A Investigative Rainfall Forecast Using atom throng optimization System

Mohammed Inamurrahman, Research Scholar in Chhatrapati Shahu Ji Maharaj University, Kanpur
Mohammed Burhanurrahman, Lecturer in College of Computer Science, KKU, Mohayail, Aseer

Abstract—Rainfall plays an important role in maintaining the water level in the earth. The prediction of rainfall becomes a great challenge in day to day life for researchers due to the sudden change in climates. The sudden change of climate really affects agriculture in India and other part of the world. In this work, we proposed PSO for feature extraction process and neural network for predicting rainfall. Data interpretation is the process that reduces the long data set and represents in another form in a useful manner. By using data interpretation we will compute how much rain interprets over the years and represent it in a precise form and then PSO is applied for feature extraction of the data. PSO optimizes a problem having a population of candidate solutions. Here PSO is used for the feature extraction of rainfall data and hence best solutions are selected. Feeding knowledge classifier is used to train input and predicting the rainfall after training and validation of data. Hence it reduces the total error by the input data to produce an output data, changing the weights. Therefore the rainfall prediction is carried successfully using our proposed technique and it has proved as an efficient method for prediction of the rainfall condition with an effective accuracy.

Index Terms—Minimum 7 keywords are mandatory, Keywords should closely reflect the topic and should optimally characterize the paper. Use about four key words or phrases in alphabetical order, separated by commas.

1 INTRODUCTION

Today world is witnessing an ever changing climate conditions. Climate changes have far reaching effects especially in the agricultural sector of a country [5]. Due to different aspects of climate change, rainfall prediction is imperative for agriculture sector. Rain is one of nature’s greatest gifts, and in third world countries like India, the entire agriculture depends upon rain. It is generally accepted that rainfall is unpredictable. Real-time water resources assessment can be defined as a rapid assessment of the water resources generated in a rainfall event or in a past period from a particular day of the year to the current rainfall event [14]. Knowing the condition of rainfall in advance can help in managing and dealing with agricultural management and disaster prevention [5].

Information regarding rainfall is important for food production plan, water resource management, and all activity plans in the nature. The occurrence of prolonged dry period or heavy rain at the critical stages of the crop growth and development may lead to significant reduce in crop yield [1]. Rainfall is being one of the most difficult elements of the hydrological cycle to forecast, and great uncertainties still affect the performances of both stochastic and deterministic rainfall prediction models [4]. Rainfall is not a regular phenomenon in all places. It has some seasonality effects. So, the rainfall prediction problem is not same as other regular atmospheric parameters like temperature, humidity, etc. Rainfall is also a time series data like atmospheric pressure, temperature, vapor pressure, relative humidity, radiation, etc [8]. A wide range of rainfall forecast methods are employed in weather forecasting at regional and national levels. Fundamentally, there are two approaches to predict rainfall. They are empirical and dynamical methods. The empirical approach is based on analysis of historical data of the rainfall and its relationship to a variety of atmospheric and oceanic variables over different parts of the world. In dynamical approach, predictions are generated by physical models based on systems of equations that predict the evolution of the global climate system in response to initial atmospheric conditions [2] [3].

The continuous changes in global climate and the uneven spatial and temporal distribution of rainfalls are the causes for severe problems like floods and droughts. For example, the state Orissa in India is facing the similar problems more often. Most of the rainfall in the region occurs during monsoon period. The rainfall received during the months, i.e., June, July, August, and September (JJAS) is considered as summer monsoon and is very crucial for the farming community. Researchers have used various approaches to study and predict the seasonal and intra-seasonal rainfall [7]. Applications of synthetic rainfall data may
then be made in such diverse fields as flood modeling, urban drainage, pesticide fate modeling, landslide modeling, desertification vulnerability, water resource assessment, and flood risk assessment [6]. In this work, we proposed PSO for feature extraction process and neural network for predicting rainfall. The rest of the paper is organized as follows, Section 1.1 briefly gives an introduction about the PSO and Section 2 reviews the recent research works related to the rainfall prediction techniques. Section 3, 4 and 5 details the steps involved in the proposed technique with necessary illustrations and mathematical formulations. Section 6 discusses about the implementation results and Section 7 concludes the paper.

