International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2014): 5.611

Arbitrary Decision Tree for Weather Prediction

Nalanda B Dudde[1], Dr. S. S. Apte[2]

[1]Walchand Institute of Technology, Solapur, India

[2]HOD Walchand Institute of Technology, Solapur, India

Abstract: ***Long before technology was developed, folks had to trust observations, patterns and their expertise to predict the weather. Here we present novel approach for predicting weather using decision tree, it is based on concepts and techniques from data mining and prediction systems. The suitability of this approach for prediction and its advantages compared with different techniques are considered here. This paper highlights on development of weather prediction model using decision tree and implementation of it. Concepts of entropy is used to find the homogeneity of classes present in the dataset, information gain used for finding the threshold values for the nodes and finally the Hoeffding bounds [1] of inequality to choose the minimum number of split examples from the dataset. It selects attributes randomly and constructs a tree which will be efficient and will improve the accuracy of classifiers.***

Keywords: Decision tree, Entropy, information Gain, weather prediction system

1. Introduction

An important technique in **machine learning is** decision trees, which are used extensively in **data mining**. It is used to turn out human-readable descriptions of trends within the underlying relationships of a dataset and might be used for classification and prediction tasks. A decision tree is predictive modeling technique used in classification, clustering and prediction tasks. It uses divide and conquer technique to split the problem search space into subsets [5]. It can be used to explain why a question is being asked. The decision tree assumes that questions are answered with a certain yes or no. In globe issues, the intuition of a personality's knowledgeable, or knowledgeable system package, is critical to see the possible finish node. Each end node represents a situation with known effective and efficient leadership styles. The technique has been used with success in many various areas, like diagnosis, plant classification, weather prediction, client selling methods etc. A decision tree is a representation of a decision procedure for determining the class of a given instance. Each node of the tree specifies either a class name or a specific test that partitions the space of instances at the node according to the possible outcomes of the test. Each subset of the partition corresponds to a classification sub problem for that subspace of the instances, which is solved by a sub tree [6][4]. Formally, one can define a decision tree to be either.

A leaf node (or answer node) represent successful/ unsuccessful answer, and a non-leaf node (or decision node) that contains an attribute test with a branch to another decision tree for each possible value of the attribute [3]. An attribute check, with a branch to a different call tree for every doable price of the attribute, the positive and negative counts for every doable price of the attribute, and therefore the set of non-test attributes at the node, every with positive and negative counts for every doable price of the attribute.

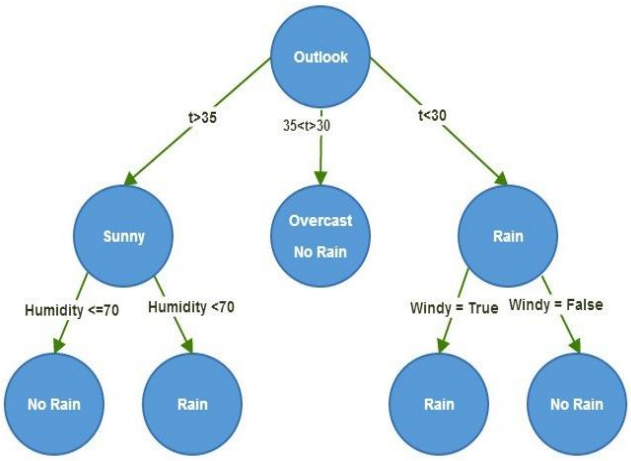

Figure 1.1: Graphical Decision tree

Weather prediction is one of the most effective environmental constraints in our routine. We adjust ourselves with respect to weather condition from our dressing to strategic organizational planning activities. There are 2 methods used for weather prediction one is empirical approach and another is dynamic approach [8]. The empirical approach is based upon the comparable cases (i.e., similar weather condition). The dynamical approach is based upon equations of the atmosphere and is commonly referred to as computer modeling.

2. Literature Review

The decision tree for numerical data [2] is proposed by Nandagaonkar S Attar V.Z, Sinha P.K. in paper Efficient Decision Tree Construction for Classifying Numerical Data which uses random strategy for building a decision tree. It also uses heuristic method to compute the information gain for obtaining the split threshold of numerical attributes.

ID3 is a simple decision tree learning algorithm developed by Ross Quinlan (1983)[3]. The basic plan of ID3 formula is to construct the choice tree by using a top-down, greedy search

through the given sets to check every attribute at each tree node.

