IJDACR
International Journal Of Digital Application & Contemporary Research

International Journal of Digital Application & Contemporary Research
Website: www.ijdacr.com (Volume 8, Issue 10, May 2020)

Rainfall Prediction System using PCA and Cultural Algorithm Optimized Neural Network Classifier

Shahista Navaz
PhD. Research Scholar
Dept. of Computer Science and Engineering
CVRU, Bilaspur, Chhattisgarh (India)
shahista.navazcse@gmail.com

Dr. S. M. Ghosh
Professor
Dept. of Computer Science and Engineering
CVRU, Bilaspur, Chhattisgarh (India)
samghosh06@rediffmail.com

Abstract – **Climatic forecasting of the rainfall field is a key aspect of meteorology. Rainfall is a variable associated with natural disasters (droughts and floods) and agricultural crops, with impacts on the tourism and transport sectors. However, this meteorological variable is difficult to predict, due to the great temporal and spatial variability (discontinuous variable). This paper uses Probability Density Function (PDF) followed by the min-max normalization for the pre-processing of the Kaggle Indian Rainfall Dataset. The selection of attributes is achieved by Principal Component Analysis (PCA) followed by the classification using Cultural Algorithm Optimized Neural Network Classifier..**

Keywords – **Cultural Algorithm, Min-Max Normalization, Neural Network, PCA, PDF, etc.**

I. Introduction

Rainfall is the amount of water that falls to the earth's surface and comes from atmospheric humidity, either in a liquid state (drizzle and rain) or in a solid state (frost, snow, hail). Rainfall is one of the most important meteorological processes for Hydrology, and together with evaporation they constitute the way in which the atmosphere interacts with surface water in the hydrological cycle of water [1].

Evaporation from the ocean surface is the main source of moisture for rainfall, it can be said that it is 90% of the rainfall that falls on the continent. However, the greatest amount of rainfall does not necessarily fall on the oceans, since atmospheric circulation transports humidity for great distances, as evidence of this, some desert islands can be observed. The location of a region with respect to atmospheric circulation, its latitude and distance to a source of humidity are mainly responsible for its climate [2].

Concerns about climate change are growing in the scientific community. Amid the great prominence that the climate and its changes have been presenting in the last decade, scientists from all over the world are trying to understand the nature of the changes that are likely to occur, as well as the possible impacts that they can cause for society in general [3].

The dependence on climatic factors motivated the development of this work, whose objective is to carry out the climatic forecast of the rainfall field using cultural algorithm optimized neural network classifier.

A. *Prediction and Estimation Techniques*

There are numerous forecasting techniques with different forecast times and valid for areas of different sizes. As already mentioned, satellite images are an important tool in these techniques and are present in most of them [4].

The rainfall prediction and estimation techniques that mainly use satellite images can be divided into four categories [5]:

- Cloud Indexing: They are based on the fact that it is relatively easy to identify cloud types in the satellite images and a rainfall intensity is assigned to each cloud type taking into account the time that it remains above the observatory [6].
- Bispectral Techniques: They combine the information from the two VIS and IR channels taking into account that cold clouds (in the IR channel) and bright clouds (in the VIS channel) are the ones that produce the most rain.
- Life Cycle Techniques: They are based on the fact that the amount of rain from a cloud, especially from a convective cloud, is a function of the stage of its life cycle in which it is located. These techniques use geostationary satellite images given their high temporal resolution [7].
- Cloud Model Techniques: They incorporate the physical information of the cloud system

and the environmental conditions that surround it. As these techniques are the most complete in terms of the amount of information used, they are the ones used in this work, bearing in mind that it is intended to make a prediction and monitoring of heavy rains in a short period of time and that in the Mediterranean area, the systems evolve quite rapidly, (between 3 and 24 hours, with an average of 11.5 hours) [8].

- Machine Learning based Techniques: The dynamic nature of statistical methods for rainfall prediction does not provide the estimated accuracy. Nonlinearity of rainfall data makes machine learning based techniques better as compared to conventional methods of rainfall forecasting [9] [10].

This paper proposes a predictive climate forecasting model for the rainfall field using PCA and cultural algorithm optimized neural network for calculating similarity score for performance evaluation.

II. Proposed Methodology

Figure 1 shows the flow of proposed research work.

