ROUGH SET THEORY

Supervised rule based classification technique

	Age	Height	Play Volleyball
x_1	0-10	1-3	No
x_2	0-10	3-4.5	Yes
x_3	10-20	4-6	Yes
x_4	10-20	4-6	No
x_5	20-30	5-6.5	Yes
x_6	20-30	5-6.5	Yes
x_7	30-40	6-7	Yes

Indiscriminate Objects:

Suppose we have $B \subseteq A$

equivalence class

Ind(B) $[x]_{Age} = \{x_1 x_2\} \{x_3 x_4\} \{x_5 x_6\} \{x_7\}$

$$[x]_{alpha} = \sum_{s_i s_j \in S} \{s_i, s_j\} \quad s_{i_{alpha}} = s_{j_{alpha}}$$

$$Ind(B) = \{ (x, x') \in U^2 \mid \forall\, a \in B,\ V_a(x) = V_a(x') \}$$

value a of x

U	Age	Height	Decision
x_1	0-10	1-3	N
x_2	0-10	3-4.5	Y
x_3	10-20	4-6	Y
x_4	10-20	4-6	N
x_5	20-30	5-6.5	Y
x_6	20-30	5-6.5	Y
x_7	30-40	6-7	Y

Cardinality = 4

$IND(Age) = \{\{x_1, x_2\}, \{x_3, x_4\}, \{x_5, x_6\}, \{x_7\}\}$

$IND(Height) = \{x_1, x_2, \{x_3, x_4\}, \{x_5, x_6\}, x_7\}$

$IND(Age, Height) = \{x_1, x_2, \{x_3, x_4\}, \{x_5, x_6\}, x_7\}$

$X_{cj} = \{x \mid c(x) = v_{cj}\}$

$X_y = [x_2\ x_3\ x_5\ x_6\ x_7]$

$X_N = \{x_1\ x_4\}$

Lower approxim^n

$B(X)_Y = \{x \mid [x]_B \subseteq X\}$

play

a decision

$Y \Rightarrow \{x_2, x_5, x_6, x_7\}$

Scanned by CamScanner

upper approximation

$$\overline{B(X)}_Y = \{x \mid \exists x \in [x]_B \subseteq X_Y\}$$

$$\{x \mid [x]_B \cap X_Y \neq \phi\}$$

Boundary between

$$\underline{\overline{B(X)}_Y} = \overline{B}(X)_Y - \underline{B}(X)_Y$$

outside $\leftarrow B^{\circ}(X)_Y = U - \overline{B}(X)_Y$

When attributes are reduced, upper boundaries expand.

When we take all attributes, $U = L$

x_1	x_2	x_3	x_4	x_5	x_6	x_7
	L	L			L	L
	U	U			U	U

(I) Roughly definable : $\underline{B(X)} \neq \phi$ & $\overline{B(X)} \neq U$

(II) Internally undefinable : $\underline{B(X)} = \phi$ & $\overline{B(X)} \neq U$

(III) Externally undefinable : $\underline{B(X)} \neq \phi$ & $\overline{B(X)} = U$

(IV) Totally undefinable : $\underline{B(X)} = \phi$ & $\overline{B(X)} = U$

(V) Crisp : $\underline{B(X)} = \overline{B(X)}$

The accuracy of approximation (α) wrt to attributes B for a class X

$$\alpha_B(X) = \frac{\underline{B(X)}}{\overline{B(X)}} \begin{array}{l} = 1 \rightarrow \text{crisp} - \text{ideal (not possible)} \\ < 1 \rightarrow \text{rough} \end{array}$$

As $\alpha \rightarrow 1$, computational complexity ↑ses
Variety ↓ses too
∴ Even though we are getting crisper data, variety decreases.

Systems should be
- → Adaptable (answer acc to situation, loc^n, emotion)
- → Support real time response

U	Dip	Exp.	French	Ref	Decision
x_1	MBA	M	Y	Exc.	A
x_2	MBA	L	Y	New	R
x_3	MCE	L	Y	Good	R
x_4	MSc	H	Y	New	A
x_5	MSc	M	Y	New	R
x_6	MSc	H	Y	Exc.	A
x_7	MBA	H	N	Good	A
x_8	MCE	L	N	Exc	R

$X_{accept} = x_1\ x_4\ x_6\ x_7$
$X_{reject} = x_2\ x_3\ x_5\ x_8$

$\underline{DEFR}(X_{accept}) = x_1\ x_4\ x_6\ x_7$
$\overline{DEFR}(X_{accept}) = x_1\ x_4\ x_6\ x_7$

