Journal of Theoretical and Applied Information Technology
31st January 2012. Vol. 35 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

PATTERN VISUALIZATION ON METEOROLOGICAL DATA FOR RAINFALL PREDICTION MODEL

[1]S. MEGANATHAN , [2]T.R. SIVARAMAKRISHNAN
[1]Department of CSE, SASTRA University, Kumbakonam, 612 001
[2]School of EEE, SASTRA University, Thanjavur, 613 401
E-mail: [1]meganathan@src.sastra.edu, [2]trsivaraman@yahoo.co.in

ABSTRACT

Regional rainfall forecast is an important task for meteorologists. While statistical methods are usually in vogue, for this, the concept of data mining is applied to see the efficacy. The objective of this work is to develop a forecast method for the rainfall of the Northeast (NE) season over the Cauvery delta region of South India. There are three main parts in this work. The first part is in loading the meteorological data into centralized server so that it can be extracted and filtered according to our needs. Secondly, graphical visualization has been applied by using gnuplot. And finally the association rules have been mined from those former outputs. From these results we can assess the monthly and seasonal rainfall for the future. Results show that overall accuracy of this approach is satisfactory.

Keywords: *Association Rule, Data Mining, Rainfall Prediction, Gnuplot.*

1. INTRODUCTION

Cauvery river flows in the peninsular India from west to east which is very vital for agricultural production in Tamilnadu state of South India. Besides this region gets the main seasonal rainfall during the northeast monsoon months of October to December. In summer months from May to August as well as September also thunderstorm rains occur here. In India the period from mid June to September is called Southwest (SW) monsoon or Summer monsoon. With the Cauvery waters getting depleted in the downstream regions due to the construction of many dams, the Cauvery delta farmers have to depend upon the seasonal rainfall to a great extent since the nineties of the last century. In the late eighties many townships also came up in delta region. Hence seasonal rainfall estimate over the delta and its prediction will go a long way to help the agriculturists, hydrologists and water managers.

Several studies have been made regarding southwest monsoon as recorded by Rao Y.P,1976[19]. For Tamil Nadu and East Coastal of India, the Northeast monsoon is more relevant as that is the main rainy season. Some works are available regarding rain assessment and forecast for Northeast monsoon also Balachandran et.al.,2006 [5], Dhar and Rakhecha 1983[7], Duraiswamy 1946[8], Kripalanli and Pankaj Kumar 2004[10], Pankaj Kumar et al 2007[16], Lau[11], Raman[17], Rao K.V 1963[18], Sivaramakrishnan [24], Sivaramakrishnan and Sridharan [26], Sridaran and Muthuchamy 1990[30] and Zubair and Roplowski 2006[31] to discuss the nature and prediction of rainfall for India based on conventional statistics. A few isolated attempts in India by Sivaramakrishnan [25], Mohanty [13] and Seetharam [21] are available to study the rainfall over single stations. But all of them use conventional synoptic correlations and they are for a country as a whole. With the vastly varying terrains in India, no clear correlation between Southwest (June to September) and Northeast (October to December) monsoon rainfall could be established. Hence at sub regional level where there is fairly a uniform terrain; study has to be conducted with latest tools and methodologies to predict the Northeast (NE) monsoon rainfall from the earlier rains. Such a study has been attempted for the Cauvery delta basin in Tamilnadu state of South India. There is no earlier study specifically for this region. Usually the sowing operations start during July to August. The SW monsoon is about 30% of annual rainfall over this region, but the NE monsoon (70% of annual rainfall) is found to show large variations. Hence the water managers have to adjust the dependable rainfall realized by August with the operation schedule allowing for certain threshold

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

rainfall by November end. This needs a model for prediction of NE monsoon rain with good chances of success.

Graphical mining and data mining technique [1],[2],[9] are some computer based tools which have been successfully used for predicting parameters in certain fields in the recent past. Same are being tried here for rainfall prediction for the first time.

