AUTOML SUMMER SCHOOL 2024, HANNOVER

Chronos

Time series forecasting in the age of pretrained models

Oleksandr Shchur

Applied Scientist, AWS

aws

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Agenda

PART 1

Forecasting basics

- What is forecasting?
- Traditional approaches

PART 2

Chronos

- Pretrained models
- Benchmarking

PART 3

The way ahead

- Open research questions
- How does AutoML fit in?

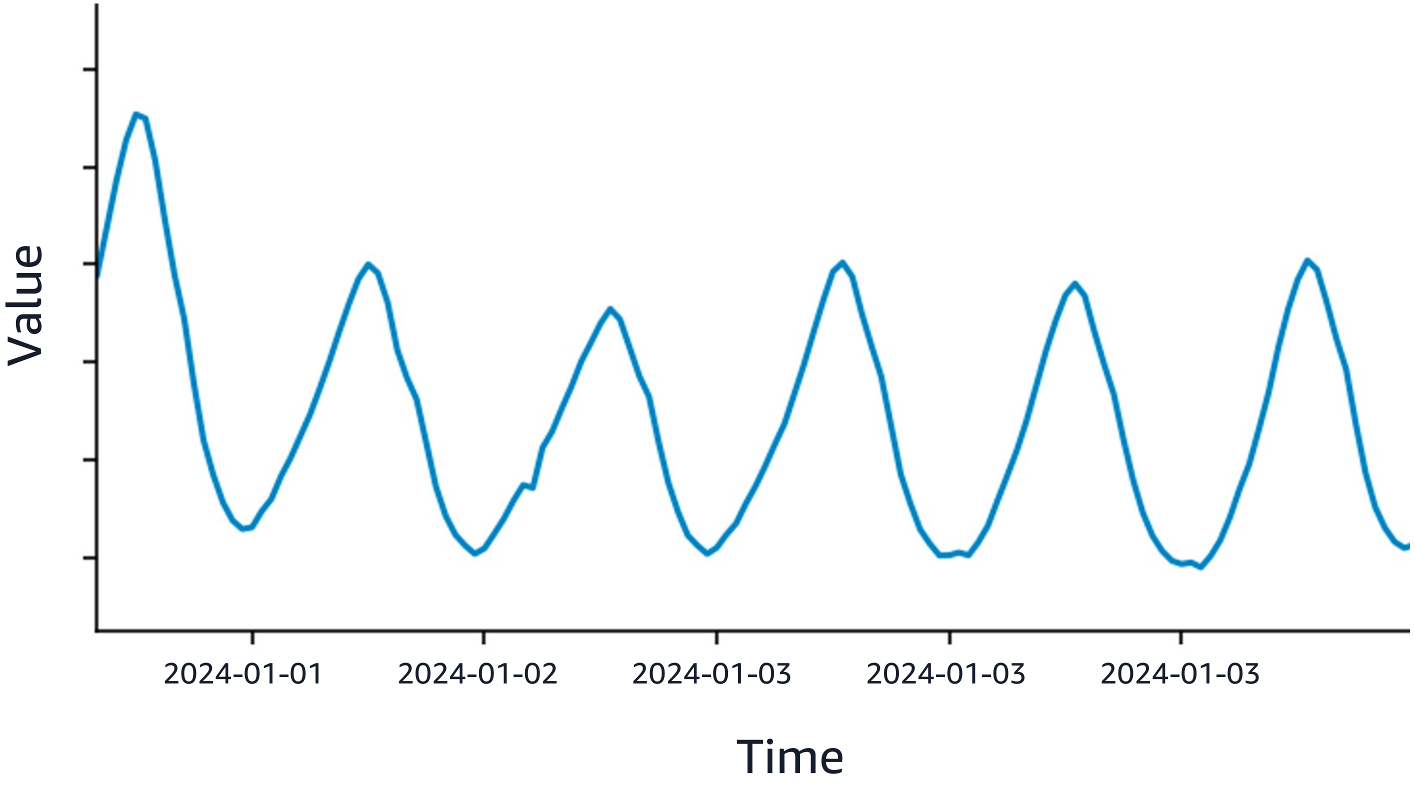
© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

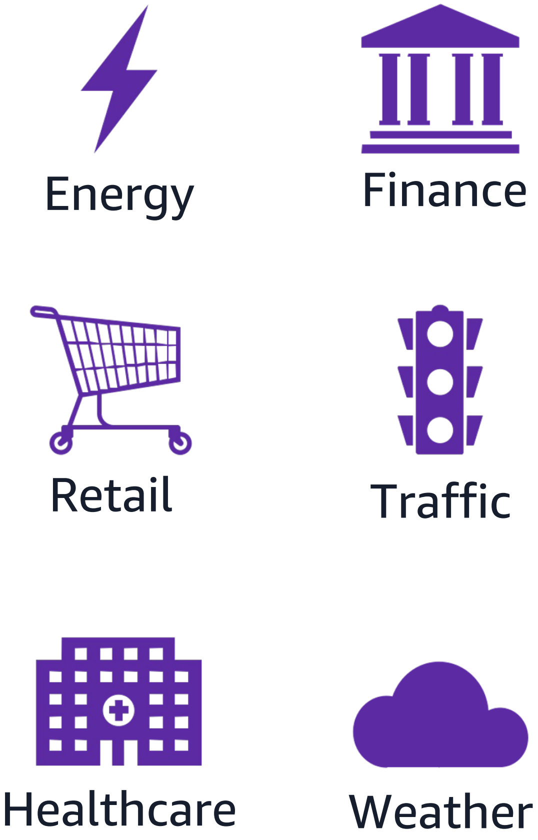
Forecasting basics

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Time series data

- Time series are measurements made at regular intervals





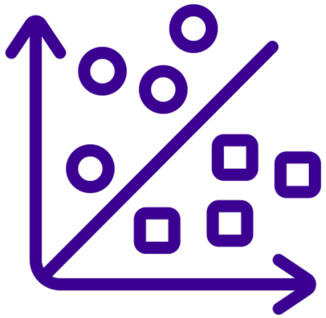
© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

ML tasks for time series data

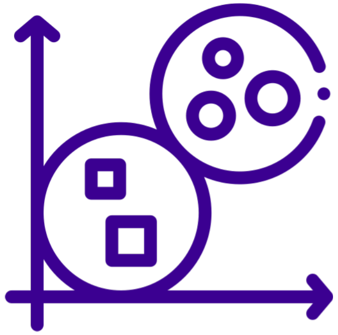
Forecasting

(focus of this talk)

Anomaly detection



Classification



Clustering

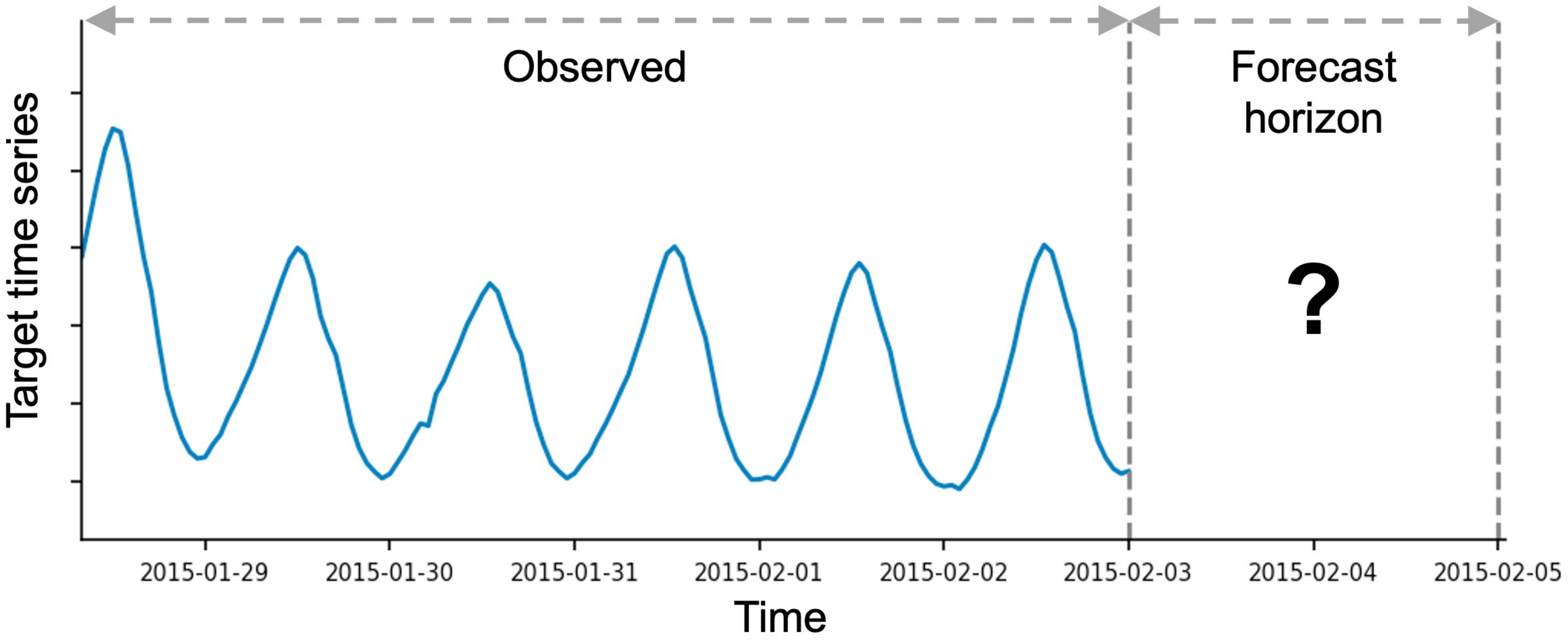
Imputation

[Icons made by Freepik/mpanicon/HAJICON from www.flaticon.com]