1.1. PARTICLE SWARM OPTIMIZATION:
PSO is introduced by Kennedy and Eberhart is one of the metaheuristics, which is inspired by the swarming behavior of animals and human social behavior [28][31]. In computer science, particle swarm optimization (PSO) is a computational method that optimizes a problem by iteratively trying to improve a candidate solution with regard to a given measure of quality. PSO optimizes a problem by having a population of candidate solutions, here dubbed particles, and moving these particles around in the search-space according to simple mathematical formulae over the particles' position and velocity. Each particle's movement is influenced by its local best known position and is also guided toward the best known positions in the search-space, which are updated as better positions are found by other particles. This is expected to move the swarm toward the best solutions [23]. The main strength of PSO is its fast convergence, which compares with many global optimization algorithms like Genetic algorithms, Simulated Annealing and other global optimization algorithms.

Particle Swarm Optimization shares many similarities with evolutionary computation techniques such as Genetic Algorithms. In the PSO assigns a randomized velocity to each potential solution, called the particle, fly through the problem space by following the current optimum particles. PSO is a global optimization algorithm for dealing with problems in which a best solution can be represented as a point or surface search in n-dimensional space[25][27]. Hypotheses are plotted in this space and seeded with an initial velocity as well as a communication channel between the particles. When a particle moves to a new location, a different problem solution is generated. This solution is evaluated by a fitness function that provides a quantitative value of the solution’s utility [29][30]. The particles were accelerated in the direction of communication grouping which have better fitness values. The main advantage of such approach great global minimization strategies such as simulated annealing is that the large number of members that make up the particle swarm formulate the technique impressively flexible to the problem of local minima[24][26].

2. RELATED WORK:
Karamouz et al. [9] have utilized field and General Circulation Models (GCM) data with the Statistical Downscaling Model (SDSM) and the Artificial Neural Network (ANN) model for long lead rainfall prediction. These models have been used for the prediction of rainfall for 5 months (from December to April) in a study area in the south eastern part of Iran. The SDSM model considers the climate change scenarios using the selected climate parameters in rainfall prediction, but the ANN models were driven by observed data and do not consider the physical relations between variables. The results have shown that SDSM outperforms the ANN model.

Pramote Luenanet al. [10] have presented the methodology of neuro-fuzzy for the rain forecasting system over the central region of Thailand. The neuro-fuzzy approach has been applied to create a classifier for rain prediction. Their objective is to demonstrate what relationship models between rain occurrence and other weather features can be developed for predicting accurate rainfall estimates to support the decisions to launch cloud seeding operations. Datasets were collected during 2004 to 2008 from the Chalermprakiat Royal Rain Making Research Center at HuaHin, PrachuapKhir Khan, and Thai Meteorological Department. A total of 179 records with 57 features have been merged and matched by unique date. There are three main parts in this work. Firstly, a correlated-based feature selection (CFS) has been used to evaluate the most important features for rain prediction and rainfall level classification. Secondly, a neuro-fuzzy algorithm, NECLASS, has been used for prediction of rain or no-rain events. Thirdly, an algorithm has also been used to classify rainfall levels into four classes as no-rain (0 mm.), light-rain (>0 - 10 mm.), moderate-rain (>10-35 mm.), and heavy rain (>35 mm.). Results have shown that the overall classification ac-
curiosity of the neuro-fuzzy classifier was satisfactory. Ladislaus B. Chang et al. [11] have described how farmers in the South-western Highland of Tanzania predict rainfall using local environmental indicators and astronomical factors. The perceptions of the local communities on conventional weather and climate forecasts have also been assessed. A study has been conducted in Rungwe and Kilolo districts in Mbeya and Iringa regions respectively. Participatory rural appraisal methods, key informant interviews and focus group discussions have been used in data collection and the collected data has been analyzed using statistical package for social science. It has been found that plant phenology is widely used by local communities in both districts in seasonal rainfall forecasting. Early and significant flowering of Mihemi (Erythrinaabyssinica) and Mikwe (Brachystegiaspeciformis) trees from July to November has been identified to be one of the signals of good rainfall season. The behavior of Dudumizi bird has been singled out as one of the best indicator for rainfall. Both Indigenous Knowledge specialists and TMA experts have predicted 2009/2010 rainfall season to feature normal to above normal rainfall. Systematic documentation and subsequent integration of indigenous knowledge into conventional weather forecasting system is recommended as one of the strategy that could help to improve the accuracy of seasonal rainfall forecasts under a changing climate.

Enreddy, Vamsidharet al. [12] have applied the back propagation neural network model for predicting the rainfall based on humidity, dew point, and pressure in the country INDIA. Two-Third of the data was used for training and One-third for testing. The numbers of training and testing patterns were 250 training and 120 testing. 99.79% accuracy for training and 94.28% accuracy for testing have been obtained. From these results, the rainfall can be predicted for the future.