2. DATA USED

Monthly rainfall data for available stations of the composite Thanjavur district of Tamilnadu in India in downstream Cauvery basin for 25 years since 1985 have been used for analysis. The area covered is shown in Fig 1. The sample stations are Grand Anaicut(1), Thanjavur(2), Papanasam(3), Kumbakonam(4), Valangaiman(5), Nannillam(6), and Nagapattinam(7).

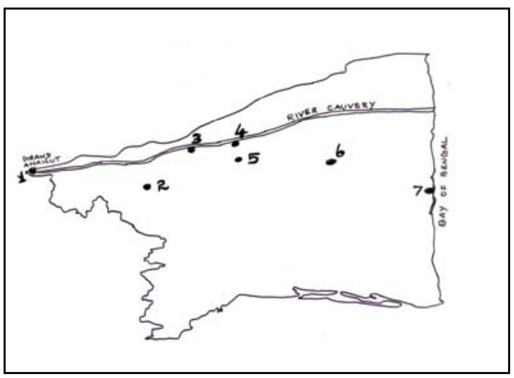

Figure 1: Rain gauge stations in the cauvery delta of composite thanjavur district in south India.

3. METHODOLOGY

3.1 Association Rules

Association rule mining [2] discovers the relationship between items from the set of transactions [3]. These relationships can be expressed by association rules such as [$i_1 \Rightarrow i_2, i_3$ support, s = 2%, confidence, c = 60%]. This association rule implies that 2% of all the transactions under analysis show that items i_1, i_2 and i_3 appear jointly. A confidence of 60% indicates that 60% of the transactions containing i_1 also contain i_2 and i_3. Associations may include any number of items on either side of the rule.

The problem of mining association rules was first introduced in the last decade by Agrawal et al in 1993 [1], Agrawal and Srikant in 1994 [2], Agrawal et al in 1995 [3], and Sarawagi et al in 2000 [20].

Let I = $\{i_1, i_2, \ldots, i_m\}$ denote a set of literals, namely, items. Moreover, let D represent a set of transactions, where each transaction T is a set of items such that $T \subseteq I$. A unique identifier, namely TID, is associated with each transaction. A transaction T is said to contain X, a set of some items in I, if $X \subseteq T$. An association rule implies the form $X \Rightarrow Y$, where $X \subset I$, $Y \subset I$ and $X \cap Y = \phi$. The rule $X \Rightarrow Y$ holds in the transaction set D with confidence, c, where c% of transactions in D that contain X also contain Y. The rule has support, s, in the transaction set D if s% of transactions in D contain $X \cup Y$. An efficient algorithm is required that restricts the search space and checks only a subset of all association rule, yet does not miss important rules.

The apriori algorithm [2] developed by Agrawal et al. (1993) and Srikant and Agrawal (1997) is such an algorithm. However, the interestingness (validity) of the rule is only based on support and confidence.

Association rule mining is a popular technique for marketing basket analysis, which typically aims at discovering buying patterns of customers in supermarkets, mail order and other types of stores. By mining association rules, marketing analysis try to find sets of products that are frequently bought together, so that certain other items can be inferred from a shopping cart containing particular items. Association rules [3] can often be used to design marketing promotions, for example, by approximately arranging products on supermarket shelves and by directly suggesting items to customers that may be of interest to them.

With the constant collection and storage of considerable quantities of business data, association rules are discovered from the domain databases and applied in many areas, such as marketing, logistics and manufacturing. In the areas of marketing, advertising and sales, corporations have found that they can benefit enormously if implicit and previously unknown buying and calling patterns of customers can be discovered from large volumes of business data.

Generally, support and confidence are taken as two measurable factors to evaluate the 'interestingness' of association rules which is discussed by Agrawal et al., 1993 [1]; and Srikant and Agrawal, 1994 [2].

