British Journal of Mathematics & Computer Science
8(6): 418-432, 2015, Article no.BJMCS.2015.175
ISSN: 2231-0851

SCIENCEDOMAIN *international*
www.sciencedomain.org

SCIENCEDOMAIN

New Approaches for Decision Making in Information Systems via Decision Diagrams

A. S. Salama[1,2*]

[1]*Department of Mathematic, Faculty of Science, Tanta University, Tanta, Egypt.*
[2]*Department of Mathematic, Faculty of Science and Human Studies, Shaqra University, Dawadmi, Saudi Arabia.*

Article Information

DOI: 10.9734/BJMCS/2015/17456
Editor(s):
(1) Vyacheslav Pivovarchik, Dept. of Applied Mathematics and Computer Science, South-Ukrainian National Pedagogical University, Ukraine.
(2) Tian-Xiao He, Department of Mathematics and Computer Science, Illinois Wesleyan University, USA.
Reviewers:
(1) Anonymous, Jordan.
(2) Harish Garg, Thapar University, Patiala, India.
(3) Anonymous, Poland.
Complete Peer review History: http://www.sciencedomain.org/review-history.php?iid=1035&id=6&aid=9099

Review Article

Received: 14 March 2015
Accepted: 17 April 2015
Published: 04 May 2015

Abstract

In this paper, new approach for data representation using decision diagrams in information systems are studied. We indicate data using binary decision diagrams and ID3-Decision tree learning algorithm to reduct information systems. Some algorithms are built and developed to refinement the process of decision making in information systems. Illustrative figures and examples are presented and real life application examples are given.

Keywords: ID3 learning algorithm; data representations; decision making; decision diagram; information system.

1 Introduction

The approach of BDD is used for binary data representation and it is very efficient for Boolean function representation such that relations, relational algebra, midterms, cubes, sets, sets of sets, state machines, partitions, set systems, graphs, covering, tables and matrices, but it is not understand for non-technical people.

There exist many methods to do decision analysis. Each method has its own advantages and disadvantages [1-17].

**Corresponding author: dr_salama75@yahoo.com, asalama@su.edu.sa;*

For data in an information system, the acquisition of knowledge and reasoning may involve vagueness, incompleteness, and granularity. In order to deal with the incomplete and vague information in classification, concept formulation, and data analysis, researchers have proposed many methods other than classical logic, for example, rough fuzzy sets [18-20], rough set theory and its generalizations [21-24], computing with words, granular computing, formal concept analysis, quotient space theory, and computational theory for linguistic dynamic systems. The advantage of the rough set method is that it does not need any additional information about the data, like probability in statistics or membership in fuzzy set theory. The main idea of applications of the rough theory in knowledge discovery comes from W. Ziarko work [25-30]. Many researchers made contributions to this theory. Applications of the rough set and fuzzy set theories can be found in [31-33].

We often get together decision making problems in our daily life or working surroundings. From time to time it is very difficult for us to make good decisions. In observe, we usually use our past understanding as a form of performing experiments costs time and money. Providentially, the developments of computer technologies and automatic learning techniques can make this easer and more resourceful. There exist many approaches of decision making techniques, such as decision trees or decision diagrams, artificial neural networks and Bayesian learning. In this paper, we focused on the decision diagrams especially on Binary Decision Diagram (BDD), Binary Decision Diagram Information System (BDDIS) and Decision Tree Learning Algorithm (DTLA) [34-39].

2 Decision Making Using Binary Decision Diagrams

The main aim of this section is to describe how data represents by Binary Decision Diagrams (BDD) and how to obtain reducts using it.

A BDD is a rooted , directed acyclic graph, with two sink nodes labeled by constants 0 and 1, and whose internal nodes has two outs. Each internal node v is labeled by a binary variable $\mathrm{var}(v) = x_i \quad ,i = 1,...,n$. It is two outgoing edges which are labeled by 0 and 1, and are usually depicted as dotted and solid lines, respectively. Let the two corresponding successor nodes of v be denoted by $low(v)$ and $high(v)$, respectively.

A BDD is ordered (OBDD) if, on all paths through the graph, the variables respect a given liner order $x_1 \prec x_2 \prec ... \prec x_n$ i.e., for each edge leading from a node labeled by x_i to a node labeled by x_j, it holds that $x_i \prec x_j$. We can say that variable must appear in ascending order along all paths.

Up to here, OBDD are not yet uniquely determined for each function. However, by further restricting the representation, a canonical form of BDDs. We observed that reduced OBDD (ROBDD) are a canonical representation of Boolean functions.

The ROBDD representation of any function is uniquely determined (up to isomorphism), so that several properties (e.g. functional equivalence, constant functions, etc.) become easily testable. Conceptually, a reduced diagram can be interpreted as the result of the repeated application of three types of transformations. One reduction rule is merge equivalent leaves. This rule means use only one terminal label for 0 and one terminal label for 1 and redirect all lines from n to the respective node.

As we mentioned before the BDD is very efficient method for represents binary data, we will introduce an example from our daily life and we will try to make a reduction for container data using ROBDD.

Example 2.1 According to Table 1 in some employee applications we want to select from 19 applicants the good ones. To finish this selection we consider the following questions system used to predict whether someone will final select (Final) based on some tests in science branches Mathematics (Math.), Physics (Phy.), Computer skills (Comp.) and Chemistry (Chem.).

Table 1. Employee information system

Employee	Comp.	Math.	Phy.	Chem.	Final
a_1	weak	good	weak	weak	good
a_2	good	weak	good	weak	weak
a_3	weak	weak	good	good	good
a_4	weak	weak	good	weak	weak
a_5	weak	weak	weak	good	weak
a_6	good	weak	weak	good	good
a_7	good	weak	weak	weak	good
a_8	weak	good	good	good	weak
a_9	weak	good	good	weak	weak
a_{10}	good	good	good	weak	good
a_{11}	good	good	weak	good	good
a_{12}	weak	weak	weak	weak	good
a_{13}	good	good	weak	weak	good
a_{14}	weak	good	weak	good	good
a_{15}	good	weak	good	good	good
a_{16}	good	good	good	good	good

By using BDD we consider the attribute value good replaced by 1 and weak by 0 , Fig. 1 is the BDD for the data given in Table 2 using the order Comp. < Math. < Phy. < Chem. Fig. 2 is the simplification BDD reduction of Fig. 1.

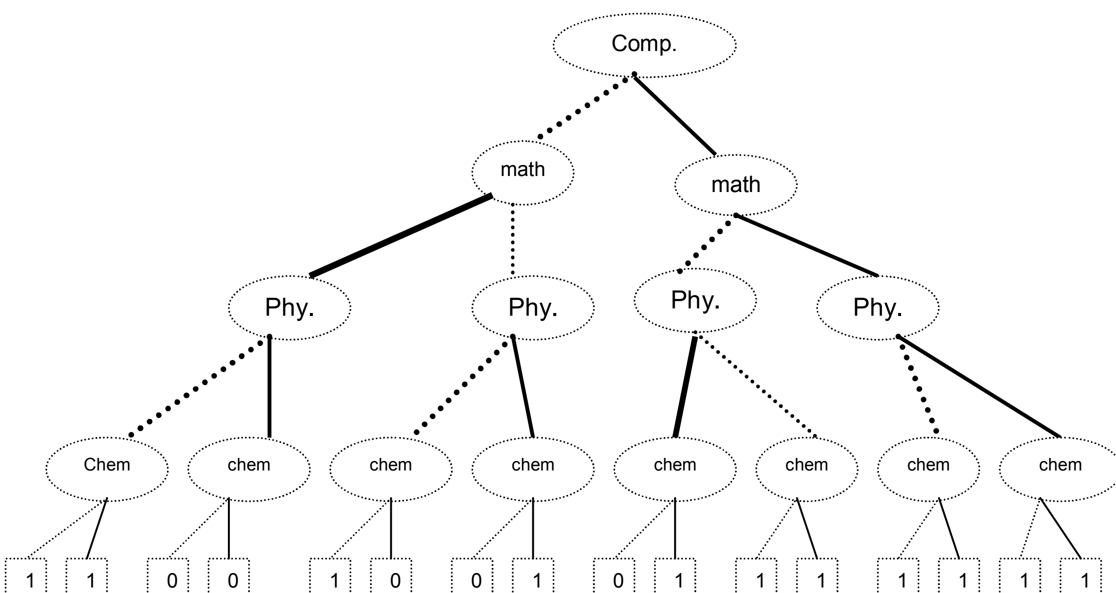

Fig. 1. BDD of employee information system

According to reduction algorithm we have the following reduct (Figs. 3, Fig. 4).

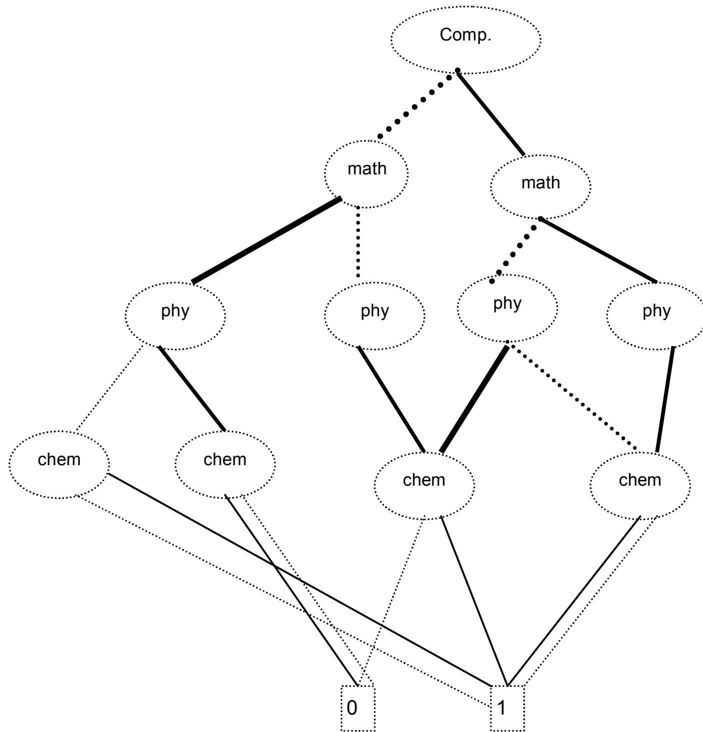

Fig. 2. Reduction of employee information system

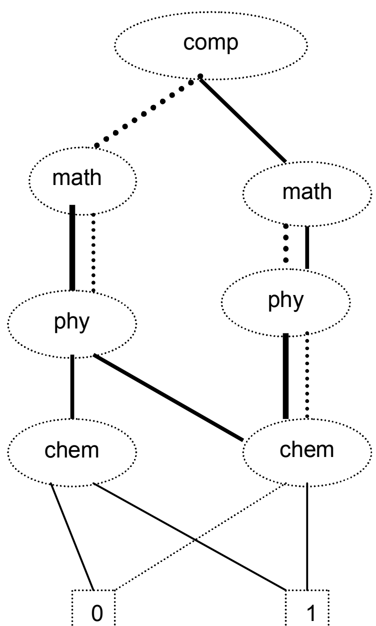

Fig. 3. Core of employee system

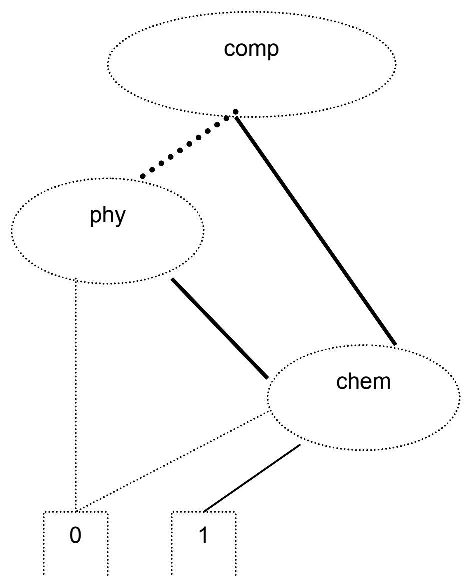

Fig. 4. Final BDD reduction of employee system

We use BDD to obtain reducts of our daily life binary data. But for the normal user, he cannot understand what is the meant by the final graph. So we can delete the redundant attributes from the table and rewrite it. By looking at the final graph the attribute Edu. can eliminated.

3 Decision Making Using BDDIS

A BDDIS has one or more root nodes each unique root node has a sub-tree that may share nodes with other objects and represents one or more indiscernible objects. The level root containing nodes which is defined as the object-top level; a top level element is referenced by $T(x)$, $x \in U$.

The following information system can be represented by BDDIS using object-top level.

Table 2. BDD information system

U	a	b	c	d
u_1	13	25	35	26
u_2	13	25	17	18
u_3	12	36	17	47
u_4	12	36	17	47

The following diagram represents the data in Table 2 using BDDIS (Fig. 5).

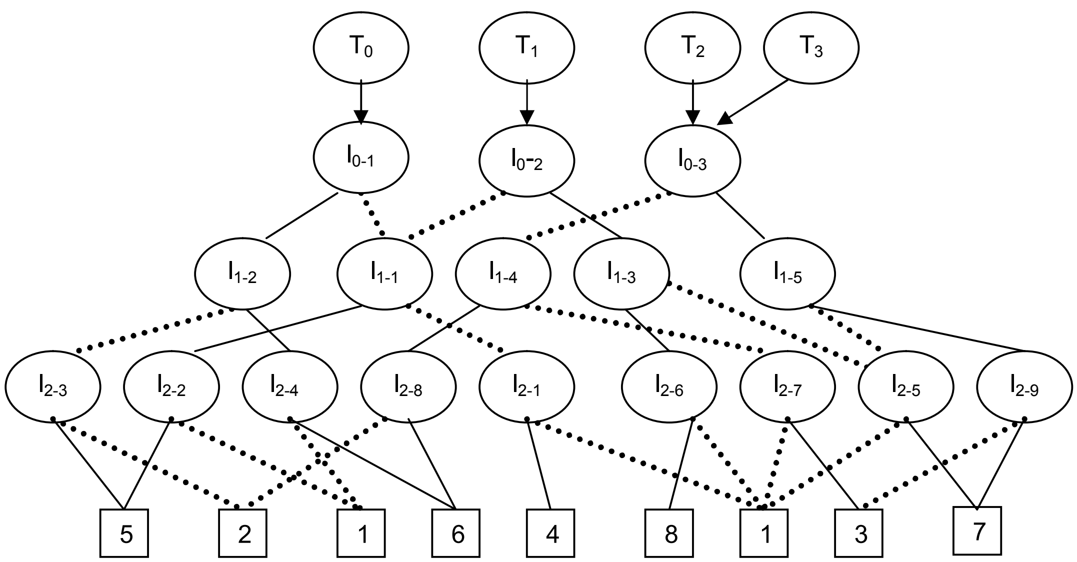

Fig. 5. Object top-levels of Table 2

Each column within the information system is represented by a path composed of a unique series of l_o and /or h_i branches arriving at attribute top level node.

In Fig. 5 the first column of Table 2 is represented by the path $l_o - l_o$, the second column is represented by the path $l_o - h_i$ and the third is represented by the path $h_i - l_o$. Each node is labeled with $l_n - x$ for level number n and node number x .

Two methods can be taken to reduce the size of the BDDIS. One can reduce the BDDIS by applying the ROBDD rules described before. Alternatively, one can create the BDDIS using a coded information system in which each attribute value V_a of the original information system is coded to value $0,1,2,...,k-1$, $k=|V_a|$; here each unique coded value represents a unique attribute values.

We can reduce the data of Table 2 by using the coded method. The corresponding coded data is given in Table 3.

Table 3. Coded information system of Table 2

U	*a*	*b*	*c*	*d*
u_1	0	0	0	0
u_2	0	0	1	1
u_3	2	2	1	2
u_4	2	2	1	2

We can represent the coded table by attribute top level using BDDIS. By make a coded table we observe that the attribute values of attribute a is the same for the attribute value for attribute b, then we can represent BDDIS without attribute b. the reduced graph of the coded data is shown in Fig. 6.

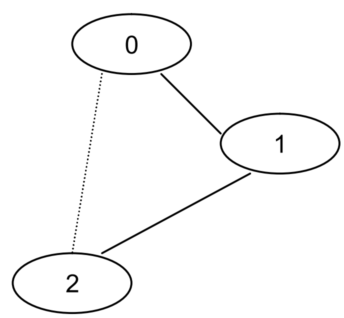

Fig. 6. Reduced graph of coded data in Table 3

4 New Approach for Decision Making Using Decision Trees

A decision tree (DT) 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 is commonly used for gaining information for the purpose of decision making. DT starts with a root node on which it is for users to take actions. From this node, users split each node recursively according to DT in which each branch represents a possible scenario of decision and its outcome.

4.1 Decision Tree Learning Algorithm:

function DECISION –TREE-LEARNING (examples, attributes, default)
returns a decision tree
inputs: examples, set of examples, attributes, set of attributes
Default, default value for the goal predicate
If (examples is empty) *then return* default
else if (all examples have the same classification) *then return* the classification.

else if (attributes is empty) *then return* MAJORITY-VALUE(examples)
else
best $\leftarrow$ CHOOSE-ATTRIBUTE(attributes, examples)
tree $\leftarrow$ a new decision tree with root test best
for each value v_i of best *do*
example s_i $\leftarrow$ {elements of examples with best = v_i }
subtree $\leftarrow$ DECISION-TREE-LEARNING(example s_i , attributes- best,
MAJORITY-VALUE(examples) add a branch to tree with label v_i and subtree sub-tree.
end

5 Remarks about ID3-decision Tree Learning Algorithm

J. R. Quinlan in [40-43] have been developed Interactive Dichotomize 3 (ID3) algorithm for a simple decision tree learning by based on the concept learning system (CLS) algorithm. It builds the decision tree from some fixed or famous figurative data in order to classify them and foresee the classification of new data. The data must have more than a few attributes with different values, this data also has to belong to varied predefined separate classes (i.e., yes/no).

The entropy measure is defined as follows:

$H(S) = \sum_{i=1}^{N} - p_i^+ \log_2(p_i^+) - p_i^- \log_2(p_i^-)$ where p_i^+ is the ratio of the positive examples S^+ with respect to the set of all examples S i.e., $p_i^+ = |S^+| / |S|$. And p_i^- is the ratio of the negative examples S^- with respect to the set of all examples S i.e., $p_i^- = |S^-| / |S|$, where $|S|$ is the cardinality of S .

For example, we assume that a sample set S has 14 members altogether ($|S| = 14$), including 9 positive ($|S^+| = 9$) and 5 negative examples. Then the entropy of S is:

$$H(S) = -(9/14)\log_2(9/14) - 5/14\log_2(5/14) = 0.94 .$$

Below we discuss the entropy in the special case of the Boolean classification.

- If all the members of a set S belong to the identical kind (have the same decision value), then the entropy is null. That means that there is no classification uncertainty.
- If the number of the positive examples equals to the number of the negative examples, then entropy takes its maximum value. In this case, the classification uncertainty is maximum.

These results express separately that the same set has no uncertainty (the decision is clear); or it is 100% uncertain for decision making. If the number of positive examples is not equal to the number of the negative examples, then the value of entropy is measured between 0 and 1.

To carry on the attribute expansion, which is based on the data of this sample set, we must define a standard measure called information gain. An information gain of an attribute is the final information content, which is a result of the reduction of the sample set entropy after using this attribute to divide the sample set. The definition of the information gain of an attribute a relates to the sample set S is:

$G(S,a) = H(S) - \sum_{v \in V_a} |S_v| / |S| \bullet H(v)$, where $S_v = \{x \in S : V_a(x) = v, v \in V_a\}$ and $H(v)$ is the entropy of the $a-$attribute value v.

Example 5.1 In this example, people decide to go to city patch or stay at home according to the weather (w ={Hot (w1),Mild (w2),Cool (w3)}), rain (r = {Strong (r1), Weak (r2)}) and traffic (t={Long (t1), Short (t2)}) situation as shown in Table 4. Just like Table 5, the set $S = \{s_1,...,s_{14}\}$ contains in total 14 examples (9 positive decisions and 5 negative ones). We can calculate the information gain of the attribute *r* as follows:

$G(S,r) = H(S) - 8/14H(r2) - 6/14H(r1)$, where
$H(r2) = -6/8\log_2(6/8) - 2/8\log_2(2/8)$ and $H(r1) = -3/6\log_2(3/6) - 3/6\log_2(3/6)$.
Then $G(S,r) = 0.048$.

Table 4. Weather information system

S	w	r	t	Decision
s_1	w1	r2	t1	No
s_2	w1	r1	t1	No
s_3	w1	r2	t1	Yes
s_4	w2	r2	t1	Yes
s_5	w3	r2	t2	Yes
s_6	w3	r1	t2	No
s_7	w3	r1	t2	Yes
s_8	w2	r2	t1	No
s_9	w3	r2	t2	Yes
s_{10}	w2	r2	t2	Yes
s_{11}	w2	r1	t2	Yes
s_{12}	w2	r1	t1	Yes
s_{13}	w1	r2	t2	Yes
s_{14}	w2	r1	t1	No

Using the same principle, we can calculate the information gain to the remaining attributes and we have: $G(S,w) = 0.029$ and $G(S,t) = 0.94$.

As a result, we may choose the attribute of highest information gain to used to expand the tree. We consider the attribute traffic as the root for the decision tree (Fig. 7).

From Fig. 7 and using ID3 algorithm the following rules will be generated:

- If *t* is long and *r* is strong then people will choose to stay at home.
- If *t* is long and *r* is weak then people will choose to go to city patch. In logical form $(t,t1) \wedge (r,r2) \Rightarrow (go\ to\ city\ \ patch, No)$.
- If *t* is short and *w* is hot then people will choose to go to city patch. In logical form $(t,t2) \wedge (w,w1) \Rightarrow (go\ to\ city\ \ patch, Yes)$.

- If t is short and w is cool then people will choose to go to city patch. In logical form $(t, t2) \wedge (w, w3) \Rightarrow (go\ to\ city\ \ patch, Yes)$.
- If t is short and w is mild then people will choose to stay at home. In logical form $(t, t2) \wedge (w, w2) \Rightarrow (go\ to\ city\ \ patch, No)$.

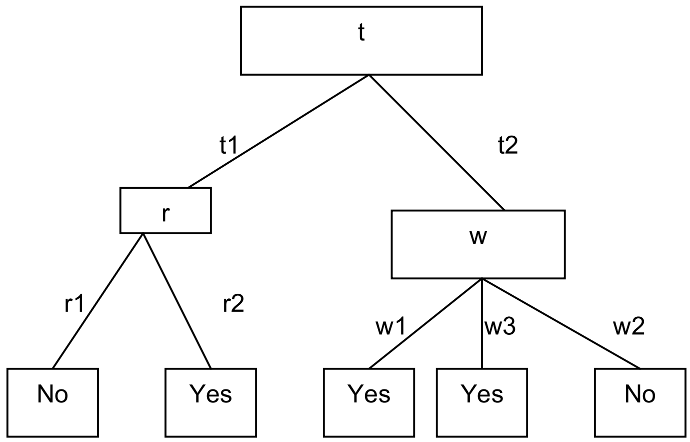

Fig. 7. Reduced graph of weather information system

In order to extend decision tree induction to a wider variety of problems, a number of issues must be addressed such as:

- Missing data: In many domains not all the attribute values will be known for every example. The values may not have been recorded, or they may be to expensive to obtain. First, given a complete decision tree, how should one classify an object that is missing one of the test attributes?
- Multi-valued attributes: When an attribute has a large number of possible values, the information gain measure gives an inappropriate indication of the attribute's usefulness.
- Multi-valued data: The multi-valued data is a generalization for a single valued one, when an attribute value has a number of valued. How we can represent this problem by the decision tree and how the information gain measure this case is an open problem.
- Continuous valued attributes: Attributes such height and weight have a large or infinite set of possible values. They are therefore not well-suited for decision tree learning algorithm.

We introduce the following example which studying the above cases.

Example 5.2 In this example, a company need some employees. It decided to accept the person if it has some languages, some scientific degrees, experience and background knowledge according to Table 5.

Table 5. Multi-valued information system

Person	Language	Degree	Experience	Background	Decision
P_1	{F,D}	{Bs, Ms, PhD}	Medium	Excellent	Accept
P_2	{H,R}	{Bs}	Low	Neutral	Reject
P_3	{F,D,S}	{Bs, Ms}	Low	Good	Reject
P_4	{F}	{Bs, Ms}	High	Neutral	Accept
P_5	{D}	{Bs}	Medium	Neutral	Reject
P_6	{F,D}	{Bs, Ms}	High	Excellent	Accept
P_7	{H,R}	{Bs}	High	Good	Accept

From the above table, $S=\{P_1,P_2,P_3,P_4,P_5,P_6,P_7\}$ is the set of examples. The set {Language, Degree, Experience, Background} is the condition attributes of the company and Decision is the decision attribute. The attribute Language has the value set {F, D, H, R, S} which is some languages in symbolic $V_{language}=\{F,D,H,R,S\}$. Also $V_{\deg ree}=\{Bs,Ms,Ph.D\}$, $V_{\exp erience}=\{low,medium,high\}$ and $V_{background}=\{neutral,good,excellent\}$.

Now, we need to represent this problem by using DTLA. If we consider the attribute values {F,D} of the attribute Language as a separately attribute value we can apply DTLA easily but the result is not good where we take more than one attribute value in the same class.

The entropy for the set S of all persons (S contains 4 positive examples and 3 negative examples) is:

$$H(S)=-4/7\log_2(4/7)-3/7\log_2(3/7)=0.985$$

Then the information gain of each attribute is:

$$G(S,language)=H(S)-2/7\,H(\{F,D\})-2/7\,H(\{H,R\})-1/7\,H(\{F,D,S\})-1/7\,H(\{F\})$$
$$-1/7\,H(\{D\})=0.985-0.2857=0.69928$$
$$G(S,\deg ree)=H(S)-1/7\,H(\{Bs,Ms,Ph.D\})$$
$$-3/7\,H(\{Bs\})-3/7\,H(\{Bs,Ms\})=0.985-3/7[-2/3\log_2(2/3)-1/3\log_2(1/3)]=0.19789$$
$$G(S,\exp eriance)=H(S)-2/7\,H(medium)-2/7\,H(low)-3/7\,H(high)=0.69928$$
$$G(S,background\ )=H(S)-2/7\,H(excellant\ )-3/7\,H(neatral)-3/7\,H(good)=0.3\,.$$

MID3 Algorithm:

function MID3 (examples, attributes, default)
returns a decision tree
inputs: examples, set of examples, attributes, set of attributes
Default, default value for the goal predicate
if (examples is empty) *then return* default
else if (all examples have the same classification) *then return* the classification.
else if (attributes is empty) *then return* MAJORITY-VALUE(examples)
else
do{
if (the attribute values have multi-valued) *then*
{ classify the attribute values into sub-classes}
best $\leftarrow$ CHOOSE-ATTRIBUTE(attributes, examples)
tree $\leftarrow$ a new decision tree with root test best
for each value v_i of best *do*
example s_i $\leftarrow$ {elements of examples with best = v_i }
subtree $\leftarrow$ MID3(examples s_i , attributes-best, MAJORITY-VALUE(examples) add a branch to tree with label v_i and subtree sub-tree}//*end do*
end
return tree

According to this algorithm Table 5 can then transformed into Table 6.

By applying the Multi-valued ID3 (MID3) algorithm for the above table then the information gain is given as:

$$G(S, language) = H(persons) - 9/23\,H(F) - 8/23\,H(D)$$
$$-2/23\,H(H) - 2/23\,H(R) - 2/23\,H(S) = 0.985 - 0.299 - 0.331977 - 4/23 = 0.1801$$
$$G(S, \deg ree) = H(persons) - 13/23\,H(Bs) - 8/23\,H(Ms)$$
$$-2/23\,H(\{PhD) = 0.985 - 13/23[-7/13\log_2(7/13) - 6/13\log_2(6/13)] = 0.09$$
$$G(S, \exp eriance) = H(persons) - 7/23\,H(medium) - 8/23\,H(low) - 8/23\,H(high) = 0.805$$
$$G(S, background) = H(persons) - 10/23\,H(excellent) - 5/23\,H(neutral) - 8/23\,H(good) = 0.49$$

The attribute experience is of highest value of information gain. Then we take it as a root of our tree as shown in Fig. 8.

Table 6. Transformed system of table 5

Person	Classes	Language	Degree	Experience	Background	Decision
P_1	P_{11}	F	Bs	Medium	Excellent	Accept
	P_{12}	F	Ms	Medium	Excellent	Accept
	P_{13}	F	PhD	Medium	Excellent	Accept
	P_{14}	D	Bs	Medium	Excellent	Accept
	P_{15}	D	Ms	Medium	Excellent	Accept
	P_{16}	D	PhD	Medium	Excellent	Accept
P_2	P_{21}	H	Bs	Low	Neutral	Reject
	P_{22}	R	Bs	Low	Neutral	Reject
P_3	P_{31}	F	Bs	Low	Good	Reject
	P_{32}	F	Ms	Low	Good	Reject
	P_{33}	D	Bs	Low	Good	Reject
	P_{34}	D	Ms	Low	Good	Reject
	P_{35}	S	Bs	Low	Good	Reject
	P_{36}	S	Ms	Low	Good	Reject
P_4	P_{41}	F	Bs	High	Neutral	Accept
	P_{42}	F	Ms	High	Neutral	Accept
P_5	P_{51}	D	Bs	Medium	Neutral	Reject
P_6	P_{61}	F	Bs	High	Excellent	Accept
	P_{62}	F	Ms	High	Excellent	Accept
	P_{63}	D	Bs	High	Excellent	Accept
	P_{64}	D	Ms	High	Excellent	Accept
P_7	P_{71}	R	Bs	High	Good	Accept
	P_{72}	H	Bs	High	Good	Accept

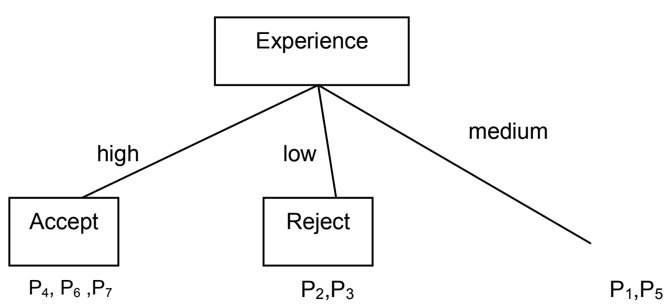

Fig. 8. Reduced graph of multi-valued system of Table 5

From the last decision tree we can deduce the following decision rules for the company example:

- The persons with high experience are accepted to work in the company.
- The persons with low experience are rejected to work in the company.
- The persons with medium experience and with excellent background are accepted to work in the company.
- The persons with medium experience and with neutral background are rejected to work in the company.

6 Conclusion and Future works

Decision trees are also a popular form of classification models. It is well known that classical trees lack the ability of modeling vagueness. By connecting fuzzy systems [44] and classical decision trees, we try to achieve classifiers that can model vagueness and are comprehensible. A popular and particularly efficient method for making a decision tree for classification from symbolic data is ID3 algorithm. Revised algorithms for numerical data have been proposed, some of which divide a numerical range into several intervals or fuzzy intervals. Their decision trees, however, are not easy to understand. In this work we proposed a new version of ID3 algorithm to generate a understandable fuzzy decision tree using fuzzy sets is fuzzy ID3.

In the near future we will generalize the approach of decision tree using topological rough sets [45,46] and topological fuzzy sets [47]. These approaches will discover new approximation class in information trees.

Competing Interests

Author has declared that no competing interests exist.

References

[1] Atkinson-Abutridy J, Mellish C, Aitken S. Combining information extraction with genetic algorithms for text mining. IEEE Intelligent System. 2004;19(3):22-30.

[2] Au WH, Chan KCC. An effective alogrithm for discovering fuzzy rules in relation databases. In proceedings of the 7th IEEE International Conference on fuzzy system. 1998;1314-1319.

[3] Babcock B, Babu S, Datar M, Motwani R, Widom J. Models and Issues Data Stream Systems. In proceedings of the 2002 ACM Symposiumon Principles of Database System (PODS 2002) (Invited Paper). Acm Press; 2002.

[4] Berghammer R, Leoniuk B, Milanese V. Implementation of relational algebra using binary decision diagrams, In de swart H, editor, 6th International Conference Relmics 2001 and 1st workshop of cost Action 274 TARSKI, Oisterwijk, The Netherlands, October 16-21, Volume 2561 of lecture notes in computer science, Berlin – Heidelberg – New York. Springer. 2002;2561:241-257.

[5] Bregler C, Omoundro SM. Non linear image interpolation using manifold learning. In Tesauro G, Touretzky DS, Leen TK. (eds.). Advances in Neural Information Processing. 1995;System 7:973-980.

[6] Bryant RE. Graph based algorithms for Boolean function mainpulation. IEEE Trans on Computers. 1986;35(8):677-691.

[7] Buntine W. Learning classification tres. Statistics and Computing. 1992;2:63-73,

[8] Carreira MA. Continuos latert variable models for dimensionality reduction and sequential data reconstruction. Phd thesis. University of Sheffield, UK; 2001.

[9] Chang RL, Pavlidis PT. Fuzzy decision tree algorithms. IEEE Transactions on systems, Man, and cybernetics. 1997;7(1):28-35.

[10] Chen S, Lee SL, Lee C. A new method for generating fuzzy ruless from numerical data for handling classification problemes. Applied Arificial Intelligence. 2001;15(7):645-664.

[11] I-Jen Chiang, Jane Yung-jen Hsu. Fuzzy classification trees for data analysis. Fuzzy Sets and Systems. 2002;130:87–99.

[12] Dash M, Choi K, Scheuermann P, Liu H. Feature selection for clustering – A filter solution. In Proceedings of IEEE International Conference on Data Mining (ICDM). 2002;155-122.

[13] Dash M, Liu H. Feature selection for classification. Intelligent Data Analysis. 1997;1(3):131-156.

[14] David M, Squire G. The ID3 decision tree induction algorithm. MONASH University Faculty of Information Technology; 2004.

[15] Devijver P, Kittler J. Pattern recognition: A statistical approach. Prentice hall; 1982.

[16] Dobra A, Gehrke J, Garofalakis M, Rastogi R. Processing complex aggregate queries over data streams. In proc. of the 2002 ACMSIG MOD Intel. Conf. On Management of Data; 2002.

[17] Fayyed UM, Keki BI. On the handing of continuous-Valued Attributes in Decision Tree Generation. Machine Learning. 1992;8:87-102.

[18] Fayyad UM, Piatetsky-Shapiro G, Smyth P. From data mining to knowledge discovery in databases. Artificial Intelligence Magazine. 1996;17(3):37-54.

[19] Friedman JH, Tukey JW. A projection pursuit algorithm for exploratory data analysis. IEEE Transaction on Computers. 1974;C-23(9):881-890.

[20] Friedman JH, Wstuetzle. Projection pursuit regression. Journal of the American Statistics Association. 1981;76:817-823.

[21] Harish Garg. An approach for analyzing the reliability of industrial system using fuzzy Kolmogorov’s differential equations, The Arabian Journal for Science and Engineering. 2015;40(3):975–987.

[22] Harish Garg. Analyzing the behavior of an industrial system using fuzzy confidence interval based methodology, National Academy Science Letters. 2014;37(4):359–370.

[23] Harish Garg. A novel approach for analyzing the behavior of industrial systems using weakest t-norm and intuitionistic fuzzy set theory, ISA Transaction. 2014;53:1199–1208.

[24] Harish Garg. Reliability, availability and maintainability analysis of industrial systems using PSO and fuzzy approach. MAPAN – Journal of Metrology Society of India. 2014;29(2):115–129.

[25] Harish Garg. Preventive maintenance scheduling of pulping unit in a paper mil. Japan Journal of Industrial and Applied Mathematics. 2013;30(2):397–414.

[26] Harish Garg. Performance analysis of complex repairable industrial systems using PSO and fuzzy confidence interval based lambda-tau methodology. ISA Transactions. 2013;52(2): 171–183.

[27] Harish Garg. Performance and behavior analysis of repairable industrial systems using Vague Lambda-Tau methodology. Applied Soft Computing. 2014;22:323–338.

[28] Kruskal JB. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika. 1964;29(1):1-27.

[29] Muir A, Lvo Dünstsch, Günther Gediga. Rough set data representation using decision diagram. RACSAM, Rev. R. Acad. Cien. Serie A. Mat. 2004;98(1):197-211.

[30] Peng Y, Flach PA. Soft discretization to enhance the continuous decision tree induetion. Integrating Aspects of Data Mining, Decision, Support and Metalearning. 2001;109-118.

[31] Qin KY, Pei Z. On the topological properties of fuzzy rough sets, Fuzzy Sets and Systems. 2005;151(3):601-613.

[32] Qin KY, Yang JL, Pei Z. Generalized rough sets based on reflexive and transitive relations. Information Sciences. 2008;178:4138-4141.

[33] Raman B, Ioerger TR. Instance-based filter for feature selection. Journal of Machine Learning Research. 2002;1:1-23.

[34] Sahinoglu M, Security Meter. A practical decision tree model to quantify risk. IEEE Security and privacy. 2005; 03(3):18-24.

[35] Slowinski R. Intelligent decision support, handbook of applications and advances of rough set theory. Kluwer Academic Publishers. ISBN, 0-7923-1923-0; 1992.

[36] Salama AS. Topological solution of missing attribute values problem in incomplete information tables. Information Sciences. 2010;180:631-639.

[37] Wu WZ. A study on relationship between fuzzy rough approximation operator and fuzzy topological spaces, in: L. Wang, Y. Jin (Eds.), FSKD 2005, LNAI, Springer-Verlag, Berlin, Heidelberg. 2005;3613:167-174.

[38] Zadeh LA, Fuzzy Sets. Information and Control. 1965;8:338-353.

[39] Zadeh LA. The concept of a liguistic variable and its application to approximate reasoning. Information Sciences. 1975;(8)199-249, 301-357;(9):43-80.

[40] Quinlan JR. Induction of decision trees. Macine learning. 1986;1:81-106.

[41] Quinlan JR. Decision trees and decision making. IEEE Transaction on System, Man and Cybernetic. 1990;20:339-346.

[42] Quinlan JR. C4.5: Programs for Machine Learing. The Morgan Kaufmann series in machine learing. Morgan kaufmann publishers, San Mateo, CA; 1993.

[43] Quinlan JR. Improved use of continuous attributes in C4.5. Journal of Artificial Intelligence Research. 1996;4:77-90.

[44] Ziarko W. The discovery, analysis and representation of data dependencies in databases. In: Piatetsky G. – Shapiro, W, Frawley J. (Eds.), Knowledge Discovery in Databases. AAAI Press / MIT press. 1991;177–195.

[45] Ziarko W. Variable precision rough set model. Journal of Computer and System Sciences. 1993;46(1):39-59.

[46] Ziarko W. Rough sets and knowledge discovery. An overview. In: Ziarko. 1994;11–15.

[47] Ziarko W. Rough Sets, Fuzzy Sets and Knowledge Discovery (RSKD'93) Workshops in Computing. Springer–Verlag & British Computer Society, London, Berlin; 1994.

Peer-review history:
The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)
www.sciencedomain.org/review-history.php?iid=1035&id=6&aid=9099