International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2018 IJSRCSEIT | Volume 3 | Issue 6 | ISSN : 2456-3307

An Efficient Way for Scrutinizing the Job Seekers Data to Select a Right Candidate

K. Prema*[1], Dr. L. Venkateswara Reddy[2]
[1]Department of CSE, Sree Venkateswara Engineering College For Women, Tirupati, Andhra Pradesh , India
[2]Department of Information Technology, Sree Vidyanikethan Engineering College, Tirupati, Andhra Pradesh, India

ABSTRACT

Decision support systems play a vital role in business, science, medicine, markets, research and many more. The advances in analytical systems of data changed the way and pace of decision making process. Data mining in general and decision trees in particular are contributing a lot to decision support systems. In this paper efforts are made to introduce a simple and useful decision support system based on decision trees. Hypothetical data is considered to explain the methodology and elevate the power of the results. The proposed process can be extended to big data sets by availing the pruning techniques for decision tree construction.

Keywords : Decision Support System, Pruning Technique, Decision Tree, Dataset.

I. INTRODUCTION

All companies are continuously participating in business competitions for increasing their net profits. The growth rate of any company is dependent on many factors such as raw materials, man power management, infrastructure, the process of recruiting best employees, material management, marketing management, and so on. The recruitment process of a company is one of the toughest tasks. No doubt employees and their knowledge, combined effort, productivity, skills, work efficiency, positive attitude towards the development of the company are some of the most desirable features and employees are the basic building blocks of the company. The employee recruitment process must be strong enough to recruit the best emplyees based on their qualification, experience; job type, stream and personal intervene and so on. Entire job recruitment task is costly.

Recruitment process is basically divided into types. The first type is recruiting freshers through written test, personal verification, group discussion, interview and so on. The second type of job recruitment policy is selecting employees through automation process. Present study proposes a machine learning based system to automate the job recruitment process for experienced job applicants. Nowadays applications of machine learning techniques are spreading rapidly in diversified fields. Machine learning is a branch of Artificial intelligence and it uses many statistical techniques frequently and machine learning techniques are popularly used in data mining and artificial neural networks. Recruitment teams, Agencies, Government, Universities, Colleges, Schools and companies are facing many difficulties in recruiting suitable potential candidates for their respective jobs. Throughout the process decision making at each step and decision making at the final step is a challenging one. Hence, better tools or systems are needed to case decision making process during job recruitment process. Present study employees a powerful machine learning tool called decision tree classification for effective decision making during employee recruitment process.

Decision tree is a classification tool and it is more interpretable and reliable. It is more frequently used for classification process, sometimes it is called classification tree because it is also used for data classification. It follows divide and conquer strategy and it's time complexity is O(log n).

II. LITERATURE REVIEW

A decision support system (DSS) is a computer-aided information system that assists the decision making process for business or organizational decision-making activities. These systems serve the all levels of management and help to make right decisions. In many business contexts decisions play an important role. The areas like medicine, economics, marketing, logistics and education have been using decision support systems. Decision support systems (DSS) contributing a lot in assisting physicians as a part in medical decision making. There are many sub areas in medicine where decisions must be made timely and effectively on which the linked things may rely. Decision trees are suitable candidates for decision support systems. In [8] authors presented the comparison study about the effectiveness of decision support systems based on decision trees. Decision trees were constructed for four models. The results were compared from statistical analysis. The application predicted MRA in children. Authors concluded that the same approach can be easily used for other kinds of medical decision-making systems.

The earlier research in decision support systems attributed to traditional statistical methods. Authors in [3] used classification trees. In [1], authors made experiments using logistic regression. Some other methods used in literature include k-nearest neighbour [5], and simple neural network models [4][2][9], and cluster analysis [2][6].

In [7] authors explained C4.5 decision tree-based Decision Support System. The process is aided by entropy-based multi interval discretization model. The design made use of the components like database, models, knowledge, and dialogue. The proposed DSS model is a generic that can provide the functionalities of decision analysis required for C4.5.

