



A SURVEY ON CLASSIFICATION TECHNIQUES USED FOR RAINFALL FORECASTING

KolluruVenkata Nagendra
Research Scholar,
Department of Computer Science
Vikrama Simhapuri University, Nellore
Andhra Pradesh, India

Dr. Maligela Ussenaiah
Assistant Professor,
Department of Computer Science,
Vikrama Simhapuri University, Nellore
Andhra Pradesh, India

Abstract: Data mining is one of the major areas of research. An extensive research is being done on classification, which is one of the functionalities of data mining. A variety of classification techniques such as Decision Tree Induction, Bayesian Classification, Naïve Bayes Classifiers, Artificial Neural Networks, Multi Layer Perceptron, Genetic algorithms, Fuzzy logic and Support Vector Machines have been developed. Many researchers have made comparative analysis of different classification techniques with respect to different applications. In this paper, we discuss different classification techniques used for rainfall forecasting. The main aim of this paper is to generate sound knowledge on various techniques of classification used for rainfall forecasting.

Keywords: Bayesian classification, Decision Tree, SVM, Fuzzy logic and Genetic Algorithm

1. INTRODUCTION

Data mining is a variety of technique to identify suggest of information or decision making knowledge in the data base and extracting these in a way that they can be put to use in areas such as decision support, predictions, forecasting and estimation. The process of discovering meaningful, new correlation patterns and trends by shifting through large amount of data stored in repositories, using patterns recognition techniques as well as statistical and mathematical techniques. Classification identifies data into a predefined groups or classes. Before examining the data the classes are determined because it is supervised learning. Instead of regression Classification technique is used to process large data. The most significant data mining techniques are follow the machine learning process. All applications of machine learning used the classification Techniques. We concentrate some of the techniques in this paper, which are used in rain fall forecasting.

Decision Tree Induction is presented in below section. Bayes classifiers and Naïve Bayes classifiers are discussed in section-3 and section-4 respectively. A detailed analysis of Artificial Neural Network is presented in section-5. Multilayer Perceptron method has been discussed in section-6. A description of Genetic algorithms and fuzzy logic is presented in section-7. Support Vector Machines has been presented in section-8. Table –I shows that categorization of various approaches of classification for Rain Fall Forecasting(RFF). At Last we conclude this work.

2. DECISION TREE CLASSIFICATION

Decision tree is a predictive modeling technique which is used in classification, prediction task and clustering. A decision tree use a technique called Divide and Conquer to split the problem in to number of subsets. The construction of decision trees contains two common things: (a) the growth of the tree to enable it to accurately categorize the training dataset, and (b) the pruning stage, we eliminate

anomalies in the training data to increase accuracy of classification.

Based on adjusted cluster analysis classification called classification by clustering was introduced by Aviad and Roy in 2011. It found similarities between instances using clustering algorithms and also selected target attributes. Then it calculated the target attributes distribution for each cluster. When a threshold for the number of instances stored in a cluster was reached, all the instances in each cluster were classified with respect to the appropriate value of the target attribute.

3. BAYESIAN CLASSIFIER

Bayesian classifier is an example of statistical-based learning scheme. In Bayesian classifier, it is assumed that all attribute have equal importance and value of one attribute does not provide the value of another attribute. In Bayesian rule, classification is done with the help of Bayesian rules.

Bayes theorem works on conditional probability. Conditional probability is the probability that something will happen, given that something else has already occurred. Using the conditional probability, we can calculate the probability of an event using its prior knowledge.

Below is the formula for calculating the conditional probability.

$$P(H/E) = \frac{P(E/H)*P(H)}{P(E)}$$

where

- P(H) is the probability of hypothesis H being true. This is known as the prior probability.
- P(E) is the probability of the evidence (regardless of the hypothesis).
- P(E|H) is the probability of the evidence given that hypothesis is true.
- P(H|E) is the probability of the hypothesis given that the evidence is there.

