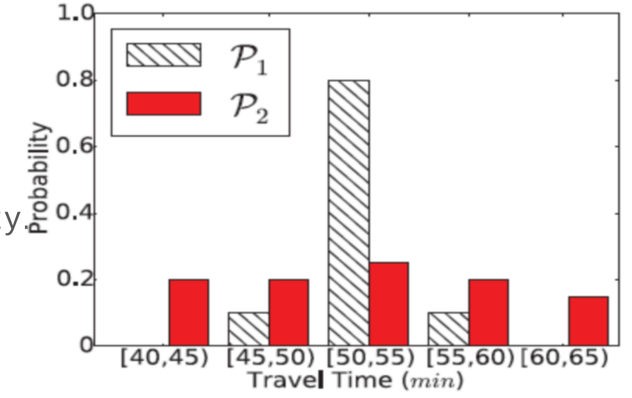
- OpenTS: A Comprehensive and Fair Benchmark for Time Series Analytics

OpenTS

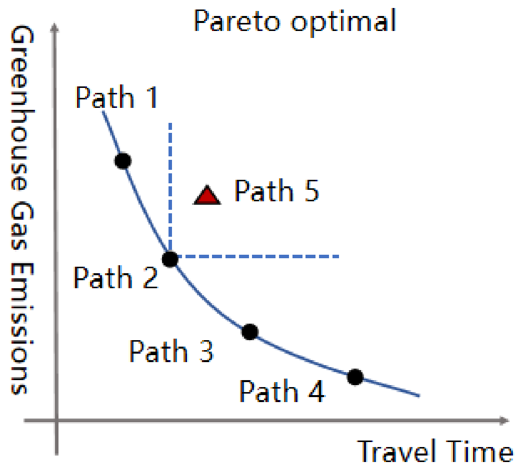
Data Science & Engineering

Data-Driven Decision Making

- Decision-making under uncertainty
  - Uncertainty introduces risks. Different decision-makers have varying preferences for risk.
    - Risk-Seeking, Risk-Averse, and Risk-Neutral
  - Risk preferences can be modeled with utility functions. Decisions aim to maximize expected utility.
- Multi-objective decision-making
  - Pareto optimal frontier
- Personalized decision-making
  - Learn personalized preferences from data, and make decisions based on these preferences
  - Preference for uncertainty (e.g., risk-seeking vs. risk-averse, customized utility function)
  - Preference for multi-objective trade-offs (e.g., 80% focus on travel time versus 20% on emissions)
- Learning to make decisions
  - Imitation learning from expert behavior (autonomous driving that mimics experienced drivers)

