Introduction to Rough Set Theory

Davide Ciucci

Dipartimento di Informatica, Sistemistica e Comunicazione
Università di Milano Bicocca

BFTA 2019

Outline

Outline

Rough Set Theory

- A set of tools starting from the notions of rough set
- Z. Pawlak, Rough Sets, International Journal of Parallel Programming 11(5): 341-356 (1982)

Rough Set Theory

- A set of tools starting from the notions of rough set
- Z. Pawlak, Rough Sets, International Journal of Parallel Programming 11(5): 341-356 (1982)
- It includes
  - Knowledge representation: set approximation in different environments

Rough Set Theory

- A set of tools starting from the notions of rough set
- Z. Pawlak, Rough Sets, International Journal of Parallel Programming 11(5): 341-356 (1982)
- It includes
  - Knowledge representation: set approximation in different environments
  - Machine Learning: feature selection, classification, rough clustering, . . .

What is a *rough* set?

More than one definition is possible... some "ingredients" are

What is a *rough* set?

More than one definition is possible... some “ingredients” are

- A set H whose elements are known (extension), but we are not able to describe it (intension)

What is a *rough* set?

More than one definition is possible... some "ingredients" are

- A set H whose elements are known (extension), but we are not able to describe it (intension)
- We are able to give (intension and extension) a pair of sets which are an approximation of H

Outline

Information Table - example

HA = Head Ache
MP = Muscle Pain

Patient	Pressure	HA	Temperature	MP
P1	Normal	yes	38–39	yes
P2	High	no	36–37	yes
P3	High	no	36-37	yes
P4	Low	yes	35–36	no
P5	Normal	yes	36–37	yes

Information Table - example

HA = Head Ache
MP = Muscle Pain

Patient	Pressure	HA	Temperature	MP
P1	Normal	yes	38–39	yes
P2	High	no	36–37	yes
P3	High	no	36-37	yes
P4	Low	yes	35–36	no
P5	Normal	yes	36–37	yes

Information Table - example

HA = Head Ache
MP = Muscle Pain

Patient	Pressure	HA	Temperature	MP
P1	Normal	yes	38–39	yes
P2	High	no	36–37	yes
P3	High	no	36-37	yes
P4	Low	yes	35–36	no
P5	Normal	yes	36–37	yes

Information Table - definition

Definition (Information Table or Information System)

$\mathcal{S}(U) = \langle U, Att, Val, F \rangle$

U set of objects

Information Table - definition

Definition (Information Table or Information System)

$\mathcal{S}(U) = \langle U, Att, Val, F \rangle$

U set of objects

Att set of attributes

Information Table - definition

Definition (Information Table or Information System)

$\mathcal{S}(U) = \langle U, Att, Val, F \rangle$

U set of objects

Att set of attributes

Val set of possible values for the attributes

Information Table - definition

Definition (Information Table or Information System)

$\mathcal{S}(U) = \langle U, Att, Val, F \rangle$
U set of objects
Att set of attributes
Val set of possible values for the attributes
$F : U \times A \mapsto V$ function that assigns to each object a value for any attribute

Information Table - definition

Definition (Information Table or Information System)

$\mathcal{S}(U) = \langle U, Att, Val, F \rangle$
U set of objects
Att set of attributes
Val set of possible values for the attributes
$F : U \times A \mapsto V$ function that assigns to each object a value for any attribute

Sometimes: Val_a with $a \in Att$

Information Table - definition

Definition (Information Table or Information System)

$\mathcal{S}(U) = \langle U, Att, Val, F \rangle$

U set of objects

Att set of attributes

Val set of possible values for the attributes

$F : U \times A \mapsto V$ function that assigns to each object a value for any attribute

Sometimes: Val_a with $a \in Att$

In the example: objects $=\{$P1, ..., P5$\}$, Attributes $= \{$Pressure, HA, Temperature, MP$\}$, Val $=\{$Yes, No, 37-38, ...$\}$

$F(P2, Pressure) = High$

Decision System - Example

HA = Head Ache
MP = Muscle Pain

Patient	Pressure	HA	Temperature	MP	Disease
P1	Normal	yes	38–39	yes	A
P2	High	no	36–37	yes	NO
P3	High	no	36-37	yes	B
P4	Low	yes	35–36	no	NO
P5	Normal	yes	36–37	yes	NO

Decision System - Example

HA = Head Ache
MP = Muscle Pain

Patient	Pressure	HA	Temperature	MP	Disease
P1	Normal	yes	38–39	yes	A
P2	High	no	36–37	yes	NO
P3	High	no	36-37	yes	B
P4	Low	yes	35–36	no	NO
P5	Normal	yes	36–37	yes	NO

Decision Classes: $U_A = \{P1\}$, $U_B = \{P3\}$, $U_{NO} = \{P2, P4, P5\}$

Decision System - Example

HA = Head Ache
MP = Muscle Pain

Patient	Pressure	HA	Temperature	MP	Disease
P1	Normal	yes	38–39	yes	A
P2	High	no	36–37	yes	NO
P3	High	no	36-37	yes	B
P4	Low	yes	35–36	no	NO
P5	Normal	yes	36–37	yes	NO

Decision Classes: $U_A = \{P1\}$, $U_B = \{P3\}$, $U_{NO} = \{P2, P4, P5\}$

P2, P3: same symptoms, different disease $\rightarrow$ the system in inconsistent

Definitions

Definition (Decision System)

$\mathcal{S}(U) = \langle U, C \cup \{d\}, Val, F\rangle$

- U set of objects
- C set of condition attributes
- d decision attribute
- Val set of possible values for the attributes

Definitions

Definition (Decision System)

$\mathcal{S}(U) = \langle U, C \cup \{d\}, Val, F \rangle$

- U set of objects
- C set of condition attributes
- d decision attribute
- Val set of possible values for the attributes
- $F : U \times C \cup \{d\} \mapsto V$ function that assigns to each object a value for any attribute