XianggenCanet et al. [13] have discussed that the continuously cloudy or rainy forecast is an important basis that is used to make choice of wheat harvest time but multiple regression weather forecast models hardly content the rate of required accuracy. Mat lab neural network toolbox is composed of a series of typical neural network activation functions that make computing network output into calling activation functions. BP artificial neural network that is based on Mat lab platform and utilizes error back propagation algorithm to revise network weight has dynamic frame characteristcs and is convenient for constructing network and programming. After it has been trained by input forecast samples, network forecast model that has three neural cells possesses very good generalization capability. After they contrast fitting rate and accuracy rate of network model with ones of regression model, network model has a distinct advantage over regression model.

PjushSamuiet al. [23] have adopted a Support Vector Machine (SVM) and Relevance Vector Machine (RVM) for prediction of rainfall in Vellore (India). SVM is firmly based on the theory of statistical learning theory. RVM is a probabilistic basis model. SVM and RVM have used air temperature (T), sunshine, humidity, and wind speed (Va) as input variables. The SVM and RVM have been used as a regression technique. Equations have also been developed for prediction of rainfall. The developed RVM gives variance of the predicted rainfall. Experiments have shown that RVM is more robust than the SVM.

YogeshShirkeet al. [24] proposed a convenient and easy tool for modeling and analysis of non-linear events. This ability ofANN to model non-linear events was important in hydrology to model various hydrological events which were dominantly non-linear in nature. It was also capable of modeling Non-linear relationships between Rainfall and Runoff as compared to other Mathematical modeling techniques. Back Propagation (BP) algorithm was used to evaluate error and back propagate it for more accurate training of ANN.

3. PROPOSED RAINFALL PREDICTION METHOD:
We intended to propose a heuristic rainfall prediction method using particle swarm optimization. In the proposed technique, we discuss about three important steps such as data interpretation, feature extraction using PSO and feeding knowledge to classifier. The proposed technique exploits the raw dataset that are collected from the renowned data resource. The acquired raw dataset can be represented as

\[ \{D\}_\text{line} : 0 \leq l \leq L - 1 ; 0 \leq m \leq M - 1 ; 0 \leq n \leq N - 1 \quad (1) \]

where, L represents the location of the data, m represents the year, n represents the month of mth year, L represents the total number of locations, M represents the total number of years and N represents the total number of months.
changing the weights along the system. All the aforementioned process can be said as offline process as this can be performed any time without the intervention of users. Once the scheduled offline process is completed, the entire setup can be used to operate in online, i.e., to predict the rainfall level in the near future.

3.1 DATA INTERPRETATION:
In our technique, we exploit sliding window operation to interpret the acquired raw data into meaningful and comfortable data and so to extract features from them. The acquired data is of size $L \times M \times N$ and so the interpretation is applied in such a way that the records are re-ordered as per the interest. An exemplary illustration about data interpretation is given in Figure 2 and the interpreted data representation is given in equation.

\[
D' = \begin{bmatrix}
    d_{11} & d_{12} & \ldots \ldots & \ldots & d_{1N} \\
    d_{21} & d_{22} & \ldots \ldots & \ldots & d_{2N} \\
    \vdots & \vdots & \ddots & \vdots & \vdots \\
    d_{M1} & d_{M2} & \ldots \ldots & \ldots & d_{MN}
\end{bmatrix}
\]

Where, $d_{11}$ represents rainfall level in the first field of the first record and the interpreted dataset is of size $M \times N$, which is quite different from the size of raw data. The size of the interpreted data can be defined as determined as
Where, \( \alpha \) is the index of the interpreted dataset, \( N \) is the number of columns and \( M \) is the number of rows, step size is the sum of increment of one after every step; \( X \) is the maximum number of step used to do the interpretation.

### 3.2 Feature Extraction Using PSO:

Generally, PSO to extract features focus on selecting the unique set of attributes which are able to distinguish the labels. However, the proposed PSO – based feature extraction is a compromise between the feature selection and the impact of feature extraction over the performance of classifier. Hence the PSO starts with a solution of feature parameters and the classifier parameters. The initial particles are generated randomly and the velocities of each particle are also generated.