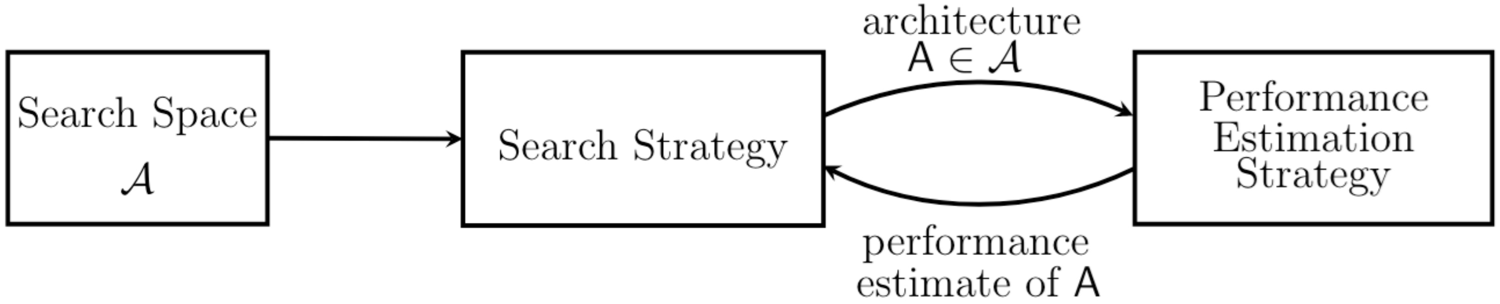


(a) $\mathcal{P}_1$ vs. $\mathcal{P}_2$



Risk-aware path selection with time-varying, uncertain travel costs: a time series approach. VLDB J. 27(2): 179-200 (2018)
Stochastic skyline route planning under time-varying uncertainty. ICDE 2014: 136-147
Toward personalized, context-aware routing. VLDB J. 24(2): 297-318 (2015)
Context-aware, preference-based vehicle routing. VLDB J. 29(5): 1149-1170 (2020)
Learning to Route with Sparse Trajectory Sets. ICDE 2018: 1073-1084

Automated Machine Learning for time series data



Neural Architecture Search: A Survey. J. Mach. Learn. Res. 20: 55:1-55:21 (2019)

- The basic AutoML includes:
  - Search Space A: Comprising all possible architectures.
  - Search Strategy: Selecting an architecture from space A.
  - Performance Evaluation Strategy: Estimating the performance of a given architecture.

- Challenges in time series data analysis include:
  - Lack of a dedicated search space for time series
  - Lack of a joint search space for both hyperparameters and neural architectures.
  - Lack of means to automatically select appropriate search spaces
  - Lack of highly-efficient performance estimation with zero-shot capabilities.

•AutoCTS: Automated Correlated Time Series Forecasting. Proc. VLDB Endow. 15(4): 971-983 (2021).
•AutoCTS+: Joint Neural Architecture and Hyperparameter Search for Correlated Time Series Forecasting. SIGMOD 2023.
•AutoCTS++: zero-shot joint neural architecture and hyperparameter search for correlated time series forecasting. The VLDB Journal 2024.

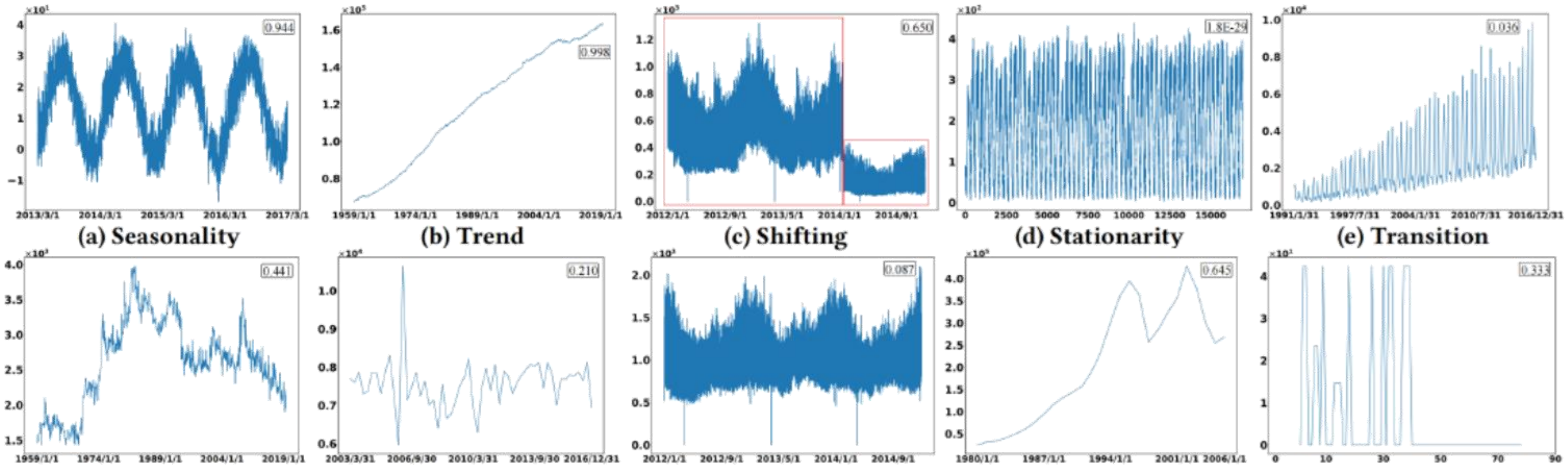
OpenTS: Benchmarking Time Series Analytics

