October 2020

A Comprehensive and Modularized Platform for Time Series Forecast and Analytics

Weijie Shen

Apratim Ganguly

Kashif Yousuf

Haoyun Wu

Kenneth C. Lichtendahl Jr.

See next page for additional authors

Inventor(s)

Weijie Shen, Apratim Ganguly, Kashif Yousuf, Haoyun Wu, Kenneth C. Lichtendahl Jr., Chris Fry, Yogishwar Maharaj, and Yiqian Lu

A Comprehensive and Modularized Platform for Time Series Forecast and Analytics

ABSTRACT

Users that work with time series data typically disaggregate time series problems into various isolated tasks and use specific libraries, packages, tools, and services that deal with each individual task. However, the tools used are often fragmented. Analysts have to load different packages for common tasks such as data preprocessing, clustering, feature extraction, forecasting, hierarchical reconciliation, evaluation, and visualization. This disclosure describes a reliable, scalable infrastructure to meet various needs of time series practitioners without adding engineering overload. The infrastructure is modularized and the modules are connected in a flow type declarative language which makes the infrastructure extensible and future proof. Practitioners can use the entire infrastructure or only certain modules, while performing other operations using first or third party libraries or pipelines.

KEYWORDS

- Forecasting
- Multivariate time series
- Hierarchical time series
- Analytics software
- Backtest
- Prediction interval
- Forecast metrics

BACKGROUND

Time series forecasting is an important business problem with applications across a wide range of domains such as weather forecasting, sales forecasting, capacity planning, inventory management, etc. In such applications, time series is the common format of structured data and time series forecasting and analytics are used as crucial tools to derive data-based business insights that drive decisions.

Time series practitioners typically disaggregate time series problems into various isolated tasks and use specific libraries, packages, tools, and services that deal with each individual task. However, the tools used are often fragmented and analysts have to load different packages for common tasks such as data preprocessing, clustering, feature extraction, forecasting, hierarchical reconciliation, evaluation, and visualization. Moreover, not all analysts have access to dedicated computing and software resources to implement and maintain pipelines or the expertise to diagnose and improve models.

Time series practitioners would be greatly benefited from a generic, coherent, reliable, and scalable time series platform that powers various time series applications and serves as a knowledge hub, guiding them through a workflow and process tailored to their use case.

DESCRIPTION

This disclosure describes a time series infrastructure that analysts and time series practitioners can rely on to conduct time series analysis and forecasting in a scalable way.