Journal of Theoretical and Applied Information Technology
31st January 2012. Vol. 35 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Association rules are regarded as interesting if their support and confidence are greater than the user-specified minimum support and minimum confidence, respectively. In data mining, it is important but difficult to determine these two thresholds of 'interestingness' (rule validity) appropriately. Data miners usually specify these thresholds in an arbitrary manner. However, this paper has attempted to employ this technique on the meteorological data.

3.2. Proposed Approach

To apply the techniques of association rules in graphical mining, it is necessary to filter the raw data and convert into the format which is accepted by visualization tool, gnuplot. The proposed approach is described in Figure 2 which emphasizes the steps involved in the knowledge discovery using graphical mining of data mining technique.

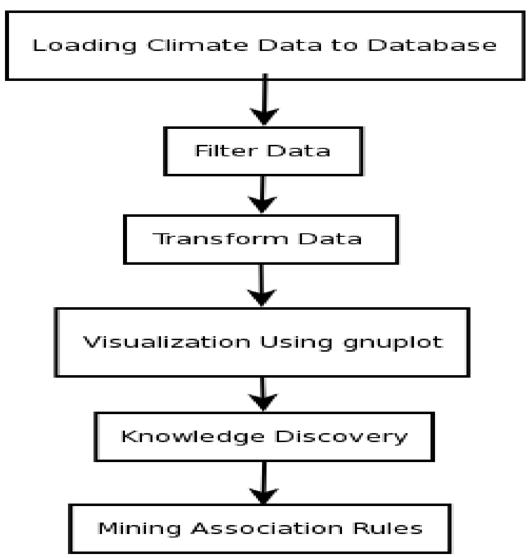

Figure 2. Steps involved in the knowledge discovery using graphical mining.

Step 1: Loading climatic data to SQL Database.

Step 2: Filtering the noises and unwanted data from the dataset.

Step 3: Transform data as required for the visualization software.

Step 4: Generating graphs using the visualization tool.

Step 5: Studying of patterns on the generated results.

Step 6: Mining rules from those patterns.

Using the graphical mining and data mining technique, rainfall realizable in the NE monsoon was assessed from the rainfall realized during June, July and August months with association rules. The running 5 year mean rainfall is shown for different stations in Figures 3(a) to 3(g).

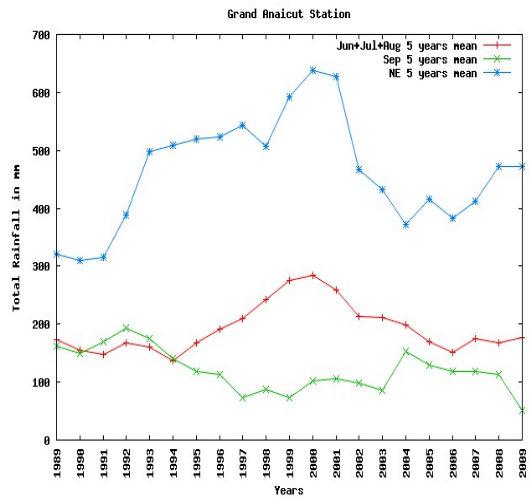

Figure 3(a): Five year mean rainfall of grand anaicut station.

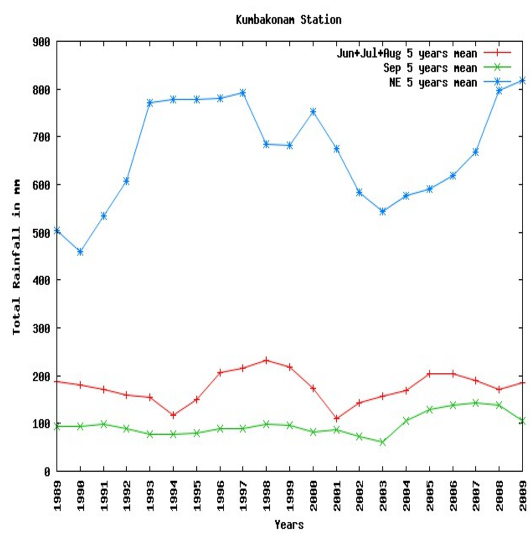

Figure 3(b): Five year mean rainfall of kumbakonam station.

