Accepted Manuscript

Title: Decomposition Genetic Programming: An Extensive Evaluation on Rainfall Prediction in the Context of Weather Derivatives

Author: Sam Cramer Michael Kampouridis Alex A. Freitas

PII: S1568-4946(18)30279-5
DOI: https://doi.org/doi:10.1016/j.asoc.2018.05.016
Reference: ASOC 4882

To appear in: *Applied Soft Computing*

Received date: 14-7-2017
Revised date: 10-5-2018
Accepted date: 11-5-2018

Please cite this article as: Sam Cramer, Michael Kampouridis, Alex A. Freitas, Decomposition Genetic Programming: An Extensive Evaluation on Rainfall Prediction in the Context of Weather Derivatives, *<![CDATA[Applied Soft Computing Journal]]>* (2018), https://doi.org/10.1016/j.asoc.2018.05.016

Decomposition Genetic Programming: An Extensive Evaluation on Rainfall Prediction in the Context of Weather Derivatives

Sam Cramer, Michael Kampouridis, Alex A. Freitas[1]

School of Computing, University of Kent

Abstract

Regression problems provide some of the most challenging research opportunities in the area of machine learning, where the predictions of some target variables are critical to a specific application. Rainfall is a prime example, as it exhibits unique characteristics of high volatility and chaotic patterns that do not exist in other time series data. Moreover, rainfall is essential for applications that surround financial securities, such as rainfall derivatives. This paper extensively evaluates a novel algorithm called Decomposition Genetic Programming (DGP), which is an algorithm that decomposes the problem of rainfall into subproblems. Decomposition allows the GP to focus on each subproblem, before combining back into the full problem. The GP does this by having a separate regression equation for each subproblem, based on the level of rainfall. As we turn our attention to subproblems, this reduces the difficulty when dealing with data sets with high volatility and extreme rainfall values, since these values can be focused on independently. We extensively evaluate our algorithm on 42 cities from Europe and the USA, and compare its performance to the current state-of-the-art (Markov chain extended with rainfall prediction), and six other popular machine learning algorithms (Genetic Programming without decomposition, Support Vector Regression, Radial Basis Neural Networks, M5 Rules, M5 Model trees, and k-Nearest Neighbours). Results show that the DGP is able to consistently and significantly outperform all other algorithms. Lastly, another contribution of this work is to discuss the effect that DGP has had on the coverage of the rainfall predictions and whether it shows robust performance across different climates.

Keywords: Weather derivatives, rainfall prediction, problem decomposition, genetic programming, genetic algorithm

1. Introduction

Regression based problems provide a unique challenge for researchers, where the prediction of outputs have a pivotal outcome in real-life problems. The complexity can be overcome through specific domain knowledge, but often this is not the case. Within complex and chaotic time series data, there is a lack of reoccurring patterns and domain knowledge can be scarce. A type of time series, which remains one of the most difficult and crucial to applications, is rainfall. This time series contains high volatility, little to no seasonality and is highly random. The effects of rainfall can lead to devastation, and unfavourable conditions can impact societies' and ecosystems' ability to survive.

The phenomenon of rainfall has a direct impact on various domains such as water resource planning, agriculture and biological systems. Within finance, predicting the level of rainfall is important for protecting an individual's income from the adverse rainfall effects. Over the years people have sought means of protecting their day-to-day income from unfavourable rainfall, but only until more recently has this been possible. Insurance from rain's adverse effects has existed for many years, but often is of little use unless the impact is of high catastrophe, causing destruction. For instance, a farmer would only be able to receive compensation if s/he could demonstrate destruction of their crop, e.g. because of a severe flood. However, such business can also be affected by unfavourable rainfall, which is not

[1]Corresponding author: Michael Kampouridis, School of Computing, Medway, ME4 4AG, UK. Tel: +44 1634 88 8837. Email: M.Kampouridis@kent.ac.uk

Preprint submitted to Applied Soft Computing *May 14, 2018*