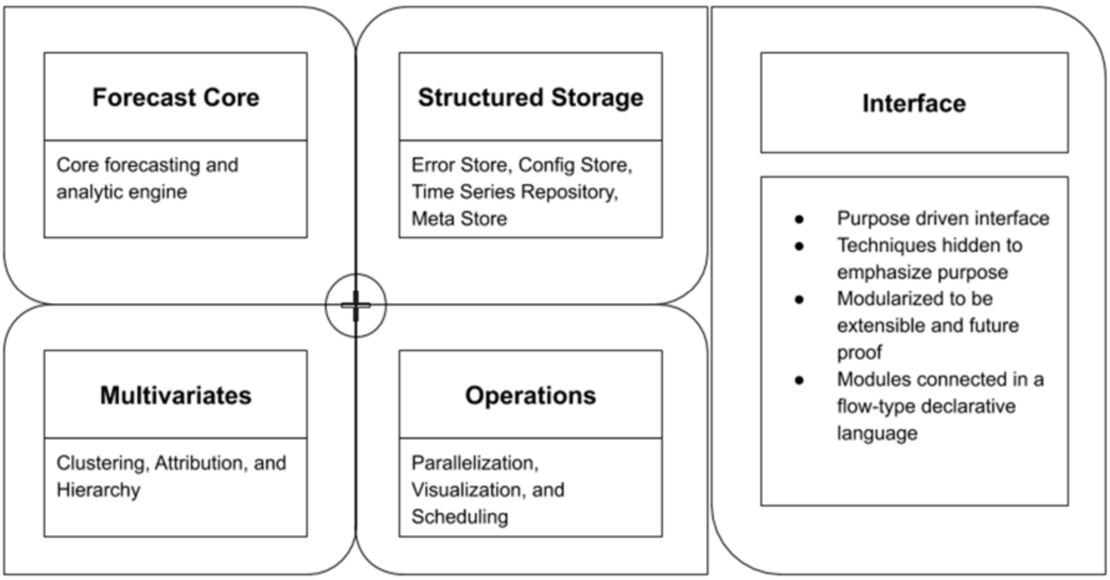

Fig. 1: Components of the Forecasting Infrastructure

Fig. 1 illustrates the infrastructure for time series analysis. The infrastructure comprises four method components and one interface. “Forecast Core” and “Multivariates” are the modeling components, while “Structured Storage” and “Operations” are the implementation components. An interface is provided to the user to hide the four method components and make the process purpose-driven. The infrastructure is modularized and the modules are connected in a flow type declarative language which makes the infrastructure extensible and future proof. Another key feature of the time series infrastructure is that practitioners can use the entire infrastructure or only certain modules, e.g., if performing other operations using third party libraries.

Individual modules of the infrastructure are described below:

- **Forecast core:** The forecast core includes core forecast and analytic engines designed for extracting information on time series. It contains multiple models for forecasting and decomposition including some techniques based on a deep learning approach. ARIMA, ETS, Prophet, N-BEATS are a few examples of the modeling techniques within the forecast core. Additionally, the module comprises a functional decomposition and downscaling adjustment tool.

- **Multivariates:** The multivariates module provides features and approaches that are specifically designed to utilize or satisfy constraints of multiple time series, such as clustering, hierarchy, attribution, etc. The module provides multiple metrics for assessing the forecast model performance along multiple dimensions such as accuracy, bias/calibration, error variance, accuracy risk, stability, etc. This module also provides the ability to tune models based on a customized set of metrics. It implements techniques that use the series information and results from forecast core to reconcile and propose answers that directly tackle key questions for multivariate time series applications. The module also provides clustering (grouping similar time series), attribution (root causing shifts in traffic), causal inference (cross correlation and counterfactuals), and hierarchy (linear constraints and natural grouping). Even for single series, tasks such as model selection and reconciling forecasts of different granularity are handled by the multivariates module.

- **Structured storage:** The structured storage module provides a structure of data type for both series and models. On the model side, it includes a config store for model configuration and error store for model performance. On the series side, it provides a time series repository for raw historical data and a meta store for series features. Config store is used as a house for all verified configuration and defaults. Error store is the main structure for performance data that is crucial for applications such as uncertainty quantification, benchmarking, tuning and reconciliation. The time series repository provides a common structure to host a set of public datasets and a number of golden datasets that are critical for users. These datasets can be used to test each new release for both quality and reliability. Meta store includes series features that characterize both business and data-driven nature of time series. It is crucial for model selection and can help with parameter tuning, and to overcome the cold-start problem. For example, one important type of metadata is the result of clustering conducted during the initial model build.
- **Operations:** The operations module provides the pipeline framework that is in charge of parallelization, visualization, scheduling and monitoring systems, and a generic interface to construct and maintain such pipelines. It also provides the links from forecast core and multivariates to structured storage. It also controls the backtesting framework.

The infrastructure also provides a comprehensive evaluation report that includes metrics for assessing the forecast model performance along multiple dimensions. The report includes a large number of metrics that span five main categories: accuracy, bias, error variance, accuracy risk, and stability. The metrics are designed to enable assessment of key attributes related to use cases such as capacity planning, budgeting, inventory management, and production planning etc. The report includes metrics for evaluation of both point forecasts as well as prediction intervals across the categories.