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Time series forecasting

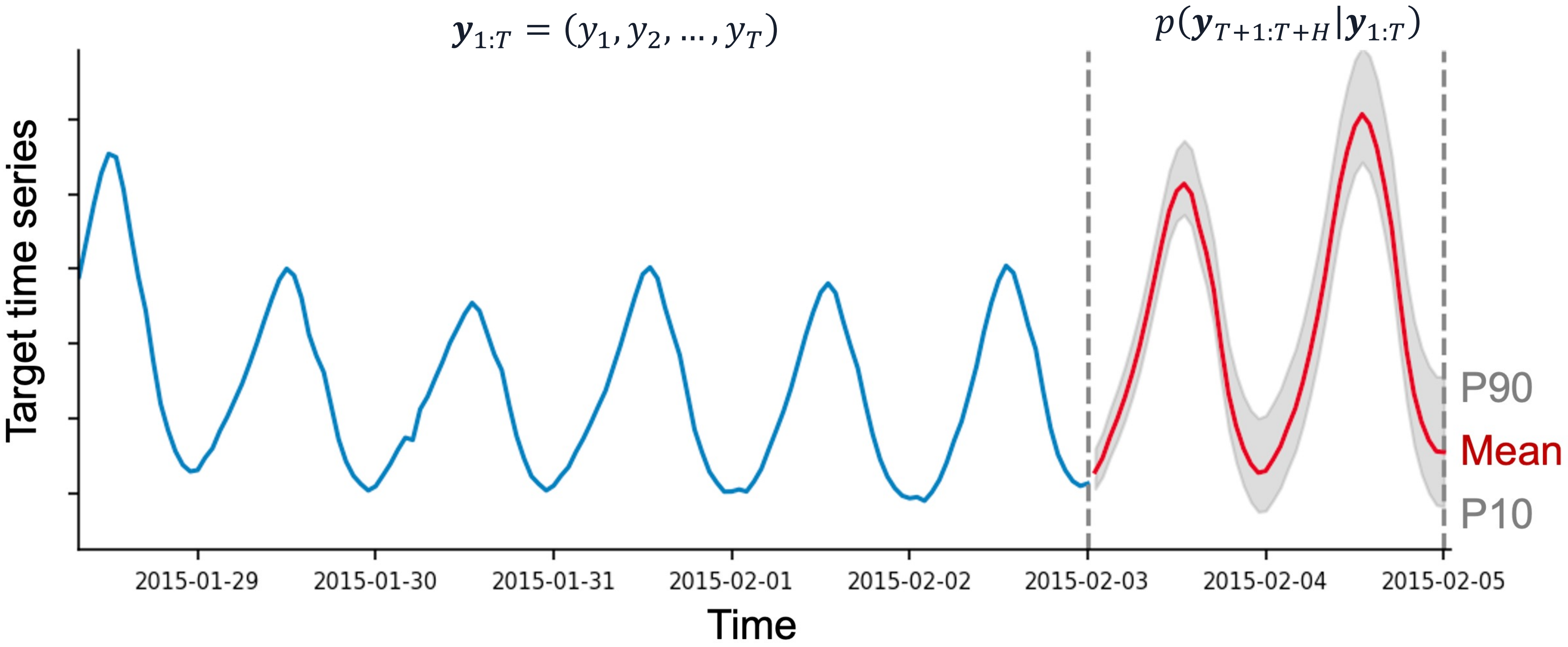
- What will happen in the future given the past?



© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Probabilistic forecasting

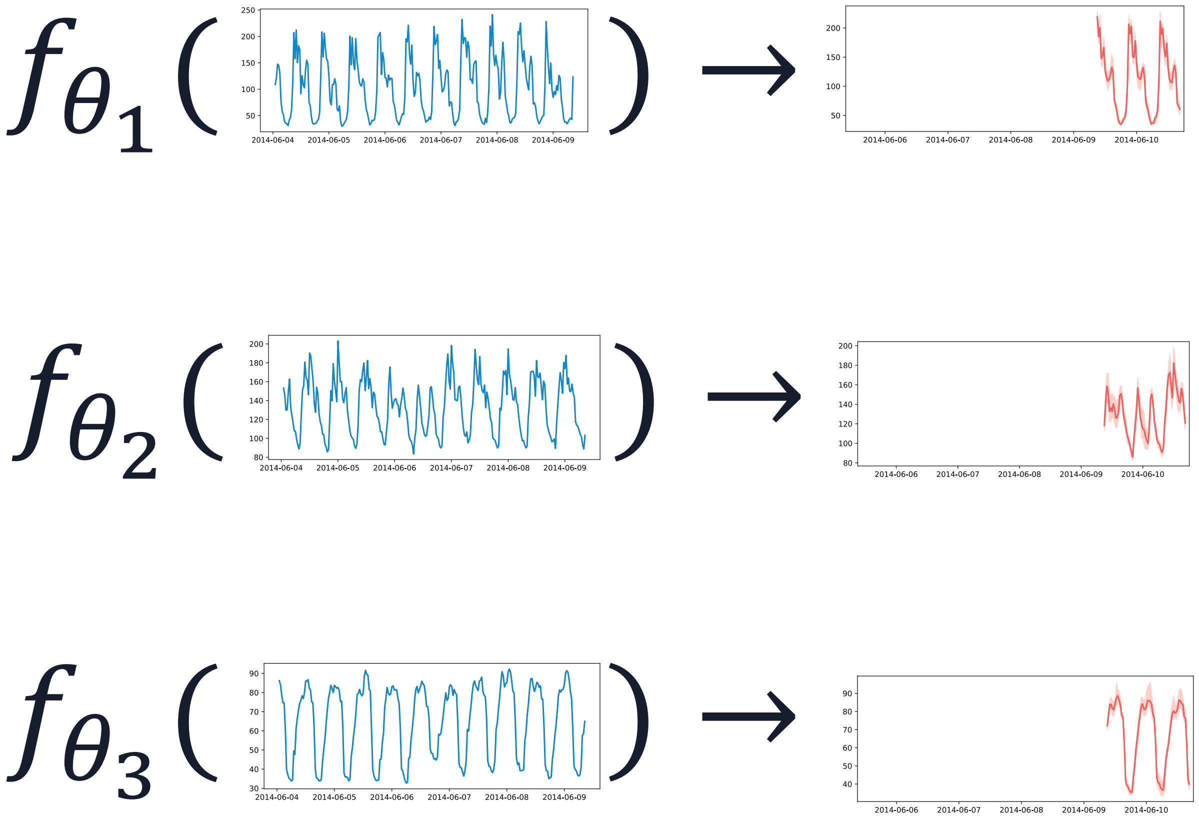
- Probabilistic forecast captures uncertainty in predictions



© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Local models

- Fit a separate model for each individual time series
- Examples: ARIMA, ETS, Theta



Strong baseline (esp. limited data)