Definitions

Definition (Decision System)

$\mathcal{S}(U) = \langle U, C \cup \{d\}, Val, F\rangle$

- U set of objects
- C set of condition attributes
- d decision attribute
- Val set of possible values for the attributes
- $F : U \times C \cup \{d\} \mapsto V$ function that assigns to each object a value for any attribute

Definition (Consistent Decision System)

There are no two objects $O_1, O_2 \in U$ with same value for condition attributes and different decision

Generalized Decision

Patient	Pressure	HA	Temperature	MP	Disease
P1	Normal	yes	38–39	yes	A
P2	High	no	36–37	yes	NO
P3	High	no	36-37	yes	B
P4	Low	yes	35–36	no	NO
P5	Normal	yes	36–37	yes	NO

- Generalized decision: $\delta_A : U \to \mathcal{P}(Val)$

Generalized Decision

Patient	Pressure	HA	Temperature	MP	Disease
P1	Normal	yes	38–39	yes	A
P2	High	no	36–37	yes	NO
P3	High	no	36-37	yes	B
P4	Low	yes	35–36	no	NO
P5	Normal	yes	36–37	yes	NO

- Generalized decision: $\delta_A : U \to \mathcal{P}(Val)$
- Example: $\delta_{ATT}(P2) = \{NO, B\}$

Generalized Decision

Patient	Pressure	HA	Temperature	MP	Disease
P1	Normal	yes	38–39	yes	A
P2	High	no	36–37	yes	NO
P3	High	no	36-37	yes	B
P4	Low	yes	35–36	no	NO
P5	Normal	yes	36–37	yes	NO

- Generalized decision: $\delta_A : U \to \mathcal{P}(Val)$
- Example: $\delta_{ATT}(P2) = \{NO, B\}$
- Definition:

$$\delta_A(x) = \{i \in Val : \exists y, x\, I_A y \text{ and } F(y,d) = i\}$$

Generalized Decision

Patient	Pressure	HA	Temperature	MP	Disease
P1	Normal	yes	38–39	yes	A
P2	High	no	36–37	yes	NO
P3	High	no	36-37	yes	B
P4	Low	yes	35–36	no	NO
P5	Normal	yes	36–37	yes	NO

- Generalized decision: $\delta_A : U \to \mathcal{P}(Val)$
- Example: $\delta_{ATT}(P2) = \{NO, B\}$
- Definition:

$$\delta_A(x) = \{i \in Val : \exists y, x \, I_A y \text{ and } F(y,d) = i\}$$

- If $\forall x \in U : |\delta_A(x)| = 1$ then the system is consistent

Generalized Decision - example

Patient	Pressure	HA	Temperature	MP	Disease	δ_{Att}
P1	Normal	yes	38–39	yes	A	A,
P2	High	no	36–37	yes	NO	B,NO
P3	High	no	36-37	yes	B	B,NO
P4	Low	yes	35–36	no	NO	NO
P5	Normal	yes	36–37	yes	NO	NO

Outline

Outline

Indiscernibility relation - example

Patient	Pressure	HA	Temperature	MP
P1	Normal	yes	38–39	yes
P2	High	no	36–37	yes
P3	High	no	36-37	yes
P4	Low	yes	35–36	no
P5	Normal	yes	36–37	yes

P2 and P3 same values for all attributes: they are indiscernible (indistinguishable, equivalent, ...)

Indiscernibility relation - example

Patient	Pressure	HA	Temperature	MP
P1	Normal	yes	38–39	yes
P2	High	no	36–37	yes
P3	High	no	36-37	yes
P4	Low	yes	35–36	no
P5	Normal	yes	36–37	yes

P2 and P3 same values for all attributes: they are indiscernible (indistinguishable, equivalent, ...)

{P2, P3} is a granule of information

Indiscernibility relation - example

Patient	Pressure	HA	Temperature	MP
P1	Normal	yes	38–39	yes
P2	High	no	36–37	yes
P3	High	no	36-37	yes
P4	Low	yes	35–36	no
P5	Normal	yes	36–37	yes

P2 and P3 same values for all attributes: they are indiscernible (indistinguishable, equivalent, ...)

{P2, P3} is a granule of information

A partition of the universe: $\Pi = \{P1\}, \{P2,P3\}, \{P4\}, \{P5\}$

Indiscernibility relation - definition

Definition (Indiscernibility)

Given a set of attributes $A \subseteq Att$

Indiscernibility relation - definition

Definition (Indiscernibility)

Given a set of attributes $A \subseteq Att$

$x, y \in U$ are indiscernible with respect to A if

Indiscernibility relation - definition

Definition (Indiscernibility)

Given a set of attributes $A \subseteq Att$
$x, y \in U$ are indiscernible with respect to A if

$$\forall a \in A \quad F(a, x) = F(a, y)$$

In this case we write $x I_A y$

Indiscernibility relation - definition

Definition (Indiscernibility)

Given a set of attributes $A \subseteq Att$
$x, y \in U$ are indiscernible with respect to A if

$$\forall a \in A \quad F(a, x) = F(a, y)$$

In this case we write xI_Ay

I_A is an equivalence relation: reflexive, symmetric, transitive
I_A partitions U in equivalence classes (the granules of information)

$$[x]_A := \{y \in U : xI_Ay\}$$

Approximations - example

Partition {P1}, {P2,P3}, {P4}, {P5}

Approximations - example

Partition {P1}, {P2,P3}, {P4}, {P5}

- The set $H = \{P1, P2, P3\}$ is the union of two equivalence classes $\{P1\} \cup \{P2, P3\}$
- The set $K = \{P1, P2\}$ is not