- Time series analytics are key functionality in a wide variety of applications.
- More and more time series analytics methods are being proposed.
- Lacks well-recognized evaluation benchmarks
  - Help researchers and practitioners to better understand and compare the strengths and weaknesses of different algorithmic models， and determine the most suitable model for specific application scenarios.
  - Facilitate fair comparisons between newly proposed algorithms and existing ones
  - Support the testing of algorithm generalization across different datasets, avoid cherry-picking
  - Support for automation of time series analytics
- OpenTS time series benchmark
  - **TFB: Time series Forecasting Benchmark**
  - **TAB: Time series Anomaly detection Benchmark**
  - **FoundTS: Benchmarking of foundation models for time series forecasting**

OpenTS

OpenTS - TFB: A Fair and Comprehensive Benchmark for Time Series Forecasting

- It encompasses a variety of time series data characteristics
  - Seasonality, trend, shift, stationarity, transience, and correlation.
- 8,068 univariate time series
- 25 sets of multivariate time series (ranging from 5 to 2,000 dimensions)

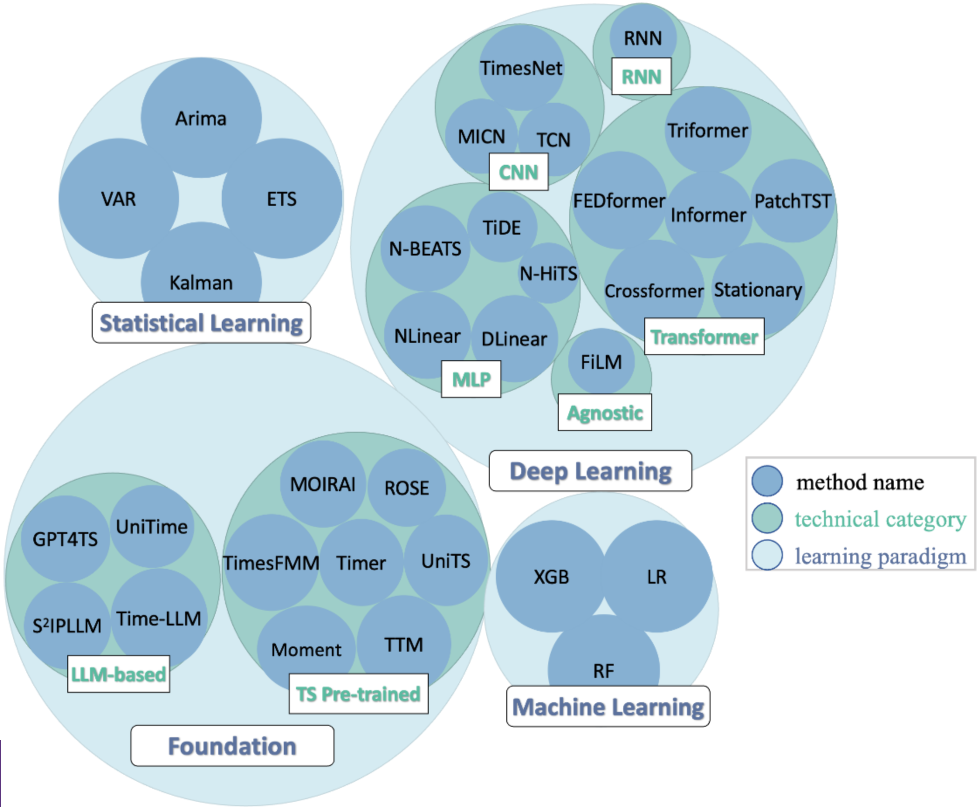


(a) Seasonality (b) Trend (c) Shifting (d) Stationarity (e) Transition
(f) Non-Seasonality (g) Non-Trend (h) Non-Shifting (i) Non-Stationarity (j) Non-Transition

