Survey Paper on Implementation of ID3 Algorithm

Kalpana T Kanade[1] Ashwini V Rajguru[2]

[1,2]Everest College of Engineering, Aurangabad

Abstract— This paper for solving the problem a decision tree algorithm based on attribute-importance is proposed. The improved algorithm uses attribute-importance to increase information gain of attribution which has fewer attributions and compares ID3 with improved ID3 by an example. ID3 builds a decision tree from a fixed set of examples. The resulting tree is used to classify future samples. The example has several attributes and belongs to a class (like yes or no). The leaf nodes of the decision tree contain the class name whereas a non-leaf node is a decision node. The decision node is an attribute test with each branch (to another decision tree) being a possible value of the attribute. ID3 uses information gain to help it decide which attribute goes into a decision node. [1].

Key words: Data Mining, Decision Tree, ID3 Algorithm, Entropy, Information Gain, Example

I. DATA MINING

Data mining is the exploration and analysis of large quantities of data in order to discover valid, novel, potentially useful, and ultimately understandable patterns in data. Generally, data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information – information that can be used to increase revenue, cuts costs, or both. Data mining software is one of the numbers of analytical tools for analyzing data. The extraction of information is not the only process we need to perform; it also involves other processes such as Data Cleaning, Data Integration, Data Transformation, Data Mining, Pattern Evaluation and Data Presentation. Once all these processes are over, we are now position to use this information in many applications such as Fraud Detection, Market Analysis, Production Control, Science Exploration etc.[2]

II. DECISION TREE

A decision tree is a tree in which each branch node represents a choice between a number of alternatives, and each leaf node represents a decision.[3]

Decision tree are commonly used for gaining information for the purpose of decision -making. Decision tree starts with a root node on which it is for users to take actions. From this node, users split each node recursively according to decision tree learning algorithm. The final result is a decision tree in which each branch represents a possible scenario of decision and its outcome. The decision tree is a structure that includes root node, branch and leaf node. Each internal node denotes a test on attribute, each branch denotes the outcome of test and each leaf node holds the class label. The topmost node in the tree is the root node.[4]

The following decision tree is for concept buy computer, that indicates whether a customer at a company is likely to buy a computer or not. Each internal node represents the test on the attribute. Each leaf node represents a class. A decision tree is a tree in which each branch node represents a choice between a number of alternatives, and each leaf node represents a decision. Decision tree are commonly used for gaining information for the purpose of decision -making. Decision tree starts with a root node on which it is for users to take actions. From this node, users split each node recursively according to decision tree learning algorithm. The final result is a decision tree in which each branch represents a possible scenario of decision and its outcome.[5]

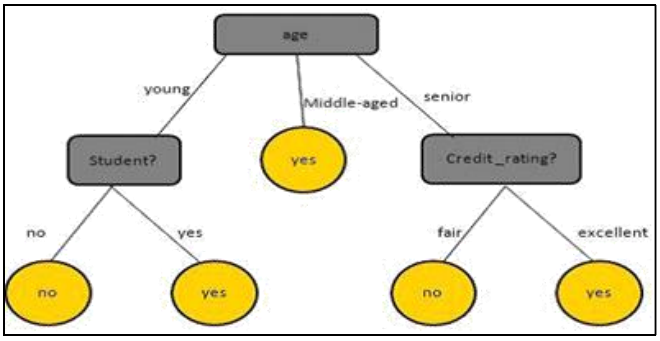

Fig. 1: Decision Tree

III. ID3 ALGORITHM

ID3 is a simple decision tree learning algorithm developed by Ross Quinlan (1983). The basic idea of ID3 algorithm is to construct the decision tree by employing a top-down, greedy search through the given sets to test each attribute at every tree node. In order to select the attribute that is most useful for classifying a given sets, we introduce a metric---information gain.[6]

To find an optimal way to classify a learning set, what we need to do is to minimize the questions asked (i.e. minimizing the depth of the tree). Thus, we need some function which can measure which questions provide the most balanced splitting. The information gain metric is such a function.