Journal of Theoretical and Applied Information Technology
31st January 2012. Vol. 35 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

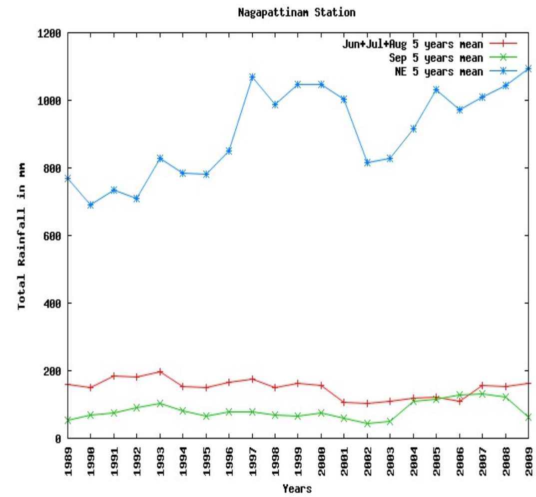

Figure 3(c): F*ive year mean rainfall of nagapattinam station.*

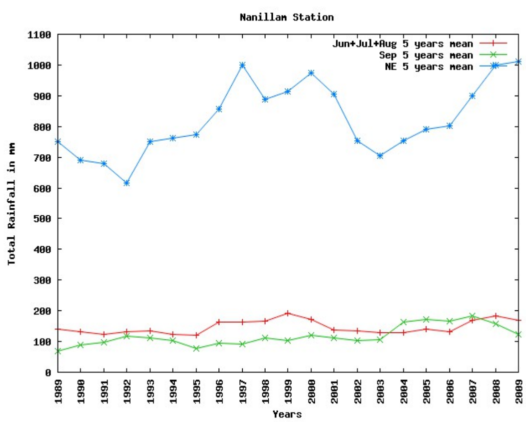

Figure 3(d): *Five year mean rainfall of nannillam station.*

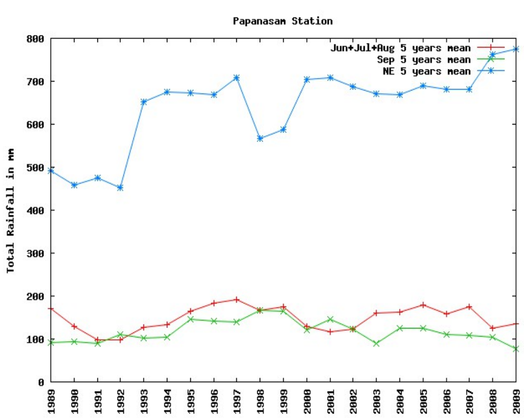

Figure 3(e): *five year mean rainfall of papanasam station.*

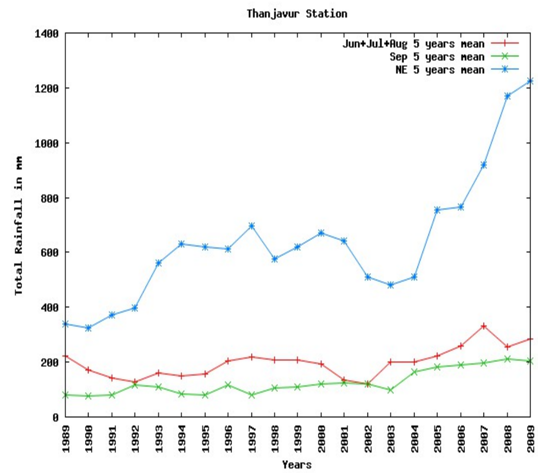

Figure 3(f): F*ive year mean rainfall of thanjavur station.*

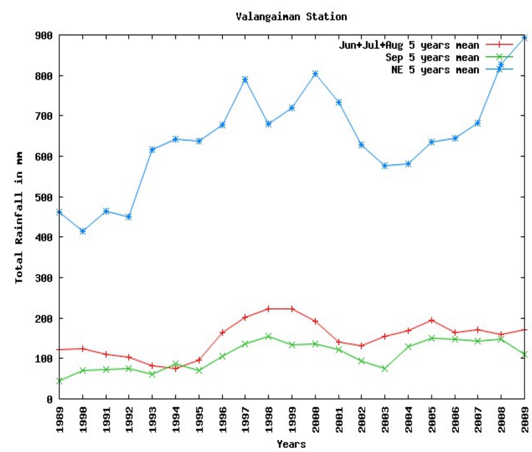

Figure 3(g): *Five year mean rainfall of valangaiman station.*

After the analysis, the following rules are mined out from the above graphs:

RF (X, " $SW_C < SW_P$ ") => NE_C (X, " > NE_P ") (3.1)

RF (X , " OCT < 200mm ") => NOV (X , " OCT R/F + 50mm ") (3.2)

RF (X , " OCT > 200mm ") => NOV (X ," >400mm ") (3.3)

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The equation 3.1 describes the rule whether the total Rainfall (RF) of current year of SouthWest (SW_C) monsoon is less than the previous year of SouthWest (SW_P) monsoon then the Rainfall(RF) of current year of North East (NE_C) monsoon is greater than the previous year of the total Rainfall(RF) of North East (NE_P) monsoon. The equation 3.2 describes the rule whether the total Rainfall(RF) of October month is less than the 200mm then the total Rainfall of the successive November month is most probably the actual October month Rainfall with an excess of 50mm. The equation 3.3 describes the rule whether the total Rainfall of October month is greater than the 200mm then the total Rainfall of the successive November month will have the precipitation of more than 400mm.

4. CONCLUSION