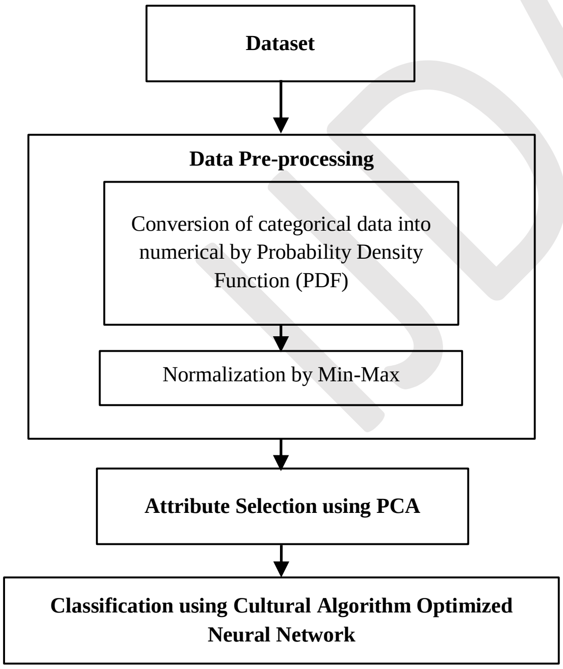
Dataset
Data Pre-processing
Conversion of categorical data into numerical by Probability Density Function (PDF)
Normalization by Min-Max
Attribute Selection using PCA
Classification using Cultural Algorithm Optimized Neural Network

Figure 1: Proposed flow diagram

A. Data Pre-processing

1) Probability Density Function (PDF)

The categorical data from the dataset is converted into the numerical form using the probability distribution function given by:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma(x-\lambda)} \exp\left\{-\frac{[\ln(x-\lambda)-\mu]^2}{2\sigma^2}\right\}, \quad x > \lambda, \sigma > 0 \qquad (1)$$

2) Normalization by Min-Max Approach

Min-Max Normalization performs a linear transformation on the original data x into the specified interval (New_{min}, New_{max}).

$$x_i = New_{min} + (New_{max} - New_{min}) \times \left(\frac{x_i - x_{min}}{x_{max} - x_{min}}\right) \qquad (2)$$

$$x_{max} = \max_{1\le i\le N} x_i, x_{min} = \min_{1\le i\le N} x_i \qquad (3)$$

This method scales the data from (x_{min}, x_{max}) to (New_{min}, New_{max}) in proportion. The advantage of this method is that it preserves all relationships of the data values exactly. It does not introduce any potential bias into the data.

B. Attributes Selection

The development of technology and the propagation of computer systems in the most varied domains of knowledge have contributed to the generation and storage of a constantly increasing amount of data at a higher speed that we are able to process. In general, the main reason for storing this huge amount of data is the use of it for the benefit of humanity. Several areas have been dedicated to research and the proposal of methods and processes to treat this data. To achieve this goal, models (hypotheses) are usually built, which can be generated with the support of different areas such as machine learning. PCA is a technique that exploits the aspect geometric and graphical representations of a set of data called observations, in order to study its variability and its dispersion in the space in which it is represented. Let P be a data set with m quantitative variables having n units (often denoted individuals), defined by equation:

$$P = \left[\begin{pmatrix} p_{11} \\ p_{21} \\ \vdots \\ p_{i1} \\ \vdots \\ p_{n1} \end{pmatrix} \begin{pmatrix} p_{12} \\ p_{22} \\ \vdots \\ p_{i2} \\ \vdots \\ p_{n2} \end{pmatrix} \cdots \begin{pmatrix} p_{1j} \\ p_{2j} \\ \vdots \\ p_{ij} \\ \vdots \\ p_{nj} \end{pmatrix} \begin{pmatrix} p_{1m} \\ p_{2m} \\ \vdots \\ p_{im} \\ \vdots \\ p_{nm} \end{pmatrix}\right] \rightarrow \begin{bmatrix} (\vec{p}_{1.})^t \\ (\vec{p}_{2.})^t \\ \vdots \\ (\vec{p}_{i.})^t \\ \vdots \\ (\vec{p}_{n.})^t \end{bmatrix} \qquad (4)$$

$$\downarrow$$

$$\vec{p}_{.1} \quad \vec{p}_{.2} \quad \cdots \quad \vec{p}_{.j} \quad \cdots \quad \vec{p}_{.m}$$

Starting from its matrix write, we can rewrite the dataset P in two vectorial forms (Equation 5): the vertical vector is composed of a set of points where each point, denoted as $\vec{p}_{i.}$, is of dimension m. The horizontal vector is the second vector form and is made up of a set of points where each, noted $\vec{p}_{.j}$, is of dimension n. Note that t denotes the transposed function. The points $\vec{p}_{.i}$ and $\vec{p}_{.j}$ can be given by the following equations:

$$\vec{p}_{.j} = \begin{pmatrix} p_{1j} \\ p_{nj} \end{pmatrix}_{(1\le j\le m)} \qquad (5)$$

And

$$\vec{p}_{.i} = \begin{pmatrix} p_{i1} \\ p_{im} \end{pmatrix}_{(1\le i\le n)} \qquad (6)$$

Where i and j indicate the individual and the variable respectively.

