

DATA MINING IN WEATHER FORECASTING

S. Sandhya ,

M.Phil Research Scholar,
Department of Computer Science,
Chikkanna Government Arts College,
Tirupur, Tamilnadu, India.

Dr. G. M. Nasira,

Head,
Department of Computer Applications,
Chikkanna Government Arts College,
Tirupur, Tamilnadu, India.

Abstract: Weather forecasting is the application of science and technology to predict the state of the atmosphere for a given location. Prediction is a challenging task and that too for weather is even more complex, because it depends on various parameters to predict the dependent variables like temperature, rainfall, humidity, wind speed and direction, which are changing from time to time and weather calculation varies with the geographical location along with its atmospheric variables. In this paper, we analyse the data mining techniques in forecasting weather.

Keywords: weather prediction, datamining, ANN, K-Nearest neighbor, decision trees.

I. INTRODUCTION

Weather forecasting is an area of meteorology that is carried by collecting dynamic data related to current state of weather like fog, rainfall, temperature, wind etc. We continually update our knowledge of the current state of the atmosphere by

- Satellites measure radiation from Earth's plane and the impression.
- Balloons and aircraft measure the bit of the air that they passing through.
- Buoys and land stations measure the lower part of the atmosphere.
- Radar systems measure the indication of emission from rain drops and snowflakes

The data collected from various states are to distorted into a numerical representation of the recent atmospheric conditions. This process is known as assimilation. Small changes in atmospheric conditions lead to very different weather patterns, so it's vital that the current state of the atmosphere is represented as accurately as possible. Data mining is a process that uses a variety of data analysis tools to discover patterns and relationships in data that may be used to make valid prediction.

II. WEATHER FORECASTING

Weather forecasting plays a significant role in meteorology. Weather forecasting remains a formidable challenge because of its data intensive and frenzied nature. Generally two methods are used to forecast weather: a) the empirical approach and b) the dynamical approach. The first approach is based on the occurrence of analogues and it is often referred to as analogue forecasting.

This approach is useful in predicting local scale weather if recorded cases are plentiful. The second case is based upon equations and forward simulations of the atmosphere and is often referred to as computer modeling. The dynamical approach is useful to predict large scale weather phenomena and may not predict short term weather efficiently. Most weather prediction systems use a combination of both the techniques.

III. DATAMINING IN WEATHER FORECASTING

Data Mining deals with what kind of patterns can be mined. Based on the kind of data to be mined, there are two kinds of functions involved in Data Mining such as Descriptive model and Predictive model. The Descriptive model identifies patterns or relationships in data and deals with general properties of data in the database. The predictive model is the process of finding a model which describes the data classes or concepts, the purpose being to be able to use this model to predict the class of objects whose class label is unknown.

Data mining techniques are mainly separated into two groups, viz. classification and clustering techniques. Classification techniques are designed for classifying unknown samples using information provided by a set of classified samples. This set is usually referred to as a training set, because, it is generally used, to train the classification technique i.e. how to perform its classification. If a training set is not available, there is no previous knowledge about the data to classify. Clustering technique is used to group the element that is particular area occupied by rainfall regions and the rainfall is predicted in a particular region. In this technique the following three techniques are discussed.

Artificial Neural Network

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. It is composed of a huge number of highly interconnected processing elements working in unison to solve specific problems. An ANN is configured for a particular application, such as pattern recognition or data classification, through a learning process. The artificial neuron is an information processing unit that is fundamental to the operation of a neural network. There are three basic elements of a neuron model. Figure 1 shows the basic elements of neuron model with the help of a perceptron model, which are, (i) a set of synapses connecting links, each of which is characterized by a weight or strength of its own, (ii) an adder for summing the input signals weighted

by the respective synapses of the neuron and (iii) an activation function for limiting the amplitude of the output of a neuron.