III. PROBLEM DEFINITION

In any company recruiting the best employees is one of the most important, time taking and challenging activities of the company. Manual existing methods for job recruitment are costly in terms of time, money, usage of resources, written tests and so on. To overcome many of the existing problems present in the existing recruitment systems there is a need to design and develop a new automated system for efficient job recruitment. There are many advantages of the proposed system for job recruitment process. Some of the potential benefits of automated recruitment process are decreased time delays, costs, risk, manpower are some of the potential benefits of the proposed automation of job recruitment process.

IV. DECISION TREE

Decision tree is a non-linear data structure. It is the most important data structure in computer science. It is frequently used for data search operation and its time complexity of search operation O (log n). This algorithmic time complexity makes the decision tree data structure suitable to use in many real time applications. Because of this sub-linear time complexity decision tree is so popular and becomes powerful weapon in the hands of research people who are professional s of machine learning and data mining fields. Decision tree converges very fast it follows data processing in top-down manner with divide and conquer strategy. Interpretability property is the main advantage of the decision tree and because of this property it stood in front of the support vector machines, and artificial neural networks techniques even though these two are more accurate but less interpretable than the decision tree counterpart.

Decision tree contains two types of nodes called internal nodes and external nodes. At each internal node a decision must be taken and this decision is based on the result of impurity measure present at that particular node. At each decision of the internal node the flow fallows the next branch selection. External nodes are leaf nodes where class labels are presented. When the decision output at the leaf are yes or no type then it is a classification problem and when the result at the leaf node is a number then it is called regression tree. If the result at the leaf node is a number then it is called regression tree. Though support vector machines and neural networks are the two state of the art and the most classification accuracy methods but they are less interpretable than decision tree classifier for classification results.

V. ALGORITHM

Algorithm-1 Create_Tree (D)

Input
D: a data set
Output
T: a decision tree

1. entropy = findNode Entropy (D)
2. if (entropy < threshoed) then
3. change the current node into leaf node by labeling with the majority class label
4. endif
5. return
6. bsAttribute = findBestSplitAttribute (D)
7. foreach distinct value, I, of bsAttribute do
8. create a sub data set, Di, for each branch
9. Create_Tree (Di)
10. Endfor

Algorithm-2 find Best split Attribute (D)

Input
D: a dataset or a sub data set
Output
Bsa: best split attribute

1. minimum Etropy = 999999
2. for attribute x_i where i = 1 to n do
3. e = findEntropy (x_i)
4. if (e < minimum_Entropy) then
5. minimum_Entropy = e
6. bsa = x_i
7. end if
8. end for
9. return bsa

Algorithm-1 creates a decision tree with divide and conquers methodology for the given training dataset. At the beginning for the root node entropy is computed, If the computed entropy is less than the threshold value then the current node is changed to the leaf node by labeling the majority class label; otherwise best split attribute is computed and then possible branches at the best split attribute are created. The process is repeated for each branch consisting of the sub dataset.

Algorithm-2 finds best split attribute for a given dataset using information measure technique. For a given set of attributes of the training dataset entropy is computed for each attribute and the attribute with minimum entropy is selected for splitting the node. In addition to the entropy there exist many other information measuring techniques at each internal node of the decision tree. No one information measure is good in all the cases. In different cases different information measuring techniques are useful.

The process of decision tree construction continuous until all the instances in the node of a specific branch belongs to the same class or the node contains only

minimal set of instances, in which case change the node to leaf node. Tree will be grown until all the data instances are perfectly classified. Sometimes there is a need to prune the tree by using special pruning techniques. Once the tree creation is completed, it represents a set of if then rules. The number of if.. then rules is equal to the number of leaf nodes of the generated decision tree. Decision tree model represents a set of if.. then rules. Decision tree model is tested by giving test instance as input. The accuracy of the decision tree is computed as the total number test tuples that are correctly classified. The decision tree classifier classifies the test tuples according to the set of already generated if…then rules in form of decision tree classifier model. In the selected hypothetical dataset the last attribute selected is class label attribute. Its values are yes or no. it is only a two class problem. Attributes of the test tuple are attributes of the applicant applied for job. The developed decision tree classifier model displays the correct decision whether the new applicant is selected or not for the concerned job.