C. Classification by Neural Network

A set of machine learning algorithm will be used for the classification:

In this work, different structures of neural networks with a hidden layer were tested, starting from a number of neurons equal to the average between the number of inputs and the number of outputs. Then the number of neurons in said layer was gradually increased until obtaining the most recommended structure for rainfall prediction. The selection of the best network structure is made considering the following evaluation measures inside and outside the sample: RMSE (Root Mean Square Error) and the MAPE (Mean Absolute Percentage Error), calculated using following equations.

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t' - y_t)^2} \qquad (7)$$

$$MAPE = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{y_t' - y_t}{y_t}\right| \qquad (8)$$

Where n is the number of observations considered, and t is the real price and y_t' is the price estimated by the model.

Fitness Function: The fitness function is a function of weight and bias with the objective of minimizing the mean square error between the predicted and target classes of the training data.

$$\min F(w, v) = \sum_{t=1}^{q}[c_t - (wx_t + v)]^2 \qquad (9)$$

Where, x_t is input and c_t is target output.

Fitness function in equation (9) is minimized using Cultural Algorithm to optimized weight and bias values.

1) Cultural Algorithm

Cultural algorithms were initially developed by Robert Reynolds [11], as an extension of evolutionary algorithms. Its operation, represented in Figure 2, is explained in the following model. The algorithm operates in two spaces, the population space and the belief space. In the first space there are a number of individuals, where each one has a set of characteristics independent of the others, each individual is evaluated according to the fitness function, as in any evolutionary algorithm. Over time, these individuals can be replaced by some of their descendants. The second space is where the knowledge acquired by individuals through the generations is stored. The information found in this space must be available to any individual. The belief space is updated after each iteration by the best individuals in the population, the knowledge categories of the belief space can affect the population component through influence, which can affect the population by altering the genome or the actions of individuals. To unite both spaces, a communication protocol is established that dictates the rules of the type of information that must be exchanged between the spaces.

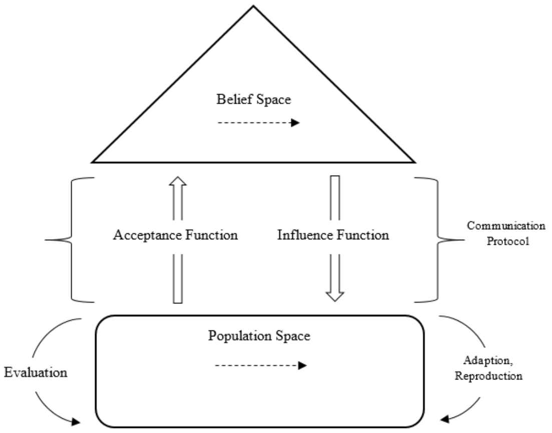

Figure 2: Structure for cultural algorithm [12]

The pseudo-code used for cultural algorithms is as follows:

```
1. Initialize the population space
2. Initialize belief space
3. Repeat until the end condition is met
    a. Carry out the actions of individuals in the population space
    b. Evaluate each individual by using the fitness function
    c. Select parents to reproduce a new generation of offspring
    d. Letting the belief space alter the offspring's genome by using the influence function
    e. Update the belief space by using the acceptance function, leaving the best individuals to affect the belief space.
```