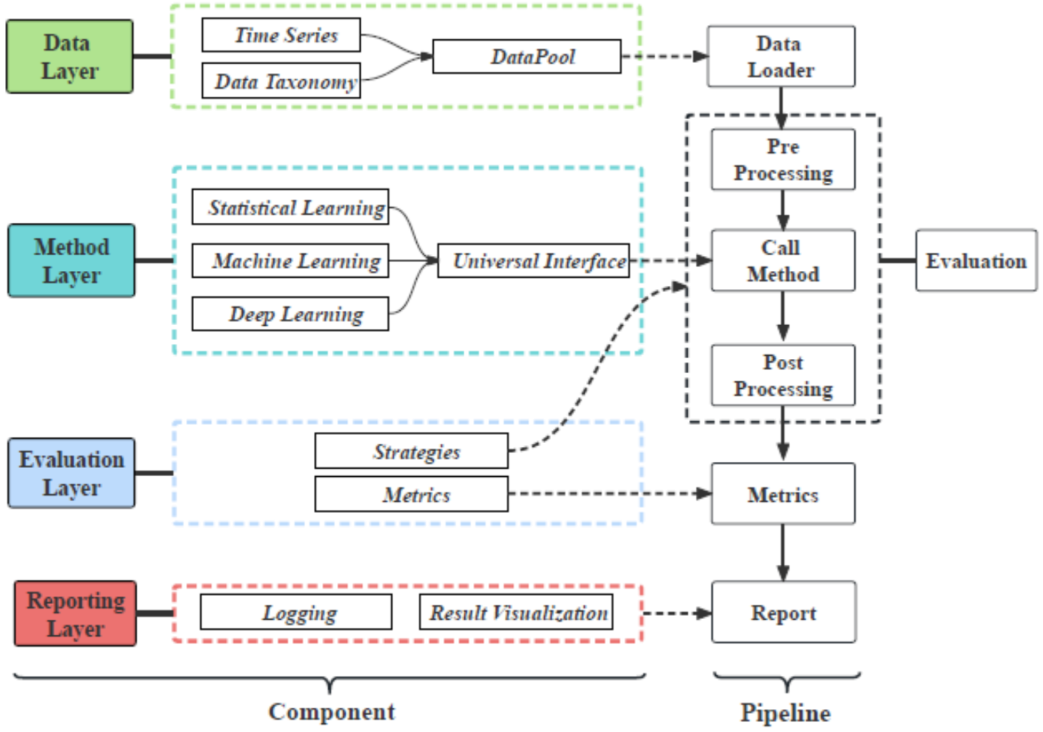
•Xiangfei Qiu, Jilin Hu, Lekui Zhou, Xingjian Wu, Junyang Du, Buang Zhang, Chenjuan Guo, Aoying Zhou, Christian S. Jensen, Zhenli Sheng, Bin Yang: TFB: Towards Comprehensive and Fair Benchmarking of Time Series Forecasting Methods. PVLDB 2024. (Best Paper Nomination)

OpenTS - TFB: A Fair and Comprehensive Benchmark for Time Series Forecasting

- Time series prediction algorithms
  - Statistical methods, machine learning, deep learning



- Unified test pipeline
  - Support for embedding third-party algorithm libraries



Data Science & Engineering

Results of Benchmark

UN GEONOW 2024
首届联合国地信周

https://decisionintelligence.github.io/OpenTS/

OpenTS

Leaderboard for multivariate time series forecasting

Metrics [all]
☑ MAE ☑ MSE

Datasets [all | profile1 | off]

