TEST Engineering & Management

May – June 2020
ISSN: 0193-4120 Page No. 1778 - 1783

An Experiment of Wet and Dry Day Prediction and Qualitative Precipitation Forecast using Data Mining Technique for Port Blair Station

R. Bala Krishnan[1], S. Meganathan[2], N. Rajesh Kumar[3], J. Senthil Kumar[4], D. Narasimhan[5]
[1,2,3,4]*Department of Computer Science and Engineering,*
[5]*Department of Mathematics,*
SASTRA Deemed University, Srinivasa Ramanujan Centre, Kumbakonam, India.
(email: meganathan@src.sastra.edu)

Article Info
Volume 83
Page Number: 1778 - 1783
Publication Issue:
May - June 2020

Article History
***Article Received**: 11August 2019*
***Revised**: 18November 2019*
***Accepted**: 23January 2020*
***Publication**: 10 May2020*

Abstract:

Short range rainfall forecast is usually given based on the synoptic situation. However point rainfall forecast is still a challenging task for the forecasters. With the recent developments of internet several new techniques have evolved in various fields. Data mining presents various rule based techniques and classifier methods for data analysis and extracting knowledge from different fields. The same is tried to forecast the rainfall in a day during various months of North East monsoon at Port Blair, a place of tourist importance station which is an island in Andaman group in Bay of Bengal. The tool is also tested for the qualitative forecast of rain and experimental results are summarized here.

Keywords: Association rule mining, classifier approach, rainfall forecast, qualitative forecast.

1. Introduction

Rainfall is predicted from the charts identifying the synoptic system and then capacity to cause rain over a location. However in practice it is found that the forecast is valid over a region as a whole and point rainfall prediction is still a serious task for operational forecasters [3]. For India cyclone and monsoons are systems with understandable rainfall pattern [4]. However for islands with larger continents far away identification of synoptic system is a difficult to predict the precipitation patterns as there would not be any observations over the surrounding oceans. Radar and satellite data of recent decades help to certain extent in assessing the rain potential of clouds for local forecast. However it is more conversion to rely on local data for a successful forecast as any effect of synoptic system has to implicitly evident in local weather. Such a method is tried for forecast of dry and wet weather for Port Blair which is an island station of Andaman in Bay of Bengal.

Data mining method is preferred in this direction as this is an era of data analytics and pattern evaluation. The potential of data mining techniques for weather forecast has been reported in recent times [5, 6, 7, 8]. Hence association rules are extracted for precipitation patterns and classified with occurrence of wet spell and dry season during the period of North East Monsoon (NME) and South West Monsoon (SWM) of island station.

2. Material and Methods

2.1 Data Preparation

Port Blair (Latitude 11°40' N / Longitude 92°44' E) is a beautiful island located in the east coast of South Andaman. The weather patterns of this island is maintained by India meteorological department since 1961. Weather forecasters are considered various atmospheric parameters such as, temperature (TMP), dew point (DEWP), speed of wind (WDSP), visibility (VISIB) precipitation (rainfall) and weather prediction. A data set consists of five atmospheric variables including TMP, DEWP, WDSP, VISIB is collected from National Climatic Data Center

May – June 2020
ISSN: 0193-4120 Page No. 1778 - 1783

which is world most active weather data center located in United States of America. This scheme concentrates over the month of North East Monsoon (October to December) and South West Monsoon (June to September).

The raw set of seasonal information is converted into nominal values (LOW, MEDIUM and HIGH) by applying unsupervised attribute of discretization algorithm. After the data preprocessing, a total of 2635 instances were obtained for analyzing weather forecast. The discretization approach provides various best fit ranges for the five atmospheric variables which can be used for the investigation of rainy day Table 1.