VI. DATA SET DESCRIPTION

The proposed algorithm is aimed to pick the right decisions among the available alternatives or options. A hypothetical dataset representing the details of job seekers is considered to go with the proposed process. The detailed data with all its attributes is presented in TABLE-1. The dataset contains only categorical attributes. A categorical attribute is an attribute which specifies category of its value.

Table 1. Job Seekers' Data

Qualification	Experience	Previous Platform	Current Platform	Class label
M.Tech(CSE)	5	java	R	1
M.Tech(ECE)	10	c	java	1
M.Tech(CSE)	10	c++	java	0
M.Tech(CSE)	10	c	java	1
M.Tech(CSE)	5	java	R	1
M.Tech(EEE)	5	c	java	0
M.Tech(CSE)	5	java	PYTHON	1
M.Tech(CSE)	10	c	java	1
M.Tech(EEE)	5	java	R	0
M.Tech(CSE)	10	c	java	1
M.Tech(CSE)	5	java	PERL	1
M.Tech(ECE)	10	c++	java	1
M.Tech(ECE)	5	c	java	0
M.Tech(CSE)	5	java	PERL	1
M.Tech(EEE)	10	c	java	1
M.Tech(EEE)	5	java	PERL	0
M.Tech(ECE)	10	c	java	1
M.Tech(CSE)	10	c	PYTHON	1
M.Tech(EEE)	5	java	R	0
M.Tech(ECE)	10	c++	java	1

Automation of job recruitment process is very much beneficial to many organizations to cope and withstand strongly for long term survival and

existence. The analysis, design and development of automated job recruitment system depend on many factors or attributes. When all the needed attributes are taken into consideration then the size of the real data set will be very high. To make the problem simple and easy to understand only a limited and important attributes are considered. Some of the most important attributes are – applicant qualification, experience, job type, company name of the previous work experience, and so on. The size of the data set must be large enough to mine and obtain sufficient good knowledge from the data. In general, the size of the data set must be high to derive useful and real inferences. In the present study only a limited set of tuples and attributes are taken into consideration. The selected data set is shown in Table 1.

VII. EXPERIMENTAL RESULTS

The proposed process consists of two algorithm components. The first algorithm constructs the decision tree representing the paths of classification. At each node the splitting attribute plays a vital role to divide the whole dataset. To find the best splitting attribute the second algorithm assists the first algorithm. For the simplicity purpose a dataset consisting of twenty job seekers is considered for the application of the proposed algorithms. The classification tree generated as a result of the process is depicted in Figure-1. The numbers in the leaf nodes represents the number of records assigned to that particular class label. For categorical attributes the branching factor of any internal node may be more than two also whereas branching factor of numeric attribute must be two only.

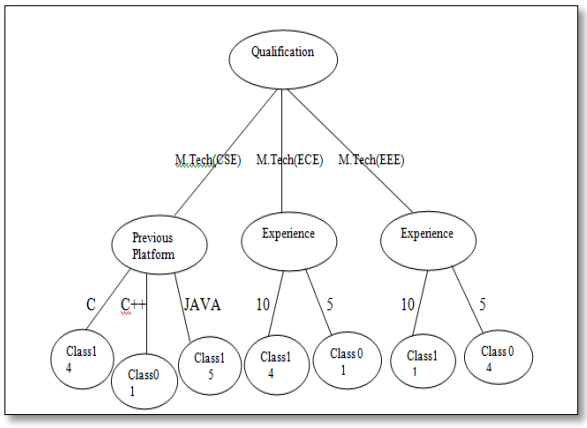

Figure 1. Decision Tree Classification.

The classification tree presented in Figure 1 guide the decision maker in deciding the right candidates to fill the job vacancy. The attribute values of the job seekers guide the classification process. The internal nodes of the decision tree present the attributes that guide the classification at node level. The leaf nodes show the result of the classification. Every path from the rot node to all leaves provides the rules for decision making.

The following sets of results show the process of execution giving the intermediate to the final results of the proposed decision tree classification process. Decision tree nodes are printed in the depth first search manner and depth first search printed results are shown for understanding purpose. With in the oval box the best splitting attribute is written and lines are drawn from splitting attribute to its children and these lines indicate branches generated at that particular internal node of the decision tree. Categorical attribute generates many branches from its corresponding internal node after splitting based on the distinct values of splitting attribute present at that internal node. The splitting criteria are different for the numerical attribute because distinct values of numerical attribute are many.

