Rough Sets: Theoretical aspects of reasoning about data

wu

July 12, 2019

Contents

1 Knowledge

1.1 Knowledge base

Given a finite set $U \neq \emptyset$ (the universe). Any subset $X \subset U$ of the universe is called a **concept** or a **category** in U. And any family of concepts in U will be referred to as **abstract knowledge** about U.

partition or **classification** of a certain universe U is a family $C = \{X_1, X_2, \ldots, X_n\}$ s.t. $X_i \subset U, X_i \neq \emptyset, X_i \cap X_j = \emptyset$ and $\bigcup X_i = U$

A family of classifications is called a **knowledge base** over U

R an equivalence relation over U, U/R family of all equivalence classes of R, referred to be **categories** or **concepts** of R, and $[x]_R$ denotes a category in R containing an element $x \in U$

By a **knowledge base** we can understand a relational system $K = (U, \boldsymbol{R})$, $\boldsymbol{R}$ is a family of equivalence relations over U

If $\boldsymbol{P} \subset \boldsymbol{R}$ and $\boldsymbol{P} \neq \emptyset$, then $\bigcap \boldsymbol{P}$ is also an equivalence relation, and will be denoted by $IND(\boldsymbol{P})$, called an **indiscernibility relation** over $\boldsymbol{P}$

$$[x]_{IND(\boldsymbol{P})} = \bigcap_{R \in \boldsymbol{P}} [x]_R$$

$U/IND(\boldsymbol{P})$ called **$\boldsymbol{P}$-basic knowledge about** U in K. For simplicity, $U/\boldsymbol{P} = U/IND(\boldsymbol{P})$ and $\boldsymbol{P}$ will be also called **$\boldsymbol{P}$-basic knowledge** . Equivalence classes of $IND(\boldsymbol{P})$ are called **basic categories** of knowledge $\boldsymbol{P}$. If $Q \in \boldsymbol{R}$, then Q is a **Q-elementary knowledge** and equivalence classes of Q are referred to as **Q-elementary concepts** of knowledge $\boldsymbol{R}$

The family of all $\boldsymbol{P}$-basic categories for all $\neq \boldsymbol{P} \subset \boldsymbol{R}$ will be called the **family of basic categories** in knowledge base $K = (U, \boldsymbol{R})$

Let $K = (U, \boldsymbol{R})$ be a knowledge base. By $IND(K)$ we denote the family of all equivalence relations defined in K as $IND(K) = \{IND(\boldsymbol{P}) : \emptyset \neq \boldsymbol{P} \subseteq \boldsymbol{R}\}$.

Thus $IND(K)$ is the minimal set of equivalence relations.

Every union of $\boldsymbol{P}$-basic categories will be **$\boldsymbol{P}$-category**

The family of all categories in the knowledge base $K = (U, \boldsymbol{R})$ will be referred to as **K-categories**

1.2 Equivalence, generalization and specialization of knowledge

Let $K = (U, \boldsymbol{P}), K' = (U, \boldsymbol{Q})$. K and K' are **equivalent** $K \simeq K', (\boldsymbol{P} \simeq \boldsymbol{Q})$ if $IND(\boldsymbol{P}) = IND(\boldsymbol{Q})$. Hence $K \simeq K'$ if both K and K' have the same set of elememtary categories. *This means that knowledge in knowledge bases K and K' enables us to express exactly the same facts about the universe.*

If $IND(\boldsymbol{P}) \subset IND(\boldsymbol{Q})$ then knowledge $\boldsymbol{P}$ is **finer** than knowledge $\boldsymbol{Q}$ (**coarser**). $\boldsymbol{P}$ is **specialization** of $\boldsymbol{Q}$ and $\boldsymbol{Q}$ is **generalization** of $\boldsymbol{P}$

2 Imprecise categories, approximations and rough sets

2.1 Rough sets

Let $X \subseteq U$. X is R-**definable** or R-**exact** if X is the union of some R-basic categories. otherwise R-**undefinable**, R-**rough**, R-**inexact** .