Approximations - example

Partition {P1}, {P2,P3}, {P4}, {P5}

- The set $H = \{P1, P2, P3\}$ is the union of two equivalence classes $\{P1\} \cup \{P2, P3\}$
- The set $K = \{P1, P2\}$ is not
- H is exact, K is rough

Approximations - example

Partition {P1}, {P2,P3}, {P4}, {P5}

- The set $H = \{P1, P2, P3\}$ is the union of two equivalence classes $\{P1\} \cup \{P2, P3\}$
- The set $K = \{P1, P2\}$ is not
- H is exact, K is rough
- K can be approximated by a pair of exact sets: {P1}, {P1,P2,P3}

$$\{P1\} \subseteq K \subseteq \{P1, P2, P3\}$$

Approximations - definition

Definition (Approximations)

Let $S(U) = \langle U, Att, val(U), F \rangle$ be an information table (a decision system)
Given a set of attributes $A \subseteq Att$, then for any set of objects $H \subseteq U$ we define

Approximations - definition

Definition (Approximations)

Let $S(U) = \langle U, Att, val(U), F \rangle$ be an information table (a decision system)
Given a set of attributes $A \subseteq Att$, then for any set of objects $H \subseteq U$ we define
the lower approximation of H:

$$L(H) := \{x : [x]_A \subseteq H\}$$

Approximations - definition

Definition (Approximations)

Let $S(U) = \langle U, Att, val(U), F \rangle$ be an information table (a decision system)
Given a set of attributes $A \subseteq Att$, then for any set of objects $H \subseteq U$ we define
the lower approximation of H:

$$L(H) := \{x : [x]_A \subseteq H\}$$

the upper approximation of H:

$$U(H) := \{x : [x]_A \cap H \neq \emptyset\}$$

Approximations - definition

Definition (Approximations)

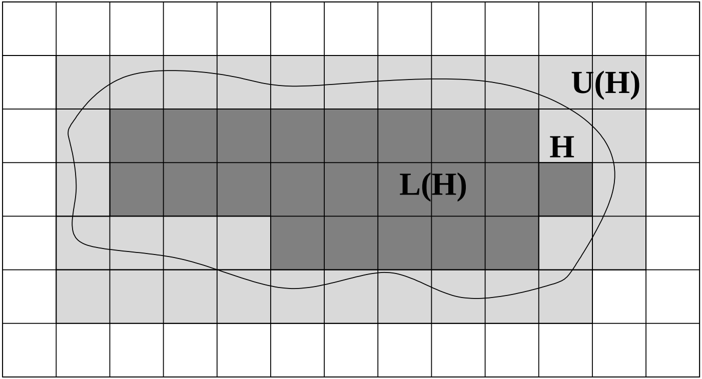
Let $S(U) = \langle U, Att, val(U), F \rangle$ be an information table (a decision system)
Given a set of attributes $A \subseteq Att$, then for any set of objects $H \subseteq U$ we define
the lower approximation of H:

$$L(H) := \{x : [x]_A \subseteq H\}$$

the upper approximation of H:

$$U(H) := \{x : [x]_A \cap H \neq \emptyset\}$$

The pair $r(H) = \langle L(H), U(H) \rangle$ is named *rough approximation* (or rough set)


U(H)
H
L(H)

Further regions

Exterior $E(H) = U^c(H) \quad L(H) \cap E(H) = \emptyset$
Rough approximation: $(L(H), E(H))$

Further regions

Exterior $E(H) = U^c(H) \quad L(H) \cap E(H) = \emptyset$
Rough approximation: $(L(H), E(H))$

Boundary $Bnd(H) = U(H) \setminus L(H)$

Further regions

Exterior $E(H) = U^c(H) \quad L(H) \cap E(H) = \emptyset$
Rough approximation: $(L(H), E(H))$

Boundary $Bnd(H) = U(H) \setminus L(H)$

Interpretation

Lower	sure belong to H
Exterior	sure not belong to H
Boundary	uncertain

Outline

Motivation

- An equivalence relation can be too strict $\rightarrow$ weaker relations
  - Similarity (no transitivity). Example: 36.5 is similar to 36.7; equality on x% attributes
  - Dominance (no symmetry). Very good $\geq$ good $\geq$ bad

Generic relation

- R a binary relation on U: $R \subseteq U \times U$

Generic relation

- R a binary relation on U: $R \subseteq U \times U$
- Granule of information $g_R(x) = \{y \in U : x \, R \, y\}$

Generic relation

- R a binary relation on U: $R \subseteq U \times U$
- Granule of information $g_R(x) = \{y \in U : x\, R\, y\}$
- Approximations

$$l_R(H) = \{x \in U : gr(x) \subseteq H\}$$
$$u_R(H) = \{x \in U : gr(x) \cap H \neq \emptyset\}$$

Generic relation

- R a binary relation on U: $R \subseteq U \times U$
- Granule of information $g_R(x) = \{y \in U : x\,R\,y\}$
- Approximations

$$l_R(H) = \{x \in U : gr(x) \subseteq H\}$$
$$u_R(H) = \{x \in U : gr(x) \cap H \neq \emptyset\}$$

Properties

- $l_R(U) = U$, $u_R(\emptyset) = \emptyset$, l, u are monotone

Generic relation

- R a binary relation on U: $R \subseteq U \times U$
- Granule of information $g_R(x) = \{y \in U : x\, R\, y\}$
- Approximations

$$l_R(H) = \{x \in U : gr(x) \subseteq H\}$$
$$u_R(H) = \{x \in U : gr(x) \cap H \neq \emptyset\}$$

Properties

- $l_R(U) = U$, $u_R(\emptyset) = \emptyset$, l, u are monotone
- R serial: $l_R(H) \subseteq u_R(H)$, $u_R(U) = U$, $l_R(\emptyset) = \emptyset$

