Global Journals LaTeX JournalKaleidoscope™
Artificial Intelligence formulated this projection for compatibility purposes from the original article published at Global Journals. However, this technology is currently in beta. *Therefore, kindly ignore odd layouts, missed formulae, text, tables, or figures.*

Decision Tree Construction: A Continues Label Support Degree Based Approach

N.Madhuri[1], T.Nagalakshmi[2] and sujatha dandu[3]

[1] Auroras Technological and Research Institute (JNTU)

Received: 25 October 2011 Accepted: 22 November 2011 Published: 2 December 2011

Abstract

Data mining and classification systems utilize decision tree algorithms since they proffer rapid speediness, advanced exactness and also simple organization of those algorithms. An ideal decision can be built only when the appropriate attributes are chosen. This paper focuses on throwing light on choosing characteristics based on the theory of attribute support degree on account of which a unique decision tree construction algorithm is proposed on the basis of rough set and granular computing theory. It is henceforth proved that the decision tree proposed by the new approach yields far more better results in terms of precision and consistency as compared to the decision trees yielded by ID3, C4.5 and DTBAS.

Index terms— rough set: decision tree: granular computing: attribute support degree: attribute selection.

1 INTRODUCTION

ecision sets can be denoted using tree structures with the help of decision tree which is a unique, spontaneous, data illustration scheme and also a competent classifier. Quinlan et al [1] proposed ID3, decision tree algorithm and hence has been persistently augmented which have been advanced to C4.5 [2]. The preeminent attribute is chosen as the existing attribute which is then recursively inflates the decision tree branches unless and until the conditional statement is achieved, which ultimately makes use of top-down greedy algorithms. There are different classification schemes that can be achieved concerning different solutions which poses two issues [3] in decision tree construction. Choosing characteristics for crafting new branches in a tree is one issue while the other one is pruning which is all about omitting and decreasing the tree. DTBAS [10] considered the Assortment of attribute as main concern, which was refined and improved by considering assortment of continuous labels also that is discussed in this paper.

Z. Pawlak et al [4] recommended rough set theory which is an expansion of set theory for studying intelligent systems which is followed up by inadequate and partial data information. There is a thriving submission of the rough set theory in the disciplines of data mining, pattern recognition, machine learning, decision analysis etc in recent times. Models are categorized into various resembling classes that houses imperceptible objects in terms of few attributes. Issues pertaining to feature selection, data reduction and pattern extraction can be amicably taken care of such that it can liberate the system of redundant data in systems containing null values or missing data.

Lin et al [5] proposed the expression of Granular Computing which spans itself covering all aspects of concerning theories, tactics, practices and means essential in solving a problem that makes use of granules. Granular Computing has witnessed vast inputs from different practices such as fuzzy sets, rough sets, shadowed sets, probabilistic sets etc.

A crucial step that needs to be taken care of while building a decision tree is choosing characteristics of nodes of a tree that houses minimum number of branches. Decision tree based on continuous label support degree (DTBLSD) algorithm is introduced which is considered as a splitting criterion on account of rough set theory and granular computing. Trial results have approved the usage of DTBLSD algorithm that assures and provides uncomplicated structures and superior categorization accuracy.

The rest of the paper is organized as follows: Section 2 discusses concepts relevant to rough set theory and granular computing. Section 3 gives a basic introduction to our new method and presents a simple example. Experimental comparison of the proposed method with 103 and C4.S is given in section 4. The final section concludes the research work of this paper.

2 II. BASIC CONCEPT

Few fundamental concepts of rough set theory [6,7] and granular computing [8] are first initiated for ease of demonstration. Definition 1 (Information System) : An information system can be labeled as wherein U is a finite set of object known as the universe; , which is a non-vacant finite group of attributes; C and D depict set of condition and decision attributes respectively as also , which says that is a value set of the attribute a and Where is the cardinal number of .(I) 2011 December (, , ,) S U A V f = A C D = U , a v v a A = ? ? U a v

When in an equivalence relation R, the granular degree of R reaches the minimum size ; When R is a domain relation, the granular degree of R attains the maximum size Definition 5 : Assume R is knowledge of repository , the granular degree of basic knowledge is defined asIII.

3 PROPOSED ALGORITHM

This section aims at familiarizing the algorithm of building a decision tree on the basis of attribute support degree.

4 a) The Principle of Label and Attribute Selection

The label that represents least average uncertainty is supposed to be chosen as the test label and then choose attribute with less uncertainty as test attribute from the class represented by the selected label, which because it makes apt decisions when compared to existing test attribute selection in different decision tree algorithms. (Attribute Support Degree) Let be an information system is a label contains subset of attributes represented as . Attribute support degree can be denoted as follows based on the definitions mentioned above.

5 Where

denotes the cardinal number of .

6 Global

I) 2011 December (f U A V = × ? (,) a f u a V ? a A ? , u U ? (( )) X BND X (, , ,) S U A V f = R A ? X U ? ( ) R X ( ) R X ( ) BND X ( ) R X Ji(x), ( ) R X ( ) BND X ( ) { | [ ] }....(1) R R X x U x X = ? ? ( ) { | [ ] }...(2) ( ) ( ) ( )....(3) R R X x U x X BND X R X R X ? = ? ? ? = ? (, , ,) S U A V f = P A ? ( ) {(,) | , (,) (.)....(4) a IND P x y U U P f x a f y a = ? × ? ? = ( ) U IND P (,) K U R = R U U ? × ( ) GD R 2 | | | | ( ) ...(5) | | | | R R GD R U U U = = × | | R R U U ? × 2 | | / | | 1/ | | U U U = 2 2 | | / | | 1 U U = (,) K U R = , 1 2 / { , ,..., } n U R X X X = 2 1 2 | | ( ) ...(6) | | n i i i x GD X U = = ? 1 2 { , ,.... } ls a a a | | 1 ( ) ( ) ....(7) | | l i i uc uc a avg l l = = ? ( ) uc avg l l l ( ) i uc a i a l (, , ,) S U A V f = , A C D = U l Q C ? ( ) | ( ) | (,) ...(8) ( ) ( ) GD Q D IND QUD S Q D GD Q IND Q = = U | ( ) | IND Q ( ) IND Q U U ? ×

Decision Tree Construction: A Continues Label Support Degree Based Approach

. Definition 5 states that whenever we get the relations among them, namely, when GD(R) is smaller, the distinguishable degree is stronger and S(Q,D) is greater, thereby Q is better sets of test attribute of D. On the contrary, the smaller S(Q,D) is, the worse we get Q as sets of test attribute of D.

7 b) The Description of DTBLSD

The basic notion of DTBLSD expresses the point that whenever label support degree with association of label level attribute support degree is made use of as a customary for choosing a test attribute concerning every node in the decision tree. The attribute reduction set assists in selecting a condition attribute that possesses the highest degree of label level attribute which can be put to use at the root of the decision tree. There will be a testing of the remaining condition attributes on each and every branch of the root node S (Q,D) D using Q can be estimated with the help of a measure addition of new sub-trees to every division until the leaf is reached.

According to the above idea, using the S(Q,D) as the splitting criterion, we propose our algorithm DTBLSD. Current sample set is depicted by T, set of labels depicted by L, condition attribute set of a label is depicted by the . depicts the number of attributes in the condition attribute set of label . All attributes of the condition attribute set are discrete and continuous values are discretized by continuous labeling. Following are the specific steps of the algorithm.

Algorithm : A decision tree is created by DTBLSD (T, attribute list) that using the given training data. The top-down recursive divide and conquer approach for construction of a decision tree wherein the recursion related division takes place only when any one criterion mentioned below is gratified. A common class contains :

8 Input

1. AII specimens for a specific branch which restores a leaf that is termed with the concerned class.

Here, a large case of voting is provisioned to change the present working node into a leaf that is termed with the concerned class that in in demand from amongst various specimens. 2. In addition, there are no more specimen test attributes and the class division specimens can be placed wherein a leaf is generated and termed with the most featured class in specimens.

IV.

9 EXPERIMENTS a) Example Analysis

Table **??** showcases a data tuple training group originated from All Electronics customer records that are implemented using polic mentioned in reference (6). The first step is to estimate the degree of attribute support for each situational attribute or characteristic. ( )

10 U IND al

= { {l,2,8,9, 11 },{3,7,12,13},{

11 U IND az

= { {1, 2, 3, 13},{4, 8, 10, 11, 12, 14}, {

(U IND a = {{1, 2, 3, 4, 8, 12, 14}, {5**3)**

12 U IND a

= { {l,3,4,5,8,9,10,13},{2,6,7,11,12,14}};

( ) Table **??** : Training data tuple from the AIIElectronics customer database Notations used in example descriptions: Age???, Income???, Student???, Credit???, Buy?? is selected as the min root of the decision tree and is tagged with age since is the maximum extent of a degree from amongst all the condition attributes as also various number of divisions which are branched in reference to a range of different attributes. In case where age=1, all the specimens that are grouped into this should belong to the same class and hence a leaf should be generated at the end of every division and should be tagged with d=yes. The figure above depicts the final decision tree that is built by DTBLSD.U IND d = { {1 ,2,**6,8,14 },{3,4,5,7,9, 10, 11,12, 13}}; (,)**

13 b) Experimentl Comparison

Experimental comparison of DTBLSD with respect to ID3 [1], C4.5 [2] and DTBAS [10] is discussed in this section. The real datasets that are used in this are approved from University of California, Irvine (UCI), and is known as the machine learning database repository where C++ design language is implemented to form the requisite algorithm. WEKA 3.7 is used for successful accomplishment of ID3 [1] and C4.5 [2] which is a compilation of machine learning algorithms used for data mining generated and procured by **??**rank

14 CONCLUSION

The paper first focuses on explaining the basic notion of label support degree and attribute support degree [10] and selecting it as a basic decisive factor on the basis of degree of involvement between condition attribute and decision attribute accordingly where a unique decision algorithm tree based on continuous label support degree and label level attribute support degree (DTBLSD) is recommended. Accordingly a suitable methodology is devised which is flexible enough to accommodate and provides lower complexity and high level of accuracy as compared to other algorithm generating methods. A disadvantage identified in [10] is issues pertaining to adjustment with adaptability of samples, which has been overcome successfully in our model. [1] [2]

[1] © 2011 Global Journals Inc. (US)

[2] Global I) 2011 December (

6

Figure 1: Definition 6 :

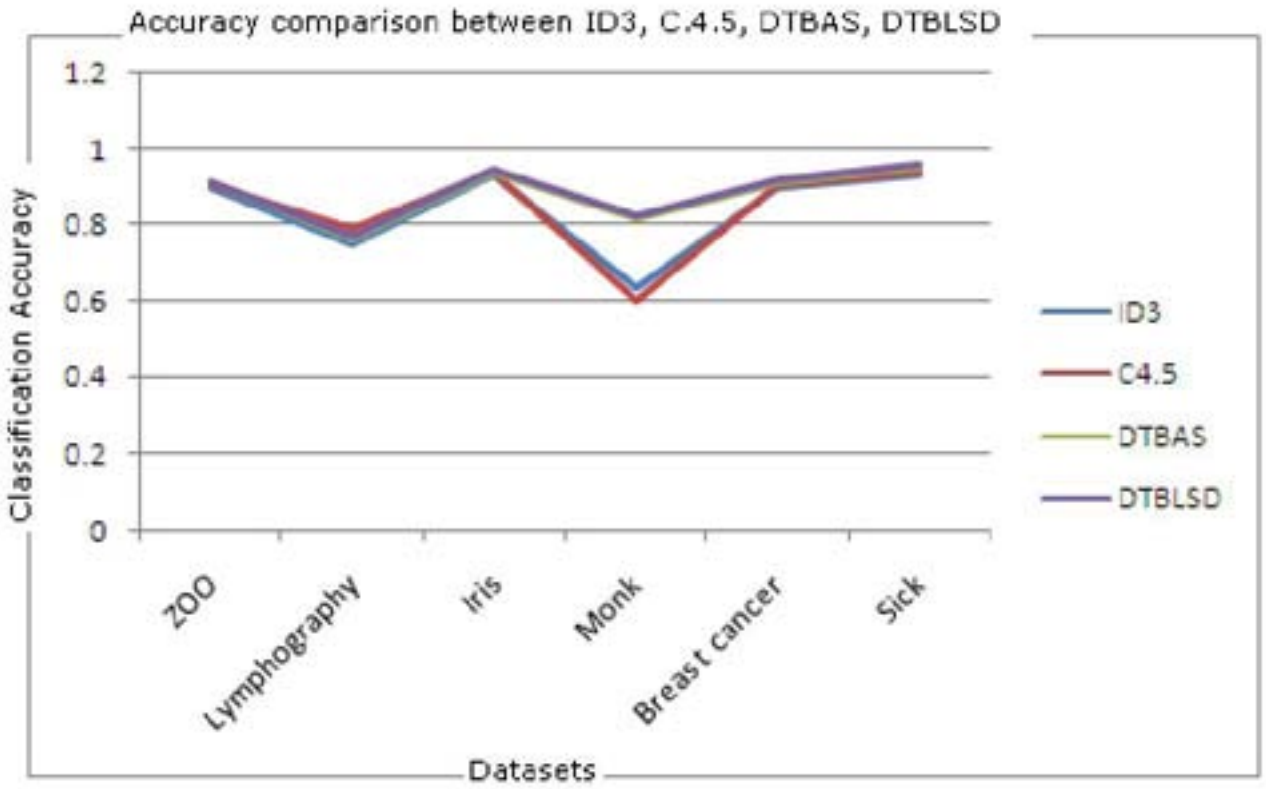

Figure 2:

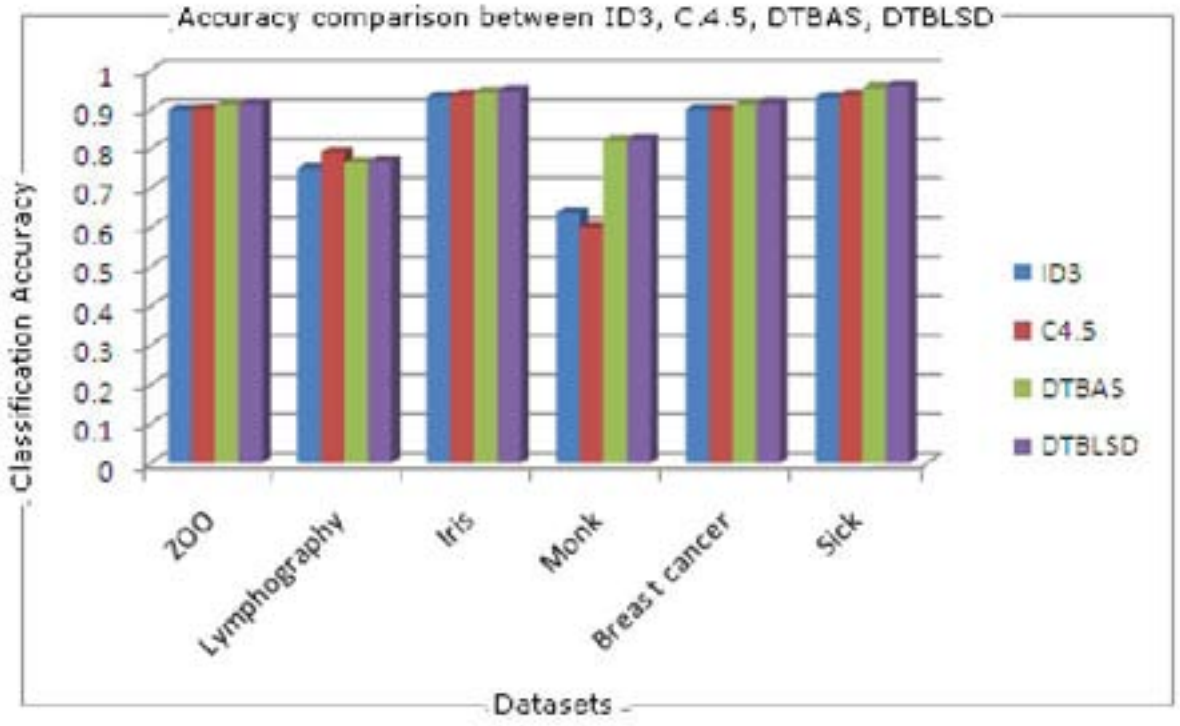

Figure 3:

	U	age	income	student	credit	
	1	<=30	high	no	fair	
	2	<=30	high	no	excellent	No
	3	31<=age<=40	high	no	fair	
	4	>40	medium	no	fair	
	5	>40	low	yes	fair	
2011	6 7 8 9 10	>40	low low	yes yes	excellent	No excellent
De-		31<=age<=40	medium	no yes		
cem-		<=30 <=30	low	yes		
ber		>40	medium			
	11	<=30	medium	yes	excellent	Yes
	12	31<=age<=40	medium	no	excellent	Yes
	13	31<=age<=40	high	yes	fair	
	14	>40	medium	no	excellent	No
	1 a					

1 (,) S a d

I)

(

Global hat involved the 10 fold cross estimation to calculate classification authenticity. All experiments were

Figure 4:

II

Dataset	ID3	? ? C4.5	DTBAS	DTBLSD
ZOO	0.899	0.901	0.911	0.914
Lymphography	0.75	0.791	0.765	0.767
Iris	0.932	0.936	0.943	0.947
Monk	0.637	0.6	0.821	0.824
Breas t cancer	0.9	0.9	0.912	0.916
Sick	0.931	0.937	0.955	0.959

Figure 5: Table II :

[Ding] , B S Ding , YQ .

[Frank et al.] , E Frank , M Hall , Weka . http://www.cs.waikato.ac.nz/ml/weka

[Zheng and Zang ()] 'A New Decision Tree Algorithm Based on Rough Set Theory'. S Y Zheng , Zang . *Proc. Asia-Pacific Conference on Information Processing*, (Asia-Pacific Conference on Information essing) 2009. p. .

[Yong and Zhou ()] 'An algorithm of decision tree construction based on attribute support degree'. Jianping Yong , ; Jun Zhou . *Educational and Information Technology (ICEIT)*, 2010.

[Chen et al. ()] 'Constructing a Multi-Valued and Multi-Labeled Decision Tree'. Y L Chen , C L Hsu , S C Chou . *Expert Systems with Apllications* 2003. 25 p. .

[December Decision Tree Construction: A Continues Label Support Degree Based Approach] *December Decision Tree Construction: A Continues Label Support Degree Based Approach*,

[Lin ()] 'Granular Computing II Infrastructure for AI-Engineering Examples, Intuitions and Modeling'. T Y Lin . *Proc. 2006 IEEE International Conference on Granular Computing*, (2006 IEEE International Conference on Granular Computing) GrC 2006. 2006. p. 2011. (I))

[Han and Kamber ()] J W Han , M Kamber . *Data Mining Concepts and Techniques*, 2001. Morgan Kaufmann Publishers.

[Quinlan ()] 'Improved Use of Continuous Attributes in C4.5'. J R Quinlan . *Journal of Artificial Intelligence Research* 1996. p. .

[Quinlan ()] 'Induction of Decision Trees'. J R Quinlan . *Machine Learning, voLl*, 1986. p. .

[International Conference on (2010)] *International Conference on*, Sept. 2010. 2 p. .

[Miao and Wang ()] D Q Miao , G Y Wang . *Granular Computing: Past, Now, Future[M), Science publishing house*, 2007.

[Qing Lin] *Qing Lin*, (Zongzhuan Ding)

[Pawlak ()] 'Rough Sets'. Z Pawlak . *International Journal of Information and Computer Science, voLlI* 1982. p. .

[Seifi et al.] *Twins Decision Tree Classification: A Sophisticated Approach to Decision Tree*, F Seifi , H Ahmadi , M Kangavari .