Often interpretable

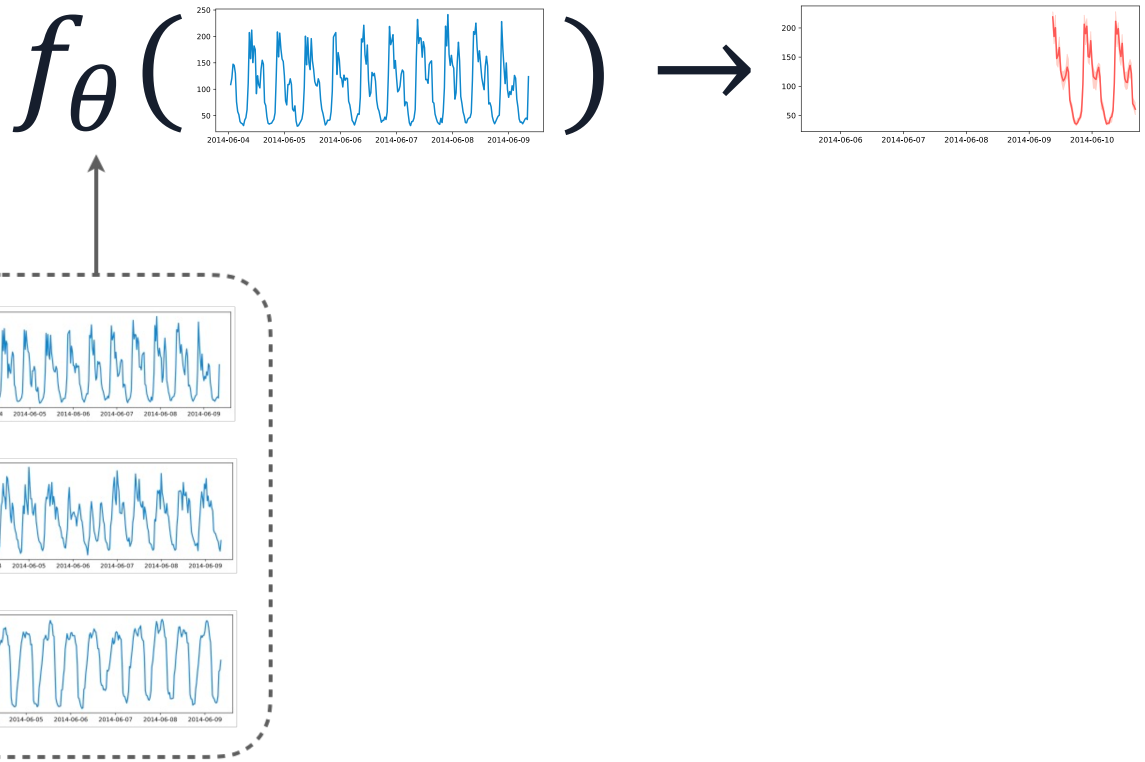
Low flexibility

Slow inference

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Global models

- Fit a single model for each task
- Examples: DeepAR, TFT, PatchTST



Training Corpus of
Related Time Series

High flexibility

Fast inference

Slow training

Data hungry

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Pretrained models

- Can we develop a single model that both
  - requires no dataset-specific training and
  - performs well on new time series tasks?

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

LLM-based forecasting models

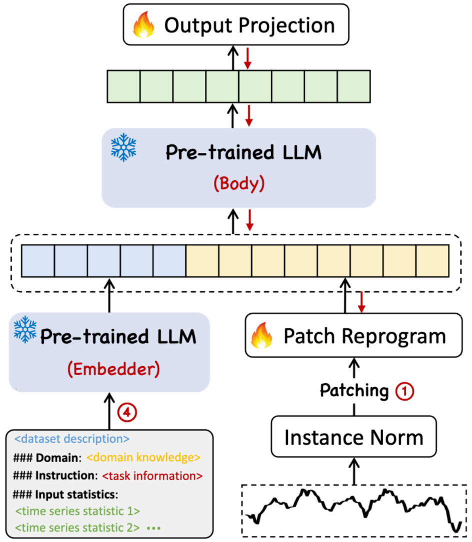
Text-based prompting ⟷ **Fine-tuning of pretrained LLMs**

Context: From August 16, 2019, Friday to August 30, 2019, Friday, the average temperature of region 110 was 78, 81, 83, 84, 84, 82, 83, 78, 77, 77, 74, 77, 78, 73, 76 degree on each day.
Question: What is the temperature going to be on August 31, 2019, Saturday?
Answer: The temperature will be 78 degree.

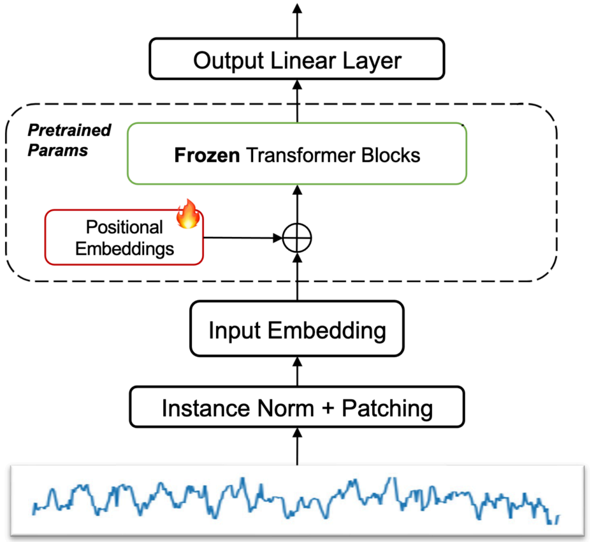
PromptCast

0.123, 1.23, 12.3, 123.0 → " 1 2 , 1 2 3 , 1 2 3 0 , 1 2 3 0 0"

LLMTime



TimeLLM



GPT4TS

❌ Extremely slow & expensive inference
Require task-specific prompt engineering & fine-tuning

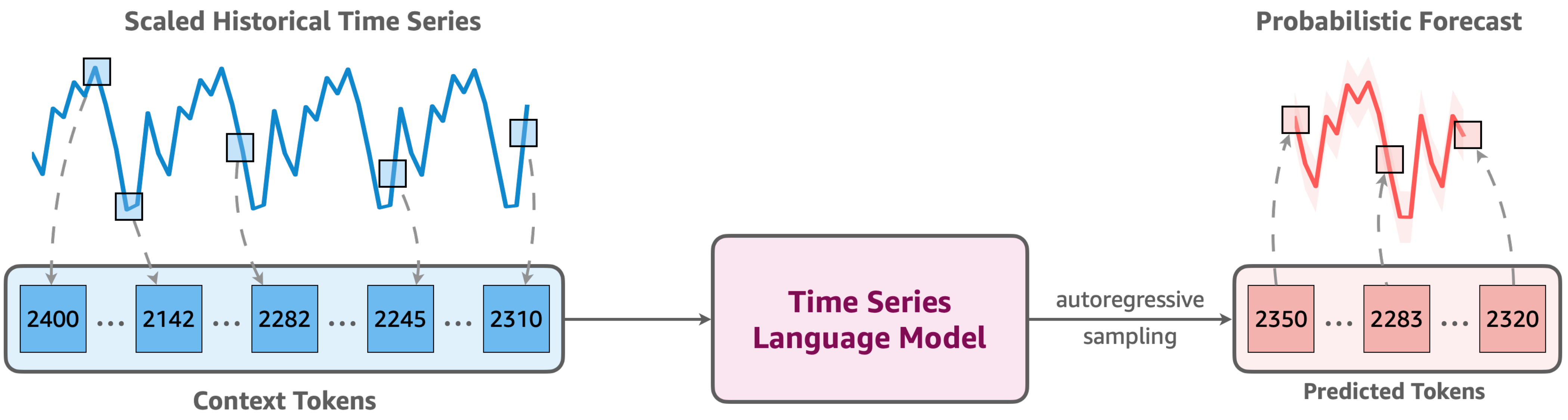
© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Chronos

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Introducing Chronos

- Main idea: Adapt LLM architectures for time series forecasting



© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Language modeling and forecasting

"Three Rings for the Elven-kings under the sky,
Seven for the Dwarf-lords in their halls of stone,
Nine for Mortal Men doomed to die,
One for the Dark Lord on his dark throne
In the Land of Mordor where the Shadows lie."

↓

Language Model

↓

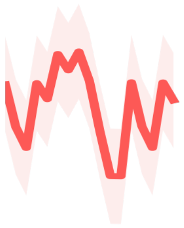
"One Ring to rule them all, One Ring to find them,
One Ring to bring them all and in the darkness bind them
In the Land of Mordor where the Shadows lie."

Predict the next sequence of words (tokens)

↓

Time Series Forecaster

↓



Predict future values conditioned on the past

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Time series tokenization

"Three Rings for the Elven-kings under the sky,
Seven for the Dwarf-lords in their halls of stone,
Nine for Mortal Men doomed to die,
One for the Dark Lord on his dark throne
In the Land of Mordor where the Shadows lie."

↓

Text Tokenizer

↓

" Three" " Ring" "s" " for" " the" " El" "ven" "-" "king" "s" …

Text language models have a discrete vocabulary

↓

?