$ER = \left.\begin{matrix} x_1 & x_2 & x_6 \\ x_5 & x_3 & x_7 \\ x_4 & x_8 & \end{matrix}\right] \rightarrow$ still all different

Indiscernability matrix

	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_1	$e \vee r$						
x_2	d						
x_3	$d \vee r$						
x_4	$d \vee e$						
x_5	$d \vee e$						
x_6	$d \vee e \vee r$						
x_7							
x_8							

e, r Different : 1
same : 0

A : set of all attributes

$$C_{ij} = \{ a \in A \mid a(x_i) \neq a(x_j) \} \quad \forall\, i,j = 1 \ldots n$$

Defining a set of attributes which are different in i^{th} & j^{th} object

discernability function :

$$\wedge \{ \vee C_{ij} \mid 1 \leq i,j \leq n \}$$

What is the min. no. of attributes such that we can discern every object from every object. This we can get through the method just discussed.

<u>Ultimate aim</u>: To ↑ the accuracy of classification

But the method discussed is not feasible. We use meta-heuristic for the same

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_1	–	–	–	–	–	–	–	–
x_2	$e \vee r$	–	–	–	–	–	–	–
x_3	$d \vee e \vee r$	$d \vee r$	–	–	–	–	–	–
x_4	$d \vee e \vee r$	$d \vee e$	$d \vee e \vee r$	–	–	–	–	–
x_5	$d \vee r$	$d \vee e$	$d \vee e \vee r$	e	–	–	–	–
x_6	$d \vee e$	$d \vee e \vee r$	$d \vee e \vee r$	r	$e \vee r$	–	–	–
x_7	$e \vee f \vee r$	$e \vee f \vee r$	$d \vee e \vee f$	$d \vee f \vee r$	$d \vee e \vee f \vee r$	$d \vee f \vee r$	–	–
x_8	$d \vee e \vee f$	$d \vee f \vee r$	$f \vee r$	$d \vee e \vee f \vee r$	$d \vee e \vee f \vee r$	$d \vee e \vee r$	$d \vee e \vee r$	–

$$a(a+b) = a$$

can be minimized using K-map

$$f_R = (e \vee r)(d \vee e \vee r)(d \vee r)(d \vee e)$$
$$(e \vee f \vee r)(d \vee e \vee f)(d \vee f \vee r)$$
$$(f \vee r)(e)(r)(d \vee e \vee f \vee r)$$
$$= er \quad (\text{using absorption})$$

⇒ If we want to discern every object from every object, we need just two features e and r.

using K-map:

de \ fr	00	01	11	10
00	0	0	0	0
01	0	1	1	0
11	0	1	1	0
10	0	0	0	0

⇓

er

Because ~~we~~ our aim is to classify objects wrt decision,
⇒ we want to discern all accepted with all rejected

	x_1	x_4	x_6	x_7	x_2	x_3	x_5	x_8
x_1	Don't care							
x_4								
x_6								
x_7								
x_2	evr	dve	dvevr	evfvr	Don't care			
x_3	dvevr	dvevr	dvevr	dvef				
x_5	dvr	~~e~~	evr	dvefvr				
x_8	dvevf	dvefvr	dvevr	dvevr				

$$f_R = (e)(dvr) = ed + er$$

⇓

If we want to discriminate the objects I reject with those I accept, this is the function.

⇒ e, d, r are important attributes
Also, since e is common, e is the core attribute.

ed.

~~Accept~~ $X_{accept} = x_1, \{x_4, x_6\}, x_7$

$X_{reject} = x_2 \{x_2, x_8\} x_5$

Every equivalence class is a rule.

<u>ex</u>

$X_{accept} = \{x_1\}, \{x_4\}\{x_6\}\{x_7\}$

$x_{reject} = \{x_2\}, \{x_3\}\{x_8\}\{x_5\}$

⇒ A total of 14 rules.

↓

No harm in considering extra rules (since it was an OR, we could have chosen just one path ed or ex, but to ~~check~~ claim support of decision strongly, we see all the rules.