III. Simulation Results

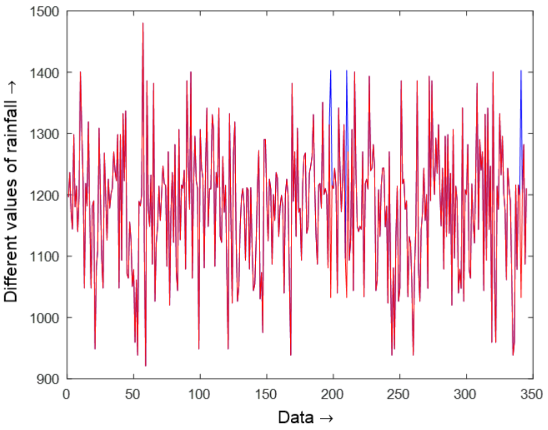

Figure 3: Rainfall prediction output using Neural Network

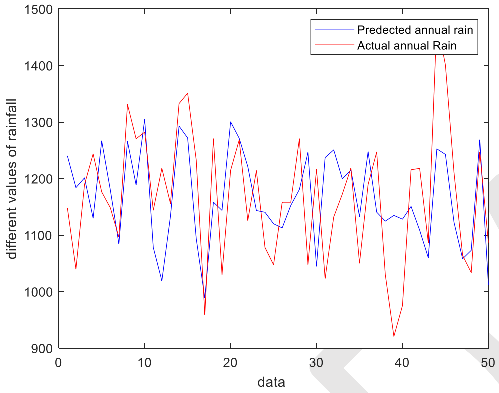

Figure 4: Comparative graph for predicted and actual rainfall using Neural Network

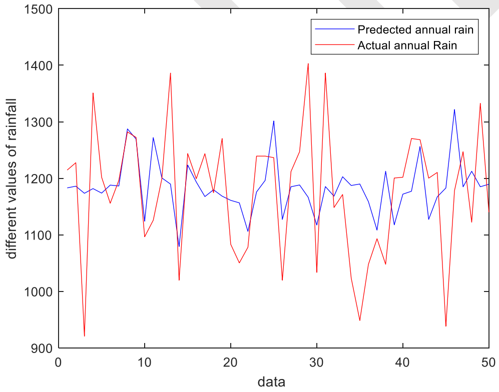

Figure 5: Comparative graph for predicted and actual rainfall using Cultural-NN

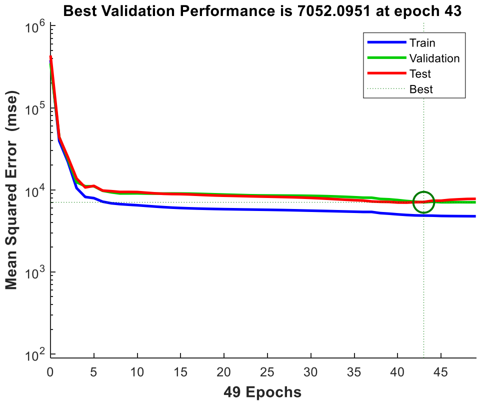

Figure 6: Mean squared error graph for Neural Network

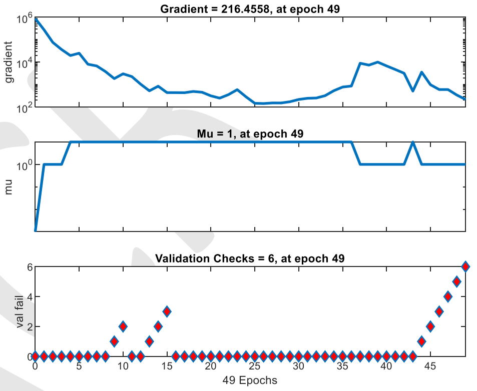

Figure 7: Gradient, mu and validation graph for Neural Network

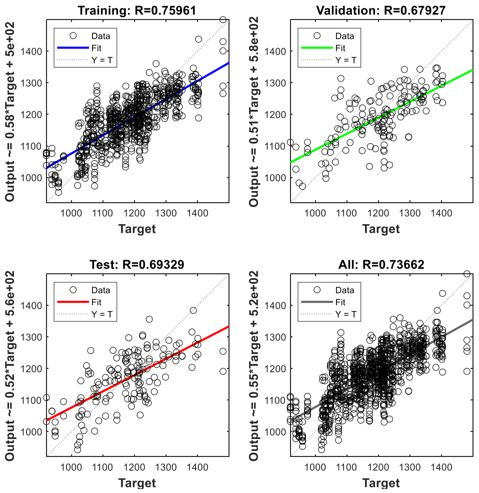

Figure 8: Neural network output

Table 1: Evaluation measures of neural network approach

MAE	MSE	RMSE	MARE	MSRE		RMSRE	MAPE	RMSPE
76.1988	9.1512e+03	95.6617	0.0650	0.0068		0.0826	6.5008	8.2552

Current MAPE (Mean Absolute Percent Error): 0.312%

IJDACR
International Journal Of Digital Application & Contemporary Research

International Journal of Digital Application & Contemporary Research
Website: www.ijdacr.com (Volume 8, Issue 10, May 2020)