The randomly generated initial particles are:

\[
C = (c_1, c_2, c_3, ..., c_N)
\]

(5)

Where, \( c_k : 0 \leq k \leq N - 3 \), represents the row index of \( D' \), i.e., feature parameters and the classifier parameters are considered on \( c_k : N - 2 \) and \( N - 1 \) are the particle parameters. In our work, we consider the primary classifier parameters such as number of hidden layers and number of hidden neurons. Each particle has a velocity which can be represented as:

\[
V = (v_1, v_2, v_3, ..., v_N)
\]

(6)

**Evaluation function:**

In PSO, each individual particle is evaluated with the aid of the evaluation function to determine the optimal solution by finding the mean of difference. In our method, we develop an evaluation model as follows:

\[
\tilde{e} = \frac{1}{e} : e = \frac{1}{|T|} \sum (T - A)
\]

(7)

Where, \( T \) is the target label and \( A \) is the actual output, while validating the developed neural network with the parameters mentioned in \( c_k : N - 2 \) and \( N - 1 \) of the evaluating particle. The maximum value of the evaluation function is stored initially in pbest (pb) value and the same obtained so far is in gbest (gb) value. In every generation, the initial velocities are updated and based on the updated velocity, the particles are also updated. The velocities and particles update are as follows:

\[
I'_{t}(l-1) = V'_{t}(l) + \gamma_1 (p_{b_t} - c_{t}^{(l)}) + \gamma_2 (g_{b_t} - c_{t}^{(l)})
\]

The particles update their position and the velocity until it reaches its termination criteria. The process is repeated until the maximum number of iterations reached. Once the maximum number of iterations is reached the process will be terminated. Here the final solution is the particle is considered as the best. Then the extracted data were used for feeding knowledge classifier. A low level diagram for feature extraction and feeding knowledge to classifier is given in Figure 3.

![Evaluation](image)

**Fig 3:** Low level diagram for proposed Feature Extraction using PSO.

### 3.3 Feeding Knowledge to Classifier:

The artificial neural network includes the input data for training and validation. The solution could not be obtained with one neuron and so we have to go for next possibility. In neural network 70% of the data used for training and 30% of the data used for validation. The training of data in neural network performs as if the rainfall depends on the state of weather conditions. It tries to improve the performance of the data set. The neural network reduces the total error by the input data to produce an output data, changing the weights along its gradient. Depending on the rainfall in the dataset, the rainfall occurs will be predicted. The neural network training intends to minimize an error function that is given as:

\[
\text{Error function} = \frac{1}{T} \sum (T - A)
\]
2.1 Review Stage
After extracting data by using PSO, the extracted data is applied to the feeding knowledge to classifier. Once the training process gets completed, the entire setup is ready to perform for online prediction.

4. RESULTS AND DISCUSSION:
The proposed technique is experimented in the working platform of MATLAB and experimented by rainfall data of various locations. In our work, we use the rainfall data collected by the department of economics and statistics, Government of Tamil Nadu. The data holds the monthly rainfall level for the period 2007-08 Chennai, Tamil Nadu. Over the data set, N-fold cross-validation is applied and the performance is evaluated. The cross-validation results and the performance graphs are given in Table and Figure.
Figure 5 shows the actual and predicted rainfall during round 1.

Figure 6 shows the actual and predicted rainfall during round 2.

Figure 7 shows the actual and predicted rainfall during round 3.

Figure 8 shows the actual and predicted rainfall during round 4.

Figure 9 shows the actual and predicted rainfall during round 5.

Figure 10 shows the actual and predicted rainfall during round 6.
Figure (5-15): shows the actual data and the predicted data over different experiments. Thus the above graphs i.e. fig. (1-11) shows the actual data and the predicted data over the year during the calculations from round (1-11). The graph is plotted for rainfall levels Vs experiments. Here 70% of the data is used for training and 30% of the data is used for validation. Thus after the training and validation of data by using neural network the rainfall measure was predicted. The table 1 shows the actual and predicted rainfall measure over the year for different experiments. Thus total of 11 rounds were performed for the prediction of rainfall measure and hence the predicted data were tabulated and plotted by using the actual and predicted rainfall level over different experiments. The figure 16 shows the comparison graph for our proposed technique and therefore our proposed technique showed to be a better technique to predict rainfall measure while comparing to existing method. The below table 2 shows the comparison over proposed and existing method.
In this paper we proposed feature extraction using PSO and training and validation of data using feeding knowledge classifier along with data interpretation technique. After computing the rainfall over N number of years in data interpretation the PSO is used for feature extraction of data. Then feeding knowledge classifier is used to minimize the error. Therefore the weights of the data are computed and hence the rainfall prediction can be done successfully using our proposed techniques. Thus our proposed method has proved as an efficient method for prediction of the rainfall with an effective accuracy.

REFERENCES:


