Prasanta Rao Jillella S S* et al.
(IJITR) INTERNATIONAL JOURNAL OF INNOVATIVE TECHNOLOGY AND RESEARCH
Volume No.3, Issue No.6, October - November 2015, 2534 – 2539.

Weather Forecasting Using Artificial Neural Networks and Data Mining Techniques

PRASANTA RAO JILLELLA S.S.
Department of CSE
Gitam University, Visakhapatnam

P BHANU SAI KIRAN
Department of CSE
Gitam University, Visakhapatnam

P. NITHIN CHOWDARY
Department of CSE
Gitam University, Visakhapatnam

B. ROHIT KUMAR REDDY
Department of CSE
Gitam University, Visakhapatnam

VISHNU MURTHY
Department of CSE
Gitam University, Visakhapatnam

Abstract—**Weather forecasts are made by collecting quantitative data about the current state of the atmosphere and using scientific understanding of atmospheric processes to project how the atmosphere will evolve. Weather prediction is basically based upon the historical time series data. The basic Data mining operations and Numerical methods are employed to get a useful pattern from a huge volume of data set. Different testing and training scenarios are performed to obtain the accurate result. To perform these kinds of predictions we are identifying the datasets. Collection of the data sets of a particular region weather report from 1901 to 2001 with 11 attributes. The collected datasets undergo pre-processing. Then clustering operation, Curve fitting and Extrapolation methods are applied, proceeding with back propagation. The Back propagation and Extrapolation results are compared. The Best future results are predicted.**

Keywords- **Weather Forecast, Artificial Neural Networks and Exploration Techniques.**

I. INTRODUCTION

Weather plays an important role in Human life. Many of our daily works and business depends upon weather conditions. Also there is a huge life and property loss due to unexpected Weather conditions. If we are able to efficiently predict the weather conditions for the future, then we can prevent or minimize these losses. The weather of the earth does not remain same at every time. We have many seasons like summer, winter, spring, autumn, Monsoon etc. Weather changes from time to time. This weather change is quite normal and regular phenomenon of earth. The earth's climate is also influenced and changed through natural causes like volcanic eruptions, ocean current, the earth's orbital changes and solar variations. Weather forecasting is also useful for various purposes like aviation, shipping, fisheries and many other special uses besides forecasts for the general public. We have different types of forecasting, we can predict weather, stock market, next ruling political parties, natural disasters like tsunamis, earthquakes, floods and so on. By predicting the fore coming disaster we can take the required precautions and security measures. Weather simply refers to the condition of air on earth at a given place and time. The application of Science and technology are to predict the state of the atmosphere in future time for a given location is so important due to its effectiveness in human life.

There are a variety of end users to weather forecasts. Weather warnings are important forecasts because they are used to protect life and property. Forecasts based on temperature and precipitation are important to agriculture, and therefore to traders within commodity markets. Temperature forecasts are used by utility companies to estimate demand over coming days. On an everyday basis, people use weather forecasts to determine what to wear on a given day. Since outdoor activities are severely curtailed by heavy rain, snow and the wind chill, forecasts can be used to plan activities around these events, and to plan ahead and survive them.

The prediction of weather should be reliable and good enough to benefit the society. The reliability of a specific prediction depends on the level of our understanding of the process, the amount and the quality of the data we have. Because the basic physical process of weather is now reasonably well understood, and high quality of data are being collected, it should be possible to make accurate predictions regarding the future weather changes.

The rest of the paper is organized as follows: Section 2 describe the literature survey. The proposed method is clearly described in

Section 3. Results are analyzed in Section 4. Section 5 concludes the work followed by future directions.

II. RELATIED WORK

Weather forecasting can be done using various techniques. The following subsections describe the same in detail:

A. Persistence

The simplest method of forecasting the weather, persistence, relies upon today's conditions to forecast the conditions tomorrow. This can be a valid way of forecasting the weather when it is in steady state, such as during the summer season in the tropics. This method of forecasting strongly depends upon the presence of a stagnant weather pattern. It can be useful in short range forecasts but may not be good for long range forecasts.