Generic relation

- R a binary relation on U: $R \subseteq U \times U$
- Granule of information $g_R(x) = \{y \in U : x\, R\, y\}$
- Approximations

$$l_R(H) = \{x \in U : gr(x) \subseteq H\}$$
$$u_R(H) = \{x \in U : gr(x) \cap H \neq \emptyset\}$$

Properties

- $l_R(U) = U$, $u_R(\emptyset) = \emptyset$, l, u are monotone
- R serial: $l_R(H) \subseteq u_R(H)$, $u_R(U) = U$, $l_R(\emptyset) = \emptyset$
- R reflexive: $l_R(H) \subseteq H \subseteq u_R(H)$

Similarity relation

Rough sets based on a similarity relation $\mathcal{R}$

- Reflexive
- Symmetric

Similarity relation

Rough sets based on a similarity relation $\mathcal{R}$

- Reflexive
- Symmetric

Similarity $S(x) := \{y \in U : x\mathcal{R}y\}$

Similarity relation

Rough sets based on a similarity relation $\mathcal{R}$

- Reflexive
- Symmetric

Similarity $S(x) := \{y \in U : x\mathcal{R}y\}$

$\Rightarrow$ A covering of the universe, not a partition

- $\bigcup_x S(x) = U$
- there can exist objects x and y such that $S(x) \cap S(y) \neq \emptyset$

Similarity: Example 1

$\mathcal{R}$ can represent a distance between objects

Similarity: Example 1

$\mathcal{R}$ can represent a distance between objects

- Similar temperature if $|Temp(P1) - Temp(P2)| \leq 0.5$

Similarity: Example 1

$\mathcal{R}$ can represent a distance between objects

- Similar temperature if $|Temp(P1) - Temp(P2)| \leq 0.5$
- P1 similar to P2 if they are equal on (at least) half of the attributes

$$\frac{|\{a_i \in Att : F(a_i, P1) = F(a_i, P2)\}|}{|Att|} \geq \frac{1}{2}$$

Similarity: Example 2

- Deal with incomplete (missing) information

Similarity: Example 2

- Deal with incomplete (missing) information

Patient	Pressure	HA	Temperature	MP	Malattia
P1	Normal	yes	38–39	yes	A
P2	High	no	36–37	*	NO
P3	High	no	*	yes	B
P4	*	yes	35–36	no	NO
P5	Normal	*	*	yes	NO

Similarity: Example 2

- Deal with incomplete (missing) information

Patient	Pressure	HA	Temperature	MP	Malattia
P1	Normal	yes	38–39	yes	A
P2	High	no	36–37	*	NO
P3	High	no	*	yes	B
P4	*	yes	35–36	no	NO
P5	Normal	*	*	yes	NO

$x\mathcal{R}_D y$ iff $\forall a_i \in D \quad F(x, a_i) = F(y, a_i)$ or $F(x, a_i) = *$ or $F(y, a_i) = *$

Similarity: Example 3

- Deal with incomplete (partial) information

Similarity: Example 3

- Deal with incomplete (partial) information

Patient	Pressure	HA	Temperature	MP
P1	{Low,Normal}	yes	38–39	yes
P2	{Normal,High}	no	36–37	yes
P3	High	no	36-37	yes
P4	Low	yes	35–36	no
P5	Normal	yes	36–37	yes

Similarity: Example 3

- Deal with incomplete (partial) information

Patient	Pressure	HA	Temperature	MP
P1	{Low,Normal}	yes	38–39	yes
P2	{Normal,High}	no	36–37	yes
P3	High	no	36-37	yes
P4	Low	yes	35–36	no
P5	Normal	yes	36–37	yes

$$x\mathcal{R}_D y \text{ iff } \forall a_i \in D \quad F(x, a_i) \cap F(y, a_i) \neq \emptyset$$

Approximation Spaces - Idea

- The granulation of the universe in given for granted and it is the starting point
- No more necessarily a partition, but it can be a covering or a partial covering, or even a more complex structure.
- Generally, it comes with a set of axioms that the granules or the induced approximations should satisfy

Definition

Definition

An approximation space is a pair $\langle X, G(X) \rangle$
$G(X) = \{G(x_1), G(x_2), \ldots, G(x_n) : x_i \in X\}$ such that
$\cup_i G(x_i) = X$

Definition

Definition

An approximation space is a pair $\langle X, G(X) \rangle$
$G(X) = \{G(x_1), G(x_2), \ldots, G(x_n) : x_i \in X\}$ such that
$\cup_i G(x_i) = X$

- $G(X)$ is a granulation of the universe, that is a collection of sets
- Often some conditions are imposed to the granulation or to the approximation
- Any binary relation forms a granulation

Approximations

- Covering approximation space are the most studied case of approximation space
- More than 30 lower and upper approximations exists on covering approximation spaces
- Here, we require:
  1. $L(X^c) = [U(X)]^c$
  2. $L(X) \subseteq X$
  3. If $X \subseteq Y$ then $L(X) \subseteq L(Y)$

Outline

Outline

Aim: feature selection + classification

- Simplify the table: eliminate “useless” attributes

Aim: feature selection + classification

- Simplify the table: eliminate “useless” attributes
- Given a decision system, found the rules:

$$\text{condition attribute} \rightarrow \text{decision}$$

Aim: feature selection + classification

- Simplify the table: eliminate "useless" attributes
- Given a decision system, found the rules:

$$\text{condition attribute} \to \text{decision}$$

Example:

If (Pressure = Normal) AND (Temp. = 38–39) THEN Disease = A

Inf. Table Reduct - example

Patient	Pressure	HA	Temperature	MP
P1	Normal	yes	38–39	yes
P2	High	no	36–37	yes
P3	High	no	36-37	yes
P4	Low	yes	35–36	no
P5	Normal	yes	36–37	yes