2.2 Approximations of set

Given $K = (U, \boldsymbol{R}), R \in IND(K)$

$$\underline{R}X = \bigcup \{Y \in U/R : Y \subseteq X\}$$
$$\overline{R}X = \bigcup \{Y \in U/R : Y \cap X \neq \emptyset\}$$

called the R-**lower** and R-**upper approximation** of X

$BN_R(X) = \overline{R}X - \underline{R}X$ is R-**boundary** of X. $BN_R(X)$ is the set of elements which cannot be classified either to X or to $-X$ having knowledge R

$$POS_R(X) = \underline{R}X, R\text{-positive region of } X$$
$$NEG_R(X) = U - \overline{R}X, R\text{-negative region of } X$$
$$BN_R(X) - R\text{-borderline region of } X$$

If $x \in POS(X)$, then x will be called an R-**positive example of** X

Proposition 2.1. *1. X is R-definable if and only if $\underline{R}X = \overline{R}X$*

2. X is rought w.r.t. R if and only if $\underline{R}X \neq \overline{R}X$

2.3 Properties of approximations

Proposition 2.2 (2.2)**.** *1. $\underline{R}X \subseteq X \subseteq \overline{R}X$*

2. $\underline{R}\emptyset = \underline{R}\emptyset = \emptyset$; $\underline{R}U = \overline{R}U = U$

3. $\overline{R}(X \cup Y) = \overline{R}X \cup \overline{R}Y$

4. $\underline{R}(X \cap Y) = \underline{R}X \cap \underline{R}Y$

5. $X \subseteq Y$ *implies* $\underline{R}X \subseteq \underline{R}Y$

6. $X \subseteq Y$ *implies* $\overline{R}X \subseteq \overline{R}Y$

7. $\underline{R}(X \cup Y) \subseteq \underline{R}X \cup \underline{R}Y$

8. $\underline{R}(-X) = -\overline{R}X$

9. $\overline{R}(-X) = -\underline{R}X$

10. $\overline{R}(-X) = -\underline{R}X$

11. $\underline{R}\underline{R}X = \overline{R}\underline{R}X = \underline{R}X$

12. $\overline{R}\overline{R}X = \underline{R}\overline{R}X = \overline{R}X$

The equivalence relation R over U uniquely defines a topological space $T = (U, DIS(R))$ where $DIS(R)$ is the familty of all open and closed set in T and U/R is a base for T. The R-lower and R-upper approximation of X in A are **interior** and **closure** operations in the topological space T

2.4 Approximations and membership relation

$$x \underline{\in}_R X \text{if and only if } x \in \underline{R}X$$
$$x \overline{\in}_R X \text{if and only if } x \in \overline{R}X$$

where $\underline{\in}_R$ read "x **surely belongs** to X w.r.t. R" and $\overline{\in}_R$ - "x **possibly belongs** to X w.r.t. R". The **lower** and **upper** membership.

Proposition 2.3. 1. $x \underline{\in} X$ *implies* $x \in X$ *implies* $x \overline{\in} X$

2. $X \subset Y$ *implies* ($x \underline{\in} X$ *implies* $x \underline{\in} Y$ *and* $x \overline{\in} X$ *implies* $x \overline{\in} Y$)

3. $x \overline{\in} (X \cup Y)$ *if and only if* $x \overline{\in} X$ *or* $x \overline{\in} Y$

4. $x \underline{\in} (X \cap Y)$ *if and only if* $x \underline{\in} X$ *and* $x \underline{\in} Y$

5. $x \underline{\in} X$ *or* $x \underline{\in} Y$ *implies* $x \underline{\in} (X \cup Y)$

6. $x \overline{\in} X \cap Y$ *implies* $x \overline{\in} X$ *and* $x \overline{\in} Y$

7. $x \underline{\in} (-X)$ *if and only if non* $x \overline{\in} X$

8. $x \overline{\in} (-X)$ *if and only if non* $x \underline{\in} X$

2.5 Numerical characterization of imprecision

accuracy measure

$$\alpha_R(X) = \frac{card\ \underline{R}}{card\ \overline{\overline{R}}}$$

2.6 Topological characterization of imprecision

Definition 2.1.