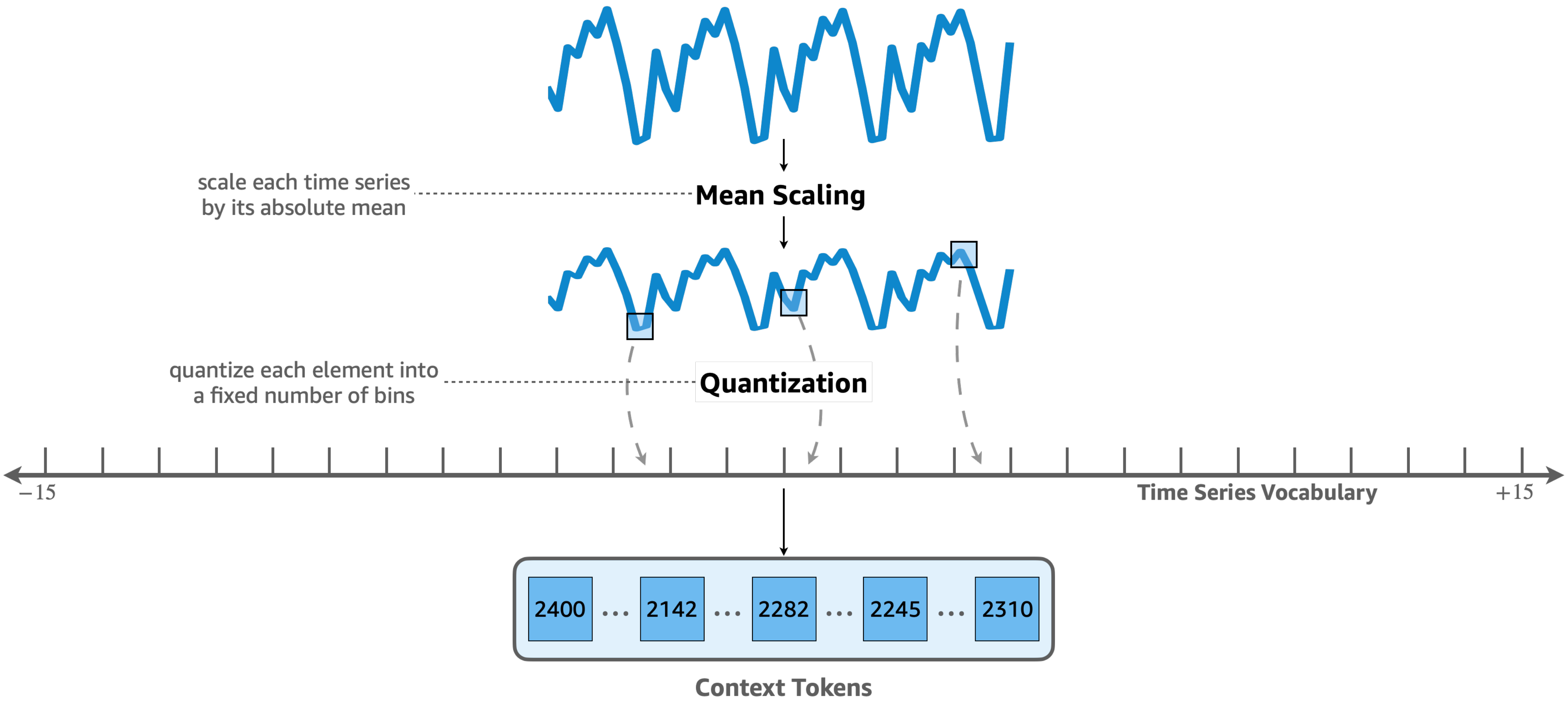
↓

?

Time series are real-valued signals

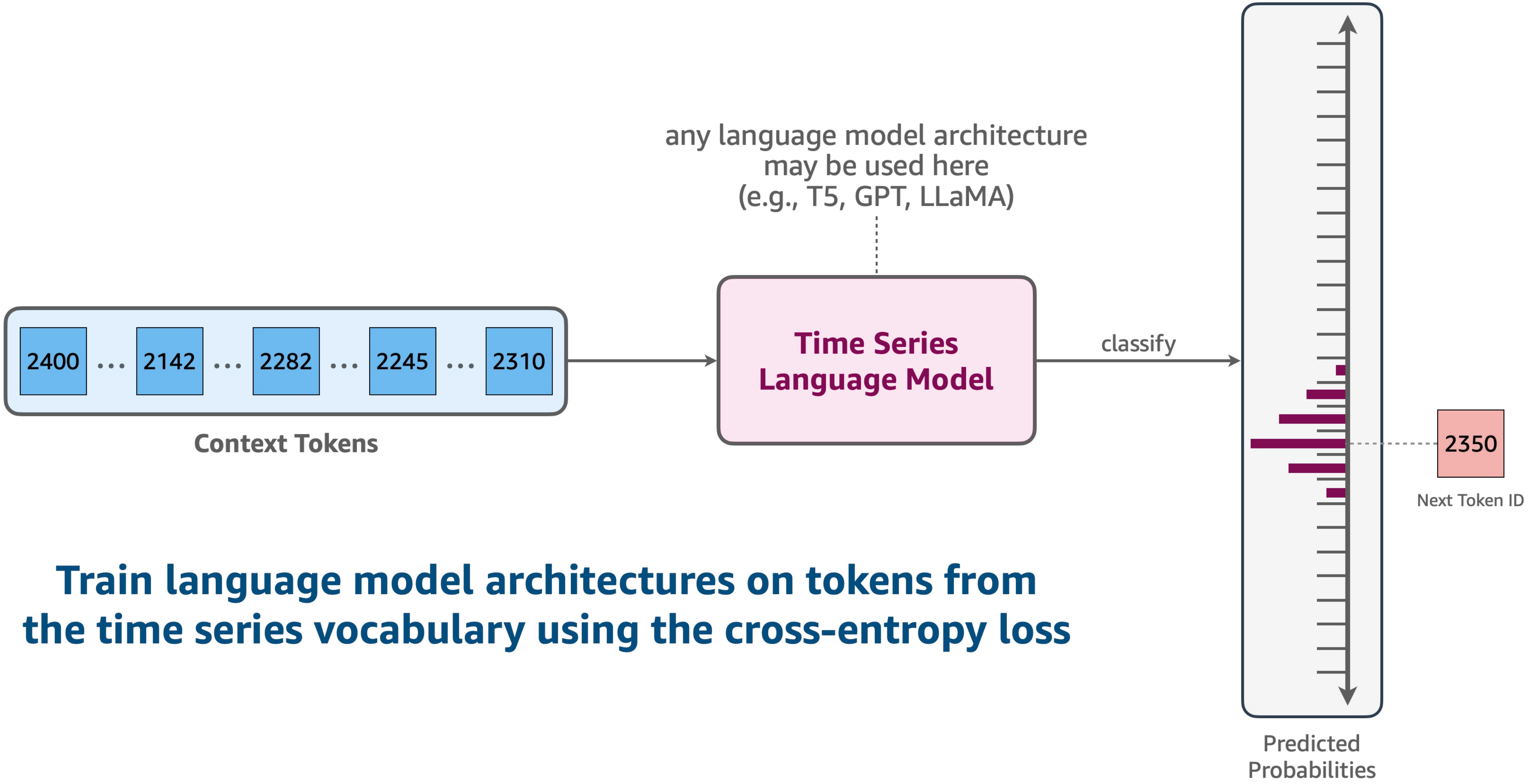
© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Time series tokenization



© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

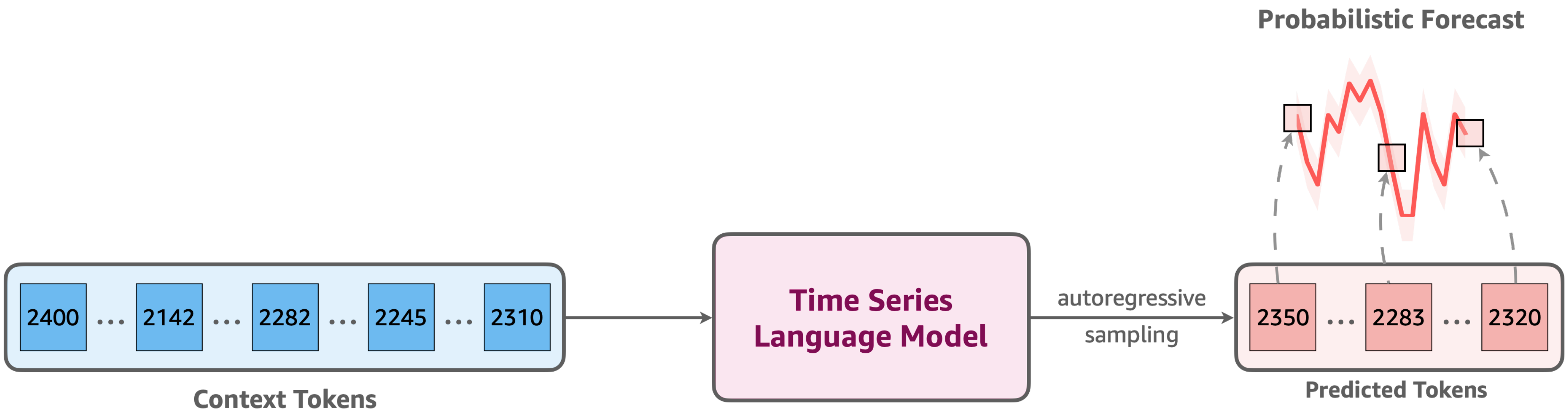
Regression via classification



Train language model architectures on tokens from the time series vocabulary using the cross-entropy loss