Π_{Att} = {P1}, {P2,P3}, {P4}, {P5}

Inf. Table Reduct - example

Patient	Pressure	HA	[redacted]	MP
P1	Normal	yes	[redacted]	yes
P2	High	no	[redacted]	yes
P3	High	no	[redacted]	yes
P4	Low	yes	[redacted]	no
P5	Normal	yes	[redacted]	yes

$\Pi_{Att} = \{P1\}, \{P2,P3\}, \{P4\}, \{P5\}$

$\Pi_{Att \setminus \{Temp\}} = \{P1,P5\}, \{P2,P3\}, \{P4\}$

Inf. Table Reduct - example

Patient	Pressure	[redacted]	Temperature	MP
P1	Normal	[redacted]	38–39	yes
P2	High	[redacted]	36–37	yes
P3	High	[redacted]	36-37	yes
P4	Low	[redacted]	35–36	no
P5	Normal	[redacted]	36–37	yes

$\Pi_{Att} = \{P1\}, \{P2,P3\}, \{P4\}, \{P5\}$

$\Pi_{Att \setminus \{HA\}} = \Pi_{Att}$

Inf. Table Reduct - example

Patient	Pressure	[redacted]	Temperature	[redacted]
P1	Normal	[redacted]	38–39	[redacted]
P2	High	[redacted]	36–37	[redacted]
P3	High	[redacted]	36-37	[redacted]
P4	Low	[redacted]	35–36	[redacted]
P5	Normal	[redacted]	36–37	[redacted]

$\Pi_{Att} = \{P1\}, \{P2,P3\}, \{P4\}, \{P5\}$

$\Pi_{Att \setminus \{HA,MP\}} = \Pi_{Att}$

Inf. Table Reduct - example

Patient	[redacted]	Temperature	[redacted]
P1	[redacted]	38–39	[redacted]
P2	[redacted]	36–37	[redacted]
P3	[redacted]	36-37	[redacted]
P4	[redacted]	35–36	[redacted]
P5	[redacted]	36–37	[redacted]

$\Pi_{Att} = \{P1\}, \{P2,P3\}, \{P4\}, \{P5\}$

$\Pi_{Temp} = \{P1\}, \{P2,P3,P5\}, \{P4\}$

Inf. Table Reduct - example

Patient	Pressure
P1	Normal
P2	High
P3	High
P4	Low
P5	Normal

[redacted]

$\Pi_{Att} = \{P1\}, \{P2,P3\}, \{P4\}, \{P5\}$

$\Pi_{Pressure} = \{P1,P5\}, \{P2,P3\}, \{P4\}$

Inf. Table Reduct - example

Patient	Pressure	[redacted]	Temperature	[redacted]
P1	Normal	[redacted]	38–39	[redacted]
P2	High	[redacted]	36–37	[redacted]
P3	High	[redacted]	36-37	[redacted]
P4	Low	[redacted]	35–36	[redacted]
P5	Normal	[redacted]	36–37	[redacted]

$\Pi_{Att} = \{P1\}, \{P2,P3\}, \{P4\}, \{P5\}$

$\Pi_{Pressure,Temperature} = \Pi_{Att}$

$\{$Pressure, Temperature$\}$ is a reduct of *Att*

Inf. Table Reduct - example

Patient	Pressure	[redacted]	Temperature	[redacted]
P1	Normal	[redacted]	38–39	[redacted]
P2	High	[redacted]	36–37	[redacted]
P3	High	[redacted]	36-37	[redacted]
P4	Low	[redacted]	35–36	[redacted]
P5	Normal	[redacted]	36–37	[redacted]

$\Pi_{Att} = \{P1\}, \{P2,P3\}, \{P4\}, \{P5\}$

$\Pi_{Pressure, Temperature} = \Pi_{Att}$

{Pressure, Temperature} is a reduct of *Att*

→ Feature Selection

Inf. Table Reduct - definition

Definition (Reduct)

$A \subseteq B \subseteq Att$

A is a reduct of B if

Inf. Table Reduct - definition

Definition (Reduct)

$A \subseteq B \subseteq Att$

A is a reduct of B if

1. $\Pi_A = \Pi_B$

Inf. Table Reduct - definition

Definition (Reduct)

$A \subseteq B \subseteq Att$

A is a reduct of B if

1. $\Pi_A = \Pi_B$
2. Minimality: $\nexists C \subset A$ and $\Pi_C = \Pi_B$

Inf. Table Reduct - definition

Definition (Reduct)

$A \subseteq B \subseteq Att$

A is a reduct of B if

1. $\Pi_A = \Pi_B$
2. Minimality: $\nexists C \subset A$ and $\Pi_C = \Pi_B$

$a \in A \subseteq Att$ is indispensable in A if $\Pi_A \neq \Pi_{A \setminus \{a\}}$

Inf. Table Reduct - definition

Definition (Reduct)

$A \subseteq B \subseteq Att$

A is a reduct of B if

1. $\Pi_A = \Pi_B$
2. Minimality: $\nexists C \subset A$ and $\Pi_C = \Pi_B$

$a \in A \subseteq Att$ is indispensable in A if $\Pi_A \neq \Pi_{A \setminus \{a\}}$

CORE = set of indispensable attributes in Att = intersection of all reducts in Att

Complexity issues

- Given n attributes, there are at most $O(\frac{3^n}{\sqrt{n}})$ reducts

Complexity issues

- Given n attributes, there are at most $O(\frac{3^n}{\sqrt{n}})$ reducts
- Find the shortest reduct is a NP^{NP} complete problem

Complexity issues

- Given n attributes, there are at most $O(\frac{3^n}{\sqrt{n}})$ reducts
- Find the shortest reduct is a NP^{NP} complete problem

Solutions

- Heuristics (Approximate reducts, genetic algorithms, ...)
- Parallel algorithms

