Time Series Forecasting Model Selection Methodology

Julian Kaljuvee (2020)

Contents

Background

The goal of this document is to propose an outline for a methodology for selecting time series forecasting models for corporate revenue data.

Methodology Overview

The proposed methodology for forecasting model selection has the following high-level steps:

- Characterise Forecasting Problem
- Model Evaluation Scheme
- Evaluate Models
- Finalise Model

We next detail each of these steps.

Characterise Forecasting Problem

Apply Time Series Forecasting Taxonomy

The initial step in the process is to characterise the forecasting problem and assessing which of the attributes described in Time Series Forecasting Taxonomy would apply:

- Inputs vs. Outputs (IO)
- Endogenous vs. Exogenous (EE)
- Regression vs. Classification (RC)
- Unstructured vs. Structured (US)
- Univariate vs. Multivariate (UM)
- Single-step vs. Multi-step (SM)
- Static vs. Dynamic (SD)
- Contiguous vs. Discontiguous (CD)

See Appendix for more detail of each of these attributes.

Approaches to Characterise Model Attributes

There are some standard approaches to best characterise the model and determine its attributes:

- **Data visualisation** - line and distribution plots to explore features of data such as trends and seasonality
- **Statistical analysis and tests** - includes tools such as autocorrelation function (ACF) correlograms, feature selection bar graphs, statistical tests for stationarity
- **Domain experts** – conversations with subject matter experts such as analysts
- **Project stakeholders** – input from project stakeholders

Attributes of Target Corporate Revenue Model

The table below describes attributes applying to a more advanced model for a corporate revenue data as a sample.

Table – Attributes of Target Corporate Revenue Model

Attribute	Corporate Target Revenue Model	Comment
IO	Inputs - both revenue history and additional regressors Outputs – revenue, or its components	
EE	Both Endogenous and Exogenous variables are present	Many of the variables are expected to endogenous (revenue history itself, various industry and customer data), and some exogenous
RC	Regression	
US	Structured	The data have systematic time-dependent patterns in a time series variable such as trend and/or seasonality
UM	Multivariate	Multivariate model including external data sets is expected to outperform a univariate model
SM	Multi-step	The business stakeholders would like to see forecasts over multiple horizons
SD	Dynamic	The model is expected to be continuously re-trained or re-calibrated given changing nature of the business and revisions of analysts
CD	Contiguous	Mos of the data are expected to be relatively high quality and few missing values

Model Evaluation Scheme

Below we present a typical time series forecasting model evaluation scheme.

Data Preparation and Pre-Processing

Some of the common data preparation or pre-processing methods include:

- Differencing to remove a trend
- Seasonal differencing to remove seasonality
- Standardize to centre
- Normalise to rescale
- Power Transform to make data normally distributed

Split Data

The data set should split the dataset into a train (calibration) and test set, for example 80% for training and 20% for testing.

Feature Importance and Selection

In cases where were have many predictors or features, we may want to first filter out the features that explain the forecast best to to reduce overfitting and improve performance. Some of the common feature selection methods include:

- **Correlation and Mutual Information** – evaluating correlation and mutual information
- **SHAP score** – a framework to explain and visualise feature importance and impact
- **Recursive Feature Selection (RFE)** - RFE works by creating predictive models, weighting features, and pruning those with the smallest weights, then repeating the process until a desired number of features are left

Training and Calibration

The model candidate will fit on the training dataset. An important consideration on this step is to ensure that any coefficients used for data preparation are estimated from the training dataset only and then applied on the test set to avoid overfitting. This might include mean and standard deviation in the case of data standardization.

Forecasting

In this step, we preform out-of-sample prediction**,** or forecasting and compare to the test set directly or using walk-forward cross validation

Evaluate Performance

Finally, we calculate various performance metrics that compare the predictions to the expected values to evaluate goodness of fit of the model

Evaluating Model Types

Below is a typical model evaluation order structured in increasing complexity from classical to modern methods.

