Use of new generation geospatial data and technology for low cost drought monitoring and SDG reporting solution

A thesis presented in partial fulfillment of the requirement for the degree of

Master of Science

in

Computer Science

at Massey University, Manawatū, New Zealand.

Mohammad Hossain Dehghan-Shoar

2018

Abstract

Food security is dependent on ecosystems including forests, lakes and wetlands, which in turn depend on water availability and quality. The importance of water availability and monitoring drought has been highlighted in the Sustainable Development Goals (SDGs) within the 2030 agenda under indicator 15.3. In this context the UN member countries, which agreed to the SDGs, have an obligation to report their information to the UN. The objective of this research is to develop a methodology to monitor drought and help countries to report their findings to UN in a cost-effective manner.

The Standard Precipitation Index (SPI) is a drought indicator which requires long-term precipitation data collected from weather stations as per World Meteorological Organization recommendation. However, weather stations cannot monitor large areas and many developing countries currently struggling with drought do not have access to a large number of weather-stations due to lack of funds and expertise. Therefore, alternative methodologies should be adopted to monitor SPI.

In this research SPI values were calculated from available weather stations in Iran and New Zealand. By using Google Earth Engine (GEE), Sentinel-1 and Sentinel-2 imagery and other complementary data to estimate SPI values. Two genetic algorithms were created, one which constructed additional features using indices calculated from Sentinel-2 imagery and the other data which was used for feature selection of the Sentinel-2 indices including the constructed features. Followed by the feature selection process two datasets were created which contained the Sentinel-1 and Sentinel-2 data and other complementary information such as seasonal data and Shuttle Radar Topography Mission (SRTM) derived information.

The Automated Machine Learning tool known as TPOT was used to create optimized machine learning pipelines using genetic programming. The resulting models

yielded an average of 90 percent accuracy in 10-fold cross validation for the Sentinel-1 dataset and an average of approximately 70 percent for the Sentinel-2 dataset. The final model achieved a test accuracy of 80 percent in classifying short-term SPI (SPI-1 and SPI-3) and an accuracy of 65 percent of SPI-6 by using the Sentinel-1 test dataset. However, the results generated by using Sentinel-2 dataset was lower than Sentinel-1 (45 percent for SPI-1 and 65 percent for SPI-6) with the exception of SPI-3 which had an accuracy of 85 percent.

The research shows that it is possible to monitor short-term SPI adequately using cost free satellite imagery in particular Sentinel-1 imagery and machine learning. In addition, this methodology reduces the workload on statistical offices of countries in reporting information to the SDG framework for SDG indicator 15.3. It emerged that Sentinel-1 imagery alone cannot be used to monitor SPI and therefore complementary data are required for the monitoring process.

In addition the use of Sentinel-2 imagery did not result in accurate results for SPI-1 and SPI-6 but adequate results for SPI-3. Further research is required to investigate how the use of Sentinel-2 imagery with Sentinel-1 imagery impact the accuracy of the models.

Acknowledgements

I would like to thank my kind supervisors Prof. Hans Guesgen, Dr. Sunil Lal and Dr. Lorenzo De Simone for their guidance and support during my study. In addition I would like to thank my family members for helping me in my time of need in particular my father and mother for their kind support.

Contents

List of Tables

List of Figures