☑**Traffic**
☑Traffic
☑PEMS04
☑PEMS-BAY
☑METR-LA
☑PEMS08

☑**Electricity**
☑Electricity
☑ETTm2
☑ETTm1
☑ETTh2
☑ETTh1

☑**Environment**
☑Weather
☑AQShunyi
☑AQWan

☑**Energy**
☑Solar
☑Wind

☑**Health**
☑ILI
☑Covid19

☑**Economic**
☑Exchange
☑FRED-MD

☑**Nature**
☑ZafNoo
☑CzeLan

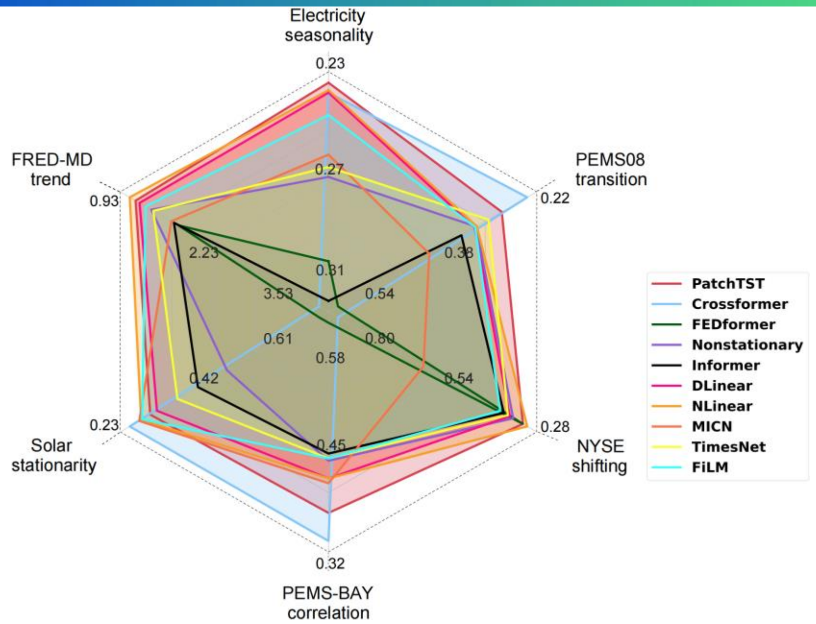
☑**Stock**
☑NASDAQ
☑NYSE

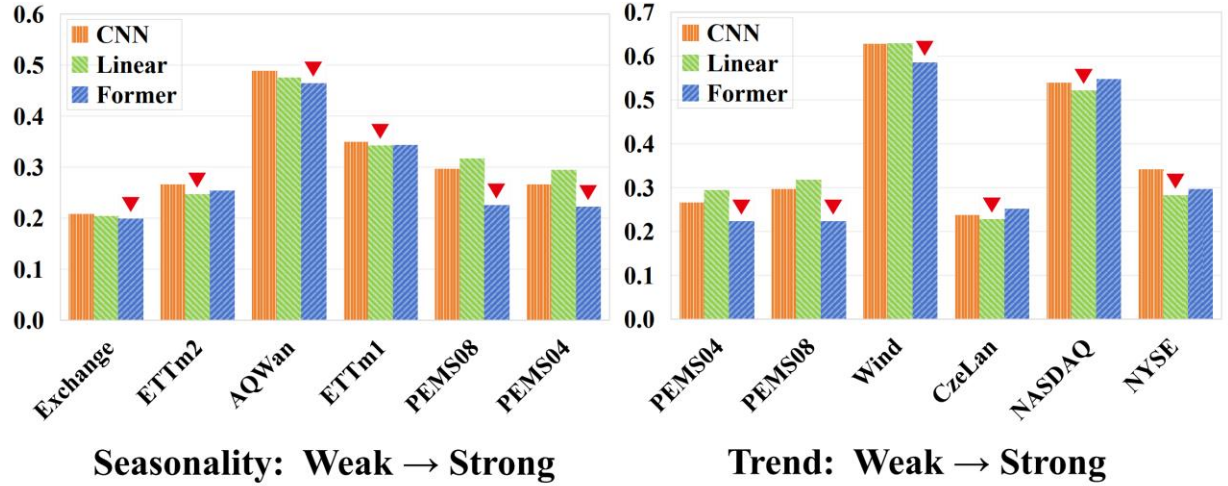
☑**Banking**
☑NN5

☑**Web**
☑Wike2000

Forecasting Horizons [all | off]
☑ 24 ☑ 36 ☑ 48 ☑ 60
☑ 96 ☑ 192 ☑ 336 ☑ 720

Rank	Model	Score	🥇	🥈	🥉	Paper	Publication	Year
1	Pathformer	50	50	14	10	paper	ICLR [bib]	2024
2	PatchTST	27	27	36	42	paper	ICLR [bib]	2023
3	NLinear	21	21	26	8	paper	AAAI [bib]	2023
4	iTransformer	19	19	18	15	paper	ICLR [bib]	2024
5	Crossformer	17	17	26	13	paper	ICLR [bib]	2023
6	Linear Regression	15	15	6	2	paper	Wiley [bib]	2005
7	TimeMixer	11	11	24	33	paper	ICLR [bib]	2024
8	DLinear	10	10	5	14	paper	AAAI [bib]	2023
9	FITS	9	9	8	7	paper	ICLR [bib]	2024
10	MICN	7	7	5	8	paper	ICLR [bib]	2022
11	VAR	4	4	0	1	paper	Economet Rev [bib]	1994
12	Informer	3	3	0	0	paper	AAAI [bib]	2021
13	TimesNet	3	3	11	13	paper	ICLR [bib]	2023
14	Triformer	2	2	7	9	paper	IJCAI [bib]	2022
15	FiLM	2	2	6	15	paper	NeurIPS [bib]	2022
16	FEDformer	0	0	6	7	paper	ICML [bib]	2022





(a) Seasonality (b) Trend

