Review on Machine Learning Algorithms for Weather Forecasting Issues

A Zakiuddin Ahmed
Research Scholar, Department of Computer Science
Jamal Mohamed College(Autonomous),Tiruchirapalli
azaki2020@gmail.com

Dr. T. Abdul Razak
Department of Computer Science
Jamal Mohamed College(Autonomous), Tiruchirappali
abdul1964@gmail.com

***Abstract*—** Machine leaning is a ground of recent research that officially focuses on the theory, performance, and properties of learning systems and algorithms. It is particularly cross disciplinary field building upon ideas from many different kinds of fields such as artificial intelligence, optimization theory, information theory, statistics, cognitive science, optimal control, and many other disciplines of science, engineering, and mathematics. Since implementation in a wide range of applications, machine learning has covered almost every scientific domain, which has brought great impact on the science and society. Machine learning techniques has been used on a variety of problems, including recommendation engines, recognition systems, informatics and data mining, and autonomous control systems. This research paper compared different machine algorithms for classification. Classification is used when the desired output is a discrete label.

Keywords- *Machine Learning, KNN, ANN, Naive Bayes, Classification*

*****

I. Introduction

Machine learning techniques have been widely adopted in a number of massive and complex data-intensive fields such as medicine, astronomy, biology, and so on, for these techniques provide possible solutions to mine the information hidden in the data. Machine learning is used to teach machines how to handle the data more efficiently. Sometimes after viewing the data, we cannot interpret the pattern or extract information from the data. However, as the time for big data is coming, the collection of data sets is so large and complex that it is difficult to deal with using traditional learning methods since the established process of learning from conventional datasets was not designed to and will not work well with high volumes of data. For instance, most traditional machine learning algorithms are designed for data that would be completely loaded into memory, which does not hold any more in the context of big data.

A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E". Learning is used when a pattern exists in the given data which we can't pin down manually. It is because of the Machine Learning that the present century has been witness to the landmark discoveries like Autonomous Helicopter, Handwriting Recognition, Natural Language Processing (NLP), Computer Vision, Speech Recognition, Recommendation Systems, Decision Support Systems (DSS), email filtering, detection of network intruders or malicious insiders working towards a data breach, optical character recognition etc. Machine Learning is at the root of the tree and has lots of branches and sub branches.

II. CLASSIFICATION OF MACHINE LEARNING TECHNIQUES

Machine learning algorithms can be broadly classified into the following four categories:

1. Supervised learning
2. Semi-supervised Learning
3. Unsupervised Learning
4. Reinforcement Learning

Supervised learning

Supervised learning is very common in classification problems, where the aim is to classify the given input data or image into predefined output classes. For example: Handwritten Digit and character recognition, classification of animals or objects depending upon the input image, medical image classification for diagnosis of various diseases, etc.. In supervised learning, there is some supervision while the algorithm is being trained. Here, labeled data is used for training. Different types of supervised learning approaches are as follows: *ƒ*

- Neural Networks
- Naive Bayes Classifier
- Decision Trees
- Linear Regression

Unsupervised learning

In unsupervised learning, no element of supervision is involved. The computer is trained with unlabeled data.

Unsupervised learning plays a vital role in those cases where the human expert doesn't know what to look for in the data. The common types of supervised learning approaches are: Association Rules, K-means clustering,etc.

Semi-supervised Learning

Semi-supervised Learning lies between the two approaches mentioned above. In this approach a combination of labeled and unlabelled data is used for training.

Reinforcement Learning

Reinforcement Learning involves the mechanism of reward and punishment for the process of learning. In this type of learning, the objective is to maximize the reward and minimize the punishment.

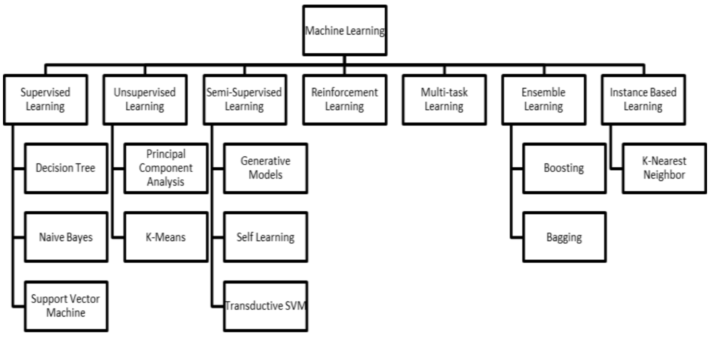

Figure 1: Machine learning Algorithms

III. RELATED WORK

In this paper we reviewed various articles which is related to machine learning classification algorithm for predict the dataset values for specific problem. In the machine learning model, a classifier was given training with already classified examples and it learns the rules for classification during this training phase.

Mark Holmstrom, Dylan Liu, Christopher Vo (2016) concluded that both linear and functional regression did not perform as well as professional weather forecasting methods but in the longer run differences in their performances decreased, suggesting that over a longer period of time, Machine learning can indeed outperform professional and traditional methods. Linear regression is a low bias and high variance algorithm and hence its accuracy can be improved by collecting further data.

PiyushKapoor and Sarabjeet Singh Bedi (2013) concluded that if we perform comparison of weather condition variation by sliding window algorithm, the results are highly accurate except for the months of seasonal change. The results can be altered by changing the size of the window. Accuracy of the unpredictable months can be increased by increasing the window size to one month.

Divya Chauhan and Jawahar Thakur (2013) made a comparison in their paper, which shows that the algorithms such as k-mean clustering and decision trees are well suited for mining data to predict future weather conditions. If we increase the size of the training set, the accuracy at first increases but then it slowly decreases after a particular period of time, depending on the size of the dataset.

Qing Yi Feng1 ,RuggeroVasile, Marc Segond , AviGozolchiani , Yang Wang , Markus Abel, ShilomoHavlin , Armin Bunde , and Henk A. Dijkstra1(2016) have made a machine-learning toolbox which is based on climate data gathered from analysis and reconstruction of complex networks. It can alsohandle data containing multiple variables from these networks. The development of predictor models in the toolbox is dynamic and data-driven.

Sanyam Gupta, Indumathy, GovindSinghal (2016)suggested and proposed an efficient and accurate weather prediction and forecasting model using linear regression concepts and normal equation model. All these concepts are a part of machine learning. The normalequation is a very efficient weather prediction model and using the entities temperature, humidity and dew-point, it can be used to make reliable weather predictions. This model also facilitates decision making in day to day life. It can yield better results when applied to cleaner and larger datasets.

Muthulakshmi A, ME (SE), Dr.S.Baghavathi Priya(2015) in their work proposed a methodology that aims at providing an efficient and accurate weatherforecasting models to predict and monitor the weather datasets to predict rainfall. In the past, the parameters of weather were recorded only for the present time. But in the future, work will be done to make a working model of selection that can be used for classifying the framework for continuous monitoring of the climatic attributes.

Aditya Grover, AshishKapoor and Eric Horvitz in their work made a weather prediction model that predicts by considering the joint influence of key weather variables. They also made a kernel and showed that interpolation of space can be done by using GPS with such a kernel, taking into account various weather phenomena like turbulence. They also performed temporal analysis within a learner based on gradient tree and augmented the system using deep neural network.