From these metrics, a full report can be generated, or a primary metric can be chosen in each category for each forecast type and included in a short summary. A summary report may include: {Weighted Mean Absolute Percentage Error (WMAPE), Weighted Mean Percentage Error (WMPE), Weighted Standard Deviation of Percentage Errors (WSDPE), Weighted Standard Deviation of Absolute Percentage Error (WSDAPE) and Weighted Mean Absolute Percentage Change (WMAPC)} for the point forecast metrics, and {Weighted Mean Percent Upper Continuous Ranked Probability Score (WMPUCRPS), Mean Absolute Weighted Coverage Deviation (MAWCD), = Weighted Mean Upper Continuous Ranked Prob. Score; Mean Absolute Weighted Coverage Devation (MAWCD), Weighted Standard Deviation of Percentage Pinball Loss (WSDPPL), and Weighted Mean Absolute Percentage Change (WMAPC)} for prediction interval metrics.

The various categories of metrics included in the framework, and examples of the metrics therein are shown in Table 1.

Category	Description	Sample metrics - point forecast evaluation	Sample metrics - prediction interval evaluation
Accuracy	How closely a model's forecasts adhere to the actual realizations of the time series.	MSE, RMSE, MAE, MAPE, MedAPE, MAE, SMAPE, SMedAPE, MALR, RMSLE, WMAPE, MASE, RMSSE	MPL, MPPL, WMPPL, MIS, MPIS, WMPIS, MCRPS, MPCRPS, WMPCRPS, WMPUCRPS
Bias/ Calibration	The degree to which a forecast is an under- or over-prediction of the actual realization of the time series.	ME, MPE, WMPE, MLE	Coverage, MACD, WCoverage, MAWCD
Error Variance	The spread of point forecasting errors around their mean.	SDE, SDPE, WSDPE, SDLE	Not Applicable
Accuracy Risk	The variability in an accuracy metric over multiple forecasts.	SDAE, SDAPE, WSDAPE, MaxAPE, SMaxAPE, MaxUPE, SMaxUPE, MaxOPE, SMaxOPE	SDPL, WSDPPL
Stability	The degree to which forecasts remain unchanged subject to minor variations in the underlying data.	MAPC, WMAPC	MAPC-PXX, WMAPC-PXX

Table 1: Categories for forecast model evaluation

The comprehensive metrics report also includes comparison to standard benchmarks such as forecasts based on Exponential Triple Smoothing (ETS) and Auto Regressive Integrated Moving Average (ARIMA). The report can be used for a pairwise comparison across models on multiple dimensions. The result of the pairwise comparison is presented along with a p-value based on Diebold-Mariano test for comparing predictive accuracy. Color coding can be used in the comprehensive metrics report to enable rapid visual assessment.

The metrics are computed and presented over time, across forecast horizons and across hierarchical levels. Additionally, the implied economic impact of quantile forecasts (e.g., stockout cost of the 90th percentile) is also presented in the comprehensive metrics report. The comprehensive metrics report can be used by analysts to make informed choices based on multiple performance metrics.

For example, consider a time series that has two models with similar accuracy. Model A has lower bias but has high error variance, whereas Model B has higher bias but low error variance. In such a situation, the analyst might choose model B over model A, since it is generally easier to correct bias compared to error variance.

Fig. 2 illustrates an application of the infrastructure for the capacity planning use case.

