A Hybrid Genetic Algorithm for Climate Input Features and Neural Network Parameters Selection

Ali Haidar
Central Queensland University, Australia
a.haidar@cqu.edu.au

Brijesh Verma
Central Queensland University, Australia
b.verma@cqu.edu.au

ABSTRACT

The climate input features and neural network parameters highly affect the overall performance of the rainfall prediction models. In this paper, a novel approach is proposed to select the input features and neural network parameters. A new hybrid genetic algorithm that combines natural reproduction and particle swarm optimization characteristics was developed to select the best climate features and network parameters. The developed model was compared against alternative models including climatology and showed a better accuracy. The aggregated time series of the proposed model showed a Root Mean Square Error (RMSE) of 141.67 mm for a location with 3553.00 mm annual average.

KEYWORDS

Genetic Algorithms, Neural Networks; Rainfall Forecasting

1. INTRODUCTION

Artificial neural networks have been widely deployed to forecast rainfall and other climate attributes for different durations in different climatic regions [1, 2, 3]. Extensive research has been conducted to link climate indices to rainfall variability in Australia and sugarcane areas [3, 4, 5].

The network architecture is normally selected using trial and error methods. Diverse networks are usually generated and the network with the best performance over test data is selected. However, a network with different parameters which has not been included in the selection could have produced better performance. Several algorithms including Genetic Algorithms (GA) with different searching criteria can be applied to select the neurons, learning algorithms, activation functions etc. In the past, some research has been conducted to set neural network parameters. In addition, several studies were conducted to select best features for prediction [5]. The aim of this study is to select both best input features and neural network parameters for forecasting monthly rainfall based on a hybrid genetic algorithm.

This paper is organized as follows: Section 2 describes the proposed approach. Section 3 presents the data collection, experiments and results. The conclusion is presented in Section 4.

GECCO '17 Companion, July 15-19, 2017, *Berlin, Germany*
ACM 978-1-4503-4939-0/17/07.
http://dx.doi.org/10.1145/3067695.3076038

2. PROPOSED APPROACH

The proposed hybrid genetic algorithm based approach consists of many steps as shown in Figure 1. The first step generates population in which each chromosome is a collection of binary numbers representing climate input features, learning algorithms, hidden neurons and activation functions. The second step is to train a neural network for each chromosome using selected parameters (1 selected, 0 not selected). After training, RMSE is calculated for each network on test data. The third step is to find and store the best chromosome in the population based on RMSE for each network. The fourth step is to conduct selection, crossover, and mutation and get a new population. The above steps are repeated until the condition is met (max iteration or RMSE= 0). Once final iteration is completed, the selected chromosome is compared with global stored chromosomes from previous iterations. The best chromosome is selected which gives the best network. The accuracy is calculated based on the final selected chromosome and network model. The novelty of this approach is that in each iteration the dynamic initial random weights are generated, neural networks are trained and best chromosome and its corresponding network is stored in a global chromosome register. This novel idea gives the GA a wider searching space since different random initial weights and biases are assigned to the same network in the next population if selected as the best to survive. The above characteristic is taken from particle swarm optimization where the global best is usually saved through iterations and particles flow towards the best solution. In proposed hybrid GA, the best chromosome is saved but other chromosomes don't follow it. While saving global chromosomes, best network with highest performance is guaranteed.

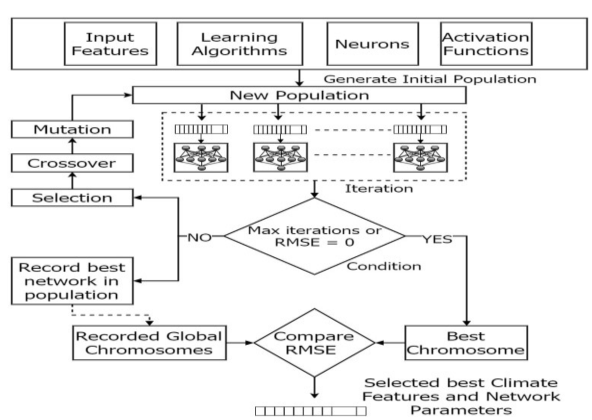

Figure 1 : Proposed Approach

3. DATA AND EXPERIMENTAL RESULTS

Innisfail, Queensland was selected to perform the study due to its closeness to sugarcane paddocks and mills. Local weather variables as rainfall (target), mean minimum temperature and mean maximum temperature were collected from Bureau of Meteorology (BOM), Australia. Global climate indices can affect rainfall variability across Australia. Southern Oscillation Index (SOI) was collected from BOM. Nino 1.2, Nino 3.0, Nino 3.4, Nino4.0, Dipole Mode Index (DMI) and sunspots were gathered from KNMI climate explorer. Tripole Index for the Interdecadal Pacific Oscillation (TPI) data were collected from Earth System Research Laboratory [6], while IPO was taken from CLIVAR International Climate of the Twentieth Century Project website.

The collected climate indices have different effects on rainfall variability over different locations and timeframes on Australia. Therefore, a climate index may have high contribution in determining the amount of rainfall. To determine the input features and best parameters for each month, the data was divided into 12 datasets representing each month. In the proposed hybrid GA approach, 75% of data for training, 15% for validation and 10% for testing were used.

