EXPLOITING CLUSTERING TECHNIQUE FOR ALMANAC BASED RAINFALL PREDICTION

K. S. Mohanasundaram

Research Scholar, Department of Computer Science,

Chikkanna Government Arts College, Tiruppur, India.

Assistant Professor in Computer Science and Application, SRM Institute of Science and Technology, Chennai, India.

Dr. G. M. Nasira

Assistant Professor & Head, Department of Computer Applications,

Chikkanna Government Arts College, Tiruppur, India.

Abstract—Data mining is the process of discovering patterns in large data sets involving methods at the intersection of machine learning, statistics and database systems. Various types of methods are available in data mining. Clustering is an unsupervised learning of data mining. Data is divided into similar groups; each similar group is called a Cluster. Most of people in Tamil Nadu are still following the activities based on astrological facts of Panchang (Almanac). Almanac has a mathematical base to predicting the meteorological occurrences. During the study, predicting the rainfall by one of the traditional Almanac or Panchang is studied in attention for one cycle of 81 Tamil year's corresponding to the Gregorian Year from 1936 to 2017 and actual annual rainfall data for Coimbatore is taken. In this work, we adopt k-mean clustering Algorithm, which is the most fashionable algorithms mainly used for cluster mining. In this work, we are trying to research the existing clustering algorithm is sufficient or not for finding the similar patterns in the almanac rainfall predictions.

Keywords: Clustering, Data Mining, K Means, Almanac or Panchang.

1. INTRODUCTION

Rainfall is important for planning the activities of builders, agriculturists, water supply engineers and all activity plans in the nature. India is an agricultural country and its economy is largely based upon crop productivity. Thus rainfall prediction becomes a significant factor in agricultural based countries like India. Rainfall Prediction is one of the most challenging tasks. Though already many algorithms have being proposed but still accurate prediction of rainfall is very difficult. In an agricultural country like India, the success or failure of the crops and water scarcity in any year is always viewed with greatest concern. Astronomy is an area where Data Mining has a immense role. Numerous techniques of Data mining have been used to solve tasks in Astrology. There has been increasing research interest in use of data mining techniques to scrutinize in the Astrology area.

Data mining is the search and analysis of large data sets, in order to discover meaningful patterns and rules. The key idea is to find effective ways to combine the computer's power to process data with the human eye's ability to detect patterns [14]. Data mining techniques have been broadly applied almost in all fields to analysis the data for pattern the rules, classification, prediction, decision trees, fuzzy rules and so on. Right now the Meteorology Department is forecasting only short term information about weather but long term forecasting is needed for planning. This can be done by two methods namely traditional forecasting and scientific weather forecasting. Traditional forecasting is based on observations and experience using animals, insects, combinations of plants, meteorological and astronomical indicators, and almanacs or panchang over a period of time. The scientific weather forecasting is based on past records of climate succeed in the area using mathematical models.

2. EXISTING APPROACH

IyerAurobind Venkatkumar [1] listed the advantages and disadvantages of clustering algorithms. K-means algorithm is better when clusters are spherical, it can be identified at beginning. But need to define number of clusters at the beginning. The BIRCH algorithm eliminates noisy data but its functioning is affected if the clusters formed are not of spherical shape. Also in DBSCAN there is no need to define number of clusters at the beginning. The complexity of STING algorithm depends on the number of grid cells and not on the number of points but the quality of the clusters highly depends on the granularity of data at the lower levels. As it is

seen K-means, DBSCAN, BIRCH are used for only numerical data. Whereas STING is used for any type of data. Gurpreet Singh [2], in his research they proposed a new EPBCA algorithm by combined the features of B tree and KMeans clustering algorithms. For this work the used WEKA data mining tool. The comparison results are showed in the tabled format. KM Archana Patel and Prateek Thakral [3], after investigating the consequences of different clustering technique he found that some clustering technique areused for vast amount of data set but some of them do not provide good result when data occur with high variance indensity. Every clustering algorithm does not give best result for all the condition. A. Raharto Condrobimo [4], Using cluster analysis in this study result with the ability to provide information quickly and efficiently for potential novice investors on the distribution map of Liquid 45 shares or blue-chip stocks in Indonesia Stock Exchange. Neelam Chaplot [5] reviews various researches done for prediction in medical, finance and weather forecasting field using artificial intelligence and machine learning. The methods used in these application ranges from Logistic Regression, K-Nearest neighbor method, Case based reasoning, Hybrid models etc. most of the methods reviewed over are supervised learning concepts. They suggested these techniques to apply in astrology. Dhawal Hirani [6] reports a detailed survey on rainfall prediction using different rainfall prediction methods extensively survey lasted 20 years. From the survey it has been found that most of the researchers used artificial neural networks for rainfall prediction and got significant results. They conclude that MLP, BPN, RBFN, SOM & SUM are suitable for predict rainfall forecasting techniques. Shoba G [7] analysis of various algorithms of data mining is used rainfall prediction model. They found difficult to found suitable particular prediction algorithm. They found sometimes when certain algorithms are combined; they performed better and are more effective.