B. Use of Barometer

Measurements of barometric pressure and the pressure tendency (the change of pressure over time) have been used in forecasting since the late 19th century. The larger the change in pressure, especially if more than 3.5 hPa (2.6 mmHg), the larger the change in weather can be expected. If the pressure drop is rapid, a low pressure system is approaching, and there is a greater chance of rain. Rapid pressure rises are associated with improving weather conditions, such as clearing skies. It cannot be used for predicting values over long range period of time.

C. Looking at the sky

Along with pressure tendency, the condition of the sky is one of the more important parameters used to forecast weather in mountainous areas. Thickening of cloud cover or the invasion of a higher cloud deck is indicative of rain in the near future. A bar can indicate a coming tropical cyclone. The use of sky cover in weather prediction has led to various weather lore over the centuries. It doesn't have any reliable approach. It is just an assumption made by looking at the sky.

D. Analog Technique

The Analog technique is a complex way of making a forecast, requiring the forecaster to remember a previous weather event which is expected to be mimicked by an upcoming event. What makes it a difficult technique to use is that there is rarely a perfect analog for an event in the future. It is not reliable and it is an inappropriate technique.

E. Weather Map

This is a process of Analysis of satellite, radar imageries and other data. The weather map depicts the distribution patterns of atmospheric pressure, wind, temperature and humidity at different levels of the atmosphere. There are two types of the basic weather map namely, the surface map and the upper-air maps. There are five standard levels of the upper-air maps that are constructed twice daily at twelve-hourly interval. The surface maps are made four times daily at six-hourly intervals. On the surface maps, the distribution patterns of rain or other forms of precipitation and cloudiness can also be delineated. Highly sophisticated and costly technology is used. Weather changes should be noticed every hour.

F. Use of forecast models

Data mining is the extraction of hidden predictive information from large databases. It is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses. Data mining tools predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions. The automated, prospective analyses offered by data mining move beyond the analyses of past events provided by retrospective tools typical of decision support systems. The manual extraction of patterns from data has occurred for centuries. Early methods of identifying patterns in data include Bayes theorem (1700s) and regression analysis (1800s). The proliferation, ubiquity and increasing power of computer technology has increased data collection, storage and manipulations. As data sets have grown in size and complexity, direct hands-on data analysis has increasingly been augmented with indirect, automatic data processing. This has been aided by other discoveries in computer science, such as neural networks, clustering, genetic algorithms (1950s), decision trees (1960s) and support vector machines (1980s). Data mining is the process of applying these methods to data with the intention of uncovering hidden patterns.

Walker [1] has done the basic work to predict Asian monsoon prognostication and it is extended in [2-4] to improve the initial model which can predict the rainfall of summer and

monsoon of India. Summer and Monsoon are considered as these two are the main seasons which gets rainfall mostly. For long range forecasting powerful regression models [5] are used by the Indian Meteorological Department (IMD). Due to the limitations of these techniques further attempts have been made to forecast the Indian Summer Monsoon Rainfall (ISMR). Hybrid principal component model (HPCM) is introduced with eight parameters to design a better system and almost forty years (1958 – 1997) dataset is used [6]. Thirty years are considered for training the model and remaining ten years for testing purpose. Later artificial intelligence based approach is attempted by [1] to formulate the relation among local rain fall of Orissa state (Orissa is one of the India state) and local climate of ocean. In [7-9] empirical studies are done to forecast the ISMR.

In weather forecasting using data mining and curve fitting techniques, we generate weather reports by plotting obtained data sets of previous weather data sets and generating a curve through this data and applying extrapolation techniques to predict the future weather conditions. This prediction can be used for all purposes and business applications basing on weather. This precursor doesn't need any sophisticated costly equipment [10]. It just needs data history and a computing machine. In order to overcome the limitations and to predict the best results, we compared the predicted results with already available data and calculated the accuracy. This project inches close to Science and further away from fiction. The weather report for a particular month is known well in advance. With prior warning of the occurrence of the rainfall and changes in temperature, the necessary steps to be taken in order to minimize the losses and damages that are caused due to bad weather. This project doesn't need any sophisticated costly equipment like the barometers, radars, satellites; it just needs the previous data sets and a computing device for the prediction.

III. PROPOSED METHODOLOGY

In this work, we made an effort to forecast the weather using both the exploration and sophisticated artificial neural networks (ANN). Figure 1 depicts the work flow and process of prediction using the proposed methodology. The following subsections explains the parts of the framework starting from Database to neural networks.