Reduct

Definition (Reduct)

A reduct is a minimal subset of conditions $C \subseteq ATT$ (features) that preserves the classification wrt the decision attribute

Reduct

Definition (Reduct)

A reduct is a minimal subset of conditions $C \subseteq ATT$ (features) that preserves the classification wrt the decision attribute

1. Consistence: same ability of the whole ATT to distinguish objects belonging to two different decision classes

Reduct

Definition (Reduct)

A reduct is a minimal subset of conditions $C \subseteq ATT$ (features) that preserves the classification wrt the decision attribute

1. Consistence: same ability of the whole ATT to distinguish objects belonging to two different decision classes
2. Minimality: any smaller subset is not consistent

Example

Patient	Pressure	HA	Temperature	MP	Disease
P1	Normal	yes	38–39	yes	A
P3	High	no	36-37	yes	B
P4	Low	yes	35–36	no	NO
P5	Normal	yes	36–37	yes	NO

Example

Patient	Pressure	HA	[redacted]	MP	Disease
P1	Normal	yes		yes	A
P3	High	no		yes	B
P4	Low	yes		no	NO
P5	Normal	yes		yes	NO

Example

Patient	Pressure	HA	[redacted]	MP	Disease
P1	Normal	yes	[redacted]	yes	A
P3	High	no	[redacted]	yes	B
P4	Low	yes	[redacted]	no	NO
P5	Normal	yes	[redacted]	yes	NO

Example

Patient	Pressure	[redacted]	Temperature	MP	Disease
P1	Normal	[redacted]	38–39	yes	A
P3	High	[redacted]	36-37	yes	B
P4	Low	[redacted]	35–36	no	NO
P5	Normal	[redacted]	36–37	yes	NO

Example

Patient	Pressure	[redacted]	Temperature	[redacted]	Disease
P1	Normal	[redacted]	38–39	[redacted]	A
P3	High	[redacted]	36-37	[redacted]	B
P4	Low	[redacted]	35–36	[redacted]	NO
P5	Normal	[redacted]	36–37	[redacted]	NO

Example

Patient	[redacted]	Temperature	[redacted]	Disease
P1	[redacted]	38–39	[redacted]	A
P3	[redacted]	36-37	[redacted]	B
P4	[redacted]	35–36	[redacted]	NO
P5	[redacted]	36–37	[redacted]	NO

Example

Patient	Pressure	[redacted]	Disease
P1	Normal		A
P3	High		B
P4	Low		NO
P5	Normal		NO

Example: rules

Patient	Pressure	[redacted]	Temperature	[redacted]	Disease
P1	Normal	[redacted]	38–39	[redacted]	A
P3	High	[redacted]	36-37	[redacted]	B
P4	Low	[redacted]	35–36	[redacted]	NO
P5	Normal	[redacted]	36–37	[redacted]	NO

Reduct $=$ {Pressure, Temperature}

Example: rules

Patient	Pressure	[redacted]	Temperature	[redacted]	Disease
P1	Normal	[redacted]	38–39	[redacted]	A
P3	High	[redacted]	36-37	[redacted]	B
P4	Low	[redacted]	35–36	[redacted]	NO
P5	Normal	[redacted]	36–37	[redacted]	NO

Reduct $=$ {Pressure, Temperature}

IF Pressure $=$ Normal AND Temp. $=$ 38–39 THEN Disease $=$ A

Example: rules

Patient	Pressure	[redacted]	Temperature	[redacted]	Disease
P1	Normal	[redacted]	38–39	[redacted]	A
P3	High	[redacted]	36-37	[redacted]	B
P4	Low	[redacted]	35–36	[redacted]	NO
P5	Normal	[redacted]	36–37	[redacted]	NO

Reduct = {Pressure, Temperature}

IF Pressure = Normal AND Temp. = 38–39 THEN Disease = A
IF Pressure = High AND Temp. = 36–37 THEN Disease = B

Example: rules

Patient	Pressure	[redacted]	Temperature	[redacted]	Disease
P1	Normal	[redacted]	38–39	[redacted]	A
P3	High	[redacted]	36-37	[redacted]	B
P4	Low	[redacted]	35–36	[redacted]	NO
P5	Normal	[redacted]	36–37	[redacted]	NO

Reduct = {Pressure, Temperature}

IF Pressure = Normal AND Temp. = 38–39 THEN Disease = A
IF Pressure = High AND Temp. = 36–37 THEN Disease = B
IF Pressure = Low AND Temp. = 35–36 THEN Disease = NO

Example: rules

Patient	Pressure	[redacted]	Temperature	[redacted]	Disease
P1	Normal	[redacted]	38–39	[redacted]	A
P3	High	[redacted]	36-37	[redacted]	B
P4	Low	[redacted]	35–36	[redacted]	NO
P5	Normal	[redacted]	36–37	[redacted]	NO

Reduct = {Pressure, Temperature}

IF Pressure = Normal AND Temp. = 38–39 THEN Disease = A
IF Pressure = High AND Temp. = 36–37 THEN Disease = B
IF Pressure = Low AND Temp. = 35–36 THEN Disease = NO
IF Pressure = Normal AND Temp. = 36–37 THEN Disease = NO

Solution 1: Generalized Decision

Patient	Pressure	HA	Temperature	MP	Disease
P1	Normal	yes	38–39	yes	A
P2	High	no	36–37	yes	NO
P3	High	no	36-37	yes	B
P4	Low	yes	35–36	no	NO
P5	Normal	yes	36–37	yes	NO

- Generalized decision: $\delta_A : U \to \mathcal{P}(Val)$

Solution 1: Generalized Decision

Patient	Pressure	HA	Temperature	MP	Disease
P1	Normal	yes	38–39	yes	A
P2	High	no	36–37	yes	NO
P3	High	no	36-37	yes	B
P4	Low	yes	35–36	no	NO
P5	Normal	yes	36–37	yes	NO

