ISSN No. 0976-5697

Volume 6, No. 5, May - June 2015

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

New Measures for Constructing Decision Tree Based on Rough Sets and Applications

Nguyen Duc Thuan
Department of Information Systems
Nha Trang University - Vietnam
Nha Trang, Vietnam

Pham Quang Tung
Department of Fundamental Sciences
Nha Trang Air Force Officer College – Vietnam
Nha Trang, Vietnam

Abstract: In a decision tree algorithm, an important factor is measure for selecting the best split the records. There are many measures that can be used to determine the splitting attributes based on Entropy, Gini index, Information Gain and so on. In this paper, two metrics from the consistency degree of two knowledges is used as the criteria for selecting the attribute that will best separate the samples into individual classes. These new measures also to define on covering rough sets for constructing decision tree on incomplete information systems.

Keywords: Decision tree, Rough set, Consistency degree, Splitting Attribute

I. INTRODUCTION

There are many criteria for selecting splitting attributes in building decision tree [5], such as ID3, C4.5 which use an entropy-based measure known as information gain as a heuristic for selecting the attribute. The measures developed for selecting the best split are often based on the degree of impurity of the child nodes. The smaller the degree of impurity, the more skewed the class distribution.

Rough set theory, proposed by Pawlak in early 1980s is a mathematical tool to deal with uncertainty and incomplete information. Nowadays, it turns out that this approach is of fundamental importance to artificial intelligence and cognitive sciences, especially in the areas of data mining, machine learning, decision analysis, knowledge management, expert systems and pattern recognition. [1-3,7-8]

In recent years, researchers proposed different effective algorithms for constructing descision tree based on rough sets, classification algorithms may be categorized as follows [7]:

Using new measures for selecting the attribute that separate the samples into individual classes (for example β-dependability, Fixed information gain,..) [2][7]

Reducing irrelevant attribute by the rough set theory then condensed the sample by removing duplicate instance [1][7].

In this paper, two metrics from the consistency degree of two knowledges based on rough set theory is used as the criteria for selecting the attribute that will best separate the samples into individual classes. This new measures also to define on covering rough sets for constructing decision tree on incomplete information systems. Furthermore, the experiments show that the proposed measures can effectively process as traditional measures.

The rest of this paper is organized as follows: In Section 2 we introduce some basic concepts and notations which will be used throughout this paper. The new measure and properties are presented in Section 3. Experimental results are summarized in Section 4. Finally, Section 5 concludes this paper.

II. BASIC CONCEPTS

We recall some basic concepts of rough set, the interested reader is referred to [8]

Definition 1: *Knowledge*: A knowledge is a pair < U,P > where U is a nonempty finite set and P is an equivalence relation on U. P will also denote the partition generated by the equivalence relation.

Definition 2. *Finer and Coarser Knowledge* : A knowledge P is said to be finer than the knowledge Q if every block of the partition P is included in some block of the partition Q. In such a case Q is said to coarser than P. We shall write it as P≤ Q.

Definition 3. Let P and Q be two equivalence relations over U. The P-positive region of Q, denoted by $POS_P(Q)$ is defined by $POS_P(Q)=\bigcup_{X\in U/Q}\underline{P}(X)$, where $\underline{P}(X)$ = {Y ∈ U/P : Y ⊆ X} called P-lower approximation of X.

Definition 4. *Dependency degree* : Knowledge Q depends in a degree k (0 ≤k ≤ 1) on knowledge P , written as $P \Rightarrow_k Q$, iff k = $\frac{|POS_P(Q)|}{|U|}$ where $|X|$ denotes cardinality of the set X.

If k = 1, we say that Q totally depends on P and we write P ⇒ Q; and if k = 0 we say that Q is totally independent of P.

Proposition 1.

1) If $R\Rightarrow_k P$ and $Q\Rightarrow_l P$ then $R\cup Q\Rightarrow_m P$, for some $m\geq \max(k,l)$

2) If $R\cup P\Rightarrow_k Q$ then $R\Rightarrow_m P$ and $Q\Rightarrow_l P$, for some $l,m\leq k$

3) If $R\Rightarrow_k Q$ and $R\Rightarrow_l P$ then $R\Rightarrow_m Q\cup P$, for some $m\leq \min(k,l)$

4) If $R\Rightarrow_k Q\cup P$ then $R\Rightarrow_m P, R\Rightarrow_l Q$, for some $l,m\geq k$

5) If $R\Rightarrow_k P$ and $P\Rightarrow_l Q$ then $R\Rightarrow_m Q$ for some $m\geq l+k-1$.

Definition 5. *Consistency degree*: Let P and Q be two knowledges such that $P \Rightarrow_a Q$ and $Q \Rightarrow_b P$. The consistency degree between the two knowledges denoted by Cons(P,Q) is given by:

$Cons(P,Q) = \dfrac{a+b+nab}{n+2}$, where n is a non negative integer.

Inconsistency degree: We now define a measure of inconsistency by: InCons(P,Q) = 1 - Cons(P,Q)

Definition 6. *Covering*: Let U is a nonempty set of objects, **C** a family of subsets of U. If none subsets in **C** is empty, and $\cup$**C** =U, then **C** is called a covering of U. The pair (U,**C**) will be called generalized approximation space.

It is clear that a partition of U is certainly a covering of U, so the concept of a covering is an extension of the concept of a partition.

Definition 7. *Tolerance space*: A tolerance space is structure <U, R> where U is a nonempty set of objects and R is a reflexive and symmetric binary relation defined on U.

Definition 8. We shall say that a covering $\mathbf{C}_1$ is finer than a covering $\mathbf{C}_2$ written as $\mathbf{C}_1 \le \mathbf{C}_2$ iff

$\forall C_j^2 \in \mathbf{C}_2$, $\exists C_{j1}, C_{j2},..,C_{jn} \in \mathbf{C}_1$ such that $C_j^2 = \bigcup_{i=1}^{n} C_{ji}$ i.e every element of $\mathbf{C}_2$ may be expressed as the union of some elements of $\mathbf{C}_1$.

For any x ∈ U the set

$$P_x^C = \{y \in U : \forall C_i (x \in C_i \Leftrightarrow y \in C_i)\}$$

will be called kernel of x. Let P be the family of all kernels (U, C) i.e. P = { $P_x^C : x \in U$} . Clearly P is a partition of U.

Definition 9. Let (U,**C**) will be called generalized approximation space and X be a subset of U. Then the lower and upper approximations are defined as follows:

$$\underline{C}(X) = \cup\{P_x^C : P_x^C \subseteq X\}$$
$$\overline{C}(X) = \cup\{P_x^C : P_x^C \cap X \neq \varnothing\}$$

III. METRIC ON THE SET OF PARTITION

Definition 10. Let U is a nonempty finite set, P,Q are partitions on U. If P $\Rightarrow_a$ Q and Q $\Rightarrow_b$ P then two distance functions between P and Q is denoted by ρ(P,Q) and $\tau(P,Q)$ defined as

1. $\rho(P,Q) = \dfrac{2-(a+b)}{2}$

2. $\tau(p,Q) = InsCons(P,Q) = 1 - Cons(P,Q)$

Proposition 2. The distances between two partitions are metrics.

Proof

1. It is clear that $0 \le \rho(P,Q) \le 1$, because $0 \le a,b \le 1$

   We have to prove 3 properties:

   a. $\rho(P,Q) = 0 \; iff \; P{=}Q$

   b. $\rho(P,Q) = \rho(Q,P)$

   c. ∀P,Q,R are partitions:
   $\rho(P,R) \le \rho(P,Q) + \rho(Q,R)$

a. $\rho(P,Q) = 0 \; iff \; P{=}Q$

If P $\Rightarrow_a$ Q and Q $\Rightarrow_b$ P then $\rho(P,Q) = 0$

$\Leftrightarrow \rho(P,Q) = \dfrac{2-(a+b)}{2} = 0 \Leftrightarrow a = b = 1$

$\Leftrightarrow P = Q$

b. $\rho(P,Q) = \rho(Q,P)$

Evidently, by the role of P,Q is symmetric in distance function.

c. ∀P,Q,R are partitions:

$$\rho(P,R) \le \rho(P,Q) + \rho(Q,R)$$

The assumption that

$$P \Rightarrow_a R, R \Rightarrow_b P, P \Rightarrow_l Q, Q \Rightarrow_m P, Q \Rightarrow_u R, R \Rightarrow_v Q$$

We proof that

$$\frac{2-(a+b)}{2} \le \frac{2-(l+m)}{2} + \frac{2-(u+v)}{2}$$

$\Leftrightarrow$ l+m+u+v-2 ≤ a+b

$\Leftrightarrow$ (l+u-1)+(m+v-1) ≤ a+b (*)

By virtue of proposition 1 (5), there are

$$a \ge l + u - 1$$
$$b \ge m + v - 1$$

This completes the proof (*).

Since a, b, c, it follows that ρ is a metric.

2. Assume, with any partitions P,R,Q

$$P \Rightarrow_p R, R \Rightarrow_q P, P \Rightarrow_l Q, Q \Rightarrow_m P, Q \Rightarrow_u R, R \Rightarrow_v Q$$

a. $\tau(P,R) = 0 \; iff \; P{=}R$

$\tau(P,R) = 0 \Leftrightarrow 1 - Cons(P,Q) = 0$

$\Leftrightarrow 1 - Cons(P,Q) \Leftrightarrow 1 - \dfrac{p+q+npq}{n+2} = 0$

$\Leftrightarrow \dfrac{(1-p)+(1-q)+n(1-pq)}{n+2} = 0 \Leftrightarrow p = q = 1$,

(because $0 \le p,q \le 1$)

$\Leftrightarrow P \Rightarrow Q \wedge Q \Rightarrow P \Leftrightarrow P = Q$

b. $\tau(P,Q) = \tau(Q,P)$

Obviously, the role of P,Q is symmetric in distance function.

c. $\tau(P,R) \le \tau(P,Q) + \tau(Q,R)$

We have,

$$\tau(P,R) \le \tau(P,Q) + \tau(Q,R) \Leftrightarrow$$

$$\frac{n+2-p-q-npq}{n+2} \le \frac{n+2-l-m-nlm}{n+2} + \frac{n+2-u-v-nuv}{n+2}$$

$$\Leftrightarrow n(lm+uv-pq-1) \le 2+p+q-(l+m+u+v) \text{ (**)}$$

From proposition 1 (5), there are

$$p \ge l+u-1$$
$$q \ge v+m-1$$

$(lm+uv-pq-1)$

Hence, $\le lm+uv-(l+u-1)(v+m-1)-1$

$= l(1-v)+u(1-m)-(1-v)-(1-m)$

$\le (1-v)+(1-m)-(1-v)-(1-m) = 0$

(because $0 \le l,u \le 1$) and

$2+p+q-l-u-v$

$\ge 2+l+u-1+v+m-1-l-m-u-v = 0)$

Thus the left hand side of (**) is negative, the right hand side of (**) is positive.

So (**) is established, i.e τ is a metric.

IV. APPLICATIONS

A. *Decision tree on complete Information*

In the decision tree algorithms, the measures for selecting the best split are often based on the degree of impurity of the child nodes. The smaller the degree of impurity, it more skewed the class distribution. In this section, we propose two measures for selecting the best splitting attribute:

$\mu(P,Q)=1-\rho(P,Q)$ and

$\chi(P,Q)=1-\tau(P,Q)=Cons(P,Q)$,

with n=2 in formula of *Cons()*, and *P,Q* are partitions corresponding to attributes P,Q.

For example, we use a sample dataset (Golf Dataset – Table 2.) to compare building of decision tree for ρ,χ.

We have,

$U=\{u_1,u_2,u_3,u_4,u_5,u_6,u_7,u_8,u_9,u_{11},u_{12},u_{13},u_{14}\}$,

$|U|=14$

The partition of U under P1, P2,P3,P4 and D respectively:

$\pi_{P1}=\{X_1,X_2,X_3\}=\{\{u_1,u_2,u_8,u_9,u_{11}\},\{u_3,u_7,u_{12},u_{13}\},\{u_4,u_5,u_6,u_{10},u_{14}\}\}$

$\pi_{P2}=\{Y_1,Y_2,Y_3\}=\{\{u_1,u_2,u_3,u_{13}\},\{u_4,u_8,u_{10},u_{11},u_{12},u_{14}\},\{u_5,u_6,u_7,u_9\}\}$

$\pi_{P3}=\{Z_1,Z_2\}=\{\{u_1,u_2,u_3,u_4,u_5,u_8,u_{10},u_{12},u_{14}\},\{u_6,u_7,u_9,u_{11},u_{13}\}\}$

$\pi_{P4}=\{W_1,W_2\}=\{\{u_1,u_3,u_4,u_5,u_8,u_9,u_{10},u_{13}\},\{u_2,u_6,u_7,u_{11},u_{12},u_{14}\}\}$

$\pi_D=\{D_1,D_2\}=\{\{u_1,u_2,u_6,u_8,u_{14}\},\{u_3,u_4,u_5,u_7,u_9,u_{10},u_{11},u_{12},u_{13}\}\}$

$POS_{P1}(D)=2/7, POS_D(P1)=0$

$\Rightarrow \mu(P1,D)=1/7$, $\chi(P1,D)=1/14$

$POS_{P2}(D)=0$, $POS_D(P2)=0$

$\Rightarrow \mu(P2,D)=0$, $\chi(P2,D)=0$

$POS_{P3}(D)=0$, $POS_D(P3)=0$

$\Rightarrow \mu(P3,D)=0$, $\chi(P3,D)=0$

$POS_{P4}(D)=0$, $POS_D(P4)=0$

$\Rightarrow \mu(P4,D)=0$, $\chi(P4,D)=0$

Thus, we choose P1 is root node, then there are 3 branches

- Sunny: $U_1=\{u_1,u_2,u_8,u_9,u_{11}\}$,
- Overcast: $U_2=\{u_3,u_7,u_{12},u_{13}\}$,
- Rain: $U_3=\{u_4,u_5,u_6,u_{10},u_{14}\}$

We now turn to Sunny class with

$\pi^1_{P2}=\{Y^1_1,Y^1_2,Y^1_3\}=\{\{u_1,u_2\},\{u_8,u_{11}\},\{u_9\}\}$

$\pi^1_{P3}=\{Z^1_1,Z^1_2\}=\{\{u_1,u_2,u_8\},\{u_9,u_{11}\}\}$

$\pi^1_{P4}=\{W^1_1,W^2_1\}=\{\{u_1,u_2,u_8\},\{u_9,u_{11}\}\}$

$\pi^1_D=\{D^1_1,D^1_2\}=\{\{u_1,u_2,u_8\},\{u_9,u_{11}\}\}$

$POS_{P2}(D^1)=3/5$, $POS_{D^1}(P2)=0$

$\Rightarrow \mu(P2,D^1)=3/10$, $\chi(P2,D^1)=3/20$

$POS_{P3}(D^1)=1$, $POS_{D^1}(P3)=1$

$\Rightarrow \mu(P3,D^1)=1$, $\chi(P3,D^1)=1$

$POS_{P4}(D^1)=1$, $POS_{D^1}(P4)=1$

$\Rightarrow \mu(P4,D^1)=1$, $\chi(P4,D^1)=1$

Thus, we can choose P3 or P4 is splitting attribute. At this time we chose P3.

There are 2 sub-branch from "Sunny" class

- Humidity-High: $\{u_1,u_2,u_8\}$
- Humidity Normal: $\{u_9,u_{11}\}$

Overcast: $U_2=\{u_3,u_7,u_{12},u_{13}\}$ with partition of decision attribute D: $\{\{u_3,u_7,u_{12},u_{13}\}\}$ (*only lass*)

Rain: $U_3=\{u_4,u_5,u_6,u_{10},u_{14}\}$

$\pi^3_{P2}=\{Y^3_1,Y^3_2\}=\{\{u_4,u_{10},u_{14}\},\{u_5,u_6\}\}$,

$\pi^3_{P4}=\{W^3_1,W^3_2\}=\{\{u_4,u_5,u_{10}\},\{u_6,u_{14}\}\}$

$\pi^3_{D^3}=\{D^3_1,D^3_2\}=\{\{u_4,u_5,u_{10}\},\{u_6,u_{14}\}\}$

$POS_{P2}(D^3)=0$, $POS_{D^3}(P2)=0$

$\Rightarrow \mu(P2,D^3)=0$, $\chi(P2,D^3)=0$

$POS_{P4}(D^3)=1$, $POS_{D^3}(P4)=1$

$\Rightarrow \mu(P4,D^3)=1$, $\chi(P4,D^3)=1$

Thus, we can choose P4 is splitting attribute (on U_3)

Finally, we have decision tree as same as decision tree which constructed by traditional measure Entropy or Gain information Figure 2.

B. *Decision tree on incomplete information*

With definition 8, we can propose a heuristic for constructing decision tree on incomplete information. The basic heuristic is as follow:

1. On incomplete information S=(U,AT), where U is a nonempty finite set of objects, and AT is a non-empty finite set of attributes (*it may happen that some of attributes values for an object are missing. To indicate such a situation a situation at distinguished value, so called null value. We will denote null value by* *). We define a binary similarity relation SIM(A), A∈AT, between objects that are possibly indiscernible in terms of values of attribute AT.

SIM(A) = {(x,y)∈UxU: A(x)=A(y) or A(x)= * or A(y)=*}

Clearly, SIM(A) is a tolerance relation. If any objects in U has not null value on A, SIM(A) is equivalence relation.

Let C^A_x note the object set {y∈U|(x,y)∈SIM(A)}. We have $C_A=\{C^A_x: x\in U\}$ is a covering of U and P be the family of all kernels (*U*, C_A) i.e. $P_A=\{P^{C_A}_x : x\in U\}$, which is a partition of U.

2. With each A∈AT, A has null values on some objects in U, let (U,P_A) be knowledge defined by A.

3. From definition above knowledge which determined by an attribute, we can consider the measures which defined in 4.1 for selecting best splitting attribute.

For example, we consider a sample incomplete information (Table.1).

We have,

$$U = \{u_1, u_2, u_3, u_4, u_5, u_6\} \quad |U| = 6$$

The covering U under P, M, S and D respectively:

$$C_P = \{\{u_1, u_3, u_4, u_5\}, \{u_2, u_3, u_5, u_6\}, \{u_3, u_5\}\}$$

$$\Rightarrow \mathrm{P_P} = \{\{u_1, u_4\}, \{u_2, u_6\}, \{u_3, u_5\}\} = \pi_P$$

$$C_M = \{\{u_1, u_2, u_3, u_4, u_5, u_6\}, \{u_2, u_3, u_4, u_5\}\}$$

$$\Rightarrow \mathrm{P_M} = \{\{u_1, u_6\}, \{u_2, u_3, u_4, u_5\}\} = \pi_M$$

$$C_S = \{\{u_1, u_2, u_4, u_5, u_6\}, \{u_3\}\}$$

$$\Rightarrow \mathrm{P_S} = \{\{u_1, u_2, u_4, u_5, u_6\}, \{u_3\}\} = \pi_S$$

$$C_A = \{\{u_1, u_2, u_6\}, \{u_3, u_4, u_5, u_6\}\}$$

$$\Rightarrow \mathrm{P_A} = \{\{u_1, u_2\}, \{u_3, u_4, u_5\}, \{u_6\}\} = \pi_A$$

$$C_D = \{\{u_1, u_2, u_4, u_6\}, \{u_3\}, \{u_5\}\} = \pi_D$$

$$POS_P(D) = 2/3,\ POS_D(P) = 1/3$$

$$\Rightarrow \mu(M, D) = 1/2,\ \chi(P, D) = 13/36$$

$$POS_M(D) = 0,\ POS_D(M) = 1$$

$$\Rightarrow \mu(M, D) = 1/2,\ \chi(M, D) = 1/4$$

$$POS_S(D) = 1/6,\ POS_D(S) = 1$$

$$\Rightarrow \mu(S, D) = 7/12,\ \chi(S, D) = 3/8$$

$$POS_A(D) = 1/2,\ POS_D(A) = 5/12$$

$$\Rightarrow \mu(A, D) = 5/12,\ \chi(A, D) = 7/24$$

Thus, we choose S is root node, then there are 2 branches

$$U_1 = \{u_1, u_2, u_4, u_5, u_6\}, U_2 = \{u_3\}$$

We now turn to U_1 class with

$$\pi_P^1 = \{\{u_1, u_4\}, \{u_2, u_6\}, \{u_5\}\}$$

$$\pi_M^1 = \{\{u_1, u_6\}, \{u_2, u_4, u_5\}\}$$

$$\pi_A^1 = \{\{u_1, u_2\}, \{u_4, u_5\}, \{u_6\}\}$$

$$\pi_{D^1}^1 = \{\{u_1, u_2, u_4, u_6\}, \{u_5\}\}$$

$$POS_P(D^1) = 1,\ POS_{D^1}(P) = 1/6$$

$$\Rightarrow \mu(P, D^1) = 7/12,\ \chi(P, D^1) = 3/8.$$

$$POS_M(D^1) = 2/6,\ POS_{D^1}(M) = 1/6$$

$$\Rightarrow \mu(M, D^1) = 1/4,\ \chi(M, D^1) = 11/72$$

$$POS_A(D^1) = 1/2,\ POS_{D^1}(A) = 1/3$$

$$\Rightarrow \mu(A, D^1) = 1/3,\ \chi(A, D^1) = 5/24$$

Thus, we can choose P is splitting attribute (on U_1), there are 3 branches

$U_4 = \{u_1, u_4\}$, $U_5 = \{u_2, u_6\}$, $U_6 = \{u_5\}$

and decision respectively

$D^4 = \{Good\}$, $D^5 = \{Good\}$, $D^6 = \{Excel\}$

In other words, we have the decision tree in Figure. 1

There are two decision rules:

(S,Compact) $\rightarrow$ Poor

(S,Full) $\rightarrow$ (D,Good)$\vee$(D,Excel)

This result is compatible with result in [3].

V. CONCLUSION

In this paper, we proposed two measures to determine the splitting attributes for constructing decision tree. This two measures developed from two metrics based on rough set. Unlike popular measures as Entropy, Information Gain, Gini index, new measures can use as heuristic on incomplete information. In our experiments, we compare constructing decision tree based on new measures and popular measures. The result shows that the decision tree which is based on new measures compatibility. However, next time we will test on many data sets to accurately assess the effectiveness of the new measures.

We believe that new measures offered here will turn out to be useful also in other tolerance information systems.

VI. REFERENCES

[1] Iftikhar U. Sikder , Toshinori Munakata, Application of Rough Set and Decision Tree for Characterization of Premonitory Factors of Low Seismic Activity, Expert Systems with Applications 36, p.102–110, 2009

[2] Longjun Huang, Minghe Huang, Bin Guo, Zhiming Zhang, A New Method for Constructing Decision Tree Based on Rough Sets Theory, IEEE International Conference on Granular Computing, p.241-244, 2007

[3] Marzena Kryszkiewicz, Rough Set Approach to Incomplete Information Systems, Information Sciences 112, p39-49, 1998.

[4] Miroslav Notvotný, Applications of Dependency Spaces, In: Incomplete Information: Rough Set Analysis, p 247-289, Springer, 1998

[5] Pang-Ning Tan, Vipin kumar, Michael Steinbach, “Introduction to Data Mining”, Pearson, 2006

[6] Shi Zhanhong, Gong Zengtai, Knowledge Reduction and Knowledge Significance Measure Based on Covering Rough sets, International Journal of Pure and Applied Mathematics, Vol 48, No.1, p 1-9, 2008

[7] Shailendra K. Shrivastava, Manisha Tantuway, A Decision Tree Algorithm Based on Rough Set Theory After Dimensionality Reduction, International Journal of Computer Applications (0975 – 8887), Vol 17–No.7, p.30-34, 2011

[8] Z.Pawlak, “Rough Sets”, Theoretical Aspects of Reasoning about Data, Kluwer Academic Publisher, 1991

Table 1. Incomplete Information

U	Price (P)	Mileage (M)	Size (S)	Max_Speed (A)	D
u_1	High	High	Full	Low	Good
u_2	Low	*	Full	Low	Good
u_3	*	*	Compact	High	Poor
u_4	High	*	Full	High	Good
u_5	*	*	Full	High	Excel
u_6	Low	High	Full	*	Good

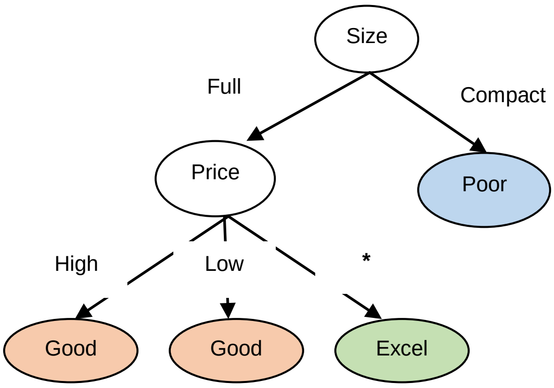

Figure 1. Decision tree from Incomplete Information

Table 2. The Gofl Dataset

U	Outlook (P1)	Temp. (P2)	Humidity (P3)	Wind (P4)	Play? (D)
u_1	Sunny	Hot	High	Weak	No
u_2	Sunny	Hot	High	Strong	No
u_3	Overcast	Hot	High	Weak	Yes
u_4	Rain	Mild	High	Weak	Yes
u_5	Rain	Cool	Normal	Weak	Yes
u_6	Rain	Cool	Normal	Strong	No
u_7	Overcast	Cool	Normal	Weak	Yes
u_8	Sunny	Mild	High	Weak	No
u_9	Sunny	Cold	Normal	Weak	Yes
u_{10}	Rain	Mild	Normal	Strong	Yes
u_{11}	Sunny	Mild	Normal	Strong	Yes
u_{12}	Overcast	Mild	High	Strong	Yes
u_{13}	Overcast	Hot	Normal	Weak	Yes
u_{14}	Rain	Mild	High	Strong	No

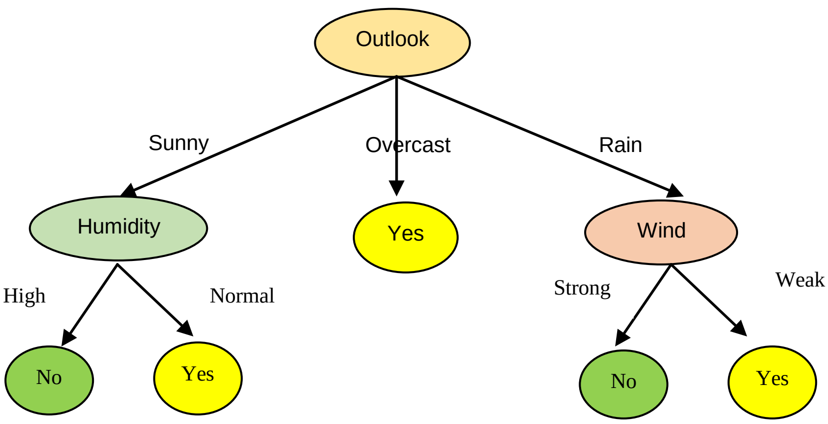

Figure2. The Decision tree of Golf Dataset