

Time-series Global Temperature Prediction and Visualization

Liuyiyan Qiao *

Department of Physics, Zhejiang University, Hangzhou, China

* Corresponding author email: 3180105905@zju.edu.cn

Abstract. Temperature prediction is a prevalent and practical research in recent years. However, due to the performance of mathematical models or the application of visualization technology, previous work either lacked accuracy and intuitively visualized results. This paper designs a pipeline for global temperature forecast, which has relatively ideal accuracy and visualized results, and makes prediction of the temperature of each country/ region from 2023 to 2100 based on the earth surface temperature dataset packaged by The Berkeley Earth Surface Temperature Study. The mathematic model used by this paper is Prophet, which is a high-performance time series forecasting model and extremely applicable for dataset with high periodicity. The forecast results are generated by using PyEcharts in the form of HTML pages, where the worldwide average temperature, the years, a drag-operable month setting bar and the colored temperature distribution map are shown intuitively. Experimental results based on the testing dataset indicate that the mean absolute error of this method is 0.735°C . Forecast results on wider dataset indicate that the worldwide average temperature in the future may continue growing at the speed of around 0.0248°C per year.

Keywords: Temperature; Prediction; Data Visualization; Time-series Forecast.

1. Introduction

The visualization and analysis of data is a prevalent computer technology, which has been used widely in various fields e.g., natural science, statistics, commercial operation and computer vision [1-3]. The greenhouse effect is now having an increasingly significant impact on both our daily lives and the future of humanity. As the environmental pollution grows, the temperature of whole world has dramatically increased, with some disaster like drought, flood and glaciers melting followed. Therefore, it is important to anticipate the world temperature for the future with some degree of accuracy. This will enable geographers and meteorologists to develop quick and efficient solutions to stop the worsening effects of climate change.

One prevalent method of predicting temperature is using the linear regression model, which is convenient to use and very fast for computer to calculate [4, 5]. However, the temperature changing has some unique features which cannot be dealt with precisely and credibly by using linear regression model. Firstly, the data of temperature changing always has some outliers, which cannot be avoided in the record from real world, and will make large influence on the prediction result. Additionally, the changing of yearly temperature is not totally linear, so the credibility of this method is questionable. Finally, as the changing of temperature has high periodicity among different years, it is hard to analyze the monthly changing of temperature by using linear regression model. In the early field of temperature prediction, the previous technology has given out some general prediction of both surface and ocean temperature using more advanced mathematical model [6, 7]. However, these traditional researches also have some disadvantages. For example, the results usually reflected a long-term trend of temperature change, rather than more subdivided one like monthly or even yearly. Besides, Although the authors provided the prediction results in form of colormaps, they described the trends of temperature changing in a period of time, rather than specific point of time like a certain year or month. Article used time-series model and gave out the prediction of monthly global average temperature in the form of graph [8], however, it did not distinguish different countries or regions, nor did it visualize the data to make forecast results more intuitive.

Considering the above issues, this study designed a predictive system to make more accurate, intuitive prediction of monthly global temperature by using Prophet [9], which is a widely-used and highly credible time-series forecast model. After that this study evaluated the accuracy of the

predictive model and analyzed the trend of climate change by using the forecast results. Finally, all the forecast results are organized and visualized in form of colormaps.

The rest of this paper is organized as follows: Section 2 provides the implemented structure of the project, including the approach method and some details about the dataset. The result of the prediction and the evaluation of the accuracy as well as the discussion will be given in Section 3. The conclusions of this work and further work are discussed in Section 4.

2. Method

The project used the earth surface temperature dataset from Kaggle [10]. The research of long-term temperature trend always has its complexity due to the uncertainty of historical means of measurement. Before the electronic thermometers were widely used, temperature data was always collected by mercury thermometers, which will be influenced by the observation time. Therefore, The Berkeley Earth Surface Temperature Study combines 1.6 billion temperature reports from 16 pre-existing archives, and packaged them to a new dataset. The dataset contains the monthly average temperature of each country or region from 18th century to 2013, the uncertainty of the temperature in each month, and has been sliced by country or region, which is convenient for the project to preprocess and analyze.

Prophet is an open-source time series prediction algorithm from Facebook, it is a piecewise linear or logistic growth curve-tipped additive regression model. It has a weekly seasonal component that is represented using dummy variables and a yearly seasonal component that is modeled using Fourier series. Therefore, Prophet is particularly helpful for datasets that contain a long time period of detailed historical observations (month or years) at hourly, daily, or weekly intervals, those that have multiple strong seasonality, include known significant but irregular events, have missing data points or large outliers, or have non-linear growth trends that are approaching a limit. Prophet is utilized in numerous Facebook applications to generate accurate forecasts for planning and goal-setting. In most instances, the prophet is found to perform better than any alternative strategy. The algorithm provides two implementations based on Python and R.

To visualize the forecast results, the PyEcharts was chose to get the colored temperature distribution map of each month. PyEcharts is a class library for generating Echarts. Echarts is a data visualization JS library open sourced by Baidu. The visualization of graphs generated by Echarts is excellent. PyEcharts is designed to interface with Python, so that it is convenient to use data directly in Python to generate graphs. Using PyEcharts, independent web pages can be generated in this study, and can also be integrated in flask and Django.