1. *If $\underline{R}X \neq \emptyset$ and $\overline{R}X \neq U$, then we say that X is* ***roughly R-definable****. We can decide whether some elements belong to X or $-X$*

2. *If $\underline{R}X = \emptyset$ and $\overline{R}X \neq U$, then we say that X is* ***internally R-undefinable****. We can decide whether some elemnts belong to $-X$*

3. *If $\underline{R}X \neq \emptyset$ and $\overline{R}X = U$, then we say that X is* ***externally R-undefinable****. We can decide whether some elements belong to X*

4. *If $\underline{R}X = \emptyset$ and $\overline{R}X = U$, then we say that X is* ***totally R-undefinable****. unable to decide*

Proposition 2.4 (2.4)**.**

1. *Set X is R-definable(roughly R-definable, totally R-undefinable) if and only if so is $-X$*

2. *Set X is externally R-undefinable if and only if $-X$ is internally R-undefinable*

Proof. 1.

$$\begin{aligned} R\text{-definable} &\Leftrightarrow \underline{R}X = \overline{R}X, \underline{R} \neq \emptyset, \overline{R} \neq U \\ &\Leftrightarrow -\underline{R}X = -\overline{R}X \\ &\Leftrightarrow \overline{R}(-X) = \underline{R}(-X) \end{aligned}$$

$$\begin{aligned} X \text{ is roughly } R\text{-definable} &\Leftrightarrow \underline{R}X \neq \emptyset \wedge \overline{R}X \neq U \\ &\Leftrightarrow -\underline{R}X \neq U \wedge -\overline{R}X \neq \emptyset \\ &\Leftrightarrow \overline{R}(-X) \neq U \wedge \underline{R}(-X) \neq \emptyset \end{aligned}$$

$\square$

2.7 Approximation of classifications

If $F = \{X_1, \dots, X_n\}$ is a family of non empty sets, then $\underline{R}F = \{\underline{R}X_1, \dots, \underline{R}X_n\}$ and $\overline{R}F = \left\{\overline{R}X_1, \dots, \overline{R}X_n\right\}$, called the R**-lower approximation** and the R**-upper approximation** of the family F

The **accuracy of approximation** of F by R is

$$\alpha_R(F) = \frac{\sum card\ \underline{R}X_i}{\sum card\ \overline{R}X_i}$$

quality of approximation of F by R

$$\gamma_R(F) = \frac{\sum card\ \underline{R}X_i}{card\ U}$$

Proposition 2.5 (2.5). *Let $F = \{X_1, \dots, X_n\}$ where $n > 1$ be a classification of U and let R be an equivalence relation. If there exists $i \in \{1, 2, \dots, n\}$ s.t. $\underline{R}X_i \neq \emptyset$, then for each $j \neq i$ and $j \in \{1, \dots, n\}$, $\overline{R}X_j \neq U$*

Proof. If $\underline{R}X_i \neq \emptyset$ then there exists $x \in X$ s.t. $[x]_R \subseteq X$, which implies $[x]_R \cap X_j = \emptyset$ for each $j \neq i$. This yields $\overline{R}X_j \cap [x]_R = \emptyset$. $\square$

Proposition 2.6 (2.6). *Let $F = \{X_1, \dots, X_n\}, n > 1$ be a classification of U and let R be an equivalence relation. If there exists $i \in \{1, \dots, n\}$ s.t. $\overline{R}X_i = U$, then for each $j \neq i$ and $j \in \{1, \dots, n\}$ $\underline{R}X_j = \emptyset$*

Proposition 2.7 (2.7). *Let $F = \{X_1, \dots, X_n\}, n > 1$ be a classification of U and let R be an equivalence relation. If for each $i \in \{1, 2, \dots, n\}$ $\underline{R}X_i \neq \emptyset$ holds, then $\overline{R}X_i \neq U$ for each $i \in \{1, \dots, n, \}$*

Proposition 2.8. *Let $F = \{X_1, \dots, X_n\}, n > 1$ be a classification of U and let R be an equivalence relation. If for each $i \in \{1, 2, \dots, n\}$ $\overline{R}X_i = U$ holds, then $\underline{R}X_i = \emptyset$ for each $i \in \{1, \dots, n\}$*

2.8 Rough equality of sets

Definition 2.2. *Let $K = (U, \boldsymbol{R})$ be a knowledge base, $X, Y \subseteq U$ and $R \in IND(K)$, then*

1. *Sets X and Y are **bottom** R**-equal** $(X \eqsim_R Y)$ if $\underline{R}X = \underline{R}Y$*
2. *Sets X and Y are **top** $R -$ **equal** $(X \simeq_R Y)$ if $\overline{R}X = \overline{R}Y$*

