Multiobjective optimization of rainfall prediction using GA based on three cross-over method

Nelson Ogbeide, ID: 22146118 [*1]

Word Count: 999

1. Problem Domain

Rainfall has several advantages for humanity, including those related to agriculture, energy generation, and the environment, all of which are necessary for human survival and influence many facets of human life. Yet, excessive or unforeseen rainfall can result in hazardous situations like flooding if it happens frequently. Almost 2.8 billion people worldwide are affected by floods, which have claimed over 200,000 lives in the past 2 decades. (**?**). Worldwide, the severity and frequency of flooding have significantly increased during the last few decades. This pattern has been influenced by urbanisation, land use changes, and climate change. Extreme precipitation events are growing more frequent and intense, according to the Intergovernmental Panel on Climate Change (IPCC), and this trend is likely to continue in the future. The 2015 Chennai flood in India is a prime example of the devastating effects of inaccurate rainfall forecasts; the city had its heaviest precipitation in almost a century, resulting in catastrophic floods and the deaths of over 400 people.(**?**).

Accurate forecasts of rainfall and quick drought-flood transitions are necessary at different times to develop management plans intended to minimize their negative effects on society and the economy. (**?**). In 2008, a forecast announced a high risk of above-normal rainfall for the July-September rainy season in West Africa, which led the International Federation of Red Cross and Red Crescent Societies(IFRC) to preposition disaster relief items, update flood contingency plans, and alert vulnerable communities, resulting in a decrease in flood-related losses compared to the previous year (**?**).

Accurate rainfall forecasting is now more challenging than ever due to significant weather changes. (**?**), Data mining and artificial intelligence techniques such as support vector machines (SVM), Artificial Neural Networks (ANN), and multilayer perceptrons (MLP) have become increasingly important for predicting rainfall occurrences because of their ability to handle large weather data and discover patterns. This study compared the effectiveness of three GA crossover methods in predicting whether rainfall will occur the following day.

[1]M.Sc. Big Data Analytics, School of Computing and Digital Technology, Birmingham City University, UK. Correspondence to: Nelson ogbeide <nelson.ogbeide@mail.bcu.ac.uk>.

Modern Optimisation Report, CMP 7213 - Modern Optimisation.

2. Problem Instance

The dataset used in this study is the "Rain in Australia" dataset available on Kaggle, which spans ten years from 2008-12-01 to 2017-06-25. The dataset contained weather observations from numerous weather stations across Australia. The dataset contains various features such as rainfall, temperature, humidity, and wind speed. The data contained 142,193 observations with 23 features. This study used a multi-objective optimization technique to determine the optimal accuracy and subset of features for predicting the occurrence of rainfall will occur or not.

Support vector machine is the ML algorithm chosen for this study owing to its implementation by researchers (**???**) in weather prediction. The fitness function used in this study maximizes accuracy while minimizing the number of features. (**?**) used a similar fitness function for minimizing classifier error and minimizing features. This fitness function is necessary because the inclusion of irrelevant features can lead to overfitting and reduced model performance.

$$Fitness = (W * Accuracy) * (1 - W)\frac{\sum_i f_i}{n} \quad (1)$$

Where *Accuracy* is the mean of the actual value compared to the predicted value, *w* is the weight attached to the accuracy and number of features selected, f_i is 1 when feature i is selected, and *n* is the total number of features in the dataset.

3. Methodology

Several techniques have been developed to solve feature selection problems using evolutionary algorithms. Some of these are genetic algorithms, particle swarm optimization, and ant colony optimization (**?**). In this study, genetic algorithm (GA) was applied to solve the feature extraction problem.

? used Genetic algorithms and feature selection techniques using the uniform cross-over method for more effective and efficient data mining. They propose a genetic algorithm approach that simultaneously performs data mining and feature selection by evolving a binary code alongside the chromosome structure. Results demonstrate that the feature selection method yields the best performance in terms of prediction accuracy and computational efficiency. (**?**) presents optimized feature selections for intrusion detection systems on the NSL-KDD Cup 99 data. Results demonstrate that the genetic algorithm using the one-point crossover on a random forest classifier produced the best result. **?** applied genetic algorithm for feature selection with two-point cross-over method in image annotation system. The system employed MPEG-7 image feature descriptors to represent low-level image features such as color, texture, and shape, which experienced a 13.6% improvement in the accuracy score. (**?**) utilized a feature selection technique for computer-aided diagnosis of skin tumours in dermatology using high-resolution skin surface profiles. The study compared the effectiveness of different feature selection methods such as heuristic strategies, greedy algorithms and genetic algorithms and ultimately determined that genetic algorithm achieved the best result. The study also compared the results of two crossover methods, two-point crossover and uniform crossover and ultimately determined that uniform cross achieved the best result with a 97.7% accuracy score.

This study compares the results of three cross-over methods Single point crossover, Two point crossover And Uniform crossover based on the fitness function.

3.1. Uniform crossover

Uniform crossover is a genetic operator that recombines the genetic material of two parents to produce new offspring solutions. It works by randomly selecting genes from each parent based on a binary mask. If the mask value is 0, the gene is copied from Parent 1, and if the mask value is 1, the gene is copied from Parent 2. from codesarts

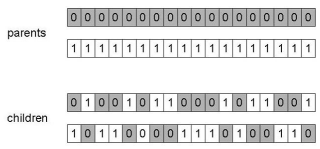

Figure 1. Figure 1: Uniform crossover

3.2. Single point crossover

A crossover method to recombine the genetic material of two-parent individuals to produce new offspring solutions. It works by selecting a random crossover point along the chromosome and swapping the genetic material between the parents. This creates two offspring with genetic material from both parents. This is better demonstrated in Figure 1 from geeksforgeeks.

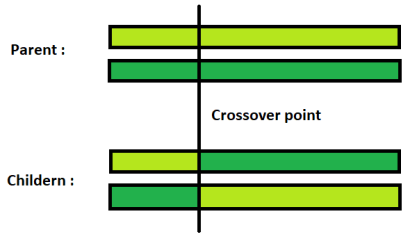

Figure 2. Figure 1: Single Point crossover

3.3. Two Point crossover

Two-point crossover is a genetic operator used in genetic algorithms for creating a new offspring solution by exchanging genetic material between two parent solutions. Two random crossover points are selected on the parent chromosomes, and the genetic material between the two crossover points is exchanged to create two new offspring solutions. from wiki

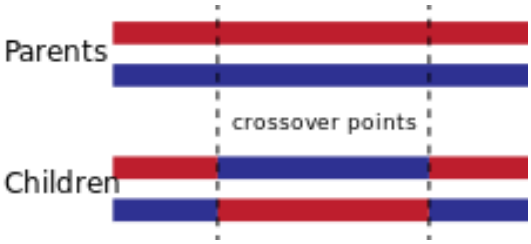

Figure 3. Figure 1: Two Point crossover