ROUGH SETS

Dt.
Pg.
B+

It is a supervised rule based classification technique.

	Age	Height	Plays 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 or
Indisceranble Objects

Assume $B \subseteq A$.

Ind[B] :

Indiscernability w.r.t. Age

Eg: $[X]_{Age} = \{\{x_1, x_2\}, \{x_3, x_4\}, \{x_5, x_6\}, \{x_7\}\}$

Defn: $[X]_{alpha} = \sum_{s_i, s_j \in B} \{s_i, s_j\}$ where $s_{i_{alpha}} = s_{j_{alpha}}$.

$Ind[B] = [X]_B$

B can be a single attribute or a set of attributes.

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

where $V_a(x)$ is the value of attribute a of object x.

$|[x]_B|$ = cardinality = no. of indiscernable sets.

$X_{Cj} = \{ x \mid C(x) = V_C(x_j) \}$

Eg. $X_Y = \{ x_2, x_3, x_5, x_6, x_7 \}$

$X_N = \{ x_1, x_4 \}$

Lower boundary /approximation : When an equivalence class lies completely inside X_c, where c is the decision variable value. In the given example c is Y or N.

$\underline{B_Y(X)} = \{ x \mid [x]_B \subseteq X_Y \}$

Eg. let B = (Age, Height) $\Rightarrow [x]_B = \{ \{x_1\}, \{x_2\}, \{x_3, x_4\}, \{x_5, x_6\} \{x_7\} \}$

then, $\underline{B_Y(X)} = \{x_2, x_5, x_6, x_7\}$

$\underline{B_N(X)} = \{x_1\}$ ($\because$ there is no equivalence class for which all members have decision N)

Upper approximation

$\overline{B_Y(x)} = \{ x_i \mid \exists : x_i \in [x]_B$ and $x_i \in X_Y \}$

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

Eg. let B = (Age, Height)

then $\overline{B_Y(X)} = \{ \cancel{x_1}, x_2, x_3, x_4, x_5, x_6, x_7 \}$

$\overline{B_N(X)} = \{ x_1, \cancel{x_2}, x_3, x_4 \}$

Dt.
Pg.
B+

Lower Approximation $\subseteq$ Upper Approximation

Boundary : $\overline{B_Y(X)}$ = Upper Approximation − Lower Approximation
$= \overline{B_Y(x)} - \underline{B_Y(x)}$

Eg. let B = (Age, Height)
$\underline{B_Y(X)} = \{\cancel{x_1}, \cancel{x_2}, x_3, x_4\}$

$\overline{B_N(x)} = \{\cancel{x_1}, \cancel{x_2}, x_3, x_4\}$

Outside Approximation : $B_Y^{\circ}(x) = U - \overline{B}_Y(x)$

a Roughly Definable : for attribute(s) B
$\underline{B_Y(x)} \neq \phi$ and $\overline{B_Y(x)} \neq U$
or $\underline{B_N(x)} \neq \phi$ and $\overline{B_N(x)} \neq U$

b Internally Undefinable : for attribute(s) B
$\underline{B_Y(x)} = \phi$ and $\overline{B_Y(x)} \neq U$

c Externally Undefinable : $\underline{B_Y(x)} \neq \phi$ and $\overline{B_Y(x)} = U$

d Totally Undefinable : $\underline{B_Y(x)} = \phi$ and $\overline{B_Y(x)} = U$

e Crisp : $\underline{B_Y(x)} = \overline{B_Y(x)}$

all the above defn are applicable for $B_N(X)$ as well.

Scanned by CamScanner

Dt.
Pg. B+

The accuracy of approximation (α) wrt attributes B for ~~aeta~~ class X

$$\alpha_B(X) = \frac{\underline{B}(X)}{\overline{B}(X)} \quad \begin{array}{l} = 1 \rightarrow \text{(crisp, ideal)} \\ < 1 \rightarrow \text{rough} \end{array}$$

As $\alpha \rightarrow 1$, computational complexity increases and variety decreases
Though we are getting more crisp data yet loss of varity is not wanted.

System should be → Adaptable (answer accurately to situations)
→ Support Real Time Response

Let At be the set of all attributes of the given decision table, except the decision variable.

We define a set C_{ij} such that

$$C_{ij} = \{ a \in At \mid a(x^{(i)}) \neq a(x^{(j)}) \}$$

where a is an attribute and $a(x^{(i)})$ is the value of attribute a of object x_i.

B+

Hence C_{ij} is the set of all atributes based on which x_i and x_j can be distinguished or discerned.

Discernibility Function : $\wedge \{ \vee C_{ij} \mid 1 \le i < j \le n \}$
This function is symetric on relation C_{ij}

The minimum set of attributes using which we can discern any object from any other object is called the reduct.

Our ultimate aim is to increase the accuracy of classification.

The above method for finding reducts is not feasible for a large decision table. For small tables we can solve using K-Maps but for large tables we need meta-heuristics for the same.