4. NAÏVE BAYES CLASSIFIERS

Naive Bayes classifier is very simple approach of Bayesian network. Naive Bayes class has no parents. In Naive Bayes each attribute had the class as its sole parent. It build the own model very easily without complex parameters estimation. From this reason only this classification methods used for large data sets. While dealing with the real world problem; this method gives an explicit model for its deployment. It uses very less training data to predict classification parameters. The researchers often use this for predicting rain fall in moderate-rain i.e. weekly.

5. ARTIFICIAL NEURAL NETWORKS

Unlike human brain the Artificial Neural Network (ANN) has heuristic knowledge. The main characteristic of such a computing system is the number of highly interconnected processing elements (neurons) working together to solve specific problems without being programmed with step-by-step instructions. Instead, ANN's are capable of learning on their own or by example through a learning process that involves adjustments to the connections that exist between the neurons. Artificial Neural Networks (ANNs) do not require restrictive assumptions and its parallel processing capability they work well on large size training samples. ANN has detected complex nonlinear relationships between dependent and independent variables and also traditional methods works on linear as well as non linear data. Due to these reasons many researchers often use ANN for rainfall forecasting.

6. MULTILAYER PERCEPTRON

The Multi Layer Perceptron (MLP) or Feed-forward network is a type of artificial neural network that consists of a non linear activation function in hidden layer. MLP network provide nonlinear mapping between input and output vectors. Neural networks have two important functions i.e. pattern classifiers and nonlinear adaptive filters. A general framework of neural network consist of three layer architecture i.e. an input layer that define the input value, one or more hidden layers define the mathematical function and an output layer define final outcome of Each layer consists of a large number of neurons that are interconnected through weights. Each neuron has mathematical function (also known as activation function) that accepts input from previous layer and produced output for next layer. So, in neural networks the prediction is defined by the activation function.

7. GENETIC ALGORITHMS & FUZZY LOGIC

An Efficient Weather Forecasting System using A Neuro fuzzy approach for daily rainfall prediction over the central regions, which classifying rain into 4 classes of no rain, light-rain, moderate-rain, and heavy-rain. Meanwhile in the research Genetic algorithm (GA)-optimized by fuzzy system method is used by categorizing rainfall based on Climatology and Geophysics (BMKG). Hybrid Fuzzy and GA are also vastly used in other sector such as researches of A fuzzy genetic algorithm for driver

scheduling and Hybrid Fuzzy-Genetic Algorithm Approach for Crew Grouping. In assessing this weather forecast, fuzzy system is used to forecast rainfall for next day such that it produce information, for example class of mild rain has degree of membership value equal to 0.65. Fuzzy system is able to do intervension or rationalization for fuzzy or blur data with linguistic variable input such as is able to do intervension or rationalization for fuzzy or blur data with linguistic variable input such as low, medium, and high temprature so the amount of linguistic, type, and function paramenter membership value should be known generally from expert's knowledge. The weather forecast information especially rainfall in Kemayoran area for the next day so this information can be used by society for the sake of desicion making in conducting their activities and works.

8. SUPPORT VECTOR MACHINES

In recent years, Support Vector Machines (SVM) with linear or nonlinear kernels have become one of the most promising learning algorithms for classification as well as for regression which are two fundamental tasks in data mining via the use of kernel mapping, Variants of SVMs have successfully incorporated effective and flexible nonlinear models Kernel-based techniques (like support vector machines, kernel principal component analysis, Bayes point machines, and Gaussian processes) represent a major development in machine learning algorithms. SVM (support vector machines) is a group of supervised learning techniques or methods, which is used to do for classification or regression. SVM (support vector machines) represents an extension to nonlinear models of the generalized portrait algorithm. The basic idea of SVM (support Vector Machines) is to map the original data X into a feature space F with high dimensionality through a non-linear mapping function and construct an optimal hyper-plane in new space. SVM can be applied to both classification and regression. In the case of classification, an optimal hyper-plane is found that separates the data into two classes. Whereas in the case of regression a hyper-plane is to be constructed or developed that lies close or near to as many points as possible.

