

https://dspace.org/

Data-driven sensitivity mitigation techniques for genetic algorithm - long short term memory water quality prediction model

Dheda, Dhruti

2021
https://hdl.handle.net/10539/33342

Downloaded from a DSpace repository. https://dspace.org/

Data-driven sensitivity mitigation techniques for Genetic Algorithm - Long Short Term Memory water quality prediction model

Dhruti Dheda, 458389



UNIVERSITY OF THE
WITWATERSRAND,
JOHANNESBURG

School of Electrical & Information Engineering
University of the Witwatersrand, Johannesburg, Gauteng

April 28, 2021

A dissertation presented for the degree of Masters of Electrical and Information Engineering

Abstract

A long short-term memory (LSTM) model developed for the prediction of water quality, based on the historical data of a particular water body, and as such a particular water quality dataset, will only be applicable to that dataset. Thus if a specific LSTM prediction model is applied to another dataset, then it is quite possible that the prediction model will fail to make an accurate prediction. These models tend to be case study specific. This research focuses on improving the tolerance (mitigating the discrepancies in model prediction capability that arise from differences in datasets) of LSTM prediction models. The two different LSTM models developed from two different water quality datasets, the Burnett and Baffle models, are optimised using the metaheuristic genetic algorithm (GA). The two hybrid GA-optimised LSTM base models, the GA-Burnett and GA-Baffle models, are fused together using a weight-based approach to form a final robust and tolerant predictive ensemble model. Both the models contribute equally to the average ensemble model. In the weighted ensemble model, the GA-Burnett model only has a 10% greater contribution than the GA-Baffle model.

Generally, the ensemble models outperform the GA-optimised hybrid LSTM models. The four models are tested on unseen and unrelated datasets and the performance of all the models are consistently similar to one another on each dataset. The consistency of performance exhibited by the different models on any particular dataset is evidence of the successful mitigation of the discrepancies of the individual LSTM models through the implementation of the linear weight

based fusion of two hybrid GA-optimised LSTM models. The models are not only applicable for the prediction of water quality, but also for domains outside of the water sector; thus asserting the relevance of the models, especially the weighted ensemble model in the wider field of LSTM and ensemble prediction.

This research involves the water quality of rivers. Water is a critical natural resource that is currently under threat, especially rivers. The models are able to successfully predict the quality of river water ahead of time, in terms of dissolved oxygen concentration. Water quality prediction aids in increasing the efficiency of water quality monitoring. Efficient water quality monitoring enables effective water management. Effective water management is necessary for the preservation of rivers.