D Angchok [8] predicted rainfall by Tibetan astrological theories with meteorological predictions was accepted. They suggested as very few scientific studies have ever been conducted in ancient Astro-science and almost all of them have reported encouraging and positive outputs, there seems to have enormous scope lying in studying ancient sciences, especially Astro - disciplinary approaches. S Sivaprakasam [9] suggested the traditional methods of forecasting rainfall may be challenged with in accuracies but they can't be ignored altogether. Rahul Shajan [10] trying to study the existing association rule mining algorithm "Apriori" is sufficient or not to find the similar patterns in the horoscope of different individuals. They found that apply some conditions in to the algorithm just before the generation of frequent item set and candidate item set, then it gives more association rules that help to analyse the horoscope of an individual. R. Raja [11] analysis 90 years (1909-1999) historical annual rainfall data of Coimbatore correlation with a particular Tamil year cycle with fourth coming Tamil cycle years. Pankaj S. Kulkarni [12] deals with converting ancient principles related to astrology

into predictions using data mining techniques. Neelam Chaplot [13] taken total 102 records, an half of the records were of persons are doctor and other half records of are not doctor by Profession, They compared various Supervised classification techniques such as Logistic, Naïve Bayes, Simple Cart, Decision Stump, Decision Table and DTNB algorithm. The better results were produced by simple logistic with 12 fold cross validation with an accuracy of 54.902%. Decision Stump algorithm with 14 fold classification gave results with an accuracy of 50%.

3. PROPOSED APPROACH

Experimental research method has been adopted for this study. A record of 81years (1936 -2017) data from almanac has taken for analysis and actual annual rainfall data for Coimbatore is taken. Through the extensive search of literature and discussion with exports the item sets and the number of transactions has been finalized. From an almanac we have considered five influencing planets for rainfall as data item King, Minister, Megathipathi, Megam (Cloud Type), Rainfall value (Marakkal), and Year, Almanac Rainfall value (MM), Actual Rainfall value (MM) for Coimbatore Agricultural University Dataeach of which has sub item sets, which as shown in Table I.

TABLE I. RAINFALL INFLUENCING ATTRIBUTE

S. No.	Attribute	Description	Domain Value
1	Year	Gregorian Year	Numeric
2	King	Ruling Planet of the year	{Sun, Moon, Mars, Mercury, Jupiter, Venus, Saturn}
3	Minister	Minister Planet of the year	{Sun, Moon, Mars, Mercury, Jupiter, Venus, Saturn}
4	Megathipathi	Planet supporting rainfall for the year	{Sun, Moon, Mars, Mercury, Jupiter, Venus, Saturn}
5	Megam(Cloud Type)	Type/Formation of the Cloud	{Aavarta, Samvarta, Pushkara, Drona, Kaala, Neela, Varuna, Vayu, Dhamo}
6	Almanac Rainfall Marakkal	Rainfall of the year as per Almanac	{Kuruni, Pathaku, Mukkuruni, Thooni}
7	Almanac Rainfall	Rainfall of the year as per Almanac in MM	Numeric
8	Actual Rainfall	Rainfall of the year as per Coimbatore AU Data	Numeric

For this study the data mining technique Simple KMeans Clustering algorithm has been in WEKA Tool. The output has been analyzed to predict rainfall which as in the next section.

4. RESULT & DISCUSSION

Result in WEKA

Simple kMeans
===========
Number of iterations: 4
Within cluster sum of squared errors: 292.5149647013334

Initial starting points (random):

Cluster0:1936,Saturn,Moon,Saturn,Kaala, Pathaku,400,1240.4
Cluster1:1996,Mercury,Saturn,Venus,Vaayu,Mukkuruni,600,743.7

Missing values globally replaced with mean/mode

Final cluster centroids:

Attribute	Cluster# Full Data (81.0)	0 (33.0)	1 (48.0)
Year	1976	1975.7576	1976.1667
King	Saturn	Saturn	Moon
Minister	Saturn	Moon	Saturn
Megathipathi	Saturn	Sun	Venus
Megam	Kaala	Kaala	Drona
Rainfall	Marakkal	Mukkuruni	Kuruni	Mukkuruni
Almanac RF (MM)	543.2099	333.3333	687.5
Actual RF (MM)	948.2259	956.197	942.7458

Time taken to build model (full training data): 0.02 seconds

=== Model and evaluation on training set ===

Clustered Instances

0   33 (41%)
1   48 (59%)

We got an output that as shown above. Time taken to build model is 0.02 seconds.

