Modelling Rainfall Prediction Using Data Mining Method - A Bayesian Approach

Mr.Chetan J. Janbandhu
Bell labs (Nokia)
New Jersey USA
Email:chetanj@gmail.com

Prof. Praful D.Meshram
Deptt. of Computer Engineering
Cummins College of Engineering
Pune University,India
Email:praful.meshram@gmail.com

Prof. Madhuri N. Gedam
Deptt. of Information Technology
Shree L. R. Tiwari College of Engineering
Mumbai University, India
Email:madhuri.gedam@gmail.com

***Abstract*—**Weather forecasting has been one of the most technically difficult problems around the globe. Weather data is meteorological data. It can be used for weather prediction. Weather data has 36 attributes but only 7 attributes are most important to rainfall prediction. Data is pre-processed to use it in this Bayesian approach. It is the data mining prediction model for rainfall prediction. The model is trained using the training data set and has been tested for accuracy on test data. The meteorological centres use high computing and supercomputing power to run weather prediction model. To address the problem of compute intensive rainfall prediction model, this paper studies data intensive model using data mining technique. This model works with efficient accuracy and uses moderate amount of compute resources for rainfall prediction. Bayesian approach is used for rainfall prediction. It works well with good accuracy.

Keywords*—*Data Mining, Bayesian, rainfall prediction, High Performance Computing.

*****

I. INTRODUCTION

Depending on the spatial and temporal scales of atmospheric systems and details of the accuracy desired, the weather forecasts are divided into the categories:

a) Now casting: Now Casting tells about the current weather and forecasts up to a few hours.
b) Short range forecasts(1 to 3 days): Short range forecasts in which the weather (mainly rainfall) in each successive 24 hrs. Intervals may be predicted up to 3 days.
c)Medium range forecasts (4 to 10 days): Medium range forecasts for 4 to 10 days.
d) Long range /Extended Range forecasts (more than 10 days to a season): Long Range Forecasting range from a monthly to a seasonal forecast.[1]

As the existing prediction models requires a supercomputing, Indian Meteorological Department (IMD) has increased its infrastructure for meteorological observations, communications, forecasting & weather services and contributed to scientific growth since its establishment in 1875. It has simultaneously nurtured the growth of meteorology and atmospheric science in India. Systematic observation of basic climate, environmental and oceanographic data is vital to capture past and current climate variability, and has the decent state of the art data capturing facilities.

Weather research and forecasting (WRF) model, General Forecasting Model, Seasonal Climate Forecasting, Global Data Forecasting Model, are currently acceptable models for weather prediction. Also, computing for these prediction models is very expensive because of compute intensive nature. On the contrary, data mining models works on historical data, it works on probability and/or similarity patterns. For all the prediction categories, the model works in similar fashion, and expects to return the moderate accuracy [2].

II. LITERATURE SURVEY

Jae-Hyun Seo et al. [3] compared the prediction models' performance using support vector machine (SVM), k-nearest neighbors algorithm (k-NN), and variant k-NN (k-VNN), which generally achieved ideal accuracy on the rain/no-rain in South Korea.

Rajesh [4] compared the following classification methods namely Decision Trees, Rule-based Methods, Neural Networks, Naïve Bayes, Bayesian Belief Networks, and Support Vector Machines. He concluded that to tap the potential of huge amount of data, decision trees can be used in predicting the dependent variables like fog and rain.

Kannan, Prabhakaran and Ramachandran [5] computed five years data and then compared with predicted data using regression approach. Here, the prediction of rainfall is by using multiple linear regression method. The predicted values lie below computed values. According to the results, it does not show accuracy but show an approximate value.

James, Bavy and Tharam [6] proposed Improved Naïve Bayes Classifier (INBC) technique and explores the use of genetic algorithms (GAs) for selection of a subset of input features in classification problems. According to the performance, two schemes is built scheme I uses all basic input parameters for rainfall prediction and scheme II uses the optimal subset of input variables which are selected by a GA. According to the results predicted INBC achieved 90% accuracy rate on the rain/no-rain classification problems

Jesada, Kok and Chung [7] proposed fuzzy inference system for monthly rainfall prediction in the northeast region of Thailand. Accordingly, the experimental results show the modular FIS is good alternative method to predict accurately. The predicted mechanism can be interpreted through fuzzy rules. The experimental results provide both accurate results and human-understandable prediction mechanism.

III. FORECASTING MODELS

Generic Forecasting Model:

The weather prediction model works with certain defined steps, which covers observation of weather parameters, collecting the weather data, plotting for analysis, making analysis, and weather prediction. The functionalities [8] of weather prediction model is described as,

1. Observation: Surface observations are made at least every three hours over land and sea. Weather stations and automatic stations observe the atmospheric parameters.
2. Collection and Transmission of Weather Data: Weather observations which are condensed into coded figures, symbols and numerals which are transmitted to designated collection centres.
3. Plotting of Weather Data: Coded messages are decoded and each set of observations is plotted in symbols or numbers.
4. Analysis of Weather Maps, Satellite and Radar Imageries and other Data: Numerical Weather Prediction Model output used plotting maps.
5. Formulation of the forecast: The preparation of forecasts starts after the analysis of all meteorological data has been completed.