The proposed hybrid GA approach was run 12 times; each time corresponds to one month. The generated forecasts for each month were compared against standard GA approach based on input feature selection only. The alternative approach did not record best chromosome in each population. The fitness function contained a neural network with 8 hidden neurons and Levenberg-Marquardt as the training algorithm. Mean Absolute Error (MAE), RMSE and correlation (r) values were calculated to measure the accuracy for each approach and they are shown in Table 1. The proposed approach that is based on selecting both input features and network parameters revealed better accuracy in terms of MAE, RMSE and r values for almost all months compared to climate features selection only. To measure the performance of the proposed approach, the outputs of each month (test data) were aggregated to create the time series between January 2005 and December 2015 (11 years).

Table 1: MAE, RMSE and r values generated by proposed hybrid GA and standard GA approaches.

Month	Proposed Hybrid GA Approach with Climate Feature & Parameter Selection			Standard GA Approach with Climate Feature Selection		
	MAE	RMSE	r	MAE	RMSE	r
1	151.14	199.73	0.622	188.82	222.69	0.483
2	235.98	276.78	0.610	288.09	321.65	0.639
3	169.67	194.63	0.841	258.74	309.71	0.365
4	125.91	150.34	0.603	147.42	182.27	0.416
5	100.40	119.46	0.697	125.78	145.09	0.459
6	70.60	100.40	0.696	89.02	116.78	0.443
7	74.63	84.30	0.655	87.11	98.77	0.475
8	37.58	50.20	0.825	63.65	81.23	0.569
9	43.76	51.28	0.837	57.65	69.30	0.657
10	100.94	118.12	0.471	85.64	126.16	0.43
11	66.58	84.30	0.893	89.93	127.06	0.782
12	70.34	81.07	0.905	81.87	116.86	0.738

The developed model was also compared to climatology which is a reference model in rainfall forecasting [7].

MAE, RMSE and r values of both proposed model and climatology are shown in Table 2. 58.65 and 36.55 mm difference in terms of RMSE was obtained when compared to climatology and climate features selection. Figure 2 represents the combined dataset generated by the 12 developed neural networks compared to actual rainfall climate features and network parameter selection. The networks have the ability to predict accurately the amount of rainfall in some locations e.g. March 2012 (actual: 1228.5 mm, forecasted: 1148.20 mm). This ability in forecasting peak rainfall is helpful for different industries including sugar.

Table 2: MAE, RMSE and r values for climatology, input feature selection based standard GA approach and input and parameter selection based proposed hybrid GA approach.

Model	MAE	RMSE	r
Climatology	147.21	200.32	0.688
Standard GA Approach with Climate Feature Selection	130.30	178.22	0.763
Proposed Hybrid GA Approach	103.81	141.67	0.859

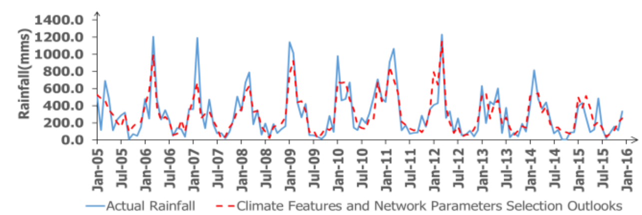

Figure 2: Combined rainfall time series produced by proposed hybrid GA approach.

4. CONCLUSION

A novel hybrid genetic algorithm based approach for selecting climate input features and network parameters was proposed and compared in this study. The aggregated time series of developed models based on both input and network parameter selection showed highest accuracy in terms of RMSE when compared to climatology and standard GA based feature selection method.

REFERENCES

[1] W. Zhi-Liang and S. Hui-Hua, Rainfall Prediction Using Generalized Regression Neural Network: Case Study Zhengzhou, International Conference on Computational and Information Sciences, 2010, pp. 1265-1268.

[2] P. T. Nastos, A. G. Paliatsos, K. V. Koukouletsos, I. K. Larissi and K. P. Moustris, Artificial Neural Networks Modeling for Forecasting the Maximum Daily Total Precipitation at Athens, Atmospheric Research, 144 (2014), pp. 141-150.

[3] F. Mekanik, M. A. Imteaz, S. Gato-Trinidad and A. Elmahdi, Multiple Regression and Artificial Neural Network for Long-term Rainfall Forecasting Using Large Scale Climate Modes, Journal of Hydrology, 503 (2013), pp. 11-21.

[4] Y. L. Everingham, A. J. Clarke and S. Van Gorder, Long Lead Rainfall Forecasts for the Australian Sugar Industry, International Journal of Climatology, 28 (2007), pp. 111-117.

[5] A. Haidar and B.Verma, A Genetic Algorithm based Feature Selection Approach for Rainfall Forecasting in Sugarcane Areas, IEEE SSCI, 2016, pp. 1-8.

[6] J. Henley, J. Gergis, D. J. Karoly, S. Power, J. Kennedy and C. K. Folland, A Tripole Index for the Interdecadal Pacific Oscillation, Climate Dynamics, 45 (2015), pp. 3077-3090.

[7] S. Hawthorne, Q. J. Wang, A. Schepen and D. Robertson, Effective Use of General Circulation Model Outputs for Forecasting Monthly Rainfalls to Long Lead Times, Water Resources Research, 49 (2013), pp. 5427-5436.