The Figure 1 below shows the framework of this project:

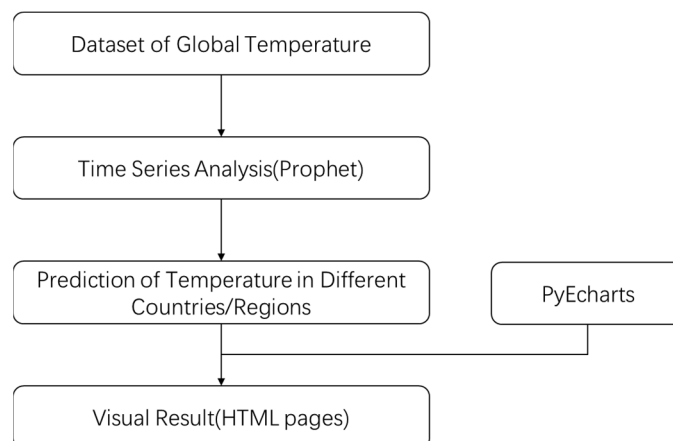


Fig 1. Project Framework

First of all, two arrays which contains the name of every country/region are created, according to the preprocessed dataset and the PyEcharts API. The elements in them form one-to-one correspondence and only have slight difference in appellation.

Considering the inaccuracy of early temperature data and the computational complexity, this study discarded the data before 1913, by only reading the last 1200 rows of data of each country or region from original csv file. Then this study used the DataFrame module of pandas to convert the data into DataFrame format for the convenience of next steps. After that, a formatted DataFrame which contains each month from 2023 to 2100 in one column was generated. It represents the target time period of prediction and can be changed to any period after 2013 theoretically. Then the prediction module of Prophet is called, whose parameter is the empty DataFrame mentioned before, to get the prediction of futural temperature from 2023 to 2100. The result is also given in the form of array whose elements are the predicted temperature of each month in that period of time. After that, a loop structure is added to the project and the forecast temperature series are saved in a 2-dimension array which represents each country/region and its corresponding forecast result. Finally, the module of PyEcharts is called to render the forecast result visually. A timeline was created to show different monthly results in each year, and the result is rendered by PyEcharts in the form of html pages.

3. Result and Discussion

A pre-experiment was designed to check the accuracy and credibility of the method used in this project. A period from 1913 to 2013 in the dataset is picked up as the test set, and it is separated to two parts: the period from 1913 to 2003 and the rest of it. The former period is used to give a forecast in the later one and compare the result with the real data from original dataset. According to the experimental data, the mean absolute deviation of the forecast result is 0.735°C shown in Figure 2.

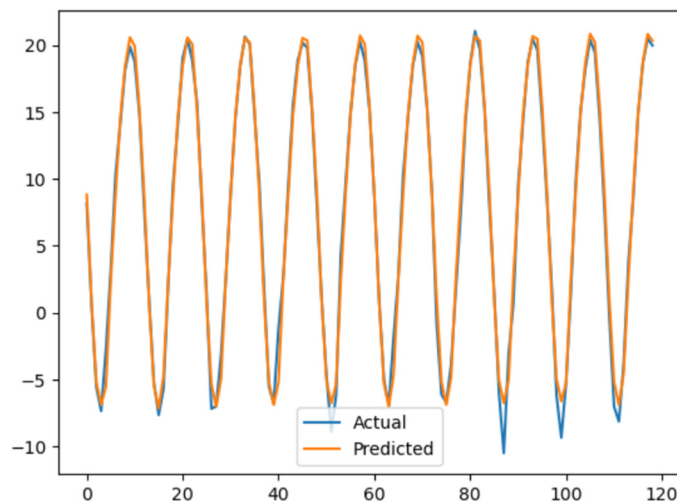


Fig 2. Result Comparison between the actual and predicted value.

After the program gave the forecast result from 2023 to 2100, the average temperature of the whole world is calculated, and is made to a line graph which shows the overall tendency of the average worldwide temperature. The average slope of global temperature is 0.0248°C per year shown in Figure 3.

Considering the aesthetics of the HTML pages shown in Figure 4, they are divided by different years, and each of them contains the monthly average temperature of each country/region. The specific year and the average worldwide temperature in that year are shown in the upper left corner of the pages, and the temperature difference among each country and region is shown by the disparity of RGB color in the distribution map. Besides, the axis below the page represents different months in one year, and can be dragged by users to adjust the month they want to see.

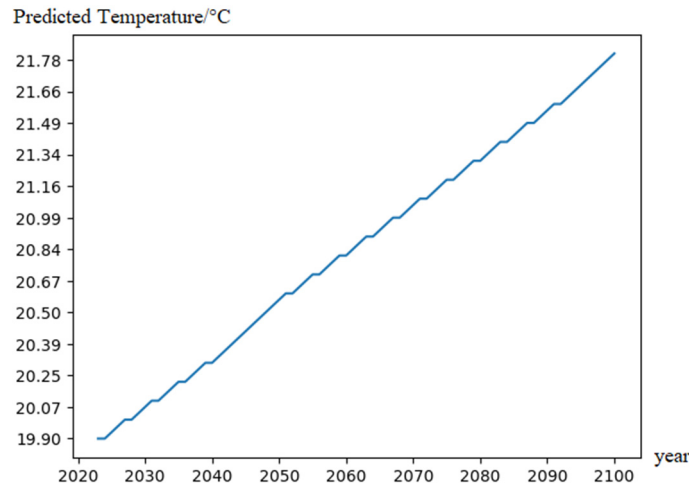


Fig 3. Worldwide Average Temperature Tendency

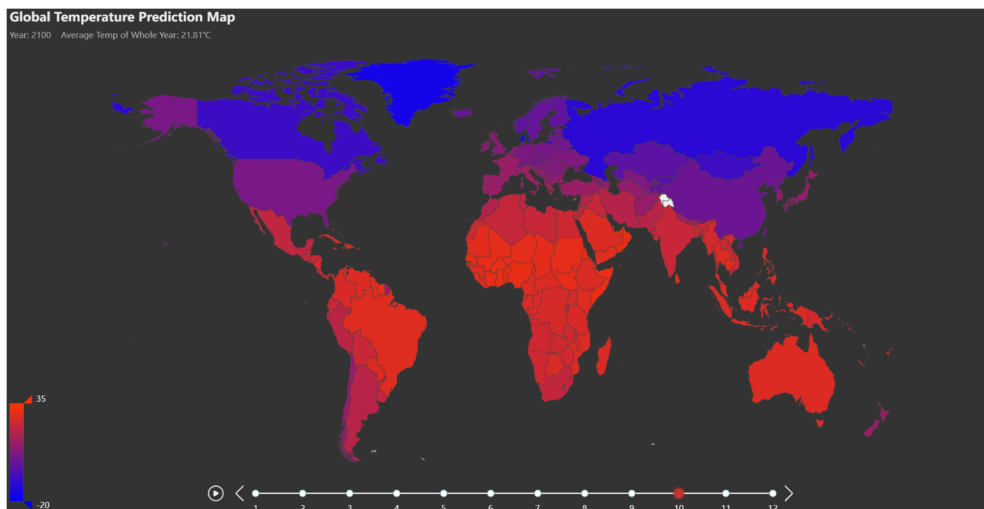


Fig 4. Colored Distribution Map

4. Conclusion

This study improves the method of temperature prediction by using the mathematical model which has more accuracy and reliability compared to the previous work, and visualize the forecast result in the form of colored distribution map, which has higher aesthetics and readability than the previous work. The project uses the Prophet, an epidemic time-series forecast model to complete the prediction part, and develops a series of HTML pages with interactive interfaces and high readability by using PyEcharts. The mean absolute deviation of the forecast result using the specific sample mentioned before is 0.735°C, and the average slope of worldwide predicted temperature is 0.0248 °C per year from 2023 to 2100. However, although the accuracy of Prophet is relatively ideal for the experimental sample, other time-series forecast models such as ARIMA, Holt-Winters, RNN, LSTM, and DeepAR haven't been compared with current one limited to the time and workload, which could be improved in the futural work.

References

- [1] Kunnathuvalappil Hariharan, Naveen. Applications of Data Mining in Finance. Naveen Kunnathuvalappil Hariharan. APPLICATIONS OF DATA MINING IN FINANCE. International Journal of Innovations in Engineering Research and Technology 5.2, 2018, 72-77.
- [2] Q, Yu et al. Pose-guided matching based on deep learning for assessing quality of action on rehabilitation training. Biomedical Signal Processing and Control 72, 2022, 103323.

- [3] Cios, Krzysztof J., and G. William Moore. Uniqueness of medical data mining. *Artificial intelligence in medicine* 26.1-2, 2002, 1-24.
- [4] Weisberg, Sanford. *Applied linear regression*. Vol. 528. John Wiley & Sons, 2005.
- [5] Seber, George AF, and Alan J. Lee. *Linear regression analysis*. John Wiley & Sons, 2012.
- [6] Keenlyside N S, Ba J. Prospects for decadal climate prediction. *Wiley Interdisciplinary Reviews: Climate Change*, 2010, 1(5): 627-635.
- [7] Hazeleger W, Guemas V, Wouters B, et al. Multiyear climate predictions using two initialization strategies. *Geophysical Research Letters*, 2013, 40(9): 1794-1798.
- [8] Ye L, Yang G, Van Ranst E, et al. Time-series modeling and prediction of global monthly absolute temperature for environmental decision making. *Advances in Atmospheric Sciences*, 2013, 30(2): 382-396.
- [9] Taylor S J, Letham B. Forecasting at scale. *The American Statistician*, 2018, 72(1): 37-45.
- [10] Kaggle. Climate Change: Earth Surface Temperature Data. <https://www.kaggle.com/datasets/Berkeley-earth/climate-change-earth-surface-temperature-data>, 2016.