Predicting Peak Bloom Date of Cherry Trees With Classification Model and Neural Network

Yanyu Chen, Hongyi Zheng, Zihan Zhang

Introduction

Our project aims to predict the exact peak blossom date of Cherry trees in the future, given the sequential daily average temperatures records in the past using machine learning techniques. We will simplify the prediction as a binary classification problem, in which the peak date is the positive label (1) and other ordinary blossom dates are negative labels (0). Our interest in the relationship between temperatures and peak cherry blossom date is intrigued by Zhang's observation – among all meteorological features, daily average temperature correlates to the first flowering date and full flowering date of ornamental plants most strongly (Zhang et al., 2010[1]). Ideally, we hope our prediction model could be useful in providing tourism guidance and pollen season alerts or possibly inspiring agricultural planting and inducing financial benefits.

Objectives

We will focus on the peak bloom date of cherry trees in Kyoto and Washington D.C., two cities renowned for their amazing cherry blossom festival. Our primary dataset is twofold for each city: peak bloom dates dataset and daily weather information dataset.

There are two datasets of peak bloom dates. First is the flowering date data provided by Yasuki Aono from Osaka Prefecture University, which records the yearly full-flowering date (peak bloom date) of cherry trees in Kyoto since 810AC. The second is the climate-indicators-cherry-blossom from the United States Environmental Protection Agency, which records the annual peak bloom date of Washington, D.C.'s famous cherry trees from 1921 to 2016. These peak bloom date data will serve as labels for our classification algorithm. (For the Kyoto flowering data we will only use the data in the near 100 years because we are unable to find temperature records in the ancient time). On the other hand, the detailed historical temperature data are from the Japan Meteorological Agency and the U.S. National Oceanic and Atmospheric Administration. They both include multiple daily weather features. To predict the bloom date, we mainly use the daily temperature data in the first 3-4 months of each year as primary input data to train our machine learning model.

Referring to Cifuentes' holistic overview of relevant techniques used in temperature forecasting (Cifuentes et al, 2020[2]), we plan to use RNN, SVM, and possibly other classification models to predict the peak bloom date of cherry trees in Kyoto and Washington D.C., then generalize the model to other

cities. Meanwhile, regarding Mi's essay (Mi et al., 2010[3]). After Mi applied the back-propagation algorithm to predict the accumulated temperature in the vegetation period for the target extension place. The observed overall performance became more accurate compared to simple multiple linear regression. So we would also try this algorithm. Generally, a comprehensive analysis of performance across different machine learning models will be conducted later.

References

[1] Aiying Zhang, Huanjiong Wang, Junhu Dai, and Deping Ding. Applicability analysis of phenological models in the flowering time prediction of ornamental plants in beijing area. *Journal of Applied Meteorological Science*, 25:483–492, 06 2014.

[2] Jenny Cifuentes, Geovanny Marulanda, Antonio Bello, and Javier Reneses. Air temperature forecasting using machine learning techniques: A review. *Energies*, 13(16), 2020.

[3] Chunqiao Mi, Jianyu Yang, Shaoming Li, Xiaodong (Frank) Zhang, and Dehai Zhu. Prediction of accumulated temperature in vegetation period using artificial neural network. *Mathematical and Computer Modelling*, 51:1453–1460, 06 2010.