- **Baseline**
    - Persistence (grid search the lag observation that is persisted)
    - Rolling moving average.
- **Autoregression**
    - ARMA for stationary data
    - ARIMA for data with a trend
    - SARIMA for data with seasonality
    - VAR/VARMA for multivariate time series
- **Exponential Smoothing**
    - Simple Smoothing
    - Holt Winters Smoothing
- **Linear Machine Learning**
    - Linear Regression
    - Ridge Regression
    - Lasso Regression
    - Elastic Net Regression
- **Nonlinear Machine Learning**
    - k-Nearest Neighbours
    - Classification and Regression Trees
    - Support Vector Regression
- **Ensemble Machine Learning**
    - Bagging
    - Boosting
    - Random Forest
    - Gradient Boosting
- **Deep Learning**
    - MLP
    - CNN
    - LSTM
    - Hybrids

Ideally, each model evaluation experiment should record results to a file so that multiple evaluation runs could be easily compared.

Note that experts often give deep learning methods lower value in this context as generally neural networks are poor at time series forecasting, but there is still a lot of room for improvement and experimentation in this area

References

- How to Develop a Skillful Machine Learning Time Series Forecasting Model - Jason Brownlee
- Feature Selection for Time Series Forecasting - Jason Brownlee
- 11 Classical Time Series Forecasting Methods - Jason Brownlee
- SHAP Feature Importance with Feature Engineering – Kaggle
- Dealing with Multicollinearity - Kaggle

Appendix - Time Series Forecasting Taxonomy

Inputs vs. Outputs

- **Inputs** - historical input data provided to the model in order to make a future forecast
- **Outputs** - prediction or forecast for a future time step beyond the data provided as input

Endogenous vs. Exogenous

- **Endogenous** - input variables that are influenced by other variables in the system and on which the output variable depends. See also Multicollinearity in the Appendix
- **Exogenous** - input variables that are not influenced by other variables in the system and on which the output variable depends.

Regression vs. Classification

- **Regression**- forecast a numerical quantity
- **Classification** - classify as one of two or more labels

A regression problem can be reframed as classification and a classification problem can be reframed as regression. Some problems, like predicting an ordinal (categorical) value, can be framed as either classification and regression. It is possible that a reframing of your time series forecasting problem may simplify it

Unstructured vs. Structured

- **Unstructured** - no obvious systematic time-dependent pattern in a time series variable.
- **Structured** - systematic time-dependent patterns in a time series variable (e.g. trend and/or seasonality)

We can often simplify the modelling process by identifying and removing the obvious structures from the data, such as an increasing trend or repeating cycle. Some classical methods even allow you to specify parameters to handle these systematic structures directly

Univariate vs. Multivariate

- **Univariate** - one variable measured over time.
- **Multivariate** - multiple variables measured over time.

In terms of inputs / outputs, we can also consider the following breakdown:

- **Univariate and Multivariate Inputs** - one or multiple input variables measured over time.
- **Univariate and Multivariate Outputs** - one or multiple output variables to be predicted.

Single-step vs. Multi-step

- **One-Step** - forecast the next time step.

- **Multi-Step** - forecast more than one future time steps

The more time steps to be projected into the future, the more challenging the problem given the compounding nature of the uncertainty on each forecast time step.

Static vs. Dynamic

- **Static** - a forecast model is fit once and used to make predictions.
- **Dynamic** - a forecast model is fit on newly available data prior to each prediction.

Contiguous vs. Discontiguous

- **Contiguous** - observations are made uniform over time.
- **Discontiguous** - observations are not uniform over time.

The lack of uniformity of the observations may be caused by missing or corrupt values, and in such cases specific data formatting may be required when fitting some models to make the observations uniform over time.

Appendix - Time Series Glossary

Decomposition Modelling

Decomposition modelling involves breaking a time series into the components of Trend, Seasonality, Cyclicity and Irregularity, described below.

Trend

Persistent over a relatively long period of time, the trend is the overall increase or decrease of the series during that time.

Seasonality

Seasonality is the presence of variations that occur at specific regular intervals; it is the component of the data and series that experiences regular and predictable changes over a fixed period.

