Advanced Data Analysis and Machine Learning

Lecture: Time Series Modelling

Lasse Lensu

2025-11-02

Outline

Time series decomposition

LUT University

- Time series decomposition refers to dividing a time series into (typically) three separate components:
  - Trend: a deterministic component that has no periodicity. The trend can be linear or non-linear.
  - Seasonal component: a deterministic seasonal component with periodicity. The period or season duration is known or can be estimated from the data.
  - Residual component: a stochastic component that is not necessarily white noise. The component can have autocorrelation and repeated patterns without clear periodicity.
- The decomposition can be based on the following:
  - Additive: a sum of the components.
  - Multiplicative, a product of the components – can be turned into a sum using a logarithm.
- The trend and seasonal components can be estimated by filters or parametric models – after the trend and seasonality have been removed, the rest is the residual.

Baseline models

- To evaluate the model performance, baseline models are used as benchmarks.
- Typical baseline models are as follows:
  - Constant Predictor
  - Naïve forecasting
  - Seasonal Naive Predictor
  - Mean Estimator

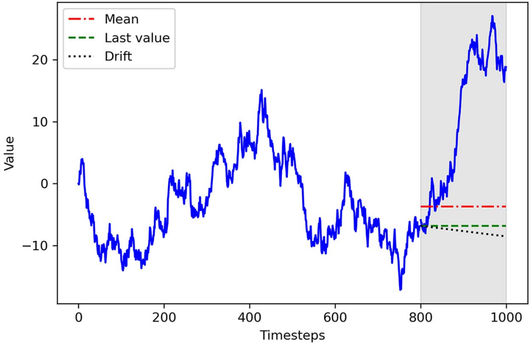

Figure: Forecasting a challenging time series, forecasting window in grey [3].

Constant predictor

LUT University

- A constant predictor is a predictor that always produces constant value for each forecasted sample.
- The model takes the constant value c and the prediction length as its parameters, then for each time series sample x, it produces a forecast of same constant value c:

$$\hat{x}(t+h|t) = c. \tag{1}$$

Naïve model

- A naïve model predicts the value based on the past history data, that is, each forecasted sample is set to the previous value.
- The parameters used in this model are the past history data, prediction horizon h and frequency of the data:
$$\hat{x}(t+h|t) = x(t). \tag{2}$$
- When data follows a random walk, a simple naive forecast is the best option.

Seasonal naïve predictor

- Seasonal naïve predictor is similar to the naïve model, but it considers the seasonality of the data.
- Each forecasted sample is equal to the last observed data of the same season.
- In the seasonal naïve predictor, for each time series x:

$$\hat{x}(T+k) = x(T+k-h) \tag{3}$$

where T is the prediction horizon, $k = 0, \ldots, T-1$, and h is the season duration.

Mean estimator

- The mean estimator assumes that all future values are equal to the average value of the context window:

$$\overline{x} = \frac{1}{T}\sum_{i=1}^{T} x_i$$
$$\hat{x}(T+n) = \overline{x} \tag{4}$$

where $\hat{x}(T+n)$ is the forecasted sample, and $\overline{x}$ is the average value of the context window of length T.

Autoregressive process

- The foundation of the autoregressive (AR) model is to express the current value of a time series $x(t)$ as a linear combination of the past data.
- When the model parameters have been set, it is possible to predict future values using the model.
- A k_{th}-order AR model, AR(k), is defined as:

$$x(t) = c + \sum_{i=1}^{k} \rho_i x(t-1) + \varepsilon_t \tag{5}$$

  where c is constant, $\rho_1, \rho_2 \ldots \rho_k$ are AR coefficients, ε_t denotes white noise.
- The model's order tells how many lagging values of the past are disturbed.
- The AR model indicates whether there is any relationship among the present $x(t)$ values and their previous values, where $x(t)$ is typically disturbed by noise ε_t.

Moving average

- Linear regression degenerating present values in contradiction to the random white noise of previous values is a moving average (MA) model.
- It is also possible to treat the MA as a model where the time series is assumed to be a MA random noise series ε_t.
- A MA model of order m, MA(m), is defined as

$$x(t) = \mu + \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \ldots + \alpha_m \varepsilon_{t-m} \quad (6)$$

  where μ is the series mean, $\alpha_1, \ldots, \alpha_p$ are parameters, and the random variable ε_t is white noise.
- It is possible to invert the MA model into an AR model.

Autoregressive moving average

- An autoregressive moving average (ARMA) model combines AR and MA models.
- In ARMA, the present value is a linear combination of the previous samples and the past noise values.
- The model ARMA(k, m) is mathematically expressed as:

$$x(t) = \varepsilon_t + \sum_{i=1}^{k} \rho_i x(t-1) + \sum_{i=1}^{m} \alpha_i \varepsilon_{t-i}. \tag{7}$$

- In the case of stationary time series, ARMA models have been successful in several applications.

