粗糙集理论与应用

代建华 教授、博士生导师
湖南师范大学信息科学与工程学院

粗糙集理论概述

1.1 粗糙集理论概况

自然界中大部分事物所呈现的信息都是：

- 不完整的、不确定的、模糊的和含糊的
- 经典逻辑无法准确、圆满地描述和解决

1904年，谓词逻辑创始人G. Frege提出：

“含糊”（Vague）——将含糊性归结到*“边界线区域”(Boundary region)*上，即在全域上存在一些个体，它既不能被分类到某一个子集上，也不能被分类到该子集的补集上。

1965年，美国数学家L. A. Zadeh提出了“模糊集”（Fuzzy sets），许多计算机科学家和逻辑学家试图通过这一理论解决G. Frege提出的“含糊”问题，但模糊集没有给出数学公式描述这一含糊概念，无法计算出它的具体的含糊元素数目。

1982年，波兰数学家Z. Pawlak针对G. Frege的“边界线区域”思想，提出了*“粗糙集”（Rough Sets）*。Pawlak把那些无法确认的个体都归属于边界线区域，而这种边界线区域被定义为：*“上近似集”与“下近似集”的差集*。由于它有确定的数学公式描述，故含糊元素的数目是可以计算的，即*在“真”、“假”二值之间的“含糊度”是可以计算的*。

粗糙集理论自诞生以来，经过许多数学家和计算机科学家的努力，其理论上日趋成熟，特别是在20世纪80年代末和90年代初，由于粗糙集理论在数据挖掘、知识发现等领域得到了成功的应用，它受到了国际上的广泛关注。

相对于其它处理不确定和模糊性的理论工具（如模糊集理论、Dempster-Shafer证据理论等）而言，*粗糙集理论有许多不可替代的优越性*。目前，它在信息科学、医药科学、工程技术、金融商业、环境科学、社会科学等领域中得到了广泛的、较为成功的应用，并且越来越受到其它更多领域的重视。

粗糙集理论在知识发现中的主要应用：

- ✓ 数据之间（精确或近似）依赖关系发现
- ✓ 数据的不确定性度量
- ✓ 评价某一分类（属性）的重要性
- ✓ 数据模式发现
- ✓ 决策规则发现
- ✓ 剔除冗余属性
- ✓ 数据集的降维

……

粗糙集理论的基本概念

近似与粗糙集

令$X \subseteq U$，R为U上一个等价关系。当R能表达成某些R基本范畴的并时，称R是可定义的；否则称R为不可定义的。

R可定义集也称R精确集，而R不可定义集也称为R非精确集或R粗糙集(rough set)。

对于粗糙集，可以使用两个精确集，即粗糙集的上、下近似来描述。

粗糙集理论的基本概念

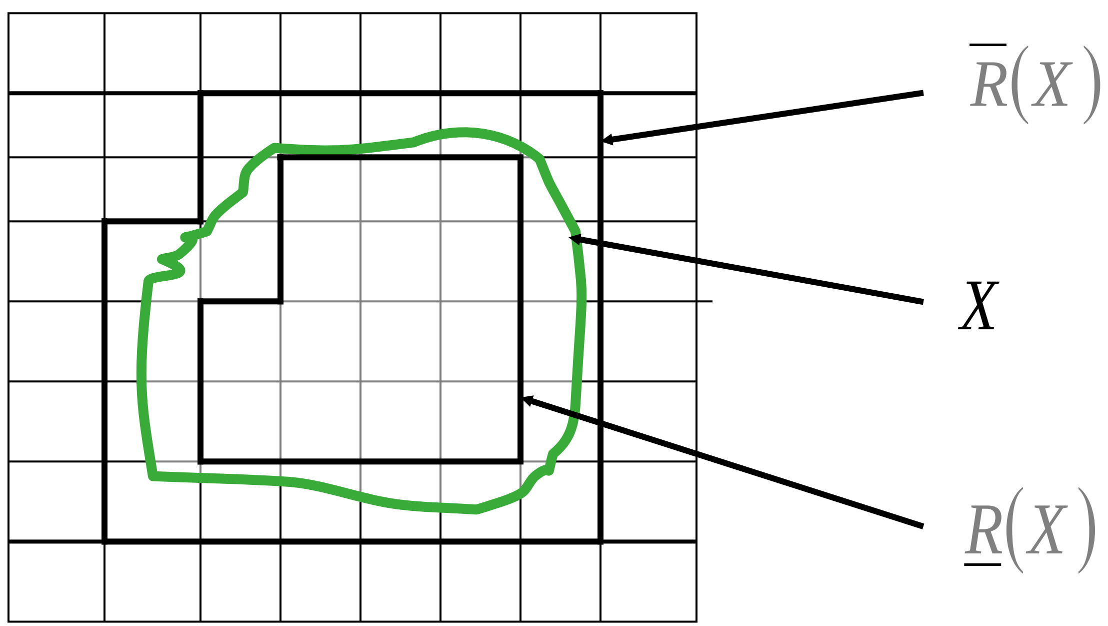
在近似空间(U, R)中，设X⊆U是个体全域上的子集，则X的下和上近似集分别为：

$$\underline{R}X = \{x \in U : [x]_R \subseteq X\}$$
$$\overline{R}X = \{x \in U : [x]_R \cap X \neq \varnothing\}$$

边界区域为：

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

集合的上、下近似概念示意图



上、下近似举例：

U	Headache	Temp.	Flu
U1	Yes	Normal	No
U2	Yes	High	Yes
U3	Yes	Very-high	Yes
U4	No	Normal	No
U5	No	High	No
U6	No	Very-high	Yes
U7	No	High	Yes
U8	No	Very-high	No

由$R = \{Headache, Temp.\}$划分出来的等价类有：

$\{u1\}, \{u2\}, \{u3\}, \{u4\}, \{u5, u7\}, \{u6, u8\}$.

$X1 = \{u \mid \mathrm{Flu}(u) = \mathrm{yes}\}$

$= \{u2, u3, u6, u7\}$

$\underline{R}X1 = \{u2, u3\}$

$\overline{R}X1 = \{u2, u3, u6, u7, \mathbf{u5, u8}\}$

$X2 = \{u \mid \mathrm{Flu}(u) = \mathrm{no}\}$

$= \{u1, u4, u5, u8\}$

$\underline{R}X2 = \{u1, u4\}$

$\overline{R}X2 = \{u1, u4, u5, u8, \mathbf{u6, u7}\}$

粗糙集上、下近似的性质：

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

(2) $\underline{R}\varnothing = \overline{R}\varnothing = \varnothing$,
$\underline{R}U = \overline{R}U = U$.

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

(6) $X \subseteq Y \Rightarrow \overline{R}X \subseteq \overline{R}Y$.

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

(8) $\overline{R}(X \cap Y) \subseteq \overline{R}X \cap \overline{R}Y$.

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

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

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

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


Thanks!