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

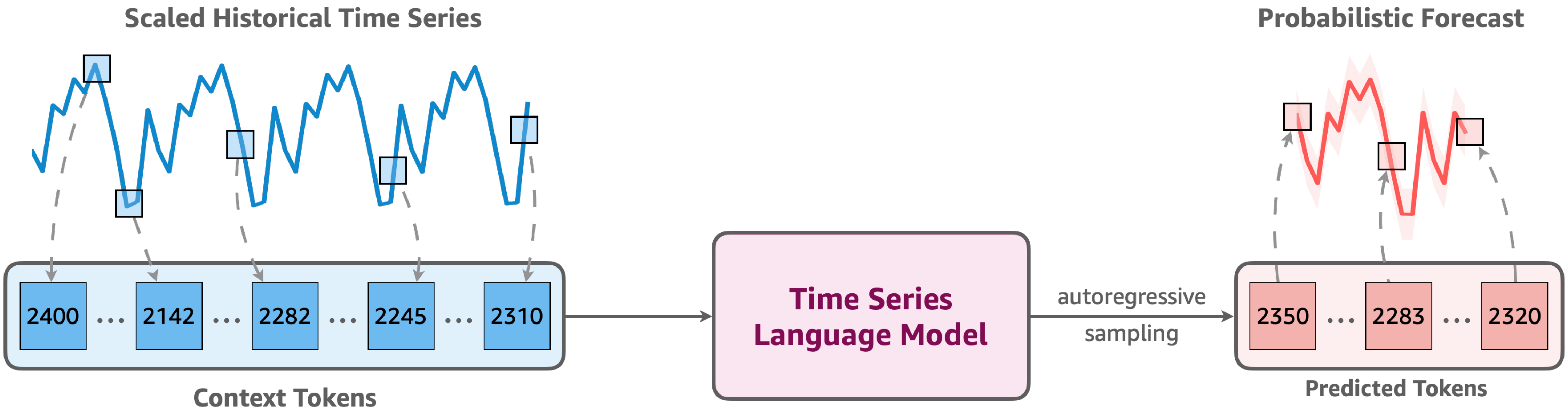
Sampling



© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

The complete Chronos framework

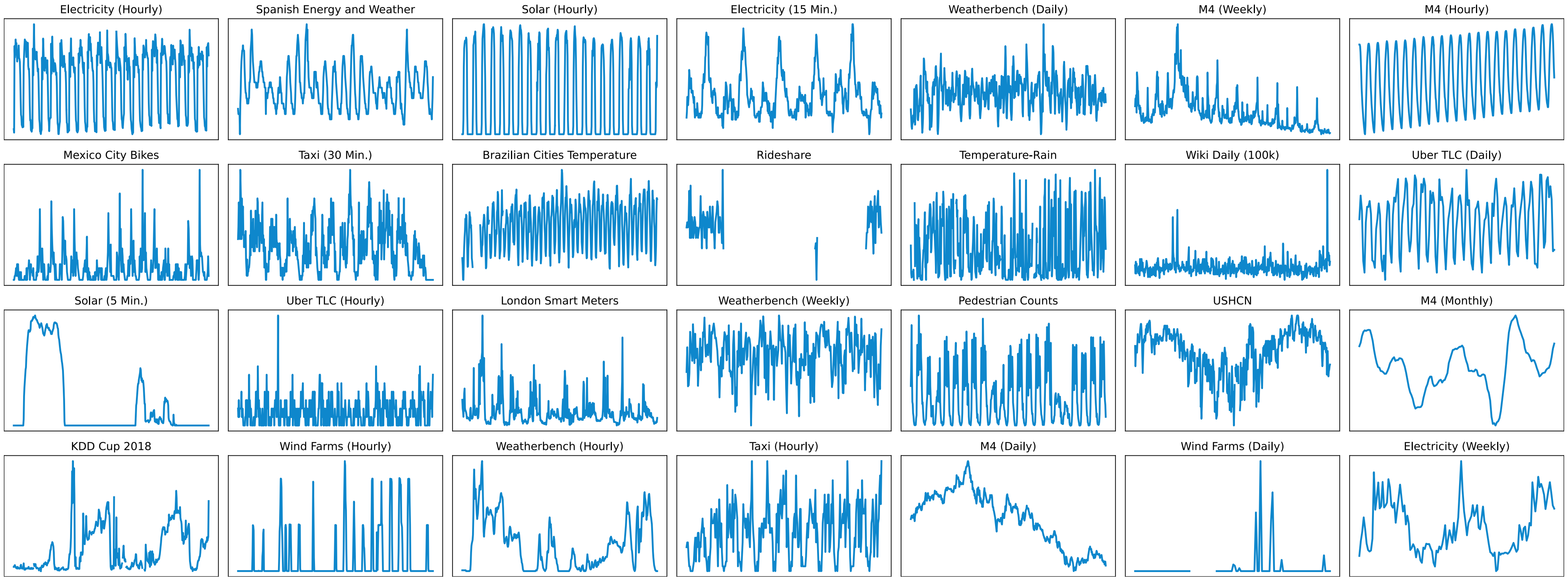
- Requires no changes to the language model architecture & training procedure
- Probabilistic by design



© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Training datasets

- 28 datasets from various domains and frequencies
- 890K time series with 84B observations



© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

TSMixup: Data augmentation for time series

Improve pattern diversity by mixing time series from different datasets

- Sample $K \sim \{1, 2, 3\}$ time series

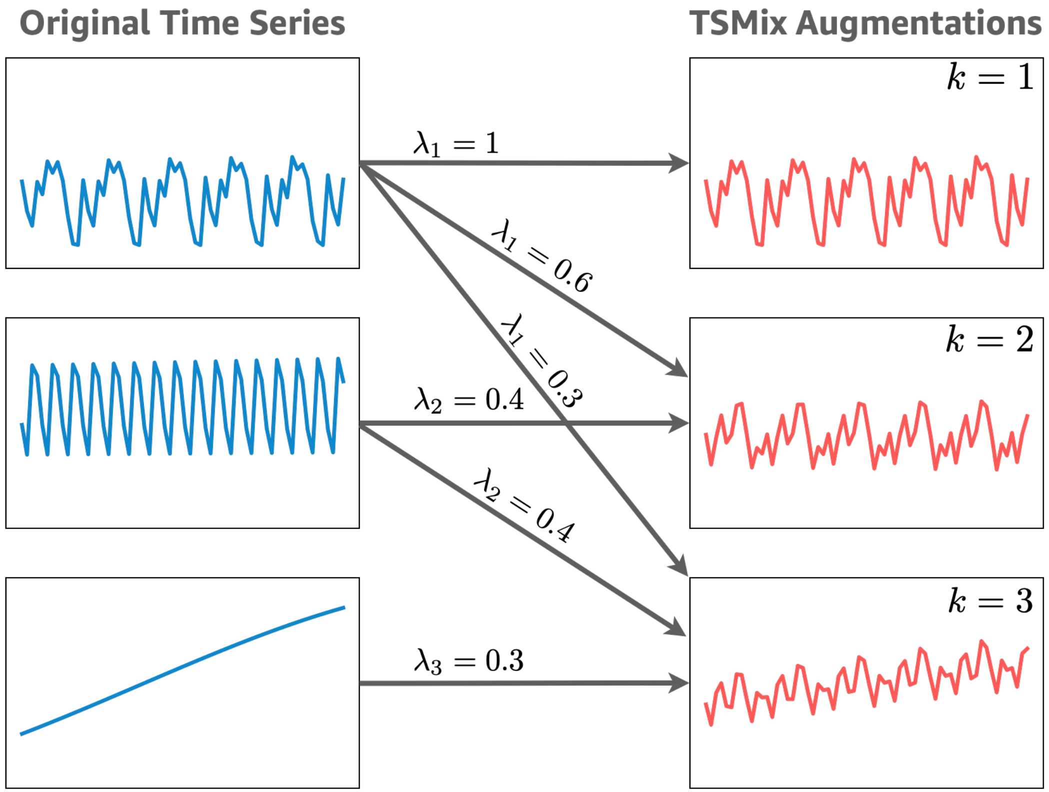
$$y_1, \dots, y_K \sim \mathcal{D}_{\text{train}}$$

- Sample weights

$$\lambda_1, \dots, \lambda_K \sim \text{Dirichlet}(\alpha)$$

- Combine time series

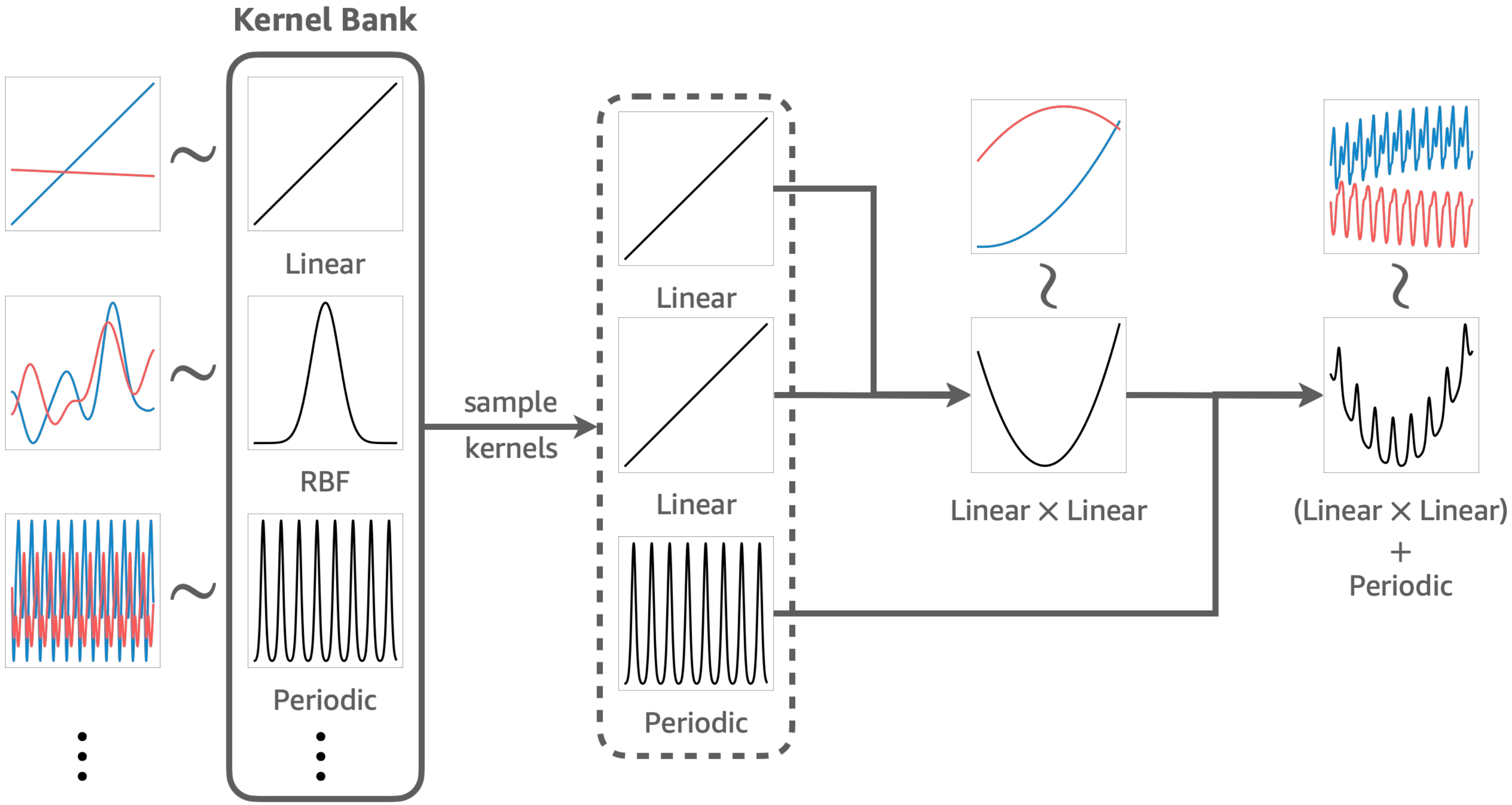
$$y = \sum_{k=1}^{K} \lambda_k y_k$$



© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

KernelSynth: Synthetic data generation

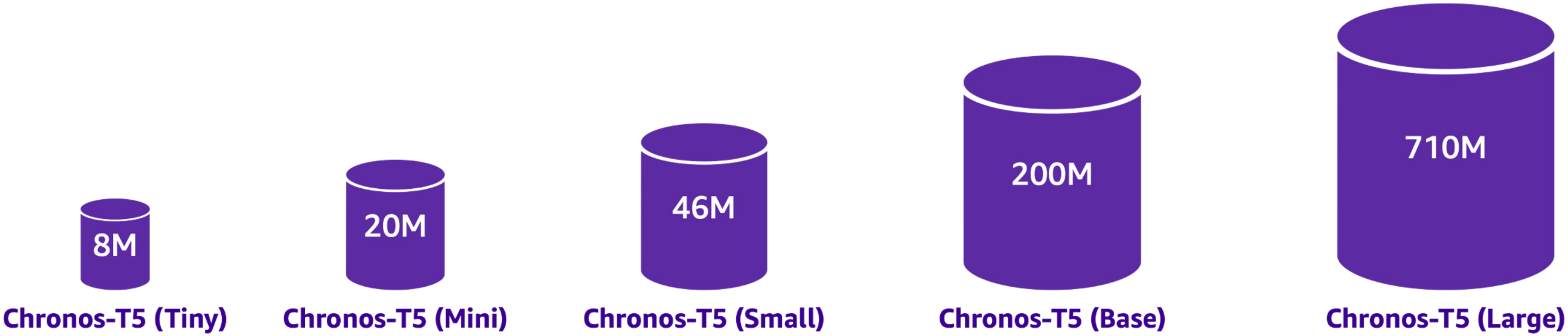
- Supplement real data with synthetic time series from Gaussian processes



© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

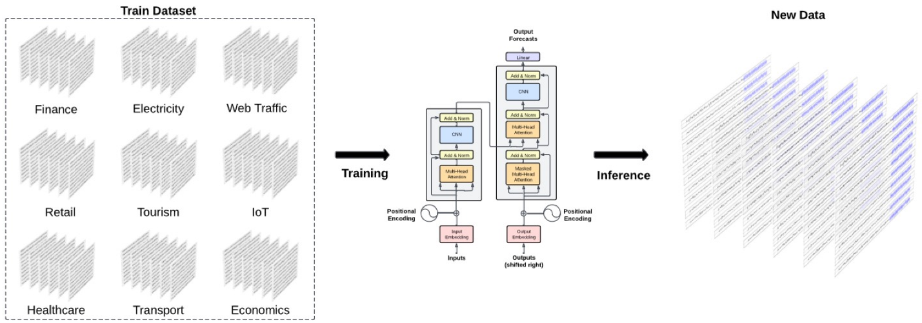
Chrons variants

Based on the T5 encoder-decoder architecture

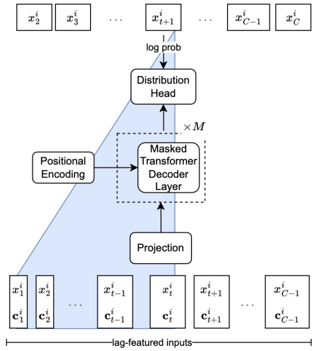


© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

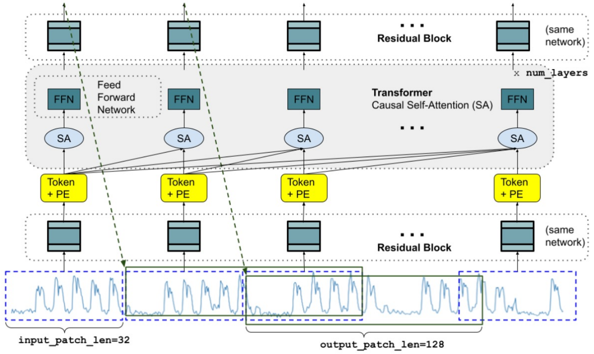
Other pretrained time series models



TimeGPT (Nixtla)

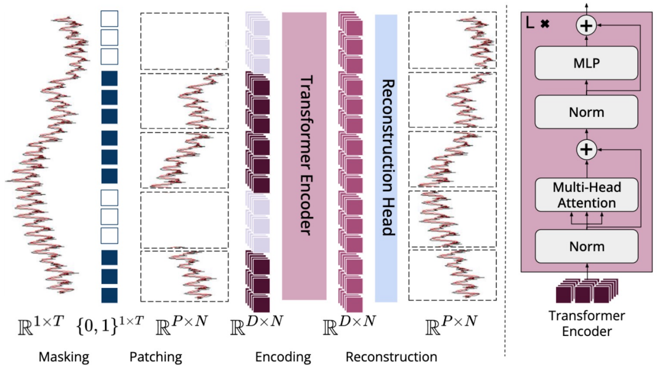


LagLlama



TimesFM (Google)