Table 1.
Nominal values for atmospheric parameters

Parameter	Description	Unit	Nominal Variable	Range
TEMP	Temperature	Celsius	LOW	< 25.2 (77.36F)
			MEDIUM	25.2-27.7 (77.36F – 82F)
			HIGH	>27.7 (82F)
DEWP	Dew Point	Celsius	LOW	<19 (66.2F)
			MEDIUM	19-22.5 (66.2F – 72.6F)
			HIGH	>22.5(72.6F)
WDSP	Wind Speed	Knots	LOW	<10.13
			MEDIUM	10.13 – 20.26
			HIGH	> 20.26
VISIB	Visibility	Miles	LOW	<4.6
			MEDIUM	4.6 – 8.5
			HIGH	> 8.5
PRCP	Rainfall	Millimeter	YES	>0
			NO	=0

2.2 Association Rule mining for weather forecast

The problem of mining association rules was first nominated and elaborated by Agrawal et al. [1] for discovering interesting patterns using association rules for business intelligence. Recently the rainfall patterns are discovered using association rule mining approach for specific stations where the coastal and inland of Tamil Nadu State of South India in a case study [9]. When the association rule is involved on meteorological data, with each record showing various atmospheric observations including wind direction, wind speed, temperature, relative humidity, rainfall, and mean sea level pressure. These climate parameters were taken at a certain period for certain region, then association rules for occurrence of wet and dry day prediction can be found.

An enhanced predictive Apriori algorithm has been used for the association rule approach with filtered dataset. This algorithm finds the support threshold value for the best 'n' rules concerning a support-based corrected confidence value.

3. Results and Discussion

3.1 Association rule for weather data

Predictive mining performs inference on the present data in order to make a prediction. Here a weather forecasting has been implemented using classification and association mining with various weather parameters such as temperature, dew point, visibility, wind speed and precipitation. The rule is defined by the expression $A \Rightarrow B$. It holds the transaction set D with support s, where s is the percentage of transactions in D that contain $A \cup B$ (i.e., the union of sets A and B, or say, both A and B). This measurement is taken for finding the probability between $P(A \cup B)$.

This term Conditional Probability (CP) for P (B|A) is stated by the following metrics.

Support (S) of $(A \Rightarrow B) = P(A \cup B)$

Confidence(C) of $(A \Rightarrow B)$ = Support_Count (SC) of $(A \cup B)$ / Support_Count (SC) of (A)

In the past the Association Rule (AR) mining techniques are implemented for the rainfall prediction for 24 hours prior by the authors Meganathan et al., 2011. The predictive Apriori mechanism generates the efficient association patterns as AR. Some of the proficient AR which are utilized to accomplish prediction from the given training dataset are the model generates AR for the incidence and non-incidence of rainfall on wet and dry days for 24 hours prior [Table 2] and 48 hours prior [Table 3]. The motivating practices are also rendered using the AR mining process for

May – June 2020
ISSN: 0193-4120 Page No. 1778 - 1783

the measurement of the rainfall prediction with class marker 'high' for the precipitation value greater than 2.56 cm (1 inch) and class marker as 'normal' for the precipitation value less than or equal to 2.56cm (1 inch) for 24 hours prior [Table 4] and 48 hours prior [Table 5].

Table 2: Weather patterns for 24 hours ahead rainfall prediction during NEM months.

AR (*A* ⇒ *B*)
[TEMP=(medium or high) ^DEWP=(low or medium or high) ^ WDSP=low ^VISIB=(low or medium)] ==> PRCP=no
[TEMP=low ^ DEWP=high ^ WDSP=medium ^ VISIB=low] ==> PRCP=yes

Table 3: Weather patterns for 48 hours ahead rainfall prediction during NEM months.

AR (*A* ⇒ *B*)
TEMP=medium or high ^ DEWP= low or medium or high ^ VISIB= medium or high ^ WDSP= low or medium ==> PRCP=no
TEMP=low or high ^ DEWP=high ^ VISIB=low or medium ^ WDSP=low or medium ==> PRCP=yes

Table 4: Weather practices for rainfall approximation at 24 hours prior during NEM months.

AR (*A* ⇒ *B*)
TEMP=high ^ DEWP= high or medium ^ VISIB=low or medium ^ WDSP=low or medium ==> PRCP=low
TEMP= high ^ DEWP = high or medium ^ VISIB=low WDSP='high 4 ==> PRCP=high

Table 5: Weather practices for 48 hours ahead rainfall estimation during NEM months.