The study reveals that in the Cauvery delta basin, the rainfall pattern of many stations obey the association rule except Nagapattinam. Nagapattinam being coastal in nature may perhaps be the reason. The rule suggests a correlation between the Southwest and Northeast monsoon rains over this area when considered in relation to previous years performance. It is also possible to assess the November month rain from the rains of October using the same rule. Frequent pattern mining leads to the discovery of associations and correlations among rainfall occurrence in large climate data set. This methodology can help to forecast rainfall during the months of Northeast monsoon for the Cauvery delta region of Thanjavur district of South India.

ACKNOWLEDGMENT

The authors thank the office of Superintending Engineer, Lower Cauvery Division, Public Works Department, Thanjavur, Tamilnadu, India for supplying the rainfall data.

REFERENCES

[1] Agrawal, R., Imielinski, T., Swami, A., 1993, "Mining Associations between Sets of Items in Massive Databases", *Proc. of the ACM-SIGMOD 1993 Int. Conference on Management of Data, Washington D.C.*.

[2] Agrawal, R., Srikant, R., 1994, "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.

[3] Agrawal, R., Mannila, H., Srikant R., Toivonen, H., and Verkamo, A. I., 1995, "Fast Discovery of Association Rules", *Advances in Knowledge Discovery and Data Mining*, Chapter 12, AAAI/MIT Press.

[4] Bayardo Jr., R. J., and Agrawal R., 1999, "Mining the Most Interesting Rules" In *Proc. of the 5th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining.*

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

[6] Chew, F H S, Piechota T C, Dracup J.A, and Mcmohan T.A., "Elnino/ Southern Oscillation and Rainfall Variations", Journal of Hydrology (1998), p 138.

[7] Dhar O.N., Rakheja R.R., 1983, "Foreshadowing Northeast monsoon rainfall over Tamil Nadu, India", Mon.Wea.Rev., 111, p 109-112.

[8] Duraiswamy I.V., 1946, Scientific note no:98, India Meteorological Department, p147.