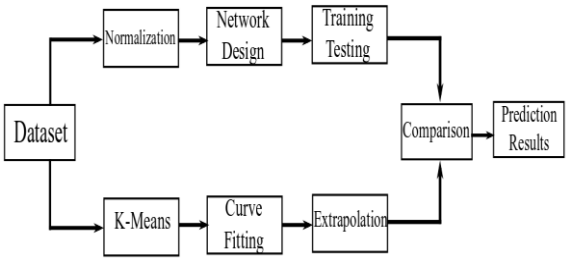

Figure. 1. Flow of proposed methodology for weather forecasting.

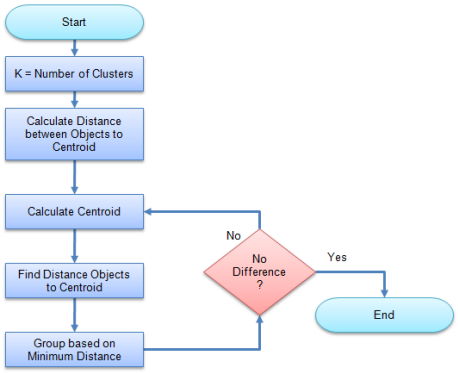

Figure 2: Process to group the dataset using k-means clustering.

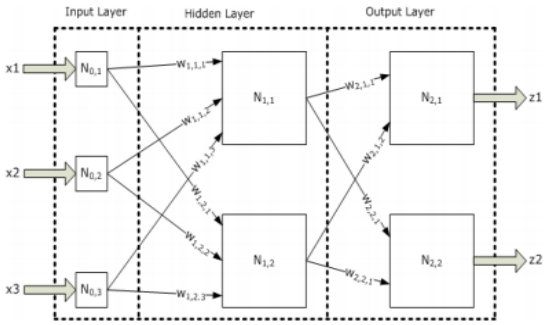

Fig. 3. Representation of Artificial Neural Networks.

G. Database

Ten years real time data is considered for reliable system with eleven dependent variables. The eleven variable are precipitation, minimum temperature, average temperature, maximum temperature, cloud cover, vapor pressure, wet day frequency, diurnal temperature range, ground frost frequency, reference crop evapotranspiration and potential evapotranspiration. The data base is collected from 1901 to 2001 for the region of Visakhapatnam, Andhra Pradesh, India. Monthly database is considered and the work is to predict the average temperature of a month.

H. Normalization

It is the process of transforming the data. To be more precise modifying the source data in to a different format which enables data mining algorithms to be applied easily. It also improves the effectiveness and the performance of the mining algorithms. Represents the data in easily understandable format for both humans and machines, supports faster data retrieval from databases and makes the data suitable for a specific analysis to be performed.

I. K-Means Clustering

The most efficient and foremost algorithm for grouping the data is k-means clustering algorithm. This algorithm aims at partitioning the data into k clusters which helps to understand the data and its category. There will be less similarity among clusters and high similarity between the data elements of clusters after applying this algorithm. To measure the similarity of an object and to find its relevant cluster, we need to find the mean value of a cluster. It helps to find the centroid of a cluster as well. If a new object is found then Euclidian distance will be applied to each object of a cluster to find the minimum distance one will be chosen as its group. Prior to that, initially, k random clusters are chosen randomly for deploying initial clusters. Then Euclidian distance is to be applied among the initial clusters and objects of space. This process helps to find initial group and then mean of all objects of a particular cluster is to be found which helps to locate new centroid. The same process is iteratively continues till the cluster values are converged. This approach is very efficient in clustering techniques and scalable as well. The time complexity is convincing for its usage and it is *O(ink)*. Where *i* is the number of iterations, *n* is the total number of objects and *k* is the total number of clusters. Figure 2 describes the process to cluster the data items using k-means clustering algorithm.

J. Curve fitting

The complexity in forecasting the weather increases using the analog method. This is due to the necessity of remembrance of the past data which is practically not possible and the approach is expensive too. Using the same past data, there are different types of techniques to fit the data and to forecast the same. One such fitting technique is discussed in this subsection.

The mathematical formulation for first degree polynomial is:

$$y = ax + b$$

It shows a line having slope *a*. It is known fact that a line is sufficient to connect two points. Hence, to connect any two distinct points, a first degree polynomial equation is sufficient. In the same way, the polynomial value gets increased depending on number of coordinate points on space. For instance, to fit four points, we need a polynomial equation with order three given below.

$$y = ax3 + bx^2 + cx + d$$

It is also true that a first order polynomial equation can be used to fit a single point and the third degree polynomial equation is also useful to fit three, two and single points respectively. Different constraints are possible with higher order polynomial equations.

Suppose *n* is the degree of polynomial equation and there are *n+1* constraints, these constraints provide a way to construct a polynomial curve. However, there is no certainty to get an exact curve. The solution to this kind of issue is least squares method to approximate the curve.

i. Least Squares

An approach to give the approximate value for the systems which are estimated inexactly or perfectly undetermined ones. This approach reduces the error value of sum of squares where it is needed. This error or residual value is the difference between the observed fit and actual fit given by the model. There are two kinds of algorithms in least squares. One is ordinary or linear and the other is non-linear. It is already known that the most of real time data falls under non-linear category. Very limited data comes under linear category, negligible one. There is a closed form solution for linear data through statistical regression analysis. It is very difficult and impossible to find closed form solution for non-linear data. The solution is to be refined through iterative process.

The general form for polynomial equation is:

$$f(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + \ldots + a_jx^j = a_0 + \sum_{k=1}^{j} a_kx^k$$

To find the residual, there is another equation given by:

$$err = \sum (d_i)^2 = (y_1 - f(x_1))^2 + (y_2 - f(x_2))^2 + (y_3 - f(x_3))^2 + (y_4 - f(x_4))^2$$

From the equation mentioned above, we can rewrite the same as:

$$err = \sum_{i=1}^{n} \left(y_i - \left(a_0 + a_1 x_i + a_2 x_i^2 + a_3 x_i^3 + \ldots + a_j x_i^j \right) \right)^2$$

Where 'n' indicates number of points, 'j' is the polynomial order and 'i' is the current data point being summed.

The above equation can be still compressed into:

$$err = \sum_{i=1}^{n} \left(y_i - \left(a_0 + \sum_{k=1}^{j} a_k x^k \right) \right)^2$$

K. Extrapolation

Prognosticating the new data point from the past behavior of data is the process inculcated in the technique of Extrapolation. The results of extrapolation led to high ambiguity and less relevance issues. Further, they are categorized into different classes based on the type of data. Some of them are linear extrapolation, polynomial extrapolation, French curve extrapolation and conic extrapolation. The use of particular technique is dependent on data. For instance, non-smooth functioning cannot be used for smooth data which results poor prediction. The divergence property can be used after predicting which generally gets accurate results.

L. Artificial Neural Networks

An Artificial Neural Network (ANN) is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information. The key element of this paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements (neurons) working in unison to solve specific problems. ANNs, like people, learn by example. An ANN is configured for a specific application, such as pattern recognition or data classification, through a learning process. Learning in biological systems involves adjustments to the synaptic connections that exist between the neurons. Neural networks, with their remarkable ability to derive meaning from complicated or imprecise data, can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. A trained neural network can be thought of as an "expert" in the category of information it has been given to analyze. Back propagation neural network (BPNN) is a special kind of ANN useful for mapping the non-linear relations efficiently. This is due to its back-ward learning technique.

IV. RESULTS AND ANALYSIS

The experimentation is done in two ways. One is using exploration techniques and the other is using artificial neural networks. Initially, experimentation is done with the extrapolation techniques. The data is plotted on two dimensional space. Later, k-means clustering technique is used to group the data according to the existing weather conditions and based on the dependents. The k-means groups the data into single clusters. Polynomial fitting algorithm is applied on the clustered data to get the curve. It will give the solution to formulate the mathematical equation to predict the next near point based on the existing data points. The mathematical equation can be any of the linear regression or other extrapolation techniques. Extrapolation technique is considered for this work as it better as compared to the other linear regression techniques. The accuracy is not up to the mark using this technique and always it falls at the range of around 60%. Hence, this method is not suitable to predict the next weather conditions.

Later, we thought about another technique to predict the weather in an efficient way. Our efforts had taken in a way to design a system which can understand the non-linear data. This is because, by looking at the data of weather came to the conclusion with some proper analysis that the data is non-linear in nature. There is no linearity on the variables. And there is a solution to map the non-linear data in an efficient way, using neural networks. Artificial neural networks are efficiently designed to map the non-linear relations. The weight factor is used for all inter connected nodes and in two levels. First level training between input layer and hidden layer and the second level training is done between hidden and output layers. Back propagation neural networks (BPNN) is used in this paper as it has

the backward learning capability. The results relaxed us from the other solutions.

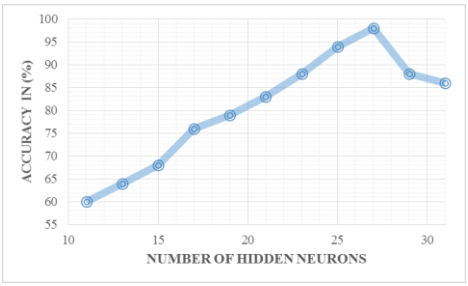

Fig. 4. Accuracy of ANN for various hidden neurons.

Experimentation is done on different values for hidden neuron to identify the number which gives best accuracy. The accuracy for different hidden neurons is shown in Fig 4. In which x axis represents the number of hidden neurons and vertical axis shows the accuracy. The best accuracy is achieved at the 1.5 times to the input neurons. The results are reached above 95% for the collected dataset. It proves that the neural networks are better to map the non-linear kind of data.

V. CONCLUSION

This paper of weather forecasting using Data mining and curve fitting is premature at this time. It offers a speculative explanation of predicting weather changes using clustering. We have given the weather report for coming few years. Comparing our results with years 2001-2005, the prediction is quite matching. Hence the predicted results are reliable but more research is certainly needed to clarify the mechanisms, as well as to accumulate the case studies and to develop the accuracy of prediction. Our attitude towards the weather prediction is optimistic. Prediction is limited by unknown conditions and the inability to know quantitative details.

No one can go against nature and fight it; no one can escape from its dreadful deeds. As the old saying goes prevention is better than cure, and people should take precautionary measures to reduce the possibilities of farms, business, property damage, and loss of life. This work is mainly aimed for future development rather than using it as if it is through coping with Geological and Meteorological department through acquiring useful information regarding weather changes.

VI. REFERENCES

[1] Vrugt, Jasper A., et al. "Multi-objective calibration of forecast ensembles using Bayesian model averaging." Geophysical research letters 33.19 (2006).

[2] Alam, Sameer, et al. "Multi-objective ant colony optimization for weather avoidance in a free flight environment." The Artificial Life and Adaptive Robotics Laboratory, University of New South Wales, Canberra, Australia (2006).

[3] Baboo, S. Santhosh, and I. Kadar Shereef. "An efficient weather forecasting system using artificial neural network." International journal of environmental science and development 1.4 (2010): 2010-0264.

[4] Chattopadhyay, Surajit. "Feed forward Artificial Neural Network model to predict the average summer-monsoon rainfall in India." Acta Geophysica 55.3 (2007): 369-382.

[5] Gibson, Emily, and Jessie Burger. "Parallel genetic algorithms: an exploration of weather prediction through clustered computing." Journal of Computing Sciences in Colleges 18.5 (2003): 272-273.

[6] Singh, Shaminder, Pankaj Bhambri, and Jasmeen Gill. "Time series based temperature prediction using back propagation with genetic algorithm technique." International Journal of Computer Science 8.5 (2011): 28-35.

[7] Tripathy, Asis Kumar, et al. "Weather Forecasting using ANN and PSO." International J. Scientific & Eng. Res 2 (2011): 1-5.

[8] Paras, Sanjay Mathur, Avinash Kumar, and Mahesh Chandra. "A feature based neural network model for weather forecasting." International Journal of Computational Intelligence 4.3 (2009).

[9] Marsh, Lawrence C., and David R. Cormier. Spline regression models. Vol. 137. Sage, 2001.

[10] Santhanam, Tiruvenkadam, and A. C. Subhajini. "An efficient weather forecasting system using radial basis function neural network." Journal of Computer Science 7.7 (2011): 962.