AR (*A* ⇒ *B*)
TEMP=high DEWP=low or medium or high
WDSP=low or medium VISIB=low ==> PRCP=low 21 acc:(0.91196)
TEMP=medium or high DEWP=medium or high WDSP=high VISIB=low => PRCP=high 3

3.2 Validation

The justification for the rainfall related data is carried out to detect the reliability of the generated results and to confirm its suitability in live environment for the prediction of dry and wet days. Validation has been done through K* methodology [2]. For the prediction of rainfall occurrence and nonoccurrence during the monsoon days, the machine learning (ML) technique K* attains an accuracy of 68% and 65% for NEM months and 75% and 73% for SWM months using cross-validation method [Table 6] and 69% and 65% for NEM months and 77% and 73% for SWM months using percentage split method [Table 8] for 24 and 48 hours prior respectively. The obtained confusion matrix (CM) for the above methods are presented in [Table 7] and [Table 9]. The substantiation outcomes are stated in [Table 10] and [Table 11] using supplied test set method for the rainfall occurrence prediction and assessment of the rainfall [Table 12] with threshold value greater than 2.56cm (1 inch) for high precipitation and less than or equal to 2.56cm (1 inch) for low precipitation and these results are reasonably accurate. The results shows that the rainfall prediction as well as rainfall estimation is more suitable in SWM months rather than NEW months in this island station. However the accuracy brings a good classification results for both NEM and SWM when the accuracy threshold is greater than 66%. It is higher than the existing synoptic method as well as numerical prediction methods.

Table 6: Correctness of classification over the rainfall data by 10-fold cross validation

Stratified cross-validation	NEM				SWM			
	24 Hours (24 Hrs) Prior		48 Hours (48 Hrs) Prior		24 Hours (24 Hrs) Prior		48 Hours (48 Hrs) Prior	
Correctly Classified Instances	1804	68%	1703	65%	2515	75%	1902	73%
Incorrectly Classified Instances	831	32%	932	35%	821	25%	720	27%

May – June 2020
ISSN: 0193-4120 Page No. 1778 - 1783

Table 7: CM of 10-fold cross validation for rainfall occurrence prediction during NEM.

24 hours prior			48 hours prior		
a=no	b=yes	<-- classified as	a=no	b=yes	<-- classified as
982	506	a = no	918	570	a = no
325	822	b = yes	362	785	b = yes

Table 8: Classification accuracy using percentage split validation for rainfall prediction.

Stratified cross-validation	NEM				SWM			
	24 Hours (24 Hrs) Prior		48 Hours (48 Hrs) Prior		24 Hours (24 Hrs) Prior		48 Hours (48 Hrs) Prior	
Correctly Classified Instances	621	69 %	581	65 %	873	77%	641	72%
Incorrectly Classified Instances	275	31 %	315	35 %	261	23%	250	28%

Table 9: Confusion matrix of percentage split validation for rainfall prediction during NEM.

24 hours ahead			48 hours ahead		
a=no	b=yes	<-- classified as	a=no	b=yes	<-- classified as
327	157	a = no	293	191	a = no
118	294	b = yes	124	288	b = yes

Table 10: Classification accuracy using supplied test set method for NEM rainy day forecasting

Testing year	24 hour ahead		48 hour ahead	
	Percentage of Instances Classified at acceptable level	Percentage of Instances Classified at unacceptable level	Percentage of Instances Classified at acceptable level	Percentage of Instances Classified at unacceptable level
2006	63 %	37 %	57 %	43 %
2007	74 %	26 %	72 %	28 %
2008	71 %	29 %	66 %	34 %
2009	60%	40 %	54 %	46 %
2010	76 %	24 %	74 %	26 %

Table 11: Classification accuracy using supplied test set method for SWM rainy day forecasting

Testing year	24 Hours (24 Hrs) Prior		48 Hours (48 Hrs) Prior	
	Percentage of Instances Classified at acceptable level	Percentage of Instances Classified at unacceptable level	Percentage of Instances Classified at acceptable level	Percentage of Instances Classified at unacceptable level
2006	89%	11%	89%	11%
2007	86%	14%	86%	14%
2008	86%	14%	86%	14%
2009	81%	19%	81%	19%
2010	84%	16%	84%	16%

May – June 2020
ISSN: 0193-4120 Page No. 1778 - 1783

Table 12: Classification accuracy using supplied test set (STS) methodology for precipitation assessment with a possible threshold values during NEM.