- Generalized decision: $\delta_A : U \to \mathcal{P}(Val)$
- Example: $\delta_{ATT}(P2) = \{NO, B\}$

Solution 1: Generalized Decision

Patient	Pressure	HA	Temperature	MP	Disease
P1	Normal	yes	38–39	yes	A
P2	High	no	36–37	yes	NO
P3	High	no	36-37	yes	B
P4	Low	yes	35–36	no	NO
P5	Normal	yes	36–37	yes	NO

- Generalized decision: $\delta_A : U \to \mathcal{P}(Val)$
- Example: $\delta_{ATT}(P2) = \{NO, B\}$
- Definition:

$$\delta_A(x) = \{i \in Val : \exists y, x\, I_A\, y \text{ and } F(y, d) = i\}$$

Solution 1: Generalized Decision

Patient	Pressure	HA	Temperature	MP	Disease
P1	Normal	yes	38–39	yes	A
P2	High	no	36–37	yes	NO
P3	High	no	36-37	yes	B
P4	Low	yes	35–36	no	NO
P5	Normal	yes	36–37	yes	NO

- Generalized decision: $\delta_A : U \to \mathcal{P}(Val)$
- Example: $\delta_{ATT}(P2) = \{NO, B\}$
- Definition:

$$\delta_A(x) = \{i \in Val : \exists y, x\, I_A\, y \text{ and } F(y, d) = i\}$$

- If $\forall x \in U : |\delta_A(x)| = 1$ then the system is consistent

Generalized Decision reduct

Definition

Given a set of attributes $A \subseteq B \subseteq ATT$, A is a reduct of B if

1. $\delta_A = \delta_B$ (I do not introduce further inconsistency)
2. Minimality: $\nexists C \subset A$ such that $\delta_C = \delta_B$

Generalized Decision Reduct - example

Patient	Pressure	HA	Temperature	MP	Disease	δ_{Att}
P1	Normal	yes	38–39	yes	A	A,
P2	High	no	36–37	yes	NO	B,NO
P3	High	no	36-37	yes	B	NO
P4	Low	yes	35–36	no	NO	NO
P5	Normal	yes	36–37	yes	NO	B,NO

Generalized Decision Reduct - example

Patient	Pressure	HA	Temperature	MP	Disease	δ_{Att}
P1	Normal	yes	38–39	yes	A	A,
P2	High	no	36–37	yes	NO	B,NO
P3	High	no	36-37	yes	B	NO
P4	Low	yes	35–36	no	NO	NO
P5	Normal	yes	36–37	yes	NO	B,NO

- Reduct $\{Pressure, Temperature\}$

Generalized Decision Reduct - example

Patient	Pressure	HA	Temperature	MP	Disease	δ_{Att}
P1	Normal	yes	38–39	yes	A	A,
P2	High	no	36–37	yes	NO	B,NO
P3	High	no	36-37	yes	B	NO
P4	Low	yes	35–36	no	NO	NO
P5	Normal	yes	36–37	yes	NO	B,NO

- Reduct $\{Pressure, Temperature\}$
- If (Pressure =High) AND (Temp=36-37) THEN (Disease = NO) OR (Disease = B)

Solution 2: Dependence

Solution 2: Dependence

Definition

Let $\mathcal{S}(U)$ be a decision system

$A \subseteq Att$ a set of attributes, X_i the decision classes

Solution 2: Dependence

Definition

Let $\mathcal{S}(U)$ be a decision system

$A \subseteq Att$ a set of attributes, X_i the decision classes

- The Positive Region: $Pos = \cup L_A(X_i)$
- The Coefficient of Dependence of decision d from A is

$$Dip(A, d) = \frac{|Pos|}{|X|}$$

Solution 2: Dependence

Definition

Let $\mathcal{S}(U)$ be a decision system
$A \subseteq Att$ a set of attributes, X_i the decision classes

- The Positive Region: $Pos = \cup L_A(X_i)$
- The Coefficient of Dependence of decision d from A is

$$Dip(A, d) = \frac{|Pos|}{|X|}$$

- Pos contains the objects that are correctly classified by the set of attributes A
- Dip(A,d) = 1 if the system is consistent

Reduct: dependence definition

Definition (Reduct)

Let $\mathcal{S}(U)$ be a decision system

Reduct: dependence definition

Definition (Reduct)

Let $\mathcal{S}(U)$ be a decision system
$A \subseteq B \subseteq Att$, A is a reduct of B if

Reduct: dependence definition

Definition (Reduct)

Let $\mathcal{S}(U)$ be a decision system
$A \subseteq B \subseteq Att$, A is a reduct of B if

1. Dip(A,d)=Dip(B,d)

Reduct: dependence definition

Definition (Reduct)

Let $\mathcal{S}(U)$ be a decision system
$A \subseteq B \subseteq Att$, A is a reduct of B if

1. Dip(A,d)=Dip(B,d)
2. Minimality: $\nexists C \subset A$ such that $Dip(C, d) = Dip(B, d)$

Reduct: dependence example

Patient	Pressure	HA	Temperature	DM	Disease
P1	Normal	yes	38–39	yes	A
P2	High	no	36–37	yes	NO
P3	High	no	36-37	yes	B
P4	Low	yes	35–36	no	NO
P5	Normal	yes	36–37	yes	NO

$L_C(X_A) = \{P1\}$, $L_C(X_{NO}) = \{P4, P5\}$, $L_C(X_B) = \emptyset$

Reduct: dependence example

Patient	Pressure	HA	Temperature	DM	Disease
P1	Normal	yes	38–39	yes	A
P2	High	no	36–37	yes	NO
P3	High	no	36-37	yes	B
P4	Low	yes	35–36	no	NO
P5	Normal	yes	36–37	yes	NO