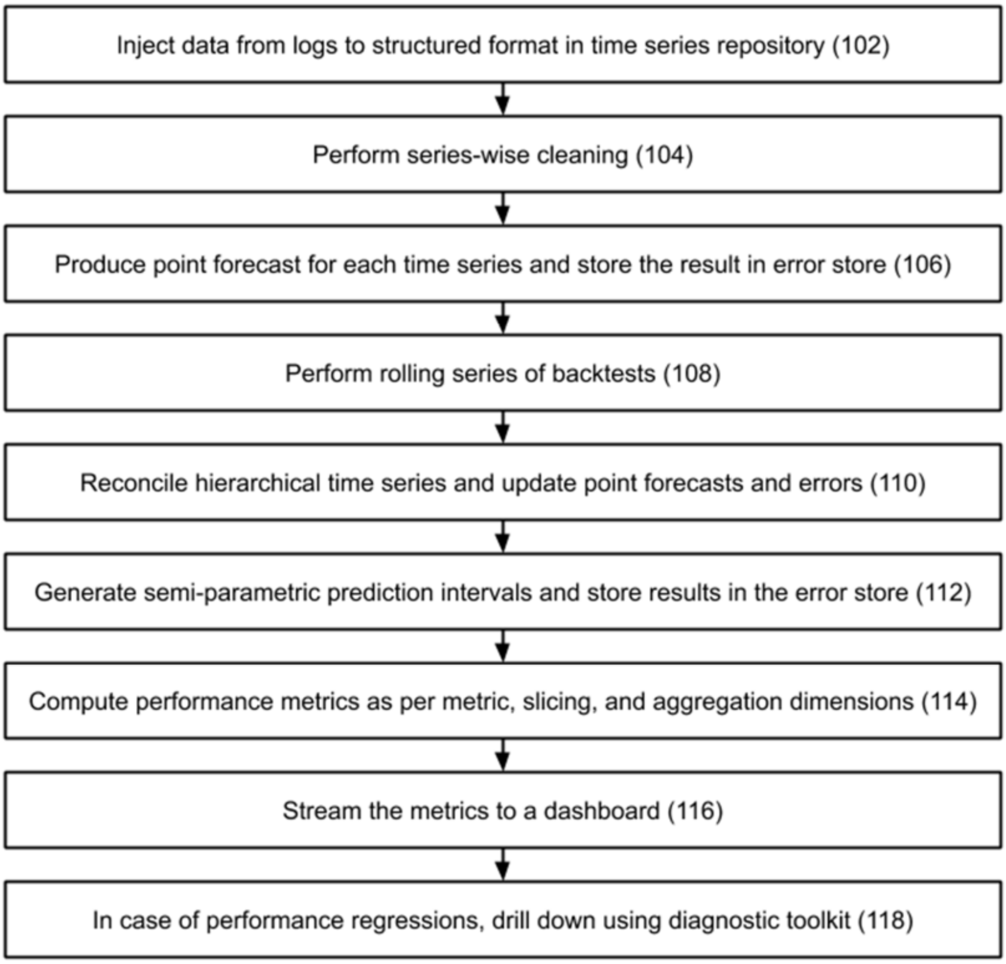

Fig. 2: A forecasting process for capacity planning

Data is injected from logs to structured format in the time series repository where the origins and hierarchy of each time series is recorded (102). Series-wise cleaning is performed including adjustments for structural breaks and inorganic changes and data validation (104). Univariate/multivariate point forecast for each time series node is produced, potentially involving hyperparameter tuning and global models. The results are stored in the error store (106).

The previous step is repeated to conduct a rolling series of backtests that look back while holding out recent actual values for comparisons and out-of-sample error calculations (108). Out-of-sample errors are used to reconcile hierarchical time series so that results satisfy hierarchical constraints. The point forecasts and errors in all backtests are updated (110). Updated out-of-sample errors are used to generate semi-parametric prediction intervals and the results are stored in the error store (112). Prediction intervals created using this approach are more robust than the model-based prediction intervals, since they are created using the results from backtesting as opposed to the original training dataset. The error store is used to customize the metric dimensions, slicing dimensions and aggregation dimensions to compute a series of metrics for point and interval forecasts (114). The metrics can be streamed to a dashboard for visualization, monitoring, and benchmarking (116). In case of performance regression, diagnostic tool sets can be used to drill down and pinpoint problems among metrics, series input, horizons, time series components (118).

The infrastructure as described herein can be utilized in different applications such as capacity planning for cloud computing infrastructure, demand forecasting in various applications, etc.

CONCLUSION

This disclosure describes a reliable, scalable infrastructure to meet various needs of time series practitioners without adding engineering overload. The infrastructure is modularized and the modules are connected in a flow type declarative language which makes the infrastructure extensible and future proof. Practitioners can use the entire infrastructure or only certain modules, while performing other operations using first or third party libraries or pipelines.