There are two clusters, which show the connection among ruling planet, Megathipathi, Rainfall Value (Markkal), Acutal Rainfall. In these we got the plants which supports the rainfalls are Venus & Moon and the plant which does not support the rainfall is Saturn& Sun.
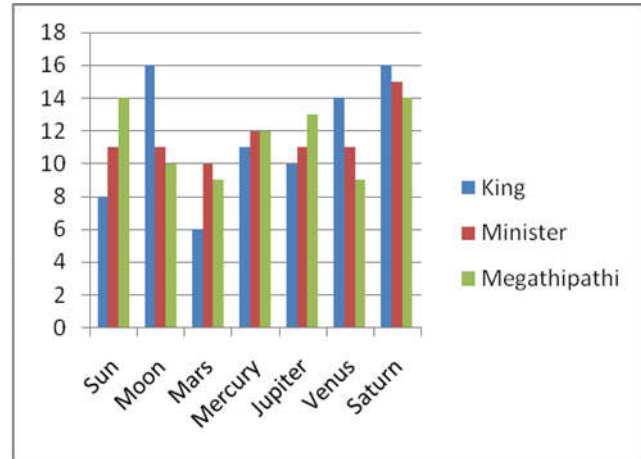

Fig. 1.   Some Selected attributes from Almanac

5. CONCLUSION

In this paper, Rainfall predicting attitudes and data sets are taken for cycle of 81 Tamil year's related to the Gregorian Year from 1936 to 2017 from Almanac or Panchang for Coimbatore. For this data set clustering mining technique Simple KMean Algorithm was applied inWEKA. Clustering which have the minimum support rainfall clustering 41% and maximum support rainfall clustering 59%.So, the existing clustering Simple KMean algorithm is sufficient to find the similar patterns in the almanac rainfall predictions.

REFERENCES

[1]   Iyer Aurobind Venkatkumar, Sanatkumar Jayantibhai Kondhol Shardaben, "Comparative study of Data Mining Clustering algorithms" ,2016 IEEE International Conference on Data Science and Engineering (ICDSE), 978-1- 5090-1281-7/16.

[2]   Gurpreet Singh, JaskaranjitKaur, MD. Yusuf Mulge, "Performance Evaluation of Enhanced Hierarchicaland Partitioning Based Clustering Algorithm(EPBCA) in Data Mining" 2015 International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT),IEEE, 978-1-4673-9223-5/15

[3]   KM Archana Patel and PrateekThakral, "The Best Clustering Algorithms in Data Mining", International Conference on Communication and Signal Processing, April 6-8, 2016, India,IEEE, 978-1-5090-0396-9/16

[4]   A. RahartoCondrobimo, BahtiarSaleh Abbas, AgungTrisetyarso, "Data Mining Technique With ClusterAnaysis Use K-Means Algorithm ForLq45

Index On Indonesia StockExchange", 2018 International Conference on Information and Communications Technology (ICOIACT), IEEE, 978-1-5386-0954-5/18

[5] NeelamChaplot, Praveen Dhyani, O.P. Rishni "A Review on Machine Learning concepts for Prediction based application" International Journal of Computational Science, Engineering & Technology, Volume I, Issue II, March 2013 [ISSN: 2320-4648]

[6] Mr. DhawalHirani, Dr. Nitin Mishra "A survey on rainfall prediction techniques", Internal Journal of Computer Application, Volume 6- No.2, March-April 2016 (2250-1797)

[7] Shoba G, Dr. ShobhaG "Rainfall prediction using data mining techniques : A Survey" International Journal of Engineering And Computer Science, Volume 3 Issue 5 May 2014 Page No. 6206-6211 [ISSN:2319-7242]

[8] D Angchok& V K Dubey "Traditional method of rainfall prediction through Almanacs in Ladakh", Indian Journal of Traditional Knowledge, Vol. 5(1), January 2006, pp. 145-150

[9] S Sivaprakasam& V Kanakasabai "Traditional almanac predicted rainfall – A case study", Indian Journal of Traditional Knowledge, Vol. 8(4), October 2009, pp. 621-625.

[10] Rahul Shajan, Gladston Raj S "Association rule mining based analysis on horoscope data – A perspective study", International Journal of Computer Engineering & Technology, Volume 8, Issue 3, May-June 2017, pp. 76-81, Article ID:IJCET_08_03_008 [ISSN Print:0976-6367 and ISSN Online:0976-6375]

[11] R. Raja, T. N. Balasubramanian and R. Karthikeyan "Almanac study on forecasting annual rainfall", The Madras Agricultural Journal 90(10-12): 596-600 October-December 2003

[12] Mr. Pankaj S. Kulkarni, Dr. S. S. Sane, Prof. N. L. Bhale "Use of Neural Networks in Horoscope prediction", ResearchGate, Conference Paper March 2012

[13] NeelamChaplot, Praveen Dhyani, O. P. Rishi "Astrological prediction for Profession Doctor using Classification techniques of Artificial Intelligence" International Journal of Computer Applications, Volume 122 – No.15, July 2015 (0975-8887)

[14] Arun k Pujari, "Data mining techniques", Hyderabad, India. Universities Press Private Ltd., ISBN 978-81-7371-380-4.