Autoregressive integrated moving average

- In the case of non-stationary time series, the autoregressive integrated moving average (ARIMA) model is a common choice.
- An ARIMA model includes the AR parameters, MA parameters, and the number of n differences carried out to (1-L), where L is a lag operator.
- A model ARIMA(k, n, m) can be defined as:

$$\left(1 - \sum_{i=1}^{k} \rho_i L^i\right)(1-L)^n x(t) = \left(1 + \sum_{j=1}^{m} \rho_j L^j\right) \varepsilon_t \tag{8}$$

where k represents the order of autoregression, m denotes the moving average, and n denotes variations that make the initial time series stationary.

Seasonal autoregressive integrated moving average

LUT University

- A variant of ARIMA is Seasonal ARIMA (SARIMA) that considers also seasonality.
- A SARIMA model is defined as SARIMA$(p, d, q) \times (P, D, Q)_m$ where
  - the first triplet of parameters is an ARIMA model and
  - the latter set of parameters represents the seasonal autoregressive, integration, and moving average components with m as the season duration.
- Seasonal ARIMA with exogenous variables (SARIMAX) model adds the possibility to use exogenous variables that complement the time series.

Evaluation criteria

- The performance of a predictive model is evaluated using error metrics.
- The metrics quantify the difference between a model output and the observed value.
- The root-mean-squared-error (RMSE) is defined as:

$$\mathrm{RMSE} = \sqrt{\frac{1}{nT}\sum_{i,t}(\hat{x}_{i,t} - x_{i,t})^2} \tag{9}$$

where $x_{i,t}$ is the actual value of the data i at time t, $\hat{x}_{i,t}$ is the mean of the predicted data, n is the number of time series and T is the prediction horizon.

Evaluation criteria

- The mean absolute error (MAE) measures the average error in a set of predictions.
- It is defined as:

$$\mathrm{MAE} = \frac{1}{n}\sum_{t=1}^{n} |x_t - \hat{x}_t| \tag{10}$$

where n is the number of time steps, x_t is the actual value and $\hat{x}_t$ is the forecast.

Evaluation criteria

- The mean absolute percentage error (MAPE) is a common metric based on MAE for quantifying forecasting accuracy.
- If the time series contains zero or close-to-zero values, MAPE produces infinite or very large values.
- To address this problem, a variant called mean arctangent absolute percentage error (MAAPE) has been proposed [2]:

$$\mathrm{MAAPE} = \frac{1}{n}\sum_{t=1}^{n} \arctan\left(\left|\frac{x_t - \hat{x}_t}{x_t}\right|\right) \tag{11}$$

Evaluation criteria

- Symmetric mean absolute percentage error (SMAPE) measures the accuracy based on the percentage of the errors.
- It is typically defined as follows:

$$\text{SMAPE} = \frac{100\%}{n} \sum_{t=1}^{n} \frac{|\hat{x}_t - x_t|}{(|x_t| + |\hat{x}_t|)/2} \tag{12}$$

where x_t is the real value and $\hat{x}_t$ is the predicted value.

Evaluation criteria

- The forecast distribution accuracy is evaluated using weighted quantile loss.
- The quantiles are in the range $[0, 1]$ and the weighted quantile loss is defined as follows:

$$wQuantileLoss[\tau] = 2\frac{\sum_{i,t} Q_{i,t}^{(\tau)}}{\sum_{i,t} |x_{i,t}|} \tag{13}$$

$$Q_{i,t}^{(\tau)} = \begin{cases} (1-\tau)|q_{i,t}^{(\tau)} - x_{i,t}|, & \text{if } q_{i,t}^{(\tau)} > x_{i,t} \\ \tau|q_{i,t}^{(\tau)} - x_{i,t}|, & \text{otherwise} \end{cases} \tag{14}$$

where $q_{i,t}^{(\tau)}$ is the τ-quantile of the forecast distribution which the model predicts [1].

Summary

- Time series decomposition refers to dividing a time series into three separate components: trend, seasonal and residual.
- Baseline models are used to solve prediction problems and as benchmarks.
- The performance of a predictive model can be evaluated using standard error metrics.

References

DeepAR Forecasting Algorithm, AmazonSageMaker.
http://docs.aws.amazon.com/sagemaker/latest/dg/deepar.html.

Sungil Kim and Heeyoung Kim.
A new metric of absolute percentage error for intermittent demand forecasts.
International Journal of Forecasting, 32(3):669–679, 2016.

Marco Peixeiro.
Time series forecasting in Python.
Manning Publications Co, Shelter Island, NY, 2022.