Testing year	24 Hours (24 Hrs) Prior		48 Hours (48 Hrs) Prior	
	Percentage of Instances Classified at acceptable level	Percentage of Instances Classified at unacceptable level	Percentage of Instances Classified at acceptable level	Percentage of Instances Classified at unacceptable level
2006	76.6667 %	23.3333 %	76.6667 %	23.3333 %
2007	81.8182 %	18.1818 %	81.8182 %	18.1818 %
2008	74.2857 %	25.7143 %	74.2857 %	25.7143 %
2009	90.6977 %	9.3023 %	90.6977 %	9.3023 %
2010	76.2712 %	23.7288 %	76.2712 %	23.7288 %

4. Conclusion

The AR mining process and instance based classifier approach (IBCA) are employed to detect the occurrence of rainfall on wet and dry days with class markers. The observed outcomes states that, the AR mining methodology based forecasting system works at a proficient rate and the predictions are reasonably accurate and can be used for predicting the occurrence and non occurrence also. The same could be deployed to estimate the rainfall with 2.56cm (1 inch) threshold value for Port Blair island station of Bay of Bengal. As per the stated outcomes, the methodology is suitable for detecting weather conditions and predicts the rainy days 48 hours prior of its occurrence and estimates the normal and high precipitation during both SWM and NEM periods. Hence the scheme prognosticates to be a valuable solitary for tropical islands.

Acknowledgements

The authors of the above stated research would like to state their heartfelt thanks to National Climatic Data Center (NCDC), Asheville, North Carolina, U.S.A., for their support to carry out this work at seamless level by their dataset and special thanks to the Indian Meteorological Department scientists Late Dr. T.R.Sivaramakrishnan, S.R.Ramanan and S.Balachandran at Chennai for useful discussions. The authors thank SASTRA University for utilizing the Discrete Mathematics Laboratory funded by Department of Science and Technology – Fund for Improvement of S&T Infrastructure in Universities and Higher Educational Institutions, Government of India (SR/FST/MSI-107/2015).

References

[1] Agrawal, R., and R. Srikant: Fast Algorithms for Mining Association Rules. Proc. of the 20th Int. Conference on Very Large Databases, Santiago, Chile, Sept. 1994. Expanded version available as IBM Research Report RJ9839.

[2] Cleary, J. G., and L. E. Trigg, K*: An Instance-based learner using an entropic distance measure. Proceedings of the 12th Int. Conf. on Machine Learning, San Francisco, Morgan Kaufmann, 1995, pp. 108-114.

[3] Balachandran, S., R. Asokan, and S. Sridaran, Global surface temperature in relation to northeast monsoon rainfall over Tamil Nadu. Journal of Earth System Science, 2006, 115, 3, pp. 349-362.

[4] Raj, Y. E. A., Inter and intra-seasonal variation of thermodynamic parameters of the atmosphere over coastal Tamil Nadu during northeast monsoon. Mausam, 1996, 47, 3, pp. 259-268.

[5] Meganathan S, Michael Raj T F, RajaKumar B, Raghuraman K, Rajesh Kumar N. Precipitation Prediction for South West Monsoon Over Karnataka using Supervised Learning Technique. International Journal of Recent Technology and Engineering, 2019; 8: pp. 4450-4454.

[6] Meganathan S, Bala Krishnan R, Sumathi A, Sheik Mohideen Shah S, Senthilkumar J. An Experiment in Weather Prediction for North East and South West Monsoon Over Chennai. International Journal of Recent Technology and Engineering, 2019; 8: pp. 4460-4465.

[7] Meganathan S, Sivaramakrishnan T.R, “A Technique for Spot Forecasting”, MAUSAM, 2015, 66(1), pp. 33-42.

[8] Varahasamy R, Meganathan S, Durga Karthik. Association Rule Mining For South West Monsoon Rainfall Prediction and Estimation over Mumbai Station. International Journal of Recent

May – June 2020
ISSN: 0193-4120 Page No. 1778 - 1783

Technology and Engineering, 2019; 8: pp. 4490-4493.

[9] Sivaramakrishnan, T.R, Meganathan, S. Association Rule Mining and Classifier Approach for Quantitative Spot Rainfall Prediction. Journal of Theoretical and Applied Information Technology, 2011, Vol. 34, No. 2, pp 173-177.

[10] Zubair, L., and Ropelewski, The strengthening relationship of ENSO and the North East Monsoon rainfall over Sri Lanka and Southern India. Journal of Climate, 2006, 19(8), 1567-1575.