Time Series Analysis

Prophet Model

Rajas Vaidya (MDS202131)
Rishika Tibrewal (MDS202135)

Assignment 2

Contents

1 Introduction to Time Series

Time series refers to a set of observations collected over a period of time i.e., time is the independent variable. These observations are typically measured at regular intervals. Examples of time series data include stock prices, weather data, economic indicators, and traffic patterns.

Time series analysis is the process of analyzing and modeling time series data to understand patterns and make predictions about future behavior. This involves techniques such as trend analysis, seasonal decomposition, autocorrelation analysis, and forecasting. It is used in different fields for time-based predictions – like Weather Forecasting models, Stock market predictions, etc. One key aspect of time series analysis is that the observations are often dependent on previous observations.

2 Components of Time Series Data

- **Trend:** The trend component represents the long-term behavior of the time series. It can be either increasing or decreasing over time, or it does not change significantly over time.
- **Seasonality:** The seasonality component represents the regular and periodic fluctuations in the time series that occur at fixed intervals of time. For example, sales of Christmas trees conditioners tend to be higher in December when compared to other months.
- **Irregularity:** The irregular component represents the unexplained or unpredictable variation in the time series that is not accounted for by the other components. It is often modeled as white noise, which has no correlation with past or future observations.
- **Cyclicity:** The cyclical component represents the non-periodic fluctuations in the time series that are associated with business cycles, economic cycles, or other long-term patterns. This component can be difficult to isolate and model accurately.
- **Holiday Effect:** This occurs on irregular schedules over a day or a period of days.

By decomposing a time series into its components, one can gain a better understanding of the underlying patterns and make more accurate forecasts for future behavior.

3 Prophet Model

Prophet is a time series forecasting model developed by Facebook's Core Data Science team. It is designed to make accurate predictions for time series data that display a variety of complex patterns, such as trends, seasonality, and holiday effects.

Prophet is based on an additive regression model that decomposes the time series into four components: **trend**, **seasonality**, **holidays**, and **an error term**.

$$y(t) = g(t) + h(t) + s(t) + \epsilon(t)$$

where $g(t)$ describes a piecewise-linear trend (or "growth term"), $s(t)$ describes the various seasonal patterns, $h(t)$ captures the holiday effects, and $\epsilon(t)$ is a white noise error term.

Using time as a regressor, Prophet is trying to fit several linear and non linear functions of time as components. So, the forecasting problem is being framed as a curve-fitting exercise rather than looking explicitly at the time based dependence of each observation within a time series.

The trend component is modeled using a piecewise linear function, while seasonality is modeled using Fourier series. The holiday component captures the impact of known holiday events, such as national holidays or major sporting events. The error term represents the random variation in the data that is not explained by the other components. Using time as a regressor, Prophet is trying to fit several linear and non linear functions of time as components. The model is estimated using a Bayesian approach to allow for automatic selection of the changepoints and other model characteristics.

3.1 Growth Term

The growth function of a time series is used to represent the general pattern (or trend) of the data. This concept is similar to that of linear or logistic functions. However, the Prophet model developed by Facebook introduces a novel idea that the growth trend can occur throughout the data or change at specific points called **changepoints**. These are moments in the time series where the direction of the data shifts.

For instance, consider the case of new COVID-19 infections. The decrease in new cases may occur due to the introduction of a vaccine, after reaching a peak. Alternatively, there may be a sudden surge in infections if a new strain of the virus is introduced into the population. Prophet is capable of automatically identifying such changepoints, but one can also manually define them. Additionally, one can adjust the degree to which changepoints affect the growth function and the extent to which automatic changepoint detection accounts for historical data.

3.1.1 Linear Growth

This is the default setting for Prophet. It uses a set of piecewise linear equations with differing slopes between change points. When linear growth is used, the growth term will look similar to the classic line, $y = mx + c$, except the slope (m) and offset (c) are variables that change values at each changepoint.

3.1.2 Logistic Growth

This option is beneficial if your time series has a limit or a minimum value, such that the modeled values saturate and cannot exceed a maximum or minimum level (like a maximum capacity). By using logistic growth, the growth component will resemble a standard equation for a logistic curve, except that the maximum capacity (C) will vary with time, and the growth rate (k) and offset (m) will be variable that change values at each changepoint.

$$g(t) = \frac{C(t)}{1 + x^{-k(t-m)}}$$

3.1.3 Flat

Lastly, one can choose a flat trend when there is no growth over time (but there still may be seasonality). If set to flat the growth function will be a constant value.

3.2 Seasonality Function

The seasonality function is simply a Fourier Series as a function of time. This series can approximate nearly any curve or in the case of Facebook Prophet, the seasonality (cyclical pattern) in our data.

$$s(t) = \sum_{n=1}^{N}(a_n \cos(\frac{2\pi nt}{P}) + b_n \sin(\frac{2\pi nt}{P}))$$

where P is the period (365.25 for yearly data, 7 for weekly data), parameters a_i, b_i are to be estimated for a given N. Prophet automatically chooses an optimal number of terms in the series, known as **Fourier order** according to the data. By default, order 10 is used for annual seasonality and order 3 is used for weekly seasonality. Note that one can choose between additive and multiplicative seasonality.

