Course: CS 836

Name: Chao Zhang

Student#: 200383834

1. Consider a finite set U. Give a definition of an equivalence on U. Give a definition of a partition of U. Show that there exists a one-to-one correspondence between the set of all equivalence classes on U and the set of all partitions of U.

For a information table T=(OB, AT, $\{Va \in AT\}, \{Ia|a \in AT\}$), where OB is a finite non-empty set of all objects called the universe, AT is a finite non-empty set of all attributes, Va is the domain of the attribute and Ia is mapping function from ob to Va. Given a subset of attributes $A \subseteq AT$, we define an equivalence relation by: Ea= $\{(x,y) \in OBxOB \mid \forall a \in A\ (Ia(x) = Ia(y))\}$. The equivalence relation induces a partition of OB, denote by $\pi_A=\{[x]_A \mid x \in OB\}$, where $[x]_A=\{y \mid yE_Ax\}$ is the equivalence class containing A. A partition of OB is also called a classification.

2. Give a definition of the refinement-coarsening relation on the set of all partitions on a set U

An equivalence relation E1 is called a refinement of E2 if $E1 \subseteq E2$

3. Give a definition of a consistent and an inconsistent decision table by using the refinement-coarsening relation in the last question.

A consistent decision table can be defined that for any pair $x,y \in OB$, $I_C(x) = I_C(y)$ implies $I_D(x)=I_D(Y)$ and for a inconsistent table we can not get that. By using the refinement-coarsening relation, for a decision table, if $IND(C) \subseteq IND(D)$, it is a consistent table and if $\neg(IND(C) \subseteq IND(D))$, it is a inconsistent table.

4. Give a definition of structured rough set approximations. Explain why the structured approximations are useful for rule mining.

$\underline{sapr}(X) = \{[x] \subseteq IND/OB \mid [x] \subseteq X\}$

$\overline{sapr}(X) = \{[x] \subseteq IND/OB \mid [x] \cap X \neq \phi\}$

For the rule mining where relations are significant, the structured rough set approximations can give a better understanding among data while standard rough set approximations just express the whole frame. It is better for structured approximations to find the connection in the rule.

5. Discuss three types of rules induced by three regions in rough set theory.

In rough set theory, for a subset $X \subseteq OB$, there exists three pair-wise disjoint regions in X which are positive, negative and boundary regions. Positive region for X is the greatest definable set contained by X.

Negative region for X is the greatest definable set contained by X^C and boundary region for X is the complement of the union of positive and negative region. From three way decision point of view, three types of rules can be induced that positive region which is description is for confirmation and acceptance; negative region is for disconfirmation and rejection; boundary region is for non-confirmation. These three rules are important three parts in rough set analysis and application.

6. Give a definition of the following concepts: superfluous attributes, core attributes.

<u>Definitions of superfluous attributes:</u>

Definition 1: In an information table with the set of attributes AT, an attribute $a \in AT$ is a superfluous attribute if $IND(AT-\{a\}) = IND(AT)$

Definition 2: In a consistent decision table with the set of attributes $AT= C \cup D$, an attribute $a \in C$ is a superfluous attribute if $IND(C-\{a\}) \subseteq IND(D)$

Definition 3: In a decision table with the set of attributes $AT = C \cup D$, an attribute $a \in C$ is a superfluous attribute if $POS(\pi_D|\pi_{C-\{a\}})=POS(\pi_D|\pi_C)$, where $\pi_D=OB/IND(D)$ is the partition induced by the set of decision attributes, $\pi_{C-\{a\}}=OB/IND(C-\{a\})$, and $\pi_C=OB/IND(C)$. Equivalently, the condition can be expressed as: $\gamma(C-\{a\}\rightarrow D)=\gamma(C\rightarrow D)$. For D*efinition 3,* it can be used both in consistent and inconsistent decision tables.

<u>Definitions of core attributes:</u>

Definition 4: In an information table with the set of attributes AT, an attribute $a \in AT$ is a core attribute if $\neg(IND(AT-\{a\}) = IND(AT))$

Definition 5: In a consistent decision table with the set of attributes AT= $C \cup D$, an attribute $a \in C$ is a core attribute if $\neg IND(C-\{a\}) \subseteq IND(D)$

Definition 6: In a decision table with the set of attributes AT = $C \cup D$, an attribute $a \in C$ is a superfluous attribute if $\neg POS(\pi_D|\pi_{C-\{a\}})=POS(\pi_D|\pi_C)$, where π_D=OB/IND(D) is the partition induced by the set of decision attributes, $\pi_{C-\{a\}}$=OB/IND(C-{a}), and π_C=OB/IND(C). Equivalently, the condition can be expressed as: $\neg\gamma(C-\{a\}\rightarrow D)=\gamma(C\rightarrow D)$. For D*efinition 6,* it can be used both in consistent and inconsistent decision tables.