9. RAINFALL FORECASTING SURVEY AND ANALYSIS

In Rain fall forecasting, lots of research had been done by the so many researchers. The rainfall forecasting ways are classified into three types depending on weekly, monthly and yearly. They called as Short term, medium term and long term. The forecasting techniques of these three terms have been discussed in the different papers [1],[2],[3],[4],[5],[6],[7],[8],[9],[10],[11] and [12]. The monthly rainfall of Chennai was predicted by R.Selvaraj and G.Geetha [13] by using Artificial Neural Network. They suggested that more parameters are needed for accurate prediction of high rainfall. In Kerala region the rain fall is predicted by K.Joseph and N.Philip [14], they explained with 12 months. In that they use 12 input nodes,7 hidden nodes and one output node. They also suggest the ANN gives high accuracy with 48 nodes. In [15], the researcher found that the average rainfall in Udipi district of Karnataka state. They applied for Back Propagation Algorithm (BPA) was better for prediction. In

[16], authors used ANN to predict the rainfall data of Thailand. The technique used here is Feed Forward Neural Network (FFNN). This research gave that 96.9% accuracy on test data.

Focused Time Delay Neural Networks (FTDNN) technique was used by the authors [17], for rain fall prediction on yearly, Quarterly and monthly basis. Most accurate results are observed on yearly basis training data. Sharma and Singh [18] observed 7 days i.e. weekly rainfall in Pantnagar. In that they used maximum temperature, humidity at 7.00 am and 2.00 PM and minimum temperature as the parameters. They achieved minimum prediction error and minimum mean difference compared to multiple linear regression models.

The following Table-I shows that the different kinds of rainfall forecasting. These different ways is based on types of Classification, region, testing period and Rainfall predicting variable.

TABLE I. RAINFALL FORECASTING ANALYSIS WITH DIFFERENT TECHNIQUES

Researcher	Area (Global/Local)	Technique Used	Rainfall predicting variable	Daily-Monthly-Yearly
Mohini P Darji Vipul K dabhi	Global (India)	Artificial Neural Network	Temperature, Humidity, And Soil	Monthly
S Renugadevi Pranay Agarwal	Global (India)	Artificial Neural Network	Cloud influence, Meteorological parameters	Daily
Ashutosh Sharma Manish Kumar Goyal	Global (India)	Bayesian Network Model Belief Network	Temperature, Humidity, Wind speed, Cloud cover	Monthly
T Matsumura Y Kito	Global (Japan)	Artificial Neural Network	River flow rate, run off ratio	Monthly
Lingzhi Wang Jaian Sheng Wu	Global (China)	Wavelet Support Vector Machines Regression	Temperature, Humidity, Wind speed	Daily
Jaian Sheng Wu LiuZhou	Global (China)	Hybrid semi Parametric Regression strategy for Artificial Neural Network	---	daily
Septian Nurcahyo Adiwijaya	Global (Indonesia)	Hybrid Genetic Algorithms & PCFNN	Wind speed, air Pressure, Humidity, length of Sun radiation	Monthly
Fhira Nhita Adiwijaya	Global (Indonesia)	Regression Smoothing & Fuzzy grammatical evolution	High humidity, wind speed	Monthly
Harshani Asanga Ratnaweera	Global (Srilanka)	Artificial Neural Network	Temperature, Humidity	Yearly
Fhira Nhita Adiwijaya	Global (Indonesia)	Fuzzy system based on Genetic Algorithm	Meteorological and Climatology parameters	Yearly
Jie Li Jiansheng Wu	Global (China)	A novel radial basic function NN based on kernel Principal component Analysis	----	Monthly
F Mekanik MA Imteaz	Global (Australia)	Linear multiple regression and non linear ANN	Temperature, High humidity, wind speed	Monthly
Adil MohammadKhadir AliAdlan	Global	Multi Layer Perceptron	Temperature, Humidity	Daily
Loris Foresti, Mikhail Kanevski	Global	Support Vector Machines	Image color, Pixcels	Monthly
Abhishek sachin	Global (India)	Artificial Neural Network, GPS	Rain fall, snow fall, heat around the earth, surface atmosphere	Daily
S.Chattopadhyay and M.Chattopadhyay	Global (India)	Multilayer perceptron	Minimum and maximum Temperature	Monthly

