Forecasting Temperature using ARIMA Models

Louis Vocatura

Departmemt of computer engineering

Wentworth Institute of technology

Boston MA

[vocatural@wit.edu](mailto:vocatural@wit.edu)

Abstract

This paper discusses the importance of data science and attempts to use an ARIMA large language model to predict weather patterns. Many experiments using these LLMs are described in this paper along with a summary of the literature and how this work contributes.

Introduction

Data is extremely important in life. I believe that within everything, analyzing relevant data creates clear benefits within that particular study. One aspect of life that is influential to everyone is the weather. Weather is important to be able to forecast as it affects many things. When we can predict the weather, it becomes much easier to prepare for disasters or traveling. Weather can be well predicted using data and algorithms. Analyzing past weather patterns, it becomes easier to predict future weather.

Now with AI, machine learning and LLMs, we can much more easily use past data to create predictions and analyze data patterns more efficiently and efficiently. Looking through the many new options the ARIMA model uses past values of the time series to forecast future values. It assumes that the current value is a linear combination of previous values. This methodology is perfect for predicting weather patterns and is being used for the project.

Methodology

The idea is to look at and graph past data and use that past data to train an LLM. Training an LLM with our weather data allows that LLM to make predictions based on the training data. Telling the LLM to predict the weather tomorrow will have the LLM look at past data and make a guess as to what the weather may be. Being able to create these raw predictions can be useful since we can compare those predictions to the predictions of the media and see how they compare. We can compare our findings to the actual weather as well after the set day. It should be possible to analyze how effective predicting the weather with LLMs is using this technique.

The raw data acquired will be daily since that will be enough to capture daily trends and long term trends as well. Using Python libraries we can manipulate the data into a more graphable form for proper analysis.

It might be difficult for an LLM to predict the weather based on the fact that weather is not an inherently numerical idea and has many contributing factors. It will be good to see how an ARIMA model can use past data to make calculations since large data sets won't be easy to manipulate by hand.

Data Sets

There are many feasible options for acquiring past weather data. Since we are going to be using python for our overall analysis, finding a good API that allows easy integration of data into a python environment is key. After some research, Tomorrow.io is a good choice for API data. Tomorrow.io allows direct downloads of weather data in many categories like temperature, rain, humidity and sun exposure. Using Tomorrow.io allows easy downloading of weather data in a desired format that will make transformation easy.

After downloading the desired CSV files for our weather data, there are many options for data transformation that we can use to better represent our data. Numpy and Pandas python libraries help our manipulation as we can convert a CSV into a Pandas data frame which is made to be used with python. Using MatPlotLib can finally let us graph our transformed data in a much more readable form. The neat thing about Tomorrow. IO is that you can download CSV weather predictions as well, and should be in a similar format. Using these various libraries along with our data sets and LLMs, showing a perfect relationship between past, future and present data having been calculated in different ways.

Having accurate forecasting data from Tomorrow IO is helpful, but that isn't the only data that will be useful for this study. Having historical data will allow us to attempt to forecast past data and compare it to what that temperature actually was at that point. Getting this data from the National Weather Service will prove worthwhile.

Data gained for this experiment will have many different pieces of information that won't be able to be inserted into an ARIMA model. We would only be able to use one type of training data for an ARIMA model such as temperature. Both historical and forecasted data will be guaranteed to have temperature which makes it most suitable for ARIMA analysis.

.