7. Give a definition of a reduct of an information table.

In a information table T, if a subset of attributes $R \subseteq AT$ satisfies the following conditions:

Existence: $DEF_{AT}(T)=DEF_{AT}(T)$

Sufficiency: $DEF_R(T)=DEF_{AT}(T)$

Minimization: $\forall a \in R\ (E_{R-\{a\}} \neq E_{AT})$

We call R an attribute reduct of AT

8. Consider the following information table:

OB	c_1	c_2	c_3	c_4	c_5	c_6
x_1	1	1	1	1	1	1
x_2	1	0	1	0	1	1
x_3	0	1	1	1	0	0
x_4	1	1	1	0	0	1
x_5	0	0	1	1	0	1
x_6	1	0	1	0	1	1
x_7	0	0	0	1	1	0
x_8	1	0	1	0	1	1
x_9	0	0	1	1	0	1

(a) Find the set of all superfluous attributes.

(b) Find the set of all core attributes.

(c) Find the set of all reducts of the table.

(a) From the definition of superflous attributes in information table: In an information table with the set of attributes AT, an attribute a $\in$ AT is a superfluous attribute if IND(AT-{a}) = IND(AT). In this table, we can get all superfluous attributes are $\{c_1\},\{c_2\},\{c_3\},\{c_4\},\{c_5\},\{c_6\}$.

(b) From the definition of core attributes in information: In an information table with the set of attributes AT, an attribute a $\in$ AT is a core attribute if $\neg$(IND (AT-{a}) = IND(AT)). In this table, we can get the core attribute is ϕ.

(c) In this table, we can get OB/IND(AT) = $\{\{c_1\}, \{c_2,c_6,c_8\}, \{c_3\}, \{c_4\}, \{c_5,c_9\}, \{c_7\}\}$. From the definition of the reducts in information table with respect to question-7, we can get all reducts of the table are {c1,c2,c5}, {c2,c4,c5}, {c2,c4,c6}, {c2,c5,c6}, {c4,c5,c6}, {c1,c2,c3,c4}.

9. Give a definition of a reduct of a decision table.

In a decision table T with AT=C $\cup$ D, if a subset of condition attributes R$\subseteq$C satisfies the following conditions:

Existence: $POS_C(OB/E_D)=POS_C(OB/E_D)$

Sufficiency: $POS_R(OB/E_D)=POS_C(OB/E_D)$

Minimization: $\forall a \in R\ (POS_{R-\{a\}}(OB/E_D) \neq POS_C(OB/E_D))$

We call R a relative attribute reduct of C

10. Consider the following decision table:

OB	c_1	c_2	c_3	c_4	c_5	c_6	d
x_1	1	1	1	1	1	1	1
x_2	1	0	1	0	1	1	1
x_3	0	1	1	1	0	0	2
x_4	1	1	1	0	0	1	2
x_5	0	0	1	1	0	1	2
x_6	1	0	1	0	1	1	3
x_7	0	0	0	1	1	0	3
x_8	1	0	1	0	1	1	3
x_9	0	0	1	1	0	1	3

(a) Find the set of all superfluous attributes with respect to D = {d}.

(b) Find the set of all core attributes with respect to D = {d}.

(c) Find the set of all reducts of the decision table.

(a) From the definition of superfluous attributes in decision table: In a decision table with the set of attributes AT = C$\cup$D, an attribute a$\in$C is a superfluous attribute if $POS(\pi_D|\pi_{C-\{a\}})=POS\pi_D|\pi_C)$, where π_D=OB/IND(D) is the partition induced by the set of decision attributes, $\pi_{C-\{a\}}=$

OB/IND(C-{a}), and π_C=OB/IND(C). In this table, according to D={d}, POS(π_D |π_C)={{x1}, {x3,x4}, {x7}}, so all superfluous attributes are {c1}, {c2}, {c3}, {c4}, {c5}, {c6}.

(b) From the definition of core attributes in decision table: In a decision table with the set of attributes AT = C ∪ D, an attribute a ∈ C is a superfluous attribute if $\neg$ POS(π_D| $\pi_{C-\{a\}}$)=POS(π_D| π_C), where π_D=OB/IND(D) is the partition induced by the set of decision attributes, $\pi_{C-\{a\}}$=OB/IND(C-{a}), and π_C=OB/IND(C).We can get the core attribute is ϕ.

(c) From the definition of reducts in decision table with respect to question-9, we can get all reducts of the decision table are {c1,c2,c5}, {c2,c4,c5}, {c2,c4,c6}, {c2,c5,c6}, {c4,c5,c6}, {c1,c2,c3,c4}.