The authors [19], compared the working results of ANN with ARIMA technique in Hyderabad, INDIA region for rainfall predicting. Past four rain falls they used and conclude that, which out-performs the ARIMA model for forecasting the rainfall. The researcher N.Philip and K.Josheph in [20] predicted yearly rainfall in Kerala region, used ABF neural network for rainfall forecasting. They concluded that ABFNN is better tool than the Fourier analysis. R.Deshpande [21] stated that for larger step ahead prediction the performance of Multilayer network is better than all other networks. P.Guhathakurta [22] predicted rainfall for 2, 5, 10, 15 years and designed model from the point of feasibility.

M.Sharma and J.Singh	Local (India)	Multiple regression model	Minimum and maximum Temperature, relative Humidity	Monthly
R.Deshpande	Local (India)	Multi Layer Perceptron	Rainfall	Daily
V.Dabhi and S.Chaudhary	Local (India)	Wavelet postfix-GP model, Wavelet ANN	Maximum Temperature, humidity and Minimum Temperature.	Daily

10. CONCLUSION

Different Classification techniques used by different researchers are presented in this paper. ANN is applied for RFF on different parameters are discussed. This survey tells that Multi Layer Perceptron method, Naïve Bayes classifiers and SVM are suitable to predict rainfall than other forecasting techniques such as statistical and numerical techniques. Neural Network gave very less performance on monthly and daily data. In the survey we identify that Feed Forward Neural Network gives good performs on monthly rainfall data. Support Vector Machines gives better performance on yearly rainfall data. In weekly rain fall, the Naïve Bayesian gives better performance for forecasting. Moreover, it is observed that the forecasting accuracy of SVM can be improved by considering other meteorological parameters like evaporation, mean temperature, humidity, and soil temperature as inputs.