U	Dip	Exp(e)	French(f)	Ref(r)	Decision
x_1	MBA	M	Y	Exc	A
x_2	MBA	L	Y	Neu	R
x_3	MCE	L	Y	Good	R
x_4	MSC	H	Y	Neu	A
x_5	MSC	M	Y	Neu	R
x_6	MSC	H	Y	Exc	A
x_7	MBA	H	N	Good	A
x_8	MCE	L	N	Exc	R

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

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

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_1	ϕ	–	–	–	–	–	–	–
x_2	er	ϕ	–	–	–	–	–	–
x_3	der	dr	ϕ	–	–	–	–	–
x_4	der	de	der	ϕ	–	–	–	–
x_5	dr	de	der	e	ϕ	–	–	–
x_6	de	der	der	r	er	ϕ	–	–
x_7	efr	efr	def	dfr	defr	dfr	ϕ	–
x_8	def	dfr	fr	defr	defr	der	der	ϕ

$$f_R = (e \vee r) \wedge (d \vee e \vee r) \wedge (d \vee r) \wedge (d \vee e) \wedge (e \vee f \vee r)$$
$$\wedge (d \vee e \vee f) \wedge (d \vee f \vee r) \wedge (f \wedge r) \wedge (e) \wedge (r)$$
$$\wedge (d \vee e \vee f \vee r)$$
$$= er$$ (using laws of boolean logic identities)

or using a K-map

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

$\Rightarrow er$

But, we want to discern objects that belong to different decision classes and not every object from every other object.

Hence we calculate discernibility function for objects that belong to different decision classes.

B+

	x_1	x_4	x_6	x_7	x_2	x_3	x_5	x_8
x_1								
x_4		DON'T						
x_6		CARE						
x_7								
x_2	er	de	der	efr				
x_3	der	der	der	def				
x_5	dr	e	er	defr		DON'T		
x_8	def	defr	der	der		CARE		

$$f_R = (e)(d \vee r) = ed + er$$

If we want to discriminate b/w the objects that are accepted vs that are rejected this is the function.

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

ed :

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

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

Every equivalence class is a rule.

er:

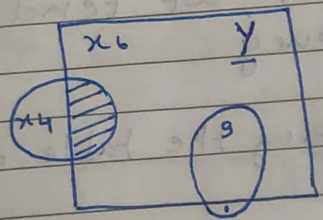
$X_{accept} = \{x_1\}, \{x_4\}, \{x_6\} \{x_7\}$

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

A total of 14 rules are possible. There is no harm in considering rules that are out of reduct. To claim the support of decision we may choose to include more rules.

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

Total no. of decisions in a decision system is called its "Rank". In the last example it was 2 (accept or reject)



Rough membership = degree of overlap

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

for x_4, rough membership $= \frac{1}{2}$

for x_6, rough membership $= 1$

For practical needs we need to define a treshold and then define lower approximations and upper approximation in terms of the treshold.

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

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

Before this point we had assumed $\pi = 1$ $\because$ our table was small.

Dt.
Pg.

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

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

C is a subset of attributes
D is the decision class.

γ is called Rough Dependency.

On increasing π, we tend to increase rough dependency

Eg. Considering the table on the next page

$\pi = 1$

$\gamma(Age, D) = \frac{\cancel{1}}{5} = \frac{3}{5}$

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

$\gamma(Age, D) = \frac{\cancel{2}}{5} = 1$

$\pi = 0.7$

$\gamma(Age, D) = \frac{\cancel{1}}{5} = \frac{3}{5}$ (Both Y & N)

only Y

use this.

Age	Upbringing	Like FB
0 - 10	Rural	Y
10 - 20	Urban	Y
10 - 20	Semi-urban	Y
10 - 20	Rural	N
>20	Urban	N

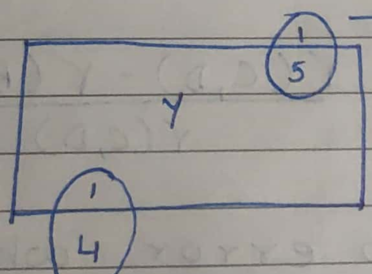
B+

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

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

$\mu_y^{Age}(\{x_2, x_3, x_4\}) = \frac{2}{3}$

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



→ If we introduced a thresold, we try to make it a rule and incorporate it in lower approximation.

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

We are trying to look out for an approximate reduct.

For significance we define 'σ' of an attribute.

$$\underset{\substack{\downarrow \\ \text{feature}}}{\sigma_{C,D}(a)} = \frac{\gamma(C,D) - \gamma(C-\{a\},D)}{\gamma(C,D)}$$

$$\underset{\substack{\downarrow \\ \text{subset of features}}}{\sigma_{C,D}(B)} = 1 - \frac{\gamma(C-B,D)}{\gamma(C,D)}$$

Error :

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

If $E=0$, no error when B is considered instead of C
We define a treshold for error too.

B is a good approximate reduct if $E=0$ or within threshold.

Through this a small number of attributes in a redeact can increase classification accuracy.

This also helps when we have missing data.

Dt.
Pg. B+

Some problems still remains:

Ranges : Should we group ~~discrete~~ individual data points into ranges?

(i) If we do not put them in ranges then we will end up with too many rules.

(ii) If we club them together, we lose information.

So losing information can spoil the classifier but we can club objects with same decisions.

eg. if the decision for all objects for ~~whose~~ which decision is Y, with their attribute a values lying b/w 0-10, can be clubbed together.

Decision discretization is an NP-Hard problem.

U	length	bredth	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

(i) Draw edges b/w all point pairs
Y-N : double edge
N-N & Y-Y : single edge

(ii) Redundant edges are removed using heuristics. Only important edges are retained.

Premium

Region Number

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

NB. We should have minimum number of cuts.

What if the data is not numerical but categorical? Since we cannot have mean for categorical data.

	Color	Decision
Eg.	B	Y
	R	Y
	G	Y
	B	N

→ Can be clubbed together.

We use new binary features instead of symbols $a_v^r(x, y) = 1$, iff $a(x)_v \neq a(y)_r$

New Attributes proportional Variable

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