NODE INFORMATION

The root node:

Split_Attribute: Qualification
Number of children = 3
Number of branches at root node is
--> M.Tech(CSE)
--> M.Tech(ECE)
--> M.Tech(EEE)
Parent of the present node is null

Node_Type 0 (internal node)
Split_Attribute: PreviousPlatform
Number of children = 3
Number of branches in present node is = 3
--> c
--> c++
--> java
parent of the present node is Qualification
Node_Type 0 (internal node)
LEAF
Class_Label 1
Parent of the present node is PreviousPlatform
Node_Type 1 (external node)
LEAF
Class_Label 0
Parent of the present node is PreviousPlatform
Node_Type 1 (external node)
LEAF
Class_Label 1
Parent of the present node is PreviousPlatform
Node_Type 1 (external node)
Split_Attribute Experience
Number of children = 2
--> 10
--> 5
parent of the present node is Qualification
Node_Type 0 (internal node)
LEAF
Class_Label 1
parent of the present node is Experience
Node_Type 1 (external node)
LEAF
Class_Label 0
parent of the present node is Experience
Node_Type 1 (external node)
Split_Attribute Experience
Number of children = 2
--> 10
--> 5
parent of the present node is Qualification
Node_Type 0 (internal node)
LEAF
Class_Label 1
parent of the present node is Experience
Node_Type 1 (external node)
LEAF
Class_Label 0
parent of the present node is Experience
Node_Type 1 (external node)

VIII. CONCLUSIONS

There are two ways one can use decision tree. The first use is for classifying the data. In such a situation it is called classification tree. The second use is for regression. In this case the tree is called regression tree, which finds numeric output label for a given input data. Present study uses a decision tree for classification purpose. In the present study decision tree is trained and used for classification purposes by taking a hypothetical example. In the future same will be extended for large training data sets with suitable pruning techniques for scalability purpose. In the future numeric attributes are also considered and dynamic data classifications are also considered.

IX. REFERENCES

[1]. Baesens B., Setiono R., Mues C, Vanthienen J.(2003): Using Neural Network Rule Extraction and Decision Tables for Credit-Risk Evaluation, Management Science, Volume 49 , Issue 3, March 2003, Pp:312 – 329.

[2]. Chi G., Hao J., Xiu Ch., Zhu Z. (2001): Cluster Analysis for Weight of Credit Risk Evaluation Index. Systems Engineering-Theory Methodology, Applications, 10(1), Pp. 64-67.

[3]. Davis, R. H., Edelman, D. B., and Gammerman, A. J.(1992):Machine Learning Algorithms for Credit Card Applications,. IMA Journal of Mathematics Applied in Business and Industry (4), Pp: 43-51

[4]. Desai V.S., Crook J.N., Overstreet G.A. Jr(1996): On comparison of neural networks and linear scoring models in the credit union environment. European Journal of Operational Research, 95(1), Pp: 24-37.

[5]. Henley W.E., Hand D.E.(1997). Construction of a k-nearest neighbor credit-scoring system. IMA Journal of Mana-gement Mathematics, 8, 305-321.

[6]. Lundy M.(1993): Cluster Analysis in Credit Sco ring. Credit Scoring and Credit Control. New York: Oxford Uni-versity Press.

[7]. Rajesri Govindaraju and Dipta Mahardhika," Web-based Decision Support System using C4.5 decision tree algorithm, International Conference on Advanced Computer Science and Information Systems, IEEE 2011.

[8]. Spela Hleb Babic, Peter Kokol, Vili Podgorelec, Milan Zorman, Matej Sprogar, and Milojka Molan Stiglic, The Art Of Building Decision Trees,Journal of medical systems,Plenum press Vol.24,Num.1,pp.43,2000.

[9]. Steenackers A.,and Goovaerts M.J.(1998): A credit scoring model for personal loans. Insurance Mathematics & Economics, 8, 31-34.

[10]. West, D.(2000): Neural Network Credit Scoring Models, Computers and Operations Research, vol. 27, no. 11-12, pp. 1131.