3. *Sets X and Y are R-**equal** ($X \approx_R Y$) if $X \simeq_R Y$ and $X \eqsim_R Y$*

Proposition 2.9 (2.9)**.** 1. *$X \eqsim Y$ iff $X \cap Y \eqsim X$ and $X \cap Y \eqsim Y$*

2. *$X \simeq Y$ iff $X \cup Y \simeq X$ and $X \cup Y \simeq Y$*

3. *If $X \simeq X'$ and $Y \simeq Y'$ then $X \cup Y \simeq X' \cup Y'$*

4. *If $X \eqsim X'$ and $Y \eqsim Y'$ then $X \cap Y \eqsim X' \cap Y'$*

5. *If $X \simeq Y$, then $X \cup -Y \simeq U$*

6. *If $X \eqsim Y$, then $X \cap -Y \eqsim \emptyset$*

7. *If $X \subseteq Y$ and $Y \simeq \emptyset$, then $X \simeq \emptyset$*

8. *If $X \subseteq Y$ and $X \subseteq U$ then $Y \subseteq U$*

9. *$X \simeq Y$ iff $-X \eqsim -Y$*

10. *If $X \eqsim \emptyset$ or $Y \eqsim \emptyset$, then $X \cap Y \eqsim \emptyset$*

11. *If $X \simeq U$ or $Y \simeq U$, then $X \cup Y \simeq U$*

Proposition 2.10 (2.10)**.** *For any equivalence relation R*

1. *$\underline{R}X$ is the intersection of all $Y \subseteq U$ s.t. $X \eqsim_R Y$*

2. *$\overline{R}$ is the union of all $Y \subseteq U$ s.t. $X \simeq_R Y$*

2.9 Rough inclusion of sets

Definition 2.3. *Let $K = (U, \boldsymbol{R})$ be a knowledge base, $X, Y \subseteq U$ and $R \in IND(K)$.*

1. *Set X is **bottom** R-**included** in Y ($X \underset{\sim}{\subset}_R Y$) iff $\underline{R}X \subseteq \underline{R}Y$*

2. *Set X is **top** R-**included** in Y ($X \overset{\sim}{\subset}_R Y$) iff $\overline{R}X \subseteq \overline{R}Y$*

3. *Set X is R-**included** in Y ($X \overset{\sim}{\underset{\sim}{\subset}}_R Y$) iff $X \overset{\sim}{\subset}_R Y$ and $X \underset{\sim}{\subset}_R Y$*

Proposition 2.11 (2.11)**.** 1. *If $X \subseteq Y$, then $X \underset{\sim}{\subset} Y, X \overset{\sim}{\subset} Y$ and $X \overset{\sim}{\underset{\sim}{\subset}} Y$*

2. *If $X \underset{\sim}{\subset} Y$ and $Y \underset{\sim}{\subset} X$, then $X \eqsim Y$*

3. *If $X \overset{\sim}{\subset} Y$ and $Y \overset{\sim}{\subset} X$, then $X \simeq Y$*

4. *If $X \overset{\sim}{\underset{\sim}{\subset}} Y$ and $Y \overset{\sim}{\underset{\sim}{\subset}} X$ then $X \approx Y$*

5. $X \tilde{\subseteq} Y$ *iff* $X \cup Y \simeq Y$

6. $X \underset{\sim}{\subseteq} Y$ *iff* $X \cap Y \eqsim X$

7. *If* $X \subseteq Y, X \eqsim X', Y \eqsim Y'$, *then* $X' \underset{\sim}{\subseteq} Y'$

8. *If* $X \subseteq Y, X \simeq X', Y \simeq Y'$, *then* $X' \tilde{\subseteq} Y'$

9. *If* $X \subseteq Y, X \approx X', Y \approx Y'$, *then* $X' \underset{\sim}{\tilde{\subseteq}} Y'$

10. *If* $X' \tilde{\subseteq} X$ *and* $Y' \tilde{\subseteq} Y$, *then* $X' \cup Y' \tilde{\subseteq} X \cup Y$

11. *If* $X' \underset{\sim}{\subseteq} X, Y' \underset{\sim}{\subseteq}$ *then* $X' \cap Y' \underset{\sim}{\subseteq} X \cap Y$