Moirai (Salesforce)



MOMENT (CMU)

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Baseline models

Pretrained models

Single pretrained model used across all tasks

- LLMTime
- ForecastPFN
- LagLlama
- Moirai
- TimesFM

Task-specific models

Separate model trained / fine-tuned for each task

- PatchTST
- DeepAR
- WaveNet
- TFT
- DLinear
- NBEATS
- NHiTS
- GPT4TS

Local models

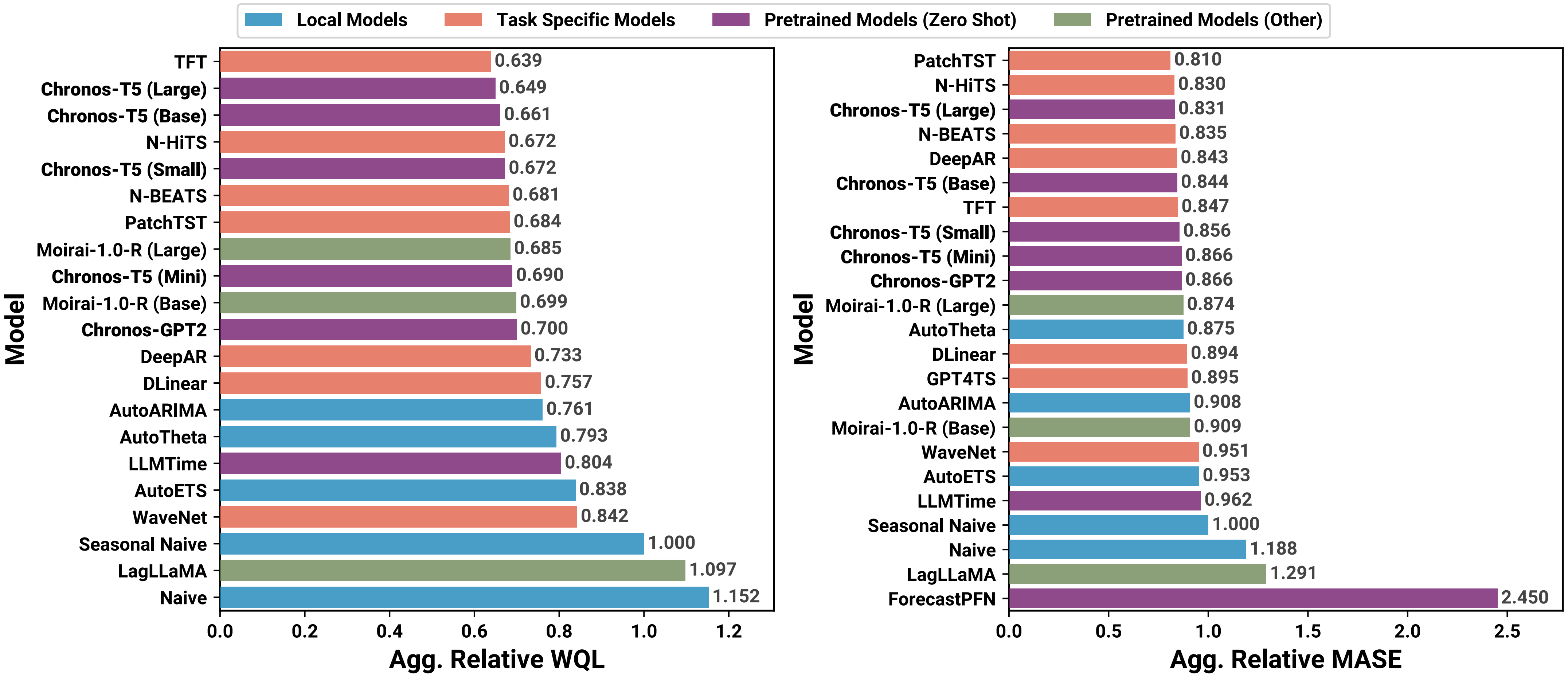
Separate model trained for each time series

- Naive
- SeasonalNaive
- AutoETS
- AutoARIMA
- AutoTheta

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Benchmarking Chronos

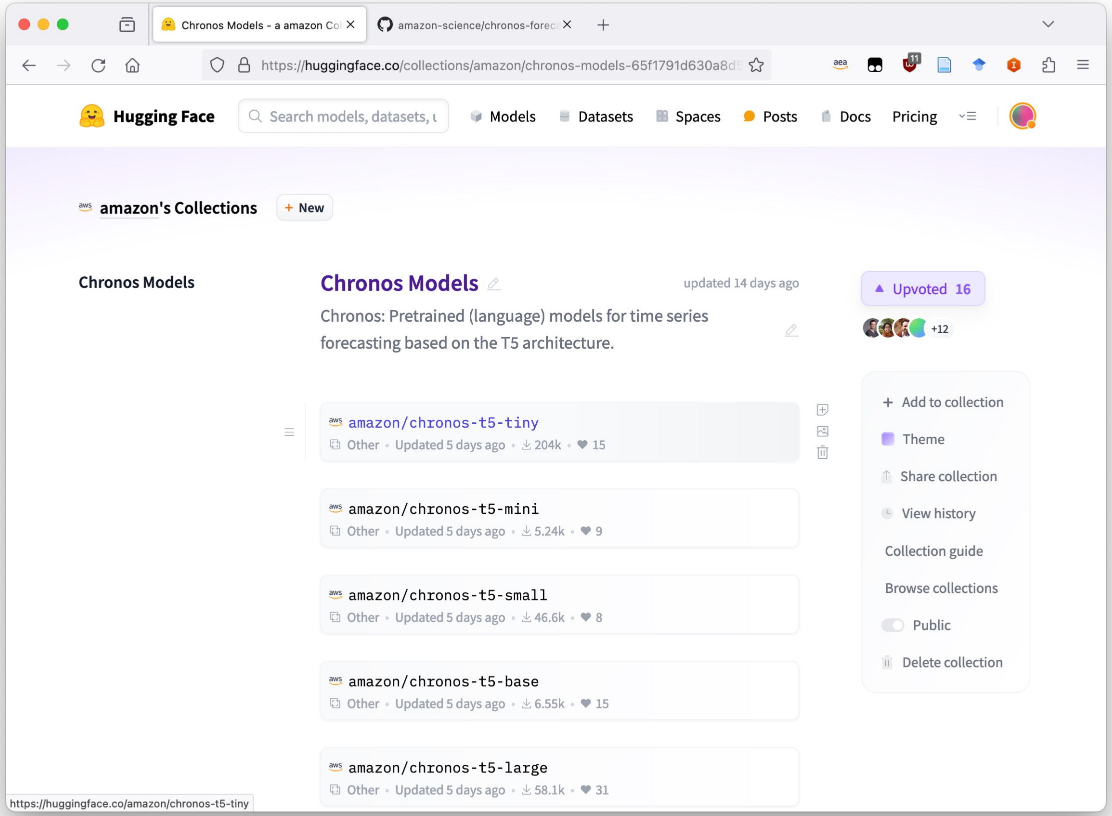
- Zero-shot probabilistic & point forecasting performance on 29 datasets



© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Try out Chronos

- Training, inference & evaluation code available on GitHub
- Model weights & training data available on Hugging Face 🤗
- Run Chronos with 1 line of code using AutoGluon



Downloaded 60M+ times on Hugging Face 🤗

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

The way ahead

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Is forecasting now “solved”?

- A powerful recipe

Lots of data + **Large transformer model** = **Accurate zero-shot forecaster**

- Can we now just follow the NLP playbook to ”solve” forecasting?
  - More data
  - Bigger models
  - ⇒ One model to rule them all?

