Conor Muldoon

Professor Goettsche

CSCI 240

September 25, 2018

Project Requirements: Time Series Analysis

Preface

In this document, the system requirements for the time series analysis algorithm will be laid out in greater detail. The current version of this algorithm is 1.0, and this document will contain the initial details for set up.

Introduction

For this software, the time series analysis will be one of the two main algorithms that we will be using to produce forecasting results. The main function of this algorithm is the estimation of coefficients for the weights of historical values and error terms. Once the coefficients for each term is calculated, we may then use the algorithm to forecast the time series *n* steps into the future, where *n* is a natural number.

The time series algorithm will retrieve data from the online database to use for the analysis. After the specific stock is selected by the user, the algorithm will extract the time series data and perform coefficient estimation. The forecasted outputs, in addition to the confidence intervals, will then be submitted to the machine learning algorithm of the software.

Glossary

Stationary Process - A process is stationary if all aspects of its behavior are independent of shifts in time.

Gaussian White Noise - If we have a series of independent normally distributed random variables, then the process is called a Gaussian white noise process.

Autoregressive Process - autoregressive processes follow a regression model where the next observation is the response variable and where the past values of the random process are the explanatory variables.

Moving Average Process - A process is a moving average process if the next observation in the series can be expressed as a weighted average of the past values of the white noise process, rather than the past values themselves.

User Requirements Definition

When the software is running, this module will not typically be interacting with the user, but if the user does wish to just perform a time series analysis, then these requirements must by documented.

In order for the time series analysis to function, the user will need to submit a .csv file of the historical values of the stock. Typically, this will include two columns, a time column and a price column. The user must also ensure that the price measurements were taken at constant time intervals, or else the assumptions held by the autoregressive process will be broken. Another input for the algorithm will be some natural number *n*, which determines the number of steps into the future to be forecasted.

The outputs of the algorithm will very simplified for the user. This will output a simple .csv file with point forecasts *n* steps into the future. This will also include a 95% confidence interval for each forecasted value.

System Architecture

This software will be composed of four separate modules, including the database, time series algorithm, machine learning algorithm, and the front-end development.

Specifically, for the time series algorithm, this function will be communicating directly with the database and the machine learning algorithm. Since the data is required for the analysis, this module will be utilizing data from the database in the form of automatic updating .csv files. From this, the output of the algorithm will be directly inputted to the machine learning algorithm, which utilizes the forecasted values with sentiment analysis. This module will have no direct communication with the front-end development since the machine learning algorithm output will contain the final forecasted values.

System Requirements Specification

The time series analysis algorithm will fit an ARIMA model to the data in order to minimize error terms. Once the coefficients are estimated, we simply utilize these coefficients with the last observed values to forecast into the future.

System Models

For this algorithm, we will be using an ARIMA model. This model utilizes autoregressive processes, moving average processes, and differencing to meet the stationary model requirements. For each time series, we will use a model estimator to determine how many steps back of every historical value that we will use in the equation. This will also help us to determine how many error terms we are willing to include in the model.

Model estimation is an important component in the time series algorithm. This is because processing time will increase exponentially as more terms are weighted in the final forecasts.

System Evolution

From the system model selection, this portion may need to evolve in order to optimize processing times. While inclusion of more Gaussian error terms and historical observations would provide more accurate predictions, the processing time is a real concern. It will take some time to determine which model type will give us the best balance of processing time and accuracy.