MACHINE LEARNING METHODS BASED RAINFALL FORECASTING MODEL

M.V.MADHURI [1], M. RADHIKA [2] PROF J.RAVISANKAR[3]

[1] PG SCHOLAR, DEPT OF CSE, KRISHNAVENI ENGINEERING COLLEGE FOR WOMEN, NARASARAOPET, AP, INDIA.
[2]ASST. PROFESSOR, DEPARTMENT OF CSE, KRISHNAVENI ENGINEERING COLLEGE FOR WOMEN, NARASARAOPET, AP, INDIA
[3]PROFESSOR, DEPARTMENT OF ECE, KRISHNAVENI ENGINEERING COLLEGE FOR WOMEN, NARASARAOPET, AP, INDIA

ABSTRACT: Rainfall prediction is the one of the important technique to predict the climatic conditions in any country. This paper proposes a rainfall prediction model using Multiple Linear Regression (MLR) for Indian dataset. The input data is having multiple meteorological parameters and to predict the rainfall in more precise. The Mean Square Error (MSE), accuracy, correlation are the parameters used to validate the proposed model. From the results, the proposed machine learning model provides better results than the other algorithms in the literature.

1. INTRODUCTION

Predicting rainfall is a major component and is essential for applications that surround water resource planning and management. Over the years numerous attempts have been made at capturing rainfall. One area where it is vital to predict the rainfall amount accurately is within rainfall derivatives. Rainfall derivatives fall under the umbrella concept of weather derivatives, which are similar to regular derivatives defined as contracts between two or more parties, whose value is dependent upon the underlying asset. In the case of weather derivatives, the underlying asset is a weather type, such as temperature or rainfall. The main difference between normal derivatives and weather derivatives is that weather is not tradeable. Hence, typical methods that exist in the literature for other derivatives are not suitable for weather derivatives. In this problem domain the underlying asset is the ac-cumulated rainfall over a given period, which is why it is crucial to predict rainfall as accurately as possible to reduce potential mispricing. Contracts based on the rainfall index are decisive for farmers and other users whose income is directly or indirectly affected by the rain. A lack or too much rainfall is capable of destroying a farmer’s crops and hence their income. Thus, rainfall derivatives are a method for reducing the risk posed by adverse or uncertain whether circumstances. Moreover, they are a better alternative than insurance, because it can be hard to prove that the rainfall has had an impact unless it is destructive, such as severe floods or drought. Similar contracts exist for other weather variables, such as temperature and wind. Within the literature rainfall derivatives is split into two main parts. Firstly, predicting the level of rainfall over a specified time and secondly, pricing the derivatives based on different contract periods/length. The latter has its own unique problem, as rainfall derivatives constitute an incomplete mar-ket1. This means the standard option pricing models such as the Black-Scholes model are incapable of pricing rainfall derivatives, because of the violation of the assumptions of the model; namely no arbitrage pricing. Thus, a new pricing framework needs to be established. This paper focuses on the first aspect of predicting the level of rainfall. Note that it is essential to have a model that can accurately predict the level of rainfall, before pricing derivatives, because the contracts are priced on the predicted accumulated rainfall over a period of time. In order to predict the level of rainfall for rainfall derivatives, the statistical approach of Markov-chain extended with rainfall prediction (MCRP) is used. As already mentioned, the use of MCRP is the most prevalent approach, due to its simplicity. The general approach of MCRP is often referred to as a 'chain-dependent process', which splits the model into capturing first the occurrence pattern, and then the rainfall intensities. The occurrence pattern is produced by calculating the probability of what the outcome of today will be given what happened in the previous day(s).The process of deciding upon what state to be in is performed by a Markov-chain, where state 0 is a dry day and state 1 is a wet day. On the other hand, the intensities are produced by generating random numbers from a distribution that fits the daily data. This step is only calculated if we are in state 1, i.e. wet day. Typically in the literature, the Gamma and Mixed- Exponential distributions provide the best fit for rain data and are most commonly used. We refer the reader to for a complete description of the MCRP approach. However, even though the MCRP approach is quite popular, it faces several drawbacks. First of all, the model is very simplistic and is heavily reliant on past information being reflective of the future. Additionally, the predicted amount is essentially the average level of rainfall observed across the study period and does not take into account annual deviations in weather patterns. Furthermore, the model for each city needs to be specifically tuned as each exhibits different statistical properties, i.e. a new model for each city. Lastly, MCR produces weak predictive models, as its only focus is on fitting the historical data. This last point is very important, as one should not only be interested in deriving models that describe past data effectively, as it currently happens; instead, we should also be focusing on producing effective predictive models, which can offer us insights on future weather trends. Due to the disadvantages highlighted above, we divert away from the use of statistical approaches and in this paper we propose using a machine learning technique called Genetic Programming (GP). Rainfall prediction has not been covered in great detail within the machine learning literature and the applications are mainly focused on the short term predictions i.e. up to a few hours. Little literature exists for the daily predictions, e.g. used a feed-forward back-propagation neural network for rainfall prediction in Sri Lanka, which was inspired by the chain-dependent approach from statistics. To the best of our knowledge, the only work that exists for daily predictions using Genetic Programming is. However, the GP performed poorly by itself, although when assisted by wavelets the predictive accuracy did improve. However, there has been no previous work in using GP in the context of rainfall weather derivatives. The goal of this paper is thus to explore whether GP is able to outperform the usual approach adopted within the rainfall derivative literature, namely MCRP. GP is chosen for this paper over other machine learning techniques, because it has the benefit of producing white box (interpretable, as opposed to black box) models, which allows us to probe the models produced. Moreover, we can capture nonlinear patterns in data without any assumptions regarding the data. This should allow us to produce a model that can reflect the ever changing process of rainfall. As a result, we could capture yearly deviations that the current MCRP is unable to replicate. Additionally, we are able to produce a more general model, which can be applied to a range of cities/climates, without having to build a new model each time. Hence, the main contribution of this paper is that we propose a new GP for the problem of rainfall prediction, and compare its predictive performance against the performance of the current state-of-the-art MCRP approach. This will be the first step towards pricing rainfall derivatives using GP.