Cyclicity

Cyclicity refers to the variation caused by circumstances, which repeat at irregular intervals. Seasonal behaviour is very strictly regular, meaning there is a precise amount of time between the peaks and troughs of the data; cyclical behaviour, on the other hand, can drift over time because the time between periods is not precise.

Irregularity

Irregularity is the unpredictable component of a time series — the 'randomness'. This component cannot be explained by any other component and includes variations which occur due to unpredictable factors that do not repeat in set patterns.

Autocorrelation

Sometimes also referred to as lagged correlation or serial correlation, autocorrelation refers to the degree of similarity between a given time series, and a lagged version of itself over successive time intervals. Mathematically it is defined as the standard Pearson correlation coefficient between S(t) and S(t-k) where S(t) is the value of the series at time t and is the lag.

Cross Correlation

Similar to autocorrelation, cross correlation represents the Pearson correlation coefficient between two time series S(i,t) and S(j,t-k), indexed by i and j where the latter is lagged by lag k.

Spurious Correlation

Spurious correlation is a mathematical relationship in which two or more events or variables are associated but not causally related. This can be due to either coincidence or the presence of a third, unseen factor, sometimes called a “common response variable”, “confounding factor”

Stationarity

A stationary time series is one in which several statistical properties — namely the mean, variance, and covariance — do not vary with time. This means that, although the values can change with time, the way the series itself changes with time does not change over time. Most classical time series forecasting methods require converting the time series into a stationary form to improve accuracy.

Mulitcollinearity

Multicollinearity (also collinearity) is a phenomenon in which one predictor variable in a multiple regression model can be linearly predicted from the others with a substantial degree of accuracy. In this situation, the coefficient estimates of the multiple regression may change erratically in response to small changes in the model or the data. Multicollinearity does not reduce the predictive power or reliability of the model as a whole, at least within the sample data set; it only affects calculations regarding individual predictors.

Granger Causality Test

Granger Causality Test determines whether one time series will be useful in forecasting another. See more at Granger Causality

Dynamic Time Warping

Dynamic Time Warping (DTW) is a family of algorithms which compute the local stretch or compression to apply to the time axes of two timeseries in order to optimally map one (query) onto the other (reference). DTW outputs the remaining cumulative distance between the two and, if desired, the mapping itself (warping function). DTW is widely used e.g. for classification and clustering tasks in econometrics, chemometrics and general timeseries mining. For Python implementation see dtw-python package.

Appendix – Key Classical Time Series Forecasting Methods

Autoregression (AR)

The autoregression (AR) method models the next step in the sequence as a linear function of the observations at prior time steps.

The notation for the model involves specifying the order of the model p as a parameter to the AR function, e.g. AR(p). For example, AR(1) is a first-order autoregression model.

The method is suitable for univariate time series without trend and seasonal components.

Moving Average (MA)

The moving average (MA) method models the next step in the sequence as a linear function of the residual errors from a mean process at prior time steps.

A moving average model is different from calculating the moving average of the time series.

The notation for the model involves specifying the order of the model q as a parameter to the MA function, e.g. MA(q). For example, MA(1) is a first-order moving average model.

The method is suitable for univariate time series **without** trend and seasonal components.

Autoregressive Moving Average (ARMA)

The Autoregressive Moving Average (ARMA) method models the next step in the sequence as a linear function of the observations and resiudal errors at prior time steps. It combines both Autoregression (AR) and Moving Average (MA) models.

The notation for the model involves specifying the order for the AR(p) and MA(q) models as parameters to an ARMA function, e.g. ARMA(p, q). An ARIMA model can be used to develop AR or MA models. The method is suitable for univariate time series **without** trend and seasonal components.

Autoregressive Integrated Moving Average (ARIMA)

The Autoregressive Integrated Moving Average (ARIMA) method models the next step in the sequence as a linear function of the differenced observations and residual errors at prior time steps.

It combines both Autoregression (AR) and Moving Average (MA) models as well as a differencing pre-processing step of the sequence to make the sequence stationary, called integration (I).

