姚一豫(Yiyu Yao). 决策粗糙集研究探讨.
李华雄, 周献中, 李天瑞, 王国胤, 苗夺谦, 姚一豫. 决策粗糙集理论及其研究进展. 北京: 科学出版社, 2011, 155~178.

第 8 章 决策粗糙集研究探讨

姚一豫
加拿大里贾纳大学

说明

关于概率粗糙集的研究，我是从 1989 年开始的。基于贝叶斯决策论(Bayesian Decision Theory)，提出了一个决策粗糙集的模型(Decision-Theoretic Rough Set Model，DTRSM)[1-5]。近几年，我觉得粗糙集研究中的一个很重要、也被忽视的问题是粗糙集的语义[6]。从这个角度出发，得出了一些新的结果。比如，决策粗糙集中属性约简，它给出了属性约简(attribute reduct)一个新的语义解释[7]。又如，基于决策粗糙集的三枝决策(three-way decision)，它给出了概率正、负和边界域的一种新解释[8-10]。最近，又提出了平凡贝叶斯粗糙集模型(naïve Bayesian rough set model)，为估计决策粗糙集中的所需的条件概率给出了一个可行的方法[11]。

从近期粗糙集研究看，文章发表的数量越来越多。大部分文章给出了非常漂亮的数学结果，也提出了很多改进的算法，而语义研究几乎没有人去做。特别是，对于一些最基本的问题没有更进一步探讨。从而，很难非常客观地评价什么是有用的和有意义的拓展，什么是没有意义的和无用的拓展。有些结果也许在数学上很完美，但缺乏一个实际的，有价值的语义解释。另一方面，很多新结果也没有同已有结果进行比较和统一，这为很多学者的研究造成了不应有的困难。

一篇文章是一个研究阶段结果的总结，它给出了研究的结果，但没有给出研究的过程。这使得很多研究中有用的中间结果及思考不可能展示给读者。而这种思考往往对新的研究有很大的指导意义。基于这种考虑，想通过一种新的形式来描述我个人在研究概率粗糙集时的困惑与思考。在本章中，尽可能地按时间关系给出概率粗糙集发展中的几个重要结果。一方面，强调不同模型是怎样发展起来的和它们之间的相互关系；另一方面，也给出自己的看法。希望这些体会能够帮助读者；也希望读者能提出新的问题，大家一起共同讨论。

由于所讨论的内容很难系统地组织起来，因此本章分为很多小节，每节尽可能地独立并讨论一个问题。这样，每节可以单独阅读。

研究方法论的几个问题

我想先谈一下对自己研究有指导性的几种观点。

Feynman[12]在 The Character of Physical Law 这本书中谈到了关于对同一个理论或同一个定律不同形式但在数学上等价的描述，及其对我们认知和理解的指导作用。用 Feynman 的

话来说，虽然不同的描述在数学上是等价的，但“psychologically they are different because they are completely unequivalent when you are trying to guess new laws”。这给出了一个非常有效的研究策略。如果一个定律可以用多个不同形式但在数学上等价的方式表示或构造出来，那么，就可以获得在认知上的优势。不同的表示给出认知上的不同解释，也对该定律的拓广暗示了不同的方向。

我在拓广粗糙集上的很多工作都受益于这个研究策略。比如，粗糙集的上、下近似有不同形式但又等价的定义，不同的定义使我们能够将粗糙集同其他的研究联系起来，同时，不同的定义可以用来获得不同的拓广。在 On generalizing rough set theory 一文中[13]，我讨论了粗糙集近似的三种定义。基于元素的定义（element-based definition）将粗糙集和模态逻辑相结合，可以获得在一般二元关系下的拓广粗糙集[14]；基于粒的定义（granule-based definition）可以将粗糙集和粒计算结合起来，可以获得在覆盖（covering）意义下的拓广粗糙集[15]；基于子系统的定义（subsystem-based definition）可以将粗糙集和拓扑等数学结构联系起来，可以获得在闭系统（closure system）、布尔代数（Boolean algebra）及偏序集（partial ordered set）上的拓广粗糙集[16]。在后面，我会谈到用不同的表示来构造概率粗糙集模型。