IV. OVERVIEW OF MACHINE LEARNING ALGORITHMS

In the following table-1, we have compared different machine learning algorithms and listed below:

ISSN: 2454-4248
94 – 98

Table 3: Interpretation of various machine learning classification algorithms

Algorithms	Advantages	Disadvantages
The *k-Means* method	• Relatively efficient • Can process large data sets.	Algorithms
k-Nearest Neighbor (*k*-NN) classifier	• Nonparametric • Zero cost in the learning process • Classifying any data whenever finding similarity measures of any given instances • Intuitive approach • Robust to outliers on the predictors	• Expensive computation for a large dataset • Hard to interpret the result • The performance relies on the number of dimensions • Lack of explicit model training • Susceptible to correlated inputs and irrelevant features • Very difficult in handling data of mixed types.
Support vector machine (SVM)	• Can utilize predictive power of linear combinations of inputs • Good prediction in a variety of situations • Low generalization error • Easy to interpret results	• Weak in natural handling of mixed data types and computational scalability • Very black box • Sensitive to tuning parameters and kernel choice • Training an SVM on a large data set can be slow • Testing data should be near the training data
Decision Trees	• Some tolerance to correlated inputs. • A single tree is highly interpretable, • Can handle missing values. • Able to handle both numerical and categorical data. • Performs well with large datasets.	• Cannot work on (linear) combinations of features. • Relatively less predictive in many situations. • Practical decision-tree learning algorithms cannot guarantee to return the globally-optimal decision tree. • Decision-tree can lead to over fitting.
Logistic regression	• Provides model logistic probability • Easy to interpret • Provides confidence interval • Quickly update the classification model to incorporate new data	• Does not handle the missing value of continuous variables • Sensitive to extreme values of continuous variables
Naïve Bayes	• Suitable for relative small training set • Can easily obtain the probability for a prediction • Relatively simple and straightforward to use • Can deal with some noisy and missing data • Can handles multiple classes	• Prone to bias when increasing the number of training sets • Assumes all features are independent and equally important, which is unlikely in real-world cases. • Sensitive to how the input data is prepared.
Neural networks	• Good prediction generally • Some tolerance to correlated inputs • Incorporating the predictive power of different combinations of inputs	• Not robust to outliers • Susceptible to irrelevant features • Difficult in dealing with big data with complex model

V. SIGNIFICANCE OF WEATHER FORECASTING

- Weather forecasting can help agricultural activities in the following ways:
- Planning for necessary inputs during the season.
- Timely land preparation to take advantage of earliest rain for timely sowing.
- Selection of crops and cultivars.
- Efficient use of fertilizers.
- Predicting pests and diseases incidence for timely action.
- Timing of weeds, pests and disease control.
- Planning for mitigation adverse effects of weather hazards.
- Adjustments in crop harvest timing to reduce the losses at harvest.

VI. METHODS OF WEATHER FORECASTING

The nature of modern weather forecasting is not only highly complex but also highly quantitative. The various methods used in forecasting the weather are as follows:

- Synoptic weather forecasting,
- Numerical methods, and
- Statistical methods.

Synoptic weather forecasting

The first method is the traditional approach in weather prediction. This primary method continued to be in use until the late 1950s. Synoptic" means that the observation of different weather elements refers to a specific time of observation. Thus, a weather map that depicts atmospheric conditions at a given time is a synoptic chart to a meteorologist. In order to have an average view of the changing pattern of weather, a meteorological centre prepares a series of synoptic charts every day. Such synoptic charts form the very basis of weather forecasts.

As stated earlier, the task of preparing synoptic charts on a regular basis involves huge collection and analysis of observational data obtained from thousands of weather stations. From the careful study of weather charts over many years, certain empirical rules were formulated. These rules helped the forecaster in estimating the rate and direction of the movement of weather systems.

Synoptic methods involved detailed analysis of current weather reports from a large area. The current weather patterns are related with the past analogous situations and forecasts are prepared on the assumption that a current weather situation will behave on the lines of the past analogous situations.

Numerical Weather Prediction (NWP)

Uses the power of computers to make a forecast. Complex computer programs, also known as forecast models, run on supercomputers and provide predictions on many atmospheric variables such as temperature, pressure, wind, and rainfall. A forecaster examines how the features predicted by the computer will interact to produce the day's weather. The NWP method is flawed in that the equations used by the models to simulate the atmosphere are not precise. If the initial state is not completely known, the computer's prediction of how that initial state will evolve will not be entirely accurate.