ID3 builds a decision tree from a fixed set of examples. The resulting tree is used to classify future samples. The example has several attributes and belongs to a class (like yes or no). The leaf nodes of the decision tree contain the class name whereas a non-leaf is a decision node. The decision node an attribute test with each branch (to another decision tree) being a possible value of the attribute. ID3 uses information gain to help it decide which attribute goes into a decision node. The advantage of learning a decision the tree is that a program, rather than a knowledge engineer, elicits knowledge from expert.[7]

J. Ross Quinlain originally developed ID3 at the University of Sydney. He first presented ID3 in 1975. ID3 is based off the concept learning system. Id3 algorithm is the best algorithm which uses an entropy based measure known as Information gain that will best separate the samples into individual classes so that in future, if any new values entered into database, the classes will be assigned to the data which helps the customer to maintain the customer relationship management and it provides the accuracy of the classification of the data. It leads us to draw decision tree where Iterative Dichotomiser 3 algorithm is used to find out which one of the attribute from the application database can made as root node. Data mining techniques basically use the ID3 algorithm as it's the basic algorithm of classification.[8]

A. The Algorithm

- Create a root node for the tree
- If all examples are positive, Return the single-node tree Root, with label = +.
- If all examples are negative, Return the single-node tree Root, with label = -.
- If number of predicting attributes is empty, then Return the single node tree Root, with label = most common value of the target attribute in the examples.
- Else
  - A = The Attribute that best classifies examples.
  - Decision Tree attribute for Root = A.
  - For each possible value, *vi*, of A,
- Add a new tree branch below Root, corresponding to the test A = *vi*.
- Let Examples(*vi*), be the subset of examples that have the value *vi* for A
- If Examples(*vi*) is empty
- Then below this new branch add a leaf node with label = most
- common target value in the examples
- Else below this new branch add the subtree ID3 (Examples(*vi*), Target_Attribute, Attributes – {A})
- End
- Return Root

B. Entropy

For decision tree construction which attribute is to split is decided by calculating entropy, information gain. And according to that which entropy is greater the node is selected.

- Formula to calculate-
- A complete homogeneous sample has an entropy of 0
- An equally divided sample as an entropy of 1
- Entropy = - p+log2 (p+) -p-log2 (p-) for a sample of negative and positive elements.

$$\text{Entropy}(S) = \sum_{i=1} \text{Pi log2Pi}$$

C. Information Gain

- Information gain is based on the decrease in entropy after a dataset is split on an attribute.
- Looking for which attribute creates the most homogeneous branches

Information gain is also important factor in constructing decision tree. As information gain is maximum the attribute is selected.[7]

Gain(S,A)= Entropy(S) -Sum for v from 1 to n of (|Sv|/|S|) * Entropy(Sv)

1) Example

Aggregate	Seminar	Project	Class Test	Theory	Practical	Lab Work	Other	Last Sem Marks	Category
Strong	Strong	Strong	Strong	Weak	Weak	Strong	Weak	Medium	Good
Strong	Strong	Strong	Strong	Strong	Strong	Strong	Strong	Strong	Good
Weak	Weak	Strong	Weak	Weak	Weak	Strong	Strong	Medium	Bad
Strong	Strong	Strong	Strong	Weak	Weak	Weak	Strong	Medium	Good
Strong	Strong	Strong	Strong	Strong	Weak	Strong	Strong	Medium	Bad
Weak	Weak	Weak	Strong	Weak	Weak	Strong	Weak	Medium	Good
Weak	Strong	Strong	Strong	Strong	Weak	Weak	Strong	Strong	Good
Weak	Weak	Weak	Weak	Strong	Weak	Strong	Weak	Weak	Bad

Table 1: Data set to calculate Entropy and Information gain using ID3 Algorithm

2) Attributes

In the above table, Aggregate, Seminar, Project, Class Test, Theory, Practical, Lab Work, Other, Last Sem Marks, Category denotes as attributes.

3) Class(C) or Classifier

Among these attributes Category refers as Class(C) or Classifier. Because based on Aggregate, Seminar, Project, ClassTest, Theory, Practical, LabWork, Other, LastSemMarks we need to decide whether Student can perform good or bad, that's why Category is a classifier to make decision.

4) Collection (S)

All the records in the table refer as Collection (S).

D. Entropy Calculation

Entropy is one kind of measurement procedure in information theory, details about Entropy is here. we will see how to calculate Entropy of given set of data. The test data used in here is Fig 1 data, [10]

$$\text{Entropy}(S) = \sum_{n=1} -p(I)\log_2 p(I)$$

p(I) refers to the proportion of S belonging to class I i.e. in the above table we have two kinds of class {Bad, Good} with {3,5} (in here 3 is the total no of No and 5 is the total no of Yes), the collection size is S=8. So the p(I) over C for the Entire collection is Bad (3/8) and Good (5/8). $\log_2 p(I)$, refers to the $\log_2(3/8)$ and $\log_2(5/8)$ over C.