不同表示的另一个好处是对问题能有一个更深的理解。每个表示可能强调一个侧面，将多种表示组合到一起可以获得任何一种表示都不可能给出的理解。关于这一点，Bateson[17] 在 Mind and Nature 这本书中有更仔细的描述。通过对不同描述，不同信息的综合，可以获得对同一个问题更深刻的认识，从而增加我们的知识。例如，通过对双眼获得的不同图像进行合成，可得到立体图像。对同一个数学问题的代数解释和几何解释会增加对该问题的理解。

在 Philosophy of Natural Science 一书中，Hempel[18]讨论了理论的一般特征和功能。这里给出其中的几段：

> “Theories are usually introduced when previous study of a class of phenomena has revealed a system of uniformities that can be expressed in the form of empirical laws. Theories then seek to explain those regularities and, generally, to afford a deeper and more accurate understanding of the phenomena in question.”
>
> “The basic entities and processes posited by a theory, and the laws assumed to govern them, must be specified with appropriate clarity and precision: otherwise, the theory cannot serve its scientific purpose.”
>
> “Broadly speaking, then, the formulation of a theory will require the specification of two kinds of principles; let us call them internal principles and bridge principles for short. The former will characterize the basic entities and processes invoked by the theory and the laws to which they are assumed to conform. The latter will indicate how the processes envisaged by the theory are related to empirical phenomena with which we are already acquainted, and which the theory may then explain, predict, or retrodict.”

Hempel 的讨论对我的研究有很重要的指导性。首先，一个理论必须能够解释自然现象，并能增加对问题更深入、更精确的理解。这是建立一个理论首先要考虑的问题。其次，在建立

一个理论时，需要精确和准确地定义所用的概念及定理。关于概率粗糙集的研究，我尽可能地贯彻这两个精神。Hempel 谈到的桥梁原理对我启发很大，促使我考虑粗糙集理论的语义问题。我早期对粗糙集的研究更偏重于“internal principles”，而近期的研究则转向“bridge principles”。这开阔了我的研究视野。Glaser 和 Strauss[19] 提出的扎根理论(grounded theory)的研究方法也强调一个理论的语义，即，怎样从所观察的数据和收集的材料中发现和建立一个基于数据的理论。

粒计算是我近十年的另一个研究方向[20-27]。其中的一个主要结果是基于多粒度和多视角的粒结构，多粒度为一个问题提供了多个层面的不同描述，多个视角为一个问题给出了多个侧面的描述。粒结构支持粒思维，粒思维是一个有效的研究方法。这种方法有其心理学和认知学的支持。

在 Human Information Processing 一书中，Lindsay 和 Norman[28]讨论了两种信息处理方法，即，自顶向下(top-down)和自底向上(bottom-up)处理方法。不同的层次具有不同的粒度，自顶向下对应于由大粒度到小粒度的细化过程，而自底向上对应于由小粒度到大粒度的抽象过程。自顶向下是一种基于概念驱动(conceptually driven)的过程，用上层的概念来指导下层概念的进一步展开，这样就可以把一个复杂问题逐步分解为简单问题。自顶向下是基于数据驱动(data driven)的过程，用下层的数据或概念支持上层概念的抽象。这样就把一些具体问题抽象成为一般问题。自顶向下和自底向上处理方法可以交互使用。

自顶向下和自底向上是计算机程序设计的两种基本方法。关于它们，Knuth[29]是这样讲的，“Top-down programming gives you a strong idea of where you are going, but it forces you to keep a lot of plans in your head; suspense builds up because nothing is really nailed down until the end. Bottom-up programming has the advantage that you continually wield a more and more powerful pencil, as more and more subroutines have been constructed; but it forces you to postpone the overall program organization until the last minute, so you might flounder aimlessly.”自底向上的方式在研究的初期非常有效。由于我们对问题还没有整体的理解，因而很难进行自顶向下处理。当我们积累了一定的知识，对问题有比较系统的认识，自顶向下的方式就能发挥其作用。特别是，自顶向下的方法对论文写作非常有指导意义。利用自顶向下的方法，我们可以对一个研究方向进行归纳分类，找到自己的研究的合理的位置，对自己所做的研究的意义及在整个学