12. $X \cap Y \underset{\sim}{\subseteq} X \tilde{\subseteq} X \cup Y$

13. *If* $X \underset{\sim}{\subseteq} Y$ *and* $X \eqsim Z$ *then* $Z \underset{\sim}{\subseteq} Y$

14. *If* $X \tilde{\subseteq} Y$ *and* $X \simeq Z$ *then* $Z \tilde{\subseteq} Y$

15. *If* $X \underset{\sim}{\tilde{\subseteq}} Y$ *and* $X \approx$ *then* $Z \underset{\sim}{\tilde{\subseteq}} Y$

3 Reduction of knowledge

3.1 Reduct and Core of Knowledge

Let $\boldsymbol{R}$ be a family of equivalence relations and let $P \in \boldsymbol{R}$. R is **dispensable** in $\boldsymbol{R}$ if $IND(\boldsymbol{R}) = IND(\boldsymbol{R} - \{R\})$. Otherwise R is **indispensable** in $\boldsymbol{R}$. The family of $\boldsymbol{R}$ is **independent** if each $R \in \boldsymbol{R}$ is indispensable in $\boldsymbol{R}$. Otherwise $\boldsymbol{R}$ is **dependent**

Proposition 3.1 (3.1)**.** *If $\boldsymbol{R}$ is independent and $\boldsymbol{P} \subseteq \boldsymbol{R}$, then $\boldsymbol{P}$ is also independent*

Proof. $IND(\boldsymbol{R}) = IND(\boldsymbol{P} \cup (\boldsymbol{R} - \boldsymbol{P})) = IND(\boldsymbol{P}) \cap IND(\boldsymbol{R} - \boldsymbol{P})$ □

$\boldsymbol{Q} \subseteq \boldsymbol{R}$ is a **reduct** of $\boldsymbol{P}$ if $\boldsymbol{Q}$ is independent and $IND(\boldsymbol{Q}) = IND(\boldsymbol{P})$

The set of all indispensable relations in $\boldsymbol{P}$ is called the **core** of $\boldsymbol{P}$ denoted by $CORE(\boldsymbol{P})$

Proposition 3.2 (3.2)**.**

$$CORE(\boldsymbol{P}) = \bigcap RED(\boldsymbol{P})$$

where $RED(\boldsymbol{P})$ is the family of all reducts of $\boldsymbol{P}$

Proof. If $\boldsymbol{Q}$ is a reduct of $\boldsymbol{P}$ and $R \in \boldsymbol{P} - \boldsymbol{Q}$, then $IND(\boldsymbol{P}) = IND(\boldsymbol{Q})$. If $\boldsymbol{Q} \subseteq \boldsymbol{R} \subseteq \boldsymbol{P}$ then $IND(\boldsymbol{Q}) = IND(\boldsymbol{R})$. Assuming $\boldsymbol{R} = \boldsymbol{P} - \{R\}$ then $R \notin CORE(\boldsymbol{P})$

If $R \notin CORE(\boldsymbol{P})$. This means $IND(\boldsymbol{P}) = IND(\boldsymbol{P} - \{R\})$ which implies that there exists an independent subset $\boldsymbol{S} \subseteq \boldsymbol{P} - \{R\}$ s.t. $IND(\boldsymbol{S}) = IND(\boldsymbol{P})$. Hence $R \notin \bigcap RED(\boldsymbol{P})$ □

3.2 Relative reduct and relative core of knowledge

Let P and Q be equivalence relations over U

P**-positive**

$$POS_P(Q) = \bigcup_{X \in U/Q} \underline{P}X$$

The P-positive region of Q is the set of all objects of the universe U which can be properly classified to classes of U/Q employing knowledge expressed by the classification U/P

Let $\boldsymbol{P}$ and $\boldsymbol{Q}$ be families of equivalence relations over U

$R \in \boldsymbol{P}$ is $\boldsymbol{Q}$**-dispensable** in $\boldsymbol{P}$ if

$$POS_{IND(\boldsymbol{P})}(IND(\boldsymbol{Q})) = POS_{IND(\boldsymbol{P}-\{R\})}(IND(\boldsymbol{Q}))$$

otherwise R is $\boldsymbol{Q}$-indispensable in $\boldsymbol{P}$

If every R