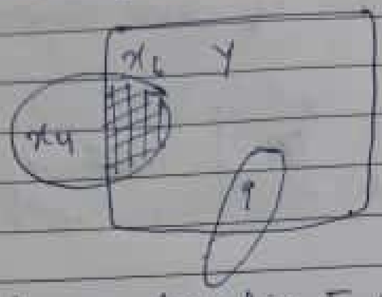
⇒ Your choice to consider all the redux attributes or a few

$$\Rightarrow \text{Total no. of rules} = \left\{ \sum_i \sum_R \left| Eq.\left(IND_{R_i}(X_R) \right) \right| \right.$$

\# Total no. of decisions in a decision system is called its "<u>Rank</u>".

eg: In the previous eg: Rank = 2 accept / reject

Soft computing ⇒



Rough membership ≡ degree of overlap

$$= \frac{|[x]_B \cap X|}{[x]_B} = \mu_X^B(x)$$

Idea of Rough sets ≡ Individual rep. by group

for x_4 Rough mem $= \frac{1}{2} = x_6$ rough mem

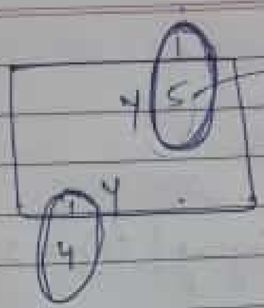
	Age	Upbringing	Like FB ?
x_1	0-10	Rural	Y
x_2	10-20	Urban	Y
x_3	10-20	Semi Urban	Y
x_4	10-20	Rural	N
x_5	>20	Urban	N

$$[x]_{Age} = \{\{x_1\}, \{x_2, x_3, x_4\}, \{x_5\}\}$$

$$\mu_Y^{Age}(x_1) = 1$$

$$\mu_Y^{Age}(\{x_2, x_3, x_4\}) = 2/3$$

$$\mu_Y^{Age}(x_5) = 0$$



If we introduce a threshold, we try to make it a cure rule and incorporate it in lower approximation

We need to define a threshold and define lower approximation & upper approxm^n wrt the threshold.

$$\underline{B}(X) = \{x \mid \mu_X^B(x) \geq \pi\}$$

$$\overline{B}(X) = \{x \mid \mu_X^B(x) > 1-\pi\}$$

If a feature is important for positive region is an important feature.

Subset of attributes

{Accept, Reject}

$$\gamma(C, D) = \frac{|POS_C(D)|}{|U|}$$

Rough Dependency

combination wrt of all the decisions

⇓

Since Pos-region is wrt combin^n of all the decisions

⇒ As we increase π, we try to increase the rough dependency

π Threshold = 1

$$\gamma(Age, D) = \frac{2}{5}$$

$$\gamma([Age, upbringing], D) = 1$$

$\pi = 0.7$

$$\gamma(Age, D) = \frac{2}{5}$$

$\pi = 0.6$

$$\gamma(Age, D) = 1$$

$\therefore$ $C \Rightarrow D$ for age when $\pi = 0.6$

A feature which is occurring in most of the redux and not all, we can say that feature is also important.

⇒ we are trying to look out for an

can be calculated using soft computing

approximate redux.

So, for this we define significance σ of an attribute:

$$\sigma_{C,D}(a) = \frac{\gamma(C,D) - \gamma(C-\{a\},D)}{\gamma(C,D)}$$

feature

subset of features

$$\sigma_{C,D}(B) = \frac{1 - \gamma(C-B,D)}{\gamma(C,D)}$$

Error:

$$\varepsilon_{C,D}(B) = \frac{\gamma(C,D) - \gamma(B,D)}{\gamma(C,D)}$$

If $\varepsilon = 0 \Rightarrow$ No error ~~can be~~ when B is considered instead of C

we also define a threshold for error too.

$\Rightarrow$ B is a good approximate redux if $\varepsilon = 0$ or within threshold.

→ Through this,
using small no. of attributes in a redux can ↑ my classificⁿ accuracy

→ This may also help when we have missing data.

One of the problems that still remain are:

→ Ranges - do they make sense?
↓
Agar alag krte to large amount of rules
Agar club krenge to are we losing info?

Not to use that info that will spoil my classification
↓
Same decision wale objects cluster krenge sath me

eg: 0-10 me all (yes) hain to vahan Y likh sakte hain
for all 1 to 10

Decision discretization is a NP-Com/ NP-Hard problem

[illegible]

U	length	breadth	Decision
1	1.2	2.5	Y
2	2.5	2.3	Y
3	3.2	3.1	N
4	4.1	3.5	Y
5	4.1	3.5	N
6	1.5	3.7	N

Draw edges between all point pairs
Y-N : double edge
N-N, Y-Y : single edge

Redundant edges are removed heuristically. Only important ones are retained.

Region no(s)

U	L	B	Decision
1	0	0	Y
2	2	0	Y
3	3	1	N
4	4	2	Y
5	4	2	N
6	1	3	N

Note: We should have minimum no. of cuts.