Statistical methods

Statistical methods are used along with the numerical weather prediction. This method often supplements the numerical method. Statistical methods use the past records of weather data on the assumption that future will be a repetition of the past weather. The main purpose of studying the past weather data is to find out those aspects of the weather that are good indicators of the future events. After establishing these relationships, correct data can be safely used to predict the future conditions. Only overall weather can be predicted in this way. It is particularly of use in projecting only one aspect of the weather at a time. At macro level, weather forecasting is usually done using the data gathered by remote sensing satellites. Weather parameters like maximum temperature, minimum temperature, extent of rainfall, cloud conditions, wind streams and their directions, are projected using images taken by these meteorological satellites to assess future trends.

The satellite-based systems are inherently costlier and require complete support system. Moreover, such systems are capable of providing only such information, which is usually generalized over a larger geographical area. The variables defining weather conditions like temperature (maximum or minimum), relative humidity, rainfall etc., vary continuously with time, forming time series of each parameter and can be used to develop a forecasting model either statistically or using some other means like artificial neural networks.

Very short range weather forecasting

Up to 12 hours description of weather variables. A relatively complete set of variables can be produced (air temperature and relative humidity, wind speed and direction, solar radiation, precipitation amount and type, cloud). Prerequisite is the availability of an efficient broadcasting

systems (e.g. frost information must be broadcasted to farmers that can activate irrigation facilities or fires or other systems of protection). Accuracy is very high and potential usefulness is moderate.

- Insect disease effects.
- Livestock protection from cold and heat.

Medium range weather forecasting

Medium range weather forecasts are for periods of 3 to 10 days. A relatively complete set of variables can be produced (air temperature and relative humidity, wind speed and direction, solar radiation, precipitation amount and type, cloud). In MRF the attention is centred on synoptic features of different meteorological fields.

MRF can be broadcasted by a wide set of media (newspapers, radio, TV, web etc.) and can represent a fundamental information for farmers. Accuracy is high or moderate until 5 days; lower after and potential usefulness is very high.

Short range weather forecasting

Short range weather forecasts are for a period of 12 hours to 72 hours. These daily forecasts are useful to irrigation engineers and farmers. A relatively complete set of variables can be produced (air temperature and relative humidity, wind speed and direction, solar radiation, precipitation amount and type, cloud). In SRF the attention is centered on meso scale features of different

Agricultural applications of short range weather forecasting

- Establish depth of sowing for optimal seedling emergence.
- Decide whether to sow or not.
- Plan irrigation based on expected rainfall.
- Ensure maximum efficiency of spraying.
- Decide to harvest or not to harvest.
- Management of labour and equipment.
- Plan for animal feed requirement.

Agricultural applications of short range weather forecasting

- Timing of field operations.
- Soil workability.
- Drying rate of soil.
- Irrigation scheduling.
- Spray applications

Extended range weather forecasting

- Extended range weather forecasts are for
- Periods of 10 to 30 days. Forecast is usually
- Restricted to Temperature and precipitation.

Table.1 Accuracy and usefulness of weather forecasting for agriculture

Table.1 Accuracy and usefulness of weather forecasting for agriculture

S. No.	Forecast type	Accuracy	Potential usefulness
1.	Now-casting (NC)	Very high	Low
2.	Very Short Range Forecast (VSRF)	Very high	Moderate
3.	Short Range Forecast (SRF)	High	High
4.	Medium Range Forecast (MRF)	High	Very high
5.	Extended Range Forecast (ERF)	Moderate	Poor
6.	Long Range Forecast (LRF)	Very low	Poor

Table.2 Terms used in rainfall forecasting

Descriptive Term used	Rainfall amount in mm
No Rain	0.0
Very light Rain	0.1- 2.4
Light Rain	2.5 – 7.5
Moderate Rain	7.6 – 35.5
Rather Heavy	35.6 – 64.4
Heavy Rain	64.5 – 124.4
Very Heavy Rain	124.5 – 244.4
Extremely Heavy Rain	>244.5
Exceptionally Heavy Rain	exceeds 12 cm.