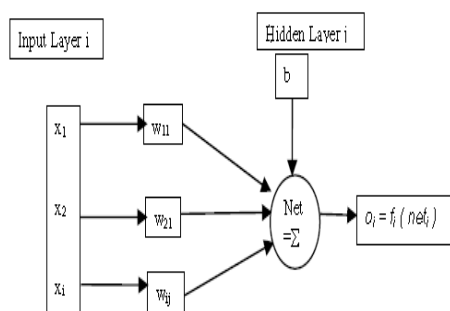


Figure1: Model of a perceptron

ANN's architecture requires determination of the number of connection weights and the way information flows through the network, this is carried out by choosing the number of layers, number of nodes in each layer and their connectivity. The numbers of output nodes are fixed by the quantities to be estimated.

Decision Trees

A Decision Tree is a flow-chart-like tree structure. Each internal node denotes a test on an attribute. Each branch represents an outcome of the test. Leaf nodes represent class distribution. The decision tree structure provides an explicit set of "if-then" rules, making the results easy to interpret. In the tree structures, leaves represent classifications and branches represent conjunctions of features that lead to those classifications. In decision analysis, a decision tree can be used visually and explicitly to represent decisions and decision making. The concept of information gain is used to decide the splitting value at an internal node. The splitting value that would provide the most information gain is chosen. Formally, information gain is defined by entropy. In order to improve the accuracy and generalization of classification and regression trees, various techniques were introduced like boosting and pruning. Boosting is a technique for improving the accuracy of a predictive function by applying the function repeatedly in a series and

combining the output of each function with weighting so that the total error of the prediction is minimized or growing a number of independent trees in parallel and combine them after all the trees have been developed. Pruning is carried out on the tree to optimize the size of trees and thus reduce overfitting which is a problem in large, single-tree models where the model begins to fit noise in the data. When such a model is applied to data that was not used to build the model, the model will not be able to generalize. Many decision tree algorithms exist and these include: Alternating Decision Tree, Logitboost Alternating Decision Tree (LAD), C4.5 and Classification and Regression Tree (CART).

K-Nearest Neighbor

The k -nearest neighbor (k -NN) method is one of the data mining techniques. It tries to classify an unknown model based on the known classification of its neighbors. Suppose that a set of samples with known classification is available, the so-called *training set*. Intuitively, each model should be classified likewise to its close samples. Therefore, if the classification of a sample is unknown, then it could be predicted by considering the classification of its nearest neighbor samples. Given an strange sample and a training set, all the distances between the unknown sample and all the sample in the guidance set can be computed. The distance with the smallest value corresponds to the sample in the training set closest to the unknown sample. Therefore, the unknown sample may be classified based on the classification of this nearest neighbor.

IV. CONCLUSION

Weather prediction is necessary so as to inform people and prepare them in advance about the current and upcoming weather condition. This helps in reduction in loss of human life and loss of resources and minimizing the mitigation steps that are expected to be taken after a natural disaster occurs. Data mining is one of the most widely used techniques for weather prediction. Data mining offers a way of analysing data statistically and extract or derive such rules that can be used for predictions. In this paper mostly used data mining techniques are discussed.

V. REFERENCES

- [1] Due R. A., A Statistical Approach to Neural Networks for Pattern Recognition, 8th edition. New York: John Wiley and Sons publication.
- [2] S. Kotsiantis and et. al., "Using Data Mining Techniques for Estimating Minimum, Maximum and Average Daily Temperature Values", World Academy of Science, Engineering and Technology.
- [3] Han & Kamber, "Data Pre-processing & Mining Algorithm, Knowledge & Data Mining & Pre-processing".
- [4] Jiawei Han, Jian Pei, Micheline Kamber, "Data Mining: Concepts and Techniques"
- [5] Han, J., Micheline K., Data Mining: Concepts and Techniques, San Fransisco, CA: Morgan Kaufmann publishers.
- [6] Arun K Pujari, "Data Mining Techniques", Universities Press(India).