→ Agar values number nahi hain balki symbols hain, so unko kaise discretize karein?

eg: Colors Red Blue can't have a mean

B Y

R Y → Can be clubbed together

G Y

B N

we use new binary features instead of symbolic values $a_2^{r}(x,y) = 1$ iff $a(x) \neq a(y)$

U	a	b	decision
1	a_1	b_1	0
2	a_1	b_2	0
3	a_2	b_3	0
4	a_3	b_1	0
5	a_1	b_4	1
6	a_2	b_2	1
7	a_2	b_1	1
8	a_4	b_2	1
9	a_3	b_4	1
10	a_2	b_5	1

New attributes

proportional variables

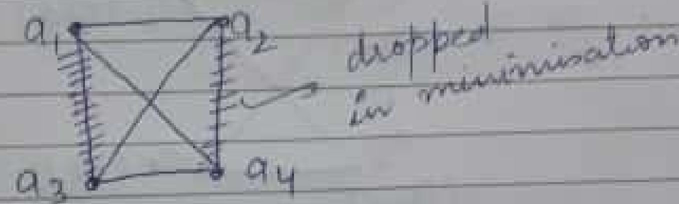
$M(R)$	u_1	2	3	4
U_5	b_4^1	b_4^2	$a_1^2\ b_4^3$	$a_1^3 \mid b_2^4$
6	$a_2^1 \mid b_2^1$	a_2^1	b_2^3	$a_2^3 \mid b_2^1$
7	a_1^2	$a_2^1 \mid b_1^2$	b_1^3	a_2^3
8	$a_4^1 \mid b_2^1$	a_4^1	$a_4^2 \mid b_2^3$	$a_4^3\ b_2^1$
9	$a_3^1 \mid b_4^1$	$a_3^1 \mid b_4^2$	$a_3^2 \mid b_4^3$	b_4^1
10	$a_2^1 \mid b_5^1$	$a_2^1 \mid b_5^2$	b_5^3	$a_2^3 \mid b_5^1$

$b_4^1 \equiv b_4^{b_1}$

→ Yesterday's fashion can also be done

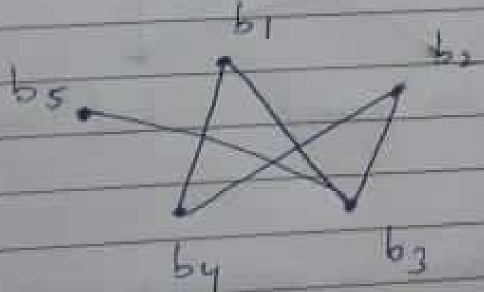
Neo4j

a - graph is coloring graph if they discern.
ie. edge if it discerns b/w 2 objects



$\Rightarrow$ b_4^1 a_1^2 b_4^2 a_4^1 b_2^3 b_1^3 b_5^3 a_2^3

$(a_4^3 \vee b_2^1)$



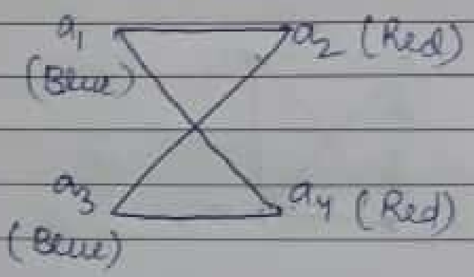
Step1: find short prime implicant of the attribute table

Step 2: Represent & graphs & color.

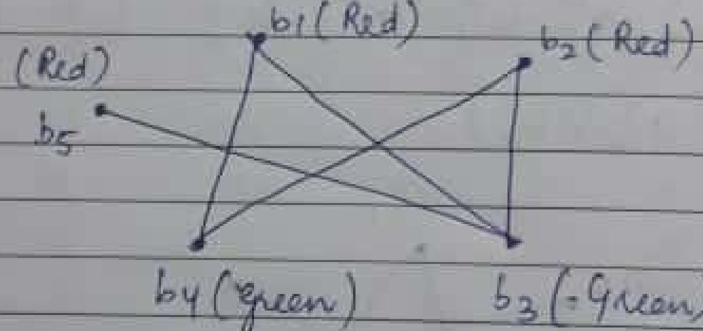
$a_2^1 \wedge a_3^2 \wedge a_4^1 \wedge$
$a_4^3 \wedge b_4^2 \wedge b_3^2$
$\wedge b_3^1 \wedge b_5^3$

Partition such that in same partition have same colour. So it is equivalent to graph coloring problem.

for a



[2 colours]



[2 colours]

Each colour is nothing but a code

$$\overset{1}{\{a_1, a_3\}} \quad \overset{2}{\{a_2\ a_4\}}$$

$$\underset{1}{\{b_1\ b_2\ b_5\}} \quad \underset{2}{\{b_3\ b_4\}}$$