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Why AutoML is the future of forecasting

- Pretrained forecasting models are fast & cheap (by LLM standards)

- Many ways to improve a single model!
  - Preprocessing, fine-tuning, calibration, conformal prediction, …

- Many ways to combine models!
  - Ensembling, stacking, boosting, …

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Pretrained models in the AutoML toolbox

Preprocessing

Preprocessing can improve accuracy

- Scaling
- Box-Cox transform
- Outlier removal
- …

Model portfolio

Collection of (small) pre-trained models

- Chronos
- TimesFM
- MOIRAI
- …

Adaptation

Adapt pretrained models to the task at hand

- Fine-tuning
- Calibration
- Conformal prediction
- …

Ensembling

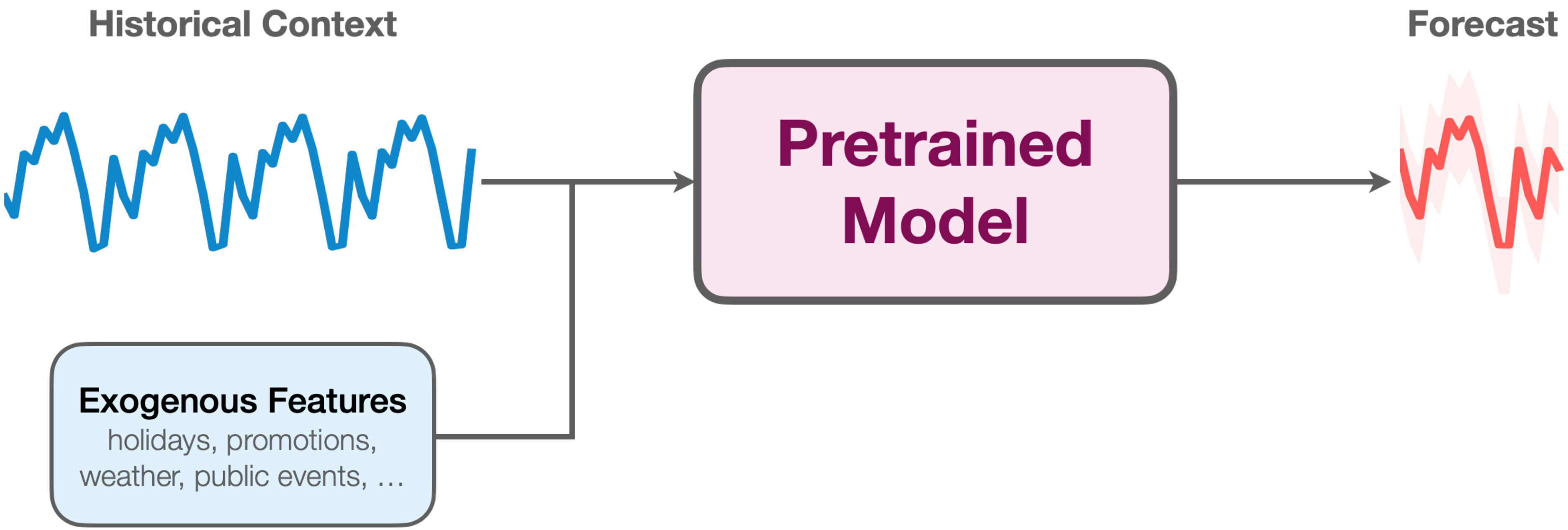
Combine several models into an ensemble

- Boosting
- Stacking
- Linear ensembles
- …

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Beyond univariate forecasting

- Covariates may provide important exogenous information

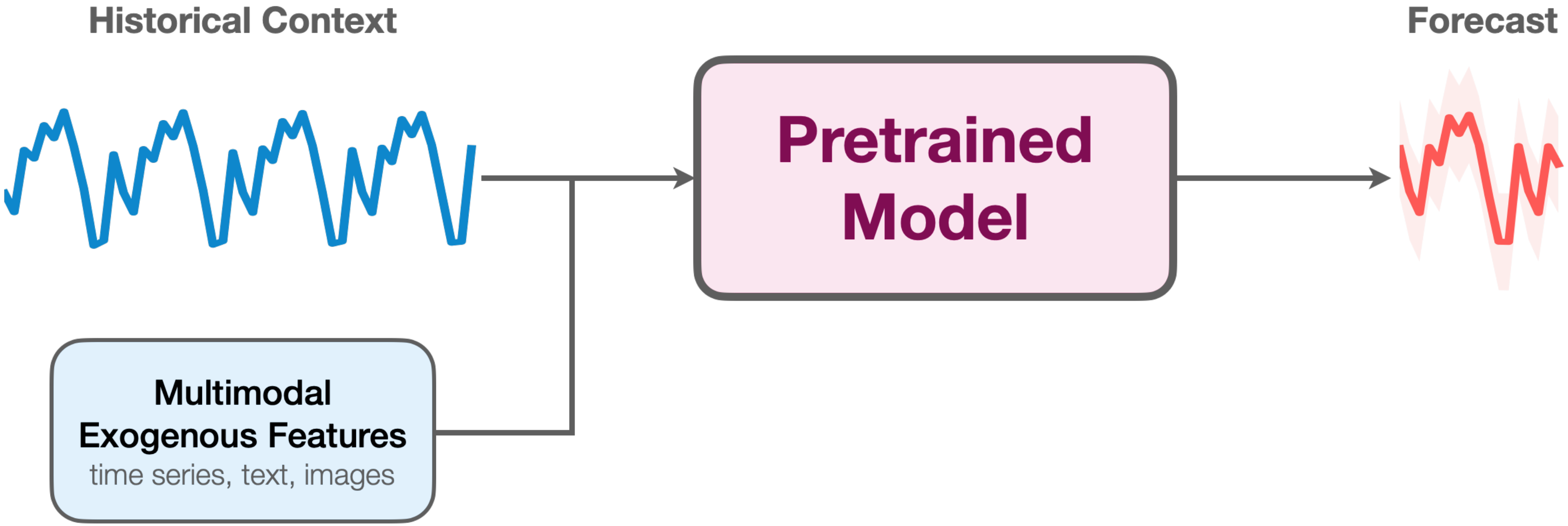


- Challenge: Number and type of covariates are not known a priori

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Multimodal forecasting

- Other modalities (e.g., text and images) can be relevant for the forecast

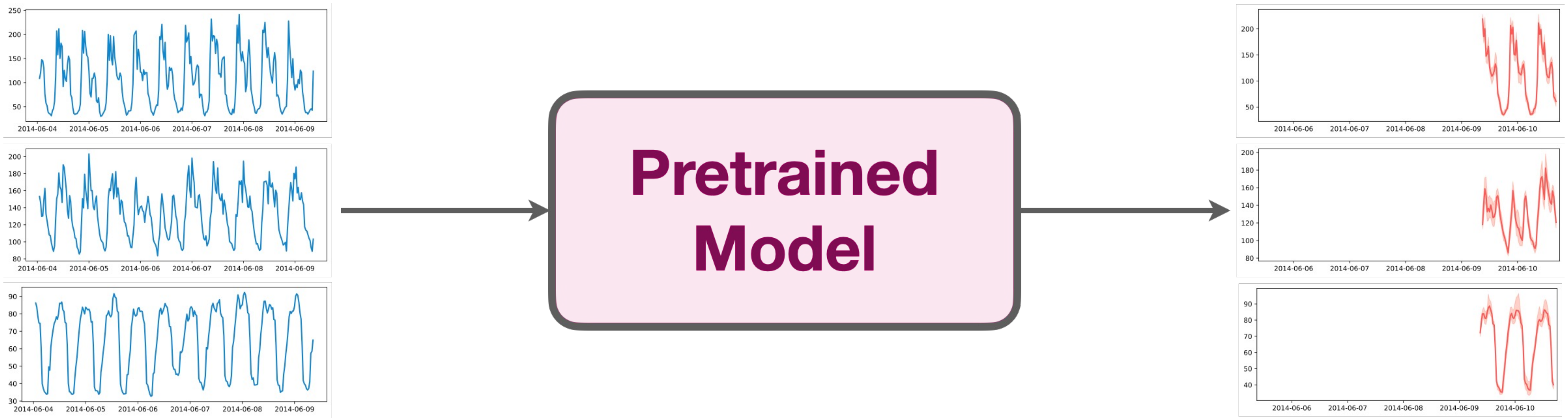


- Challenge: Public multimodal time series datasets are extremely scarce

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Multivariate forecasting

- Joint modeling of multi-dimensional time series



- Challenge: Number of dimensions and their interactions not known a priori

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Data & benchmarks

- High-quality datasets are essential for continued progress

- Many important questions on the data side
  - How to quantify the quality and diversity of time series data?
  - Is synthetic data all you need?
  - How to correctly benchmark time series models?

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Chronos team


Abdul Fatir Ansair


Lorenzo Stella


Caner Turkmen


Xiyuan Zhang


Pedro Mercado Lopez


Huibin Shen


Syama Rangapuram


Sebastian Arango


Shubham Kapoor


Jasper Zschienger


Danielle Robinson


Andrew Wilson


Kari Torkkola


Michael Mahoney


Michael Bohlke-Schneider


Bernie Wang

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Summary

- Pretrained models can make accurate zero-shot forecasts
- Chronos turns forecasting into next-token prediction via scaling & quantization
- Lots of exciting open research questions in this space

    … and AutoML is likely the answer to some of them!

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.