3.3 Holiday Function

The holiday function in Facebook Prophet enables adjustments to the forecasting model when a holiday or significant event might alter the prediction. It involves specifying a set of dates (either pre-defined US holidays or custom dates), and incorporating them into the forecast by modifying the growth and seasonality terms based on historical data from those particular dates. Additionally, users can define a range of days around holidays, such as the period between Christmas and New Year's Day, or holiday weekends like Thanksgiving.

4 Where to use Prophet

Prophet is especially useful for datasets that:

- Contain an extended time period (months or years) of detailed historical observations (hourly, daily, or weekly)

- Have multiple strong seasonalities
- Include previously known important, but irregular, events
- Have missing data points or large outliers
- Have non-linear growth trends that are approaching a limit

5 Python Prophet API

Prophet follows the sklearn model API. An instance of the Prophet class is created and then fit and predict methods are called. The input to Prophet is always a dataframe with two columns: ds and y. The ds (datestamp) column should be of a format expected by Pandas, ideally YYYY-MM-DD for a date or YYYY-MM-DD HH:MM:SS for a timestamp. The y column must be numeric, and represents the measurement we wish to forecast.

Click on this link. to see a demo of Prophet model.

6 Advantages of Prophet Model

- **Handles various types of seasonality:** Prophet can handle multiple types of seasonal patterns in the data, including weekly, monthly, and yearly seasonality, as well as holiday effects. This makes it particularly useful for datasets with complex seasonal patterns.
- **Easy to use:** Prophet is designed to be easy to use, even for users with limited experience in time-series forecasting. It requires minimal data preparation and tuning, and the model can be fit with just a few lines of code.
- **Fast computation:** Prophet is built for speed, and can quickly generate forecasts for large datasets with thousands of time series.
- **Robust to missing data:** Prophet is able to handle datasets with missing values or gaps in the time series, allowing for more flexible data preparation.
- **Provides uncertainty intervals:** Prophet generates uncertainty intervals around its predictions, allowing users to assess the confidence of the forecasts and make informed decisions.
- **Interpretable:** While Prophet is a complex model, it is designed to be transparent and interpretable. Users can easily examine the trend and seasonal components of the forecast to better understand the model's behavior.
- **Open-source:** Prophet is an open-source model, which means that the code is publicly available and can be freely modified and customized to suit the needs of different users.

7 Prophet Over ARIMA

Prophet and ARIMA (AutoRegressive Integrated Moving Average) are both popular time-series forecasting models, but they differ in several ways. Here are some ways in which Prophet is better than ARIMA:

- **Handles seasonality more effectively:** ARIMA models are designed to handle stationary time series data and struggle to handle non-linear trends and seasonal patterns, while Prophet is designed to effectively capture multiple seasonalities and non-linear trends.
- **More user-friendly:** ARIMA models require more technical expertise and data preparation than Prophet, which can be more user-friendly for users with limited experience in time-series analysis.
- **Requires less parameter tuning:** ARIMA models require extensive parameter tuning, including selecting the order of the model, whereas Prophet requires minimal tuning, making it more accessible to users with limited experience.
- **Handles missing data better:** Prophet can handle missing values and gaps in the time series more effectively than ARIMA, which requires complete data.

- **Provides uncertainty intervals:** Prophet generates uncertainty intervals around its predictions, which can help users to assess the reliability of the forecast and make informed decisions.
- **Handles outliers and change points better:** Prophet can handle sudden changes in the trend and outliers more effectively than ARIMA.

Overall, while both models have their strengths and weaknesses, Prophet is often preferred over ARIMA for datasets with complex seasonal patterns and non-linear trends, where its ease of use and ability to handle missing data, outliers, and sudden changes in the trend make it a more reliable forecasting model.

8 Limitations of Prophet Model

- **Limited input data:** Prophet is best suited for datasets that have regular time intervals and a minimum of one year of historical data. It may not perform as well when there are missing values, outliers, or irregular time intervals.
- **Assumes trend changes are linear or non-existent:** Prophet assumes that any changes in trends are either linear or non-existent. This means that it may not be able to capture sudden changes or non-linear trends in the data.
- **Limited handling of categorical features:** Prophet does not handle categorical variables very well, and any such variables must be transformed into numeric values before using them as inputs to the model.
- **Limited model transparency:** Prophet is a black-box model, meaning that it can be difficult to understand how the model is making its predictions. This can make it challenging to interpret the results and make adjustments to the model if necessary.
- **Limited ability to handle seasonality interactions:** While Prophet is designed to handle multiple seasonality, it cannot capture interactions between them, which can limit its performance in some cases.
- **Limited extrapolation capability:** Prophet is not designed to extrapolate beyond the range of the training data, and its predictions may not be reliable when forecasting far into the future.