EAST CHINA NORMAL UNIVERSITY

DaSE Data Science & Engineering

Conclusions

- Data-driven Decision intelligence with time series and spatiotemporal data
- Data - Governance - Analysis - Decision
  - Data: Multi-scale, multi-modal spatio-temporal data
  - Governance: Enhancing data quality and facilitating efficient data accesses
  - Analytics: 5-5-1
    - Five types of analytics: forecasting, anomaly detection, clustering, classification, and recommendation
    - Five desired characteristics: automation, generality, robustness, interpretability, and lightweight
    - One comprehensive benchmark
  - Decision: various data-driven decision-making strategies.

专栏 中国计算机学会通讯 第20卷 第4期 2024年4月

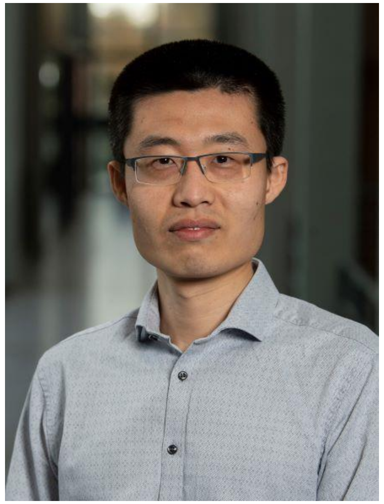
学术观点

时间序列和时空数据驱动的决策智能

杨彬 郭晨娟 胡吉林 树扬
华东师范大学

关键词：时间序列 时空数据 决策智能 时序时空决策智能

Acknowledgement



Bin Yang
Chair Professor

PhD@Fudan Uni.
Full professor,
Aalborg University
(AAU), Denmark

Chenjuan Guo
Professor

PhD@Mancheter Uni.
Associate professor,
AAU, Denmark

Jilin Hu
Professor

PhD@AAU
Associate professor,
AAU, Denmark

Yang Shu
Assistant Professor

PhD@Tinsghua

Christian S. Jensen
Academia Europaea
ACM/IEEE Fellow
Professor@AAU

We are hiring!

Decision Intelligence Lab
decisionintelligence.github.io

UN GEONOW 2024
首届联合国地信周

THANK YOU