VII. CONCLUSION

Machine learning generates a lot of buzz because it's applicable across such a wide variety of use cases. Classification techniques is a supervised learning method in which the computer program learns from the data input given to it and then uses this learning to classify new observation. Choosing an algorithm is a critical step in the machine learning process, so it’s important that it truly fits the use case of the problem.

REFERENCES

[1] X. Pan, Y. Luo, Y. Xu, “K-nearest neighbor based structural twin support vector machine,” Knowledge Based Systems, vol. 88, pp. 34-44, 2015

[2] J. Xu, Q. Wu, J. Zhang, Z. Tang, “Exploiting Universum data in AdaBoost using gradient descent,” Image and Vision Computing, vol. 32, pp. 550-557, 2014

[3] Kalai Magal. R, Shomona Gracia Jacob, “Improved Random Forest Algorithm for Software Defect Prediction through Data Mining Techniques”, International Journal of Computer, pp.0975–8887, Vol.No.23, May2015

[4] L Jiang, D Wang, Z Cai, and XYan: Survey of improving naive bayes for classification.Advanced Data Mining and Applications 2017:134-145.

[5] Cover T, Hart P. Nearest neighbor pattern classification. IEEE Transactions on Information Theory. Available from: http://ieee.org/document/1053964/. Accessed in 2017 (Feb 20).

[6] Elder, J. (n.d). Introduction to Machine Learning and Pattern Recognition, IEEE, Available at LASSONDE University EECS Department York website: http://www.eecs.yorku.ca/ course_archive/ 2011-12/F / 4404, 2015

[7] Neocleous C. & Schizas C. Artificial Neural Network Learning: A Comparative Review. In: Vlahavas I.P., Spyropoulos C.D. (eds) Methods and Applications of Artificial Intelligence,. Springer, Berlin, Heidelberg, doi: 10.1007/3-540-46014-4_27 pp.300-313,2012

[8] Taiwo, O. A. Types of Machine Learning Algorithms, New Advances in Machine Learning, Yagang Zhang (Ed.), ISBN: 978-953-307 -034-6, InTech, University of Portsmouth United Kingdom. Pp 3 –31, 2014

[9] Klassen, “A frame work for search forms classification” In Systems, Man, and Cybernetics (SMC), 2012 IEEE International Conference on , PP.1029-1034 Seoul, Korea, 14-17 Oct. 2012

[10] Anagnostopoulos, C. Anagnostopoulos, V. Loumos and E. Kayafas, “Classifying Web pages employing a probabilistic neural network”, Probabilistic Software, IEEE ,Vol.151, No.3,June 2014, PP.139-150

[11] M. Nayrolles and A. Hamou-Lhadj, "BUMPER: A Tool for Coping with Natural Language Searches of Millions of Bugs and Fixes," Software Analysis, Evolution, and Reengineering (SANER), IEEE pp. 649-652. 2016

[12] T. Cheng, K. Chakrabarti, S. Chaudhuri, V. Narasayya and M. Syamala, "Data services for E-tailers leveraging web search engine assets," Data Engineering (ICDE),IEEE, 2013, pp. 1153-1164

[13] Jebaraj Ratnakumar, “An Implementation of Web Personalization Using Web Mining Techniques”, Journal Of Theoretical And Applied Information Technology, 2014 JATIT

[14] Tsuyoshi, M and Saito, K., “Extracting User‟s Interest for Web Log Data”, IEEE 2016, pp. 343-346, ISBN: 0-7695-2747-7

[15] Lidong Wang Data Mining, Machine Learning and Big Data Analytics International Transaction of Electrical and Computer Engineers System. 2017, 4(2), 55-61. DOI: 10.12691/iteces-4-2-2ublished online: July 24, 2017

[16] https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithm