What Is Time Series Forecasting?

by **Jason Brownlee** on December 2, 2016 in **Time Series**

Last Updated on August 15, 2020

Time series forecasting is an important area of machine learning that is often neglected.

It is important because there are so many prediction problems that involve a time component. These problems are neglected because it is this time component that makes time series problems more difficult to handle.

In this post, you will discover time series forecasting.

After reading this post, you will know:

- Standard definitions of time series, time series analysis, and time series forecasting.
- The important components to consider in time series data.
- Examples of time series to make your understanding concrete.

Kick-start your project with my new book Time Series Forecasting With Python, including *step-by-step tutorials* and the *Python source code* files for all examples.

Let's get started.

What is Time Series Forecasting?
Photo by Dennis Kleine, some rights reserved.

Time Series

A normal machine learning dataset is a collection of observations.

For example:

```
1  observation #1
2  observation #2
3  observation #3
```

Time does play a role in normal machine learning datasets.

Predictions are made for new data when the actual outcome may not be known until some future date. The future is being predicted, but all prior observations are almost always treated equally. Perhaps with some very minor temporal dynamics to overcome the idea of "*concept drift*" such as only using the last year of observations rather than all data available.

A time series dataset is different.

Time series adds an explicit order dependence between observations: a time dimension.

This additional dimension is both a constraint and a structure that provides a source of additional information.

> *A time series is a sequence of observations taken sequentially in time.*

— Page 1, Time Series Analysis: Forecasting and Control.

For example:

```
1  Time #1, observation
2  Time #2, observation
3  Time #3, observation
```

Describing vs. Predicting

We have different goals depending on whether we are interested in understanding a dataset or making predictions.

Understanding a dataset, called *time series analysis*, can help to make better predictions, but is not required and can result in a large technical investment in time and expertise not directly aligned with the desired outcome, which is forecasting the future.

> *In descriptive modeling, or time series analysis, a time series is modeled to determine its components in terms of seasonal patterns, trends, relation to external factors, and the like. … In contrast, time series forecasting uses the information in a time series (perhaps with additional information) to forecast future values of that series*

— Page 18-19, Practical Time Series Forecasting with R: A Hands-On Guide.

Time Series Analysis

When using classical statistics, the primary concern is the analysis of time series.

Time series analysis involves developing models that best capture or describe an observed time series in order to understand the underlying causes. This field of study seeks the "*why*" behind a time series dataset.

This often involves making assumptions about the form of the data and decomposing the time series into constitution components.

The quality of a descriptive model is determined by how well it describes all available data and the interpretation it provides to better inform the problem domain.

> The primary objective of time series analysis is to develop mathematical models that provide plausible descriptions from sample data

— Page 11, Time Series Analysis and Its Applications: With R Examples

Time Series Forecasting

Making predictions about the future is called extrapolation in the classical statistical handling of time series data.

More modern fields focus on the topic and refer to it as time series forecasting.

Forecasting involves taking models fit on historical data and using them to predict future observations.

Descriptive models can borrow for the future (i.e. to smooth or remove noise), they only seek to best describe the data.

An important distinction in forecasting is that the future is completely unavailable and must only be estimated from what has already happened.

> The purpose of time series analysis is generally twofold: to understand or model the stochastic mechanisms that gives rise to an observed series and to predict or forecast the future values of a series based on the history of that series

— Page 1, Time Series Analysis: With Applications in R.

The skill of a time series forecasting model is determined by its performance at predicting the future. This is often at the expense of being able to explain why a specific prediction was made, confidence intervals and even better understanding the underlying causes behind the problem.

Components of Time Series

Time series analysis provides a body of techniques to better understand a dataset.

Perhaps the most useful of these is the decomposition of a time series into 4 constituent parts:

1. **Level**. The baseline value for the series if it were a straight line.
2. **Trend**. The optional and often linear increasing or decreasing behavior of the series over time.
3. **Seasonality**. The optional repeating patterns or cycles of behavior over time.
4. **Noise**. The optional variability in the observations that cannot be explained by the model.

All time series have a level, most have noise, and the trend and seasonality are optional.

> *The main features of many time series are trends and seasonal variations … another important feature of most time series is that observations close together in time tend to be correlated (serially dependent)*

— Page 2, Introductory Time Series with R

These constituent components can be thought to combine in some way to provide the observed time series. For example, they may be added together to form a model as follows:

```
1  y = level + trend + seasonality + noise
```

Assumptions can be made about these components both in behavior and in how they are combined, which allows them to be modeled using traditional statistical methods.

These components may also be the most effective way to make predictions about future values, but not always.

In cases where these classical methods do not result in effective performance, these components may still be useful concepts, and even input to alternate methods.

Concerns of Forecasting

When forecasting, it is important to understand your goal.

Use the Socratic method and ask lots of questions to help zoom in on the specifics of your predictive modeling problem. For example:

1. **How much data do you have available and are you able to gather it all together?** More data is often more helpful, offering greater opportunity for exploratory data analysis, model testing and tuning, and model fidelity.
2. **What is the time horizon of predictions that is required? Short, medium or long term?** Shorter time horizons are often easier to predict with higher confidence.
3. **Can forecasts be updated frequently over time or must they be made once and remain static?** Updating forecasts as new information becomes available often results in more accurate predictions.
4. **At what temporal frequency are forecasts required?** Often forecasts can be made at a lower or higher frequencies, allowing you to harness down-sampling, and up-sampling of data, which in turn can offer benefits while modeling.

Time series data often requires cleaning, scaling, and even transformation.

For example:

- **Frequency**. Perhaps data is provided at a frequency that is too high to model or is unevenly spaced through time requiring resampling for use in some models.
- **Outliers**. Perhaps there are corrupt or extreme outlier values that need to be identified and handled.
- **Missing**. Perhaps there are gaps or missing data that need to be interpolated or imputed.

Often time series problems are real-time, continually providing new opportunities for prediction. This adds an honesty to time series forecasting that quickly flushes out bad assumptions, errors in modeling and all the other ways that we may be able to fool ourselves.

Examples of Time Series Forecasting

There is almost an endless supply of time series forecasting problems.

Below are 10 examples from a range of industries to make the notions of time series analysis and forecasting more concrete.

- Forecasting the corn yield in tons by state each year.
- Forecasting whether an EEG trace in seconds indicates a patient is having a seizure or not.

- Forecasting the closing price of a stock each day.
- Forecasting the birth rate at all hospitals in a city each year.
- Forecasting product sales in units sold each day for a store.
- Forecasting the number of passengers through a train station each day.
- Forecasting unemployment for a state each quarter.
- Forecasting utilization demand on a server each hour.
- Forecasting the size of the rabbit population in a state each breeding season.
- Forecasting the average price of gasoline in a city each day.

I expect that you will be able to relate one or more of these examples to your own time series forecasting problems that you would like to address.

Summary

In this post, you discovered time series forecasting.

Specifically, you learned:

- About time series data and the difference between time series analysis and time series forecasting.
- The constituent components that a time series may be decomposed into when performing an analysis.
- Examples of time series forecasting problems to make these ideas concrete.

Do you have any questions about time series forecasting or about this post?
Ask your questions in the comments below.