11. REFERENCES

- [1] P. Goswami and Srividya, "A novel Neural Network design for long range prediction of rainfall pattern," *Current Sci.(Bangalore)*, vol. 70, no. 6, pp. 447-457, 1996.
- [2] S. K. Nanda, D. P. Tripathy, S. K. Nayak, and S. Mohapatra, "Prediction of rainfall in India using Artificial Neural Network (ANN) models," *Int. J. of Intell. Syst. and Applicat.*, vol. 5, no. 12, pp. 1-22, 2013.
- [3] G. Shrivastava, S. Karmakar, and M. K. Kowar, "BPN model for longrange forecast of monsoon rainfall over a very small geographical region and its verification for 2012," *Geofizika*, vol. 30, no. 2, pp. 155-172, 2013.
- [4] A. K. Sahai, M. K. Soman, and V. Satyan, "All India summer monsoon rainfall prediction using an Artificial Neural Network," *Climate dynamics*, vol. 16, no. 4, pp. 291-302, 2000.
- [5] D. R. Nayak, A. Mahapatra, and P. Mishra, "A Survey on rainfall prediction using Artificial Neural Network," *Int. J. of Comput. Applicat.*, vol. 72, no. 16, pp. 32-40, 2013.
- [6] B. K. Rani and A. Govardhan, "Rainfall prediction using Data Mining techniques-A Survey," *Comput. Sci. and Inform. Technology*, pp. 23-30, 2013.
- [7] Shoba G and Shobha G., "Rainfall prediction using Data Mining techniques: A Survey," *Int. J. of Eng. and Comput. Sci.*, vol. 3, no. 5, pp. 6206-6211, 2014.
- [8] R. S. Sangari and M. Balamurugan, "A Survey on rainfall prediction using Data Mining," *Int. J. of Comput. Sci. and Mobile Applicat.*, vol. 2, no. 2, pp. 84-88, 2014.
- [9] K. C. Luk, J. E. Ball, and A. Sharma, "An application of Artificial Neural Networks for rainfall forecasting," *Mathematical and Comput. modelling*, vol. 33, no. 6, pp. 683-693, 2001.
- [10] J. Abbot and J. Marohasy, "Application of Artificial Neural Networks to rainfall forecasting in Queensland, Australia," *Advances in Atmospheric Sci.*, vol. 29, no. 4, pp. 717-730, 2012.
- [11] V. K. Dabhi and S. Chaudhary, "Hybrid Wavelet-Postfix-GP model for rainfall prediction of Anand region of India," *Advances in Artificial Intell.*, pp. 1-11, 2014.
- [12] G. Zhang, B. E. Patuwo, and M. Y. Hu, "Forecasting with Artificial Neural Networks:: The state of the art," *Int J. of forecasting*, vol 14, no. 1, pp. 35-62, 1998.
- [13] G. Geetha and R. S. Selvaraj, "Prediction of monthly rainfall in Chennai using Back Propagation Neural Network model," *Int. J. of Eng. Sci. and Technology*, vol. 3, no. 1, pp. 211-213, 2011.
- [14] N. S. Philip and K. B. Joseph, "On the predictability of rainfall in Kerala-An application of ABF neural network," *Computational Science- ICCS, Springer Berlin Heidelberg*, pp. 1-12, 2001.
- [15] A. Kumar, A. Kumar, R. Ranjan, and S. Kumar, "A rainfall prediction model using artificial neural network," *Control and Syst. Graduate Research Colloq. (ICSGRC)*, pp. 82-87, 2012.
- [16] N. Chantasut, C. Charoenjit, and C. Tanprasert, "Predictive mining of rainfall predictions using artificial neural networks for Chao Phraya River," *4th Int Conf. of the Asian Federation of Inform. Technology in Agriculture and the 2nd World Congr. on Comput. in Agriculture and Natural Resources, Bangkok, Thailand*, pp. 117-122, 2004.
- [17] K. K. Htike and O. O. Khalifa, "Rainfall forecasting models using Focused Time-Delay Neural Networks," *Comput. and Commun. Eng. (ICCCE), Int. Conf. on IEEE*, 2010.
- [18] M. A. Sharma and J. B. Singh, "Comparative Study of rainfall forecasting models," *New York Sci. J.*, pp. 115-120, 2011.
- [19] V. K. Somvanshi, O. P. Pandey, P. K. Agrawal, N.V.Kalanker1, M.Ravi Prakash, and Ramesh Chand, "Modeling and prediction of rainfall using Artificial Neural Network and ARIMA techniques," *J. Ind. Geophys. Union*, vol. 10, no. 2, pp. 141-151, 2006.
- [20] N. S. Philip and K. B. Joseph, "On the predictability of rainfall in Kerala-An application of ABF neural network," *Computational Science- ICCS, Springer Berlin Heidelberg*, pp. 1-12, 2001.
- [21] R. R. Deshpande, "On the rainfall time series prediction using Multilayer Perceptron Artificial Neural Network," *Int. J. of Emerging Technology and Advanced Eng.*, vol. 2, no. 1, pp. 148-153, 2012.
- [22] P.Guhathakurta, "Long lead monsoon rainfall prediction for meteorological sub-divisions of India using deterministic Artificial Neural Network model," *Meteorology and Atmospheric Physics* 101, pp. 93-108, 2008.