An expert system for weather prediction based on animal behavior by Najat O. Alsaiari[7] it is an expert system for predicting short term weather based on the behavior of both birds and animals like sea- crab, which have been observed to have certain reactions before weather changes.

Weather prediction expert system approaches[8](Ceng-568 Survey) by Bulent Kiskac and Harun Yardimci it's a survey paper where Hybrid systems are promising for integration of current expert systems on large scale also many web weather report and forecast service systems enable implementations to fetch some weather information, hybrid system can reduce intensive forecast computations. Instead of computing some detailed information, here we can easily pick up knowledge about a satellite post-processing reports and comments.

Performance Analysis on decision tree algorithms [11] by Nalanda Dudde and Dr Mrs S S Apte is a review article in which the various algorithm`s performance with respect to complexity and accuracy is analyzed,which will be helpful for the current work.

3. Methodology / Implementation

"Arbitrary Decision Tree for Weather Prediction" The aim of this work is to study, propose, and implement Arbitrary Decision tree Algorithm for weather prediction. This system will provide a model for weather prediction with much more accuracy rate.

The basic framework for ADT is as follow:

Input: Training set ADTTR = {TR_1, $TR_{2,\ldots}$, TR_m};
Testing set ADTTE = {TE_1, $TE_{2,\ldots}$, TE_m};
Attribute set A = {A_1…., $A_{M(Att)}$};
Initial height of tree h_0;
The number of minimum split-examples n_{min};
The parameter in Hoeffding bounds inequality ∂;
The threshold τ;
Split estimator function $H(\bullet)$;
The number of ADT N;
The size of training window WinSize;
The checked period for constructing the alternative subtree CheckPeriod;
The size of testing window for the alternative subtree TestSize;
The classification error counts of N-ADT Errors

Output:
An Arbitrary Decision Tree – ADT.
Procedure of ADT Algorithm {ADTTR, ADTTE, A, h_0, n_{min} , ∂, τ, $H(\bullet)$, N, WinSize, CheckPeriod, TestSize , Errors}
1. Create the framework of arbitrary decision tree
   a) Generate the root of single tree;
   b) Choose an available attributes A_j from A as the split-attribute at the current node;
   if A_j is a discrete attribute
   Generate m + 1 child nodes (where m is the number of different values in A_j
   else
   Generate two child branches, i.e., < and ≥ a cut-point value of A_j;
   for each child node
   Go to step b) until the height of tree = h_0 ;

2. Update the node counts and determine the split-threshold of nodes with the training data
   for each training data records TR_i (i=1, . . . ,m)
   from DSTR
   a. Sort TR_i into a leaf of ADT;
   b. if the current node with A_j is a numerical attribute && the count of examples at
   this node ≥ n_{min}
   {Compute the values of ϵ by Hoeffding bounds inequality;}
   If $(((H (Value_x(A_j)) – H(Value_y(A_j))) > \varepsilon \,||$
   $(H (Value_x(A_j)) – H(Value_y(A_j))) \leq \varepsilon < \tau))$
   { Set $Value_x(A_j)$ with the highest gain value as the split point ;}

3. Classify the testing data for each testing data record TE_i (i=1…., n) from ADTTE
   a.Travel the tree from the root to leaves and increase the counts of class labels at each passed node;
   b. Classify the class label of this testing data records by the judging function of class labels;

Return ADT;

4. Experimental Analyses

Input:
The proposed framework takes input of weather details of London as historical data from http://en.tutiempo.net/climate which is a authenticate weather analytical organization. We have downloaded the weather details of London city of year 2014 from above mentioned URL. Now the data has been divided into 2 parts.

Training Data:
It is the first 10 months data (i.e. from January 2014 to October 2014). This dataset is used to prepare decision tree and also used to train the tree.

Testing Data:
It is the last 2 months data (i.e. from November 2014 to December 2014). This dataset is used for classification and finding out the accuracy of tree.

Result:
The proposed framework which is based on Arbitrary Decision Tree gives above 80% accuracy for predicting weather of different cities of different countries, which is very good accuracy in machine learning classification. To perform classification in this stage we have used testing dataset i.e. we have 61 number of weather records and out of 61 records over 90% of records are hitting on their classified

leaf nodes. We also found that it takes less time and memory (i.e. System Resources).

Table 4.1: Analysis of Various DT's Algorithms

Name of Approach	Input Dataset	Input Dataset Type	Accuracy
ID3[10]	Diabetes	Static	57.5%
ID3[11]	CPU Performance	Static	68.84%
C4.5[10]	Diabetes	Continuous and Discrete	74.8%
VFDTc[10]	Weather	Only Streaming data	62.4%
ADT (Proposed Methodology)	Weather	Continuous ,discrete and Streaming data	Above 80%

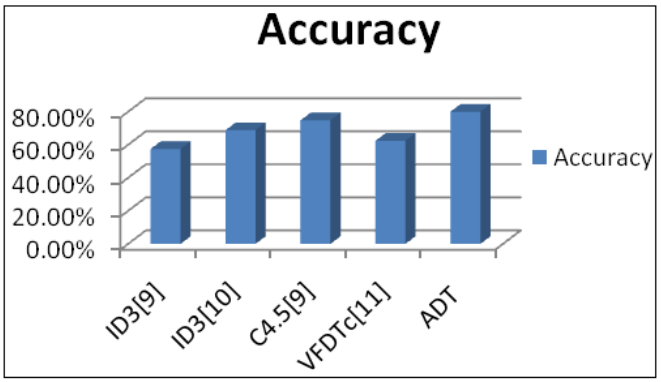

Figure 4.1: Comparison Graph

The table 6.1 and fig 6.1 shows the detailed analysis of prediction techniques which were adopted decision tree framework. Using the approaches mentioned in table 6.1 we can say that, our proposed methodology will give better results in terms of time, space and accuracy because,

a) The input data may reside on web or it's a streaming data so our proposed approach takes less space.
b) The proposed methodology not support windowing approach and backtracking so it can run in less amount of time as compare to ID3, C4.5 and VFDTc[1].
c) The proposed methodology will give better accuracy because
   - It uses Maximus classifier instead of Navie bayes classifier.
   - The weather data is a continuous data which changes timely, so the dataset is very huge for training and testing the tree. The maximum amount of data is directly proportional to the accuracy.
   - The ADT can prune decision trees to stop them over fitting the training data.

5. Conclusion

The implemented Arbitrary Decision tree for weather prediction as shown in results can used to deal with categorical attributes. Till now the prediction and forecasting of weather is done using weather maps, behavior of animals, satellite images, clouds positioning etc, but our approach helps in classification and prediction. Our approach proved that Information Gain, entropy and the use of the split Evaluator (Hoeffding Bound) for accurate decision tree and it can prune decision trees to stop them over fitting the training data, such things leads to better accuracy in predictions as well as smaller trees for predicting the weather.

References

[1] Nalanda Dudde and Dr Mrs S. S. Apte "Performance Analysis on decision tree algorithms" International Journal of Advanced Research in Computer Science, 5 (7), Sept–Oct, 2014,199-201
[2] Nandagaonkar S., Attar V.Z. ; Sinha P.K. ," Efficient Decision Tree Construction for Classifying Numerical Data", Advances in Recent Technologies in Communication and Computing, 2009. , Page(s): 761 – 765
[3] Quinlan J.R., "INDRODUCTION OF DECISION TREES" Machine learning. VOL 1986, 81-106
[4] J. E. Gehrke, R. Ramakrishnan, and V. Ganti, "Rain-Forest - A framework for fast decision tree construction of large datasets," Data Mining and Knowledge Discovery, Vol. 4, No. 2/3, Jul. 2000, pp. 127-162.
[5] J. Han and M. Kamber, Data Mining: Concepts and Techniques, Morgan Kaufmann Publish, 2001
[6] Kalles D, Morris T. "Efficient incremental induction of decision tress", machine learning,1996,24(3); 231~242
[7] Najat O. Alsaiari. "An expert system for weather prediction based on animal behavior".
[8] Bulent Kiskac and Harun Yardimci. "Weather prediction expert system approaches"(Ceng-568 Survey)
[9] D. Lavanya & Dr. K. Usha Rani, Sri padmavathi mahila visvavidyalayam, Tirupati , AP. "Performance Evaluation of decision tree classifiers on medical datasets"
[10] Ricardo Rocha Projecto Mathematica Ensino Departamento de Mathematica 3810 Aveiro, Portugal. "Accurate Decision tree for mining high speed data streams"
[11] Anuja Priyama, Abhijeeta, Rahul Guptaa, Anju Ratheeb and Saurabh Srivastava "Comparative Analysis of Decision Tree Classification Algorithms" International Journal of Current Engineering and Technology, Vol.3, No.2 (June 2013)