Singapore Temperature Predictor Report

Submitted as part of the requirements for Assessment 1 of the ACSE-5 module.

Developed by Deborah Pelacani Cruz and John C Walding.

Introduction

The aim of this project was to implement a numerical model to predict temperature variations for a specific city using available data from the internet. Singapore was chosen for two main reasons: 1) its significant increase in global economic presence and 2) the online availability of temporal data in a friendly format. Temperature predictions are made using a multivariable linear regression of average surface temperature and population data in the city of Singapore. The data used for this project was obtained from the Changi Climate Station [1] in Singapore, for years 1982 - 2018 and the World Bank Data[2], for years 1982 -2017.

Software Structure

The solution is a C++ executable, organised into functions held in singapore_temperature.cpp and using four classes: LinearRegression, PopRegression, GNUPlot and Interface. GNUPlot plots an arbitrary number of data sets saved in .csv format, whilst Interface handles the UI, a branching path interface. Linear and Pop Regression handle the linear regression calculations for simple and multivariable regressions, respectively. The function main() initialises the Interface class and from there the other classes and their respective methods are called in order to produce the final model. The functions in singapore_temperature are utility functions, and are not related to the model.

Using the Software

Prior to operation, the datasets used must be in a particular .csv format. As a result this software is not compatible with new data without prior preparation. The UI presented by the Interface class is very self explanatory, and guides the user through operation at runtime. The primary option is whether to take population data into account for a multivariable regression, or to perform a simple linear regression. After this, there is the option to choose the years over which the data should be analysed, and the years for which the data should be extrapolated. Finally, there are several ways the software can output this data, and the raw data can always be found in the processed_data/ folder in the code directory.

Model Analysis

The model used in this solution is linear regression, either simple or multivariable. In the simple case, the software will calculate the regression from historical surface air temperature in Singapore. For the multivariable case, this dataset will be supplemented by Singaporean population data, as it is suspected that there will be a correlation between the two, due to an increased population leading to increased greenhouse gas emissions and population density. In both cases the algorithm used is simple least-squares regression.

Limitations and Further Work

The veracity of the predictions provided is limited due to the crudity of linear regression as a predictor. Whilst a levelling-off of the temperature is expected, a linear model will provide an infinite increase of temperature with time. The efficiency of the software is also not optimised: there is opportunity to remove unnecessary code repetition and to improve memory storage. However, the effect of this is attenuated by the limited size of the datasets. Another possible improvement would be to add emissions and global temperature variation data to correlate with the current model.

References

1. 2019. *Weather.Gov.Sg*. http://www.weather.gov.sg/home/.
2. 2019. *World Bank Open Data*. https://data.worldbank.org/.