The notation for the model involves specifying the order for the AR(p), I(d), and MA(q) models as parameters to an ARIMA function, e.g. ARIMA(p, d, q). An ARIMA model can also be used to develop AR, MA, and ARMA models.

The method is suitable for univariate time series **with** trend and **without** seasonal components.

Seasonal Autoregressive Integrated Moving-Average (SARIMA)

The Seasonal Autoregressive Integrated Moving Average (SARIMA) method models the next step in the sequence as a linear function of the differenced observations, errors, differenced seasonal observations, and seasonal errors at prior time steps. It combines the ARIMA model with the ability to perform the same autoregression, differencing, and moving average modeling at the seasonal level.

The notation for the model involves specifying the order for the AR(p), I(d), and MA(q) models as parameters to an ARIMA function and AR(P), I(D), MA(Q) and m parameters at the seasonal level, e.g. SARIMA(p, d, q)(P, D, Q)m where "m" is the number of time steps in each season (the seasonal period). A SARIMA model can be used to develop AR, MA, ARMA and ARIMA models.

The method is suitable for univariate time series **with** trend and/or seasonal components.

Seasonal Autoregressive Integrated Moving-Average with Exogenous Regressors (SARIMAX)

The Seasonal Autoregressive Integrated Moving-Average with Exogenous Regressors (SARIMAX) is an extension of the SARIMA model that also includes the modelling of exogenous variables.

Exogenous variables are also called covariates and can be thought of as parallel input sequences that have observations at the same time steps as the original series. The primary series may be referred to as endogenous data to contrast it from the exogenous sequence(s). The observations for exogenous variables are included in the model directly at each time step and are not modelled in the same way as the primary endogenous sequence (e.g. as an AR, MA, etc. process).

The SARIMAX method can also be used to model the subsumed models with exogenous variables, such as ARX, MAX, ARMAX, and ARIMAX.

The method is suitable for univariate time series **with** trend and/or seasonal components and exogenous variables.

Vector Autoregression (VAR)

The Vector Autoregression (VAR) method models the next step in each time series using an AR model. It is the generalization of AR to multiple parallel time series, e.g. multivariate time series.

The notation for the model involves specifying the order for the AR(p) model as parameters to a VAR function, e.g. VAR(p).

The method is suitable for multivariate time series **without** trend and seasonal components.

Vector Autoregression Moving-Average (VARMA)

The Vector Autoregression Moving-Average (VARMA) method models the next step in each time series using an ARMA model. It is the generalization of ARMA to multiple parallel time series, e.g. multivariate time series.

The notation for the model involves specifying the order for the AR(p) and MA(q) models as parameters to a VARMA function, e.g. VARMA(p, q). A VARMA model can also be used to develop VAR or VMA models.

The method is suitable for multivariate time series **without** trend and seasonal components.

Vector Autoregression Moving-Average with Exogenous Regressors (VARMAX)

The Vector Autoregression Moving-Average with Exogenous Regressors (VARMAX) is an extension of the VARMA model that also includes the modelling of exogenous variables. It is a multivariate version of the ARMAX method. Exogenous variables are also called covariates and can be thought of as parallel input sequences that have observations at the same time steps as the original series. The primary series(es) are referred to as endogenous data to contrast it from the exogenous sequence(s). The observations for exogenous variables are included in the model directly at each time step and are not modelled in the same way as the primary endogenous sequence (e.g. as an AR, MA, etc. process). The VARMAX method can also be used to model the subsumed models with exogenous variables, such as VARX and VMAX.

The method is suitable for multivariate time series without trend and seasonal components with exogenous variables.

Simple Exponential Smoothing (SES)

The Simple Exponential Smoothing (SES) method models the next time step as an exponentially weighted linear function of observations at prior time steps. The method is suitable for univariate time series **without** trend and seasonal components.

Holt Winter's Exponential Smoothing (HWES)

The Holt Winter's Exponential Smoothing (HWES) also called the Triple Exponential Smoothing method models the next time step as an exponentially weighted linear function of observations at prior time steps, taking trends and seasonality into account. The method is suitable for univariate time series **with** trend and/or seasonal components.