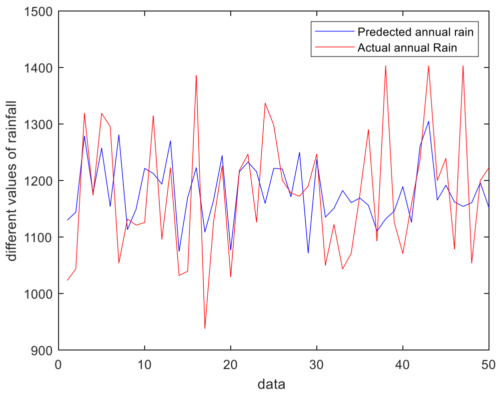

Figure 9: Comparative graph for predicted and actual rainfall using PCA-NN

Table 2: Evaluation measures of PCA-NN approach

MAE	MSE	RMSE	MARE	MSRE	RMSRE	MAPE	RMSPE
75.2673	9.9006 e+03	99.5019	0.0644	0.0075	0.0867	6.4415	8.6709

Current MAPE (Mean Absolute Percent Error): 6.441%

IV. Conclusion

The model presented took a greater step in research in the area of rainfall estimation, by exploring more than one model of neural network in various combinations of configurations of input variables, seeking the best set of input variables that minimizes the error of the input rainfall data of the Kaggle Database [13]. Thus, the NN and Cultural Algorithm optimized neural networks were evaluated.

In general, it is known that, as the forecast horizon increases, the predictive capacity of any model deteriorates. Therefore, in the proposed model it is natural to expect that, for a given dataset, the mean absolute percentage error (MAPE) will increase as the forecast horizon grows.

These results confirm that the CA-optimized neural network outperforms the other approach.

It is intended, as future work, to carry out tests with hybrid, neural-statistical models, with the objective of obtaining even more accurate predictions.

Reference

[1] Mishra SK, Sharma N (2018) Rainfall forecasting using backpropagation neural network. In: Innovations in computational intelligence. Springer, Singapore, pp 277–288

[2] Pankratz A (2018) Forecasting with univariate box-jenkins method. Wiley, NY

[3] Vuille M, Bradley RS, Werner M, Healy R, Keimig F (2018) Modeling c18o in precipitation over the tropical Americas: 1. Interannual variability and climatic controls. J Geophys Res 108(106)

[4] WMO (2018) Calculation of monthly and annual 30-year standard normals, wmo-td/no. 341. World Climate Programme Data, Washington, DC.

[5] Xavier, Alexandre C., Roberto A. Cecílio, Fernando F. Pruski, and Julião S. de S. Lima. "Methodology for spatialization of intense rainfall equation parameters." Engenharia Agrícola 34, no. 3 (2014): 485-495.

[6] Calzadilla A, Rehdanz K, Betts R, Falloon P, Wiltshire A, Tol RS (2013) Climate change impacts on global agriculture. Clim Change 120:357–374.

[7] Feng G, Cobb S, Abdo Z, Fisher DK, Ouyang Y, Adeli A, Jenkins JN (2016) Trend analysis and forecast of precipitation, reference evapotranspiration, and rainfall deficit in the blackland prairie of eastern Mississippi. J Appl Meteorol Climatol 55:1425–1439

[8] Partal T, Cigizoglu HK, Kahya E (2015) Daily precipitation predictions using three different wavelet neural network algorithms by meteorological data. Stoch Environ Res Risk Assess 29:1317–1329

[9] Rivero CR, Patiño HD, Pucheta JA (2015) Short-term rainfall time series prediction with incomplete data, in neural networks (IJCNN). Int Joint Conf IEEE 2015:1–6

[10] Singh P, Borah B. Indian summer monsoon rainfall prediction using artificial neural network. Stochastic Environ. Res. Risk Assess. 2013;3:1436-3240.

[11] Reynolds, R.G., 1994, February. An introduction to cultural algorithms. In Proceedings of the third annual conference on evolutionary programming (Vol. 131139). Singapore.

[12] Reynolds, R.G. and Peng, B., 2004, November. Cultural algorithms: modeling of how cultures learn to solve problems. In Tools with Artificial Intelligence, 2004. ICTAI 2004. 16th IEEE International Conference on (pp. 166-172). IEEE.

[13] Kaggle. Indian Rainfall Dataset. Available online at: https://www.kaggle.com/rajanand/rainfall-in-india