[9] Jiawei Han and Micheline Kamber, 2008, "Data Mining – Concepts and Techniques", Morgan Kaufmann Publications.

[10] Kripalani R.H, Pankaj Kumar., 2004, International Journal of Climatology, pp 1267-1282.

[11] K. M. Lau, 1992, "East Asian Summer Monsoon Rainfall Variability and Climate teleconnections", Journal of Meteorological Society of Japan, 70, p 211.

[12] Meganathan, S., Sivaramakrishnan, T.R. and Chandrasekhara Rao, K. , 2009, "OLAP operations on the multidimensional climate data model: A theoretical approach", Acta Ciencia Indica, XXXVM, 4, p 1233.

[13] Mohanty V.C., "Forecast of precipitation over Delhi during SW Monsoon", Mausam, 45 (1994), p 87.

[14] Nacheketa Acharya, Kar S.C, Kulkarni M.A, Mohanty U.C, Sahoo L.N., 2011, Journal of Earth System Science, 120, No. 5, p 795.

[15] Nayagam Lorna R., 2009, "Variability and teleconnectivity of northeast monsoon

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

rainfall over India", Journal of Global and Planetary Change, 69, pp 225-231.

[16] Pankaj Kumar, Rupa Kumar K, Rajeevan M, Sahai A.K., 2007, Climate Dynamics, 28, p 649.

[17] Raman Kzyszie, 2001, "The case for probabilistic forecast in hydrology.", Journal of Hydrology 249, p 2.

[18] Rao K.V,, 1963, "A Study of the Indian northeast monsoon season", Indian Journal of Meteorololgy, Hydrology and Geophysics, 14, pp 143-155.

[19] Rao Y.P., 1976, "Southwest Monsoon", Met. Monograph, India Meteorological Department, pp 354-364.

[20] Sarawagi, S., Thomas, S., Agrawal R., 2000, "Integrating Association Rule Mining with Databases: Alternatives and Implications", *Data Mining and Knowledge Discovery Journal*, **4**.

[21] Seetharam. K , "Arima Model of Rainfall prediction over Gangtok", Mausam, 60 (2009), p 361.

[22] Shukla, J. and Mooley, D.A. , "Empirical Prediction of Summer monsoon rainfall in India", Monthly Weather Rev. (1987), p 695.

[23] Shukla, J. and Pavolino, D.A., "Southern Oscillation and long range forecast of summer monsoon rainfall in India", Monthly Weather Rev., 111 (1983), p 1830.

[24] Sivaramakrishnan, T.R, 1989, "Annual Rainfall over Tamil Nadu", Hydrology Journal, IAH, p 20.

[25] Sivaramakrishnan, T.R, et al, 1983, "A study of rainfall over Madras", Vayumandal, p 69.

[26] Sivararamakrishnan, T.R, and Sridharan, S., 1987, "Occurrence of heavy rain episodes over Madras, Proceedings of National symposium on Hydrology, NIH, Roorkee, P VI 54.

[27] Sivaramakrishnan, T.R, Meganathan, S., and Sibi, P., "An analysis of northeast monsoon rainfall for the Cauvery delta of Tamil Nadu", Proceedings of INEMREC – 2011, pp 105-107.

[28] Sivaramakrishnan,T.R, Meganathan,S., August 2011, " Point Rainfall Estimation using Association Rule Mining – A Case Study", proceedings of the National Conference on Environmental Science & Technologies for Sustainable Development, Chennai, pp 81-85.

[29] Sivaramakrishnan.T.R, Meganathan.S, December 2011, "Association Rule Mining and Classifier Approach for Quantitative Spot Rainfall Prediction", Journal of Theoretical and Applied Information Technology.

[30] Sridharan S, Muthuchami A., 1990, "Northeast monsoon rainfall in relation to El NLino, QBO and Atlantic hurricane frequency", Vayumandal, 20,

3-4, 1.08-1.11.

[31] Zubair L, Ropelewski, 2006, J.Climate, 19, pp 1567-1575.