2. LITERARY WORKS SURVEY.

1) Elucidating The Role Of Topological Pattern Discovery And Support Vector Machine In Generating Predictive Models For Indian Summer Monsoon Rainfall

AUTHORS: Manojit Chattopadhyay, Surajit Chattopadhyay

The present paper reports a study, where growing hierarchical self-organising map (GHSOM) has been applied to achieve a visual cluster analysis to the Indian rainfall dataset consisting of 142 years of Indian rainfall data so that the yearly rainfall can be segregated into small groups to visualise the pattern of clustering behaviour of yearly rainfall due to changes in monthly rainfall for each year. Also, through support vector machine (SVM), it has been observed that generation of clusters impacts positively on the prediction of the Indian summer monsoon rainfall. Results have been presented through statistical and graphical analyses. Behavior of systems with many interdependent components that lead to organized as well as irregular features is referred to as complexity. In such systems the knowledge of the parts does not necessarily lead to the predictable behaviour of the entire system. Complexities associated with meteorological and geophysics processes have been reviewed in Sharma et al (2012). Modelling complexity of atmospheric phenomena and generating prediction schemes accordingly has long been an area of major concentration for the meteorologists over the globe (Kondratyev and Varotsos, 1995; Varotsos 2005, 2013, Blackwell, 2014). In view of importance of the estimation of the future projected precipitation and rainfall on short- and long-term basis detrended fluctuation analysis has been implemented by Efstathiou and Varotsos (2012) in rainfall time series to explore the intrinsic properties of their temporal variability. In another recent study, Chattopadhyay and Chattopadhyay (2013) explored the association between solar activity and Indian summer monsoon rainfall through spectral analysis after carrying out Box-Cox transformation. Association between SST and ENSO over the tropics has been discussed in a recent study by Varotsos et al. (2014), where they suggested that the warming in the sea surface temperature (SST) since 1900, did not occur smoothly and slowly, but with two rapid shifts in 1925/1926 and 1987/1988, which are more obvious over the tropics and the northern midlatitudes.

2) A Rainfall Prediction Model using Artificial Neural Network AUTHORS: Kumar Abhishek, Abhay Kumar, Rajeev Ranjan, Sarthak Kumar

The multilayered artificial neural network with learning by back-propagation algorithm configuration is the most common in use, due to of its ease in training. It is estimated that over 80% of all the neural network projects in development use back-propagation. In back- propagation algorithm, there are two phases in its learning cycle, one to propagate the input patterns through the network and other to adapt the output by changing the weights in the network. The back-propagation-feed forward neural network can be used in many applications such as character recognition, weather and financial prediction, face detection etc. The paper implements one of these applications by building training and testing data sets and finding the number of hidden neurons in these layers for the best performance. In the present research, possibility of predicting average rainfall over Udupi district of Karnataka has been analyzed through artificial neural network models. In formulating artificial neural network based predictive models three layered network has been constructed. The models under study are different in the number of hidden neurons.

3) A Short-Term Rainfall Prediction Model using Multi-Task Convolutional Neural Networks

AUTHORS : Minghui Qiu, Peilin Zhao, Ke Zhang, Jun Huang, Xing Shi, Xiaoguang Wang, Wei Chu

Precipitation prediction, such as short-term rainfall prediction, is a very important problem in the field of meteorological service. In practice, most of recent studies focus on leveraging radar data or satellite images to make predictions. However, there is another scenario where a set of weather features are collected by various sensors at multiple observation sites. The observations of a site are sometimes incomplete but provide important clues for weather prediction at nearby sites, which are not fully exploited in existing work yet. To solve this problem, we propose a multi-task convolutional neural network model to automatically extract features from the time series measured at observation sites and leverage the correlation between the multiple sites for weather prediction via multi-tasking. To the best of our knowledge, this is the first attempt to use multi-task learning and deep learning techniques to predict short-term rainfall amount based on multi-site features. Specifically, we formulate the learning task as an end-to-end multi-site neural network model which allows to leverage the learned knowledge from one site to other correlated sites, and model the correlations between different sites. Extensive experiments show that the learned site correlations are insightful and the proposed model significantly outperforms a broad set of baseline models including the European Centre for Medium-range Weather Forecasts system (ECMWF).

3. EXISTING SYSTEM:

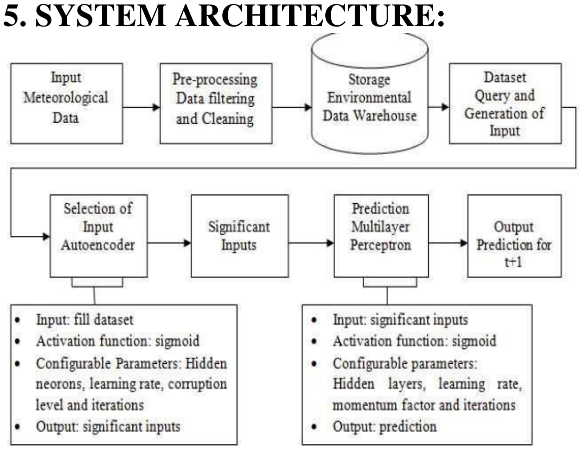
In the Existing system used back propagation neural network for rainfall prediction. This model used by Xianggen Gan and he was tested using the dataset from 1970 to 2000 which has 16 meteorological parameters. During network training the target error is set as 0.01 and learning rate is set as 0.01. This model implemented on mat lab neural network. Genetic Programming (GP) and MCRP were compared on 21 different datasets of cities across Europe. Daily rainfall data for 10 years were taken as training data and one year rainfall data were taken as testing data.
The disadvantage of MCRP is that it predicts accurate only for annual rainfall when compared with monthly rainfall prediction. The assumptions which are made by the multiple linear regression are: linear relationship between the both the descriptive and independent variables, the highly correlated variables are independent variables, yi is calculated randomly. Weather is extremely difficult to forecast correctly. It is expensive to monitor-so many variables from so many sources. The computers needed to perform the millions of calculations necessary are expensive.

4. PROPOSED SYSTEM:

The proposed method is based on the multiple linear regression. The data for the prediction is collected from the publically available sources and the 70 percentage of the data is for training and the 30 percentage of the data is for testing. Multiple regression is used to predict the values with the help of descriptive variables and is a statistical method. It is having a linear relationship between the descriptive variable and the output values. The number of observation is indicated by *n*. The dependent variable is *yi* and the descriptive variable is *xi*. *β0* and *β p* are the constant y intercept and slop of descriptive variable respectively.

The error free prediction provides better planning in the agriculture and other industries.The linear relationship between the both the descriptive and independent variables, the highly correlated variables are independent variables, yi is calculated randomly and the mean and variance are 0 and σ. The ability to determine the relative influence of one or more predictor variables to the criterion value Ability to identify outliers or anomalies

5. SYSTEM ARCHITECTURE:



6. IMPLEMENTATION

User:

The User can register the first. While registering he required a valid user email and mobile for further communications. Once the user register then admin can activate the customer. Once admin activated the customer then customer can login into our system. After login he can search the weather report based on city. For searching the weather report we use open weather map api. The rest connection we are sending to the server and it will return the json data. The json data we are parsing and required information converted into python dictionary sent to the user side. Once it done then the user can perform the preprocess operations. The hist diagram plot diagram are displayed based on the data.

Admin:

Admin can login with his credentials. Once he login he can activate the users. The activated user only login in our applications. The admin can set the data set by Indian metrological weather report. In this report the data has consider as monthly wise and yearly quarterly wise. The admin can add new data to the dataset. So this data user can perform the testing process. The admin can view data based on paginations. The total paginations will be displayed on the browser.

Data Preprocess:

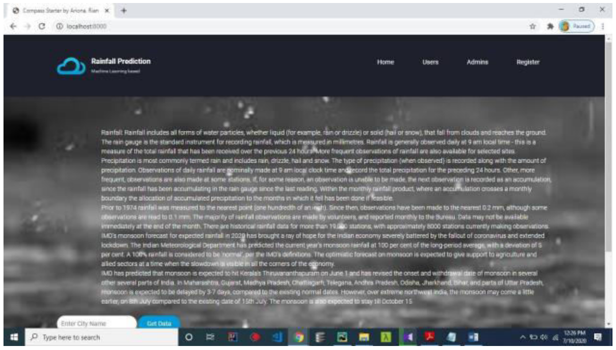
The admin provided data has been stored in the sqlite database. To process our methodology we need to perform data cleaning process. By using pandas data frame we can fill the missing values with its mean type. Once data cleaned the hist diagram will be displayed. This hist diagram, is monthly wise will be displayed. Alter monthly wise bar graph and quarter wise bar graph will be displayed. Then we are calculating the confusion matrix of the selected model. Once its done the user can see the cleaned data in the browser.

Multiple Linear Regression:

The number of observation is indicated by n. The dependent variable is yi and the descriptive variable is xi. β0 and βp are the constant y intercept and slop of descriptive variable respectively. Model error is indicated by €. In the proposed model multiple meteorological parameters are necessary to predict the rain fall, it is better to use the multiple linear regression instead of simple linear regression.

7. SCREEN SHORT

HOME PAGE



PREDICTION RESULT:

8.CONCLUSION

Rain fall prediction plays the major role in agriculture production. The growth of the agricultural products is based on the rainfall amount. So it is necessary to predict the rainfall of a season to assist farmers in agriculture. The proposed method predicts the rainfall for the Indian dataset using multiple linear regression and provides improved results in terms of accuracy, MSE and correlation.

REFERENCES

[1] Manojit Chattopadhyay, Surajit Chattopadhyay, “Elucidating the role of topological pattern discovery and support vector machine in generating predictive models for Indian summer monsoon rainfall”, Theoretical and Applied Climatology, pp. 1-12, July 2015, DOI: 10.1007/s00704-015-1544-5

[2] Kumar Abhishek, Abhay Kumar, Rajeev Ranjan, Sarthak Kumar,” A Rainfall Prediction Model using Artificial Neural Network”, 2012 IEEE Control and System Graduate Research Colloquium (ICSGRC 2012), pp. 82-87, 2012.

[3] Minghui Qiu, Peilin Zhao, Ke Zhang, Jun Huang, Xing Shi, Xiaoguang Wang, Wei Chu, “A Short-Term Rainfall Prediction Model using Multi-Task Convolutional Neural Networks”, IEEE International Conference on Data Mining, pp. 395-400, 2017, DOI 10.1109/ICDM.2017.49.

[4] Aswin S, Geetha P and Vinayakumar R, “Deep Learning Models for the Prediction of Rainfall”, International Conference on Communication and Signal Processing, April 3-5, 2018, India, pp. 0657-0661.

[5] Xianggen Gan, Lihong Chen, Dongbao Yang, Guang Liu, “The Research Of Rainfall Prediction Models Based On Matlab Neural Network”, Proceedings of IEEE CCIS2011, pp. 45- 48.