$L_C(X_A) = \{P1\}$, $L_C(X_{NO}) = \{P4, P5\}$, $L_C(X_B) = \emptyset$

$Dip(Att, Disease) = \frac{3}{5}$

Reduct: dependence example

Patient	Pressure	HA	Temperature	DM	Disease
P1	Normal	yes	38–39	yes	A
P2	High	no	36–37	yes	NO
P3	High	no	36-37	yes	B
P4	Low	yes	35–36	no	NO
P5	Normal	yes	36–37	yes	NO

$L_C(X_A) = \{P1\}$, $L_C(X_{NO}) = \{P4, P5\}$, $L_C(X_B) = \emptyset$

$Dip(Att, Disease) = \frac{3}{5}$
$Dip(\{Pressure, Temperature, DM\}, Disease) = \frac{3}{5}$

Reduct: dependence example

Patient	Pressure	HA	Temperature	DM	Disease
P1	Normal	yes	38–39	yes	A
P2	High	no	36–37	yes	NO
P3	High	no	36-37	yes	B
P4	Low	yes	35–36	no	NO
P5	Normal	yes	36–37	yes	NO

$L_C(X_A) = \{P1\}$, $L_C(X_{NO}) = \{P4, P5\}$, $L_C(X_B) = \emptyset$

$Dip(Att, Disease) = \frac{3}{5}$
$Dip(\{Pressure, Temperature, DM\}, Disease) = \frac{3}{5}$
$Dip(\{Pressure, Temperature\}, Disease) = \frac{3}{5}$

Reduct: dependence example

Patient	Pressure	HA	Temperature	DM	Disease
P1	Normal	yes	38–39	yes	A
P2	High	no	36–37	yes	NO
P3	High	no	36-37	yes	B
P4	Low	yes	35–36	no	NO
P5	Normal	yes	36–37	yes	NO

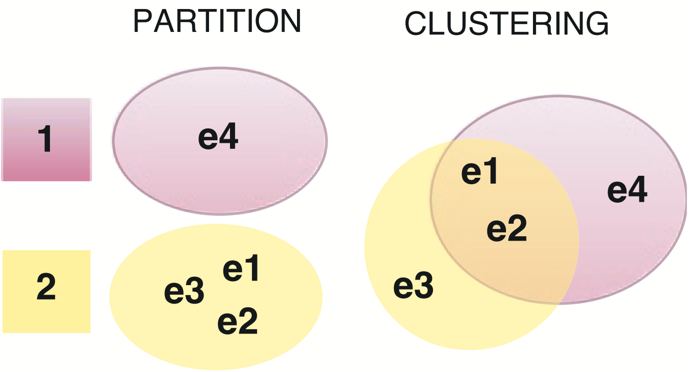
$L_C(X_A) = \{P1\}$, $L_C(X_{NO}) = \{P4, P5\}$, $L_C(X_B) = \emptyset$

$Dip(Att, Disease) = \frac{3}{5}$
$Dip(\{Pressure, Temperature, DM\}, Disease) = \frac{3}{5}$
$Dip(\{Pressure, Temperature\}, Disease) = \frac{3}{5}$

IF (Pressure=High AND Temp= 36–37) THEN (Disease =NO OR Disease =B)

Outline

Example of Soft Clustering



Rough K-means

- A soft clustering technique
- Inspired by rough sets
  - each cluster C_i is made by a lower region and an uncertain region, named boundary
- An element x is assigned to the boundary of two (or more) clusters if it is approximately to the same distance to the centroid's clusters

Rough k-means (Peters,Lingras)

1. the lower approximation is contained in the upper approximation

Rough k-means (Peters,Lingras)

1. the lower approximation is contained in the upper approximation
2. x belongs at most to one lower approximation

Rough k-means (Peters,Lingras)

1. the lower approximation is contained in the upper approximation
2. x belongs at most to one lower approximation
3. If x does not belong to any lower approximation, then, it belongs to at least two upper approximations

Rough Clustering (Yao, Yu)

- Known as Three-way clustering or Interval clustering

Rough Clustering (Yao, Yu)

- Known as Three-way clustering or Interval clustering
- $\forall x \in U$, there *exists* at most one lower approximation containing x

Rough Clustering (Yao, Yu)

- Known as Three-way clustering or Interval clustering
- $\forall x \in U$, there *exists* at most one lower approximation containing x
- The lower approximation cannot be empty

Rough Clustering (Yao, Yu)

- Known as Three-way clustering or Interval clustering
- $\forall x \in U$, there *exists* at most one lower approximation containing x
- The lower approximation cannot be empty
- Each object belongs to at least one upper approximation

Software

Free software based on rough sets

Software

Free software based on rough sets

- Rosetta (2001), limited to tables with 500 objects and 20 attributes
  http://www.lcb.uu.se/tools/rosetta
- Rose2 (2002), uses also similarity and variable precision
  http://idss.cs.put.poznan.pl/site/61.html
- Rough Set Exploration System (2005), also with dynamic reducts
  http://logic.mimuw.edu.pl/~rses/

Software

Free software based on rough sets

- Rosetta (2001), limited to tables with 500 objects and 20 attributes
  http://www.lcb.uu.se/tools/rosetta
- Rose2 (2002), uses also similarity and variable precision
  http://idss.cs.put.poznan.pl/site/61.html
- Rough Set Exploration System (2005), also with dynamic reducts
  http://logic.mimuw.edu.pl/~rses/
- R package "RoughSets: Data Analysis Using Rough Set and Fuzzy Rough Set Theories" (2015)
  https://cran.r-project.org/web/packages/RoughSets/index.html
- R package "Soft Clustering" (2015)
  https://cran.r-project.org/web/packages/SoftClustering/index.html