IV. EXISTING FORECASTING MODELS

1. The Weather Research and Forecasting (WRF) model: It is a numerical weather prediction (NWP) and atmospheric simulation system designed for both research and operational applications [9].
2. Global Forecasting System: The Global Forecast System (GFS) is a weather forecast model produced by the National Centers for Environmental Prediction (NCEP).
3. Seasonal Climate Forecasting: Seasonal climate forecasting was developed for Pacific Island nations. It provides a seasonal climate prediction system.

V. PROPOSED MODELS

The proposed model works in following steps:

1. Data Collection and Pre-processing: Data is obtained from Indian Meteorological Department. But the meteorological data is poor in quality. So data has to preprocess carefully to obtain correct results. Out of 36 features, only following 7 features have been used.

TABLE I. FEATURES OF DATA

Attribute	***Description***
Temp	Temp is in deg. C
Station Level Pressure	SLP in hpa
Mean Sea Level Pressure	MSLP in hpa
Relative Humidity	RH in percentage
Vapor Pressure	VP in hpa
Wind Speed	Wind speed in kmph
Rainfall	Rainfall in mm

2. Bayesian Rainfall Prediction Model: The Bayesian classifier is a supervised learning method for classification. Bayes classification can be viewed as both a descriptive and a predictive type of approach. The probabilities are descriptive and are used to predict the class membership for a target tuple. The prediction model actually builds using training dataset. The build model we tested using test dataset. The data set is divided into 70:30 standard ratios for training data and test data respectively.
3. Bayesian Prediction Algorithm:

   Input: Weather data set for all 7 attributes

   Output: Rainfall prediction for the input query

   Algorithm: Bayesian Prediction model

   1. Input raw weather data
   2. Apply filters and transformations
   3. Store ‘targetdata’ for further processing
   4. BuildModel (targetdata)

   For all classes Ci
   Compute prior probability P(Ci)
   For all attribute features, Fj
   Find probability P(Fj|Ci)
   EndFor
   EndFor

   Prediction (InputQuery)

   a. Multiply, $P(Fj|Ci) = \prod P(Fj \mid Ci)$
      Calculate, $d(f) = P(Ci) * \prod P(Fj \mid Ci)$

   b. Select class, d(fi) with highest probability value to classify the input query

VI. RESULTS AND DISCUSSIONS

For building the model, four data sets have used, out of which three datasets are of actual cities data. We have used monsoon period data. The model observed to be more accurate if the training dataset is very large. The data set and the obtained results are shown below

TABLE II. ACCURACY MEASUMENT

Dataset	Data Set (Records)		Correctly Classified	Accuracy
	Training	Test		
Sample Data (Small)	432	60	49	81.66%
Pune City	4400	494	462	93.52%
Mumbai City	4840	494	448	90.69%
Delhi City	4888	494	475	96.15%

VII. CONCLUSION

Data mining approach for rainfall prediction model is data intensive model instead of compute intensive model which are being used in prediction centres. This model is nearly accurate model in comparison with compute intensive models. Because of data mining approach, computing power is reduced. The model returns good prediction results when the training dataset is large,. The negative part of model is, when a predictor category is not present in the training data, the model assumes that a new record with that category has zero probability. . The performance of the model can also be improved by designing the model for scalable platforms, either for vertical scalability or for horizontal scalability.

REFERENCES

[1] Ganesh P. Gaikwad, V. B. Nikam, “Different Rainfall Prediction Models And General Data Mining Rainfall Prediction Model”, International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181, Vol. 2 Issue 7, July – 2013

[2] Folorunsho Olaiya, Owo, “Application of Data Mining Techniques in Weather Prediction and Climate Change Studies”, IJIEEB Vol.4, No.1, February 2012

[3] J.-H. Seo, Y. H. Lee, and Y.-H. Kim, "Feature selection for very short-term heavy rainfall prediction using evolutionary computation," Advances in Meteorology, vol. 2014, 2014.

[4] Rajesh Kumar, “Decision tree for the weather forecasting”, International Journal of Computer Applications (0975 – 8887) vol.2, August 2013.

[5] M.Kannan, S.Prabhakaran and P.Ramachandran, “Rainfall Forecasting Using Data Mining Technique”, International Journal of Engineering and Technology, Vol.2 (6), 397-401, 2010.

[6] James N.K.Liu, Bavy N.L.Li, and Tharam S.Dillon, “An Improved Naïve Bayesian Classifier Technique Coupled with a Novel Input Solution Method”, IEEE Transactions on systems, man, and Cybernetics – Part C: Applications and Reviews, Vol.31, No.2, May 2001.

[7] Jesada Kajornrit, Kok Wai Wong and Chun Che Fung, “Rainfall Prediction in the Northeast Region of Thailand Using Modular Fuzzy Inference System”, World Congress on Computational Intelligece, 10-15, June 2012.

[8] Andrew Kusiak,Xiupeng Wei,Anoop Prakash Verma, Evan Roz “Modeling and Prediction of Rainfall Using Radar Reflectivity Data: A Data-Mining Approach” IEEE Transactions On Geoscience and Remote Sensing,Volume 54, Issue 4, April 2013.

[9] http://wrf-model.org/index.php