∑ is over c i.e. summation of all the classifier items.In this case summation of $-p(Bad/S)\log_2 p(Bad/S)$ and $-p(Good/S)\log_2 p(Good/S) = -p(Bad/S)\log_2 p(Bad/S)+-p(Good/S)\log_2 p(Good/S)$

So,

$\text{Entropy}(S) = \sum -p(I)\log_2 p(I)$

$\text{Entropy}(S)=-p(Bad/S)\log_2 p(Bad/S)+-p(Good/S)\log_2 p(Good/S)$

$\text{Entropy}(S) = ((-(3/8)\log_2(3/8)) + (-(5/8)\log_2(5/8)))$

$\text{Entropy}(S) = (-0.375 \times -0.425) + (-0.625 \times -0.204)$

$\text{Entropy}(S) = 0.530 + 0.423 = 0.940$

So the Entropy of S is 0.953

E. Information Gain G(S, A)

Information gain is the procedure to select a particular attribute to be a decision node of a decision tree. Please see here for details about information gain. [11]

F. Information gain is G(S, A)

Where S is the collection of the data in the data set and A is the attribute for which information gain will be calculated over the collection S. So if Gain(S, Aggregate) then it refers gain of Aggregate over S.

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum((|S_v|/|S|) \times \text{Entropy}(S_v))$$

Where,
S is the total collection of the records.
A is the attribute for which gain will be calculated.
v is all the possible of the attribute A, for instance in this case if we consider Aggregate attribute then the set of v is { Weak, Strong }.
S_v is the number of elements for each v for instance Sweak = 4 and Sstrong = 4.
$\sum$ is the summation of (($|S_v|/|S|$) x Entropy(S_v)) for all the items from the set of v i.e.

$$((|S_{weak}|/|S|) \times \text{Entropy}(S_{weak})) + ((|S_{strong}|/|S|) \times \text{Entropy}(S_{strong}))$$

So if we want to calculate information gain of Wind over the collection set S using following formula,

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum((|S_v|/|S|) \times \text{Entropy}(S_v))$$

Then the it will be as below,

1) *Gain(S, Aggregate)*
$= \text{Entropy}(S) - ((|S_{weak}|/|S|) \times \text{Entropy}(S_{weak})) - ((|S_{strong}|/|S|) \times \text{Entropy}(S_{strong}))$

From the above table for the Wind attribute there are two types of data (Weak, Strong) So new data set will be divided into following two subsets as below,

So, the set of Aggregate attribute is (Weak, Strong) and total number of elements set is (4, 4).

2) *Gain(S, Aggregate)*
$= 0.953 - ((3/8) \times \text{Entropy}(S_{weak})) - ((5/8) \times \text{Entropy}(S_{strong}))$

As mentioned earlier, How to calculate Entropy, now we will calculate Entropy of Weak, there two classifier Bad and Good, for Weak the set of classifier is (Bad, Good) with number of elements set is (1,3) and Strong has set of (Bad, Good) is (2,2).

3) *Entropy(S_{weak})*
$= \sum -p(I)\log_2 p(I) = (-((1/4)\times\log_2(1/4))) + (-((3/4)\times\log_2(3/4)))$

4) *Entropy(S_{weak})*
= calculated value Value of S_{weak} using entropy calculation procedure mentioned earlier.

and

5) *Entropy(S_{strong})*
$= \sum -p(I)\log_2 p(I) = (-((2/4)\times\log_2(2/4))) + (-((2/4)\times\log_2(2/4)))$

6) *Entropy(S_{strong})*
= calculated value ValueOf S_{strong} using entropy calculation procedure mentioned earlier.

So Gain(S,Aggregate) will be as below, Gain(S, Aggregate) = $0.953 - ((3/8) \times \text{Entropy}(S_{weak})) - ((5/8) \times \text{Entropy}(S_{strong}))$

7) *Gain(S, Aggregate)*
= 0.953 - Value of S_{weak} - Value of S_{strong}

8) *Gain(S, Aggregate) = 0.548*

So the information gain of Aggregate over S is 0.548. Using the same procedure it is possible to calculate information gain for Seminar, Project, Class Test, Theory, Practical, LabWork, Other, LastSemMarks.

Attributes	**Information Gain**
Aggregate	0.542
Seminar	0.17
Project	0.0568
ClassTest	0.794
Theory	0.048
Practical	0.091
LabWork	0.203
Other	0.007
LastSemMarks	0.093

Table 2: Information gain calculation

The information Gain of all attributes is shown in above table.

Based on the ID3 algorithm highest gained will be selected for the decision node, in here ClassTest.

ClassTest has two different values Strong & Weak which will be used as decision nodes.

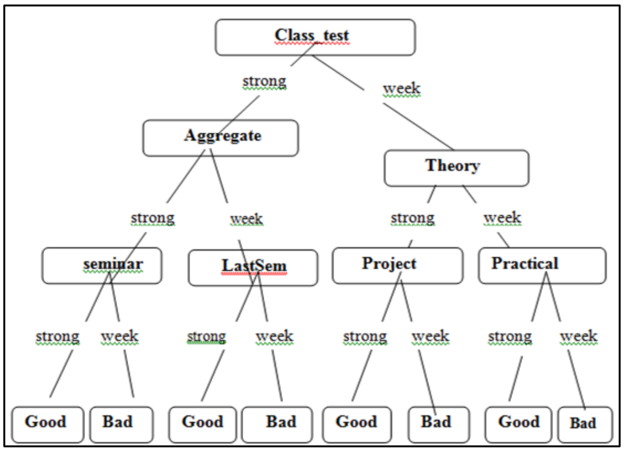

Fig. 2: ID3 Decision Tree

1) If C.T = strong and Aggregate = strong and seminar =strong then Result=Good
2) If C.T = strong and Aggregate =strong and seminar =week then Result=Bad
3) If C.T = strong and Aggregate = week and lastsem =strong then Result=Good
4) If C.T = strong and Aggregate = week and lastsem =week then Result=Bad
5) If C.T = week and theory = strong and project =strong then Result=Good
6) If C.T = week and theory = strong and project =week then Result=Bad
7) If C.T = week and theory = week and practical =strong then Result=Good
8) If C.T = week and theory = week and practical =week then Result=Bad

IV. Conclusion

The discussion and examples given show that ID3 is easy to use. Its primary use is replacing the expert who would normally build a classification tree by hand. As industry has shown, ID3 has been effective. The studies and their implementation conducted here conclude that the decision tree learning algorithm ID3 works well on any classification problems having dataset with the discrete values. Related to the research work it concludes that thus the classification tree built using ID3 algorithm is shown below. It tells if the weather was amenable to play. In this seminar we studied the concept of data mining, decision tree, and ID3 algorithm with introduction, working with example. This algorithm is widely used in every field. It is easy to use and understand.

REFERENCES

[1] International Journal of Advanced Research in Computer Science and Software Engineering
[2] Tom M. Mitchell, (1997), Machine Learning, Singapore, McGraw-Hill.
[3] Paul E. Utgoff and Carla E. Brodley, (1990), 'An Incremental Method for Finding Multivariate Splits for Decision Trees', Machine Learning: Proceedings of the Seventh International Conference, (pp.58),Palo Alto, CA: Morgan Kaufmann.
[4] Wei Peng, Juhua Chen and Haiping Zhou, of ID3,' An Implementation Decision Tree Learning Algorithm', University of New South Wales, School of Computer Science & Engineering, Sydney, NSW 2032, Australia.
[5] Quinlan, J.R. 1986, Induction of Decision trees, Machine Learning.
[6] Ruijuan Hu,(2011),' Medical Data Mining Based on Decision Tree Algorithm', Computer and Information Science, Vol. 4, No. 5; September 2011, www.ccsenet.org/cis. [7] Anand Bahety,' Extension and Evaluation of ID3 – Decision Tree Algorithm'. University of Maryland, College Park.
[7] Vıctor.H.Garcıa, Raul Monroy and MaricelaQuintana,‟ Web Attack Detection Using ID3‟.
[8] http://homepage.cem.itesm.mx/raulm/pub/ifip06.pdf.
[9] Sonika Tiwari and Prof. Roopali Soni,' Horizontal partitioning ID3 algorithm A new approach of detecting network anomalies using decision tree', International Journal of Engineering Research & Technology (IJERT)Vol. 1 Issue 7, September – 2012.
[10] Mary Slocum,‟,Decision making using ID3,RIVIER ACADEMIC JOURNAL, VOLUME 8, NUMBER 2, FALL 2012
[11] Kumar Ashok, Taneja H C, Chitkara Ashok K and Kumar Vikas,' Classification of Census Using Information Theoretic Measure Based ID3 Algorithm' . Int. Journal of Math. Analysis, Vol. 6, 2012, no. 51, 2511 – 2518.
[12] L. Sathish Kumar and A. Padmapriya,' Prediction for Common Disease using ID3 Algorithm in Mobile Phone and Television'. International Journal of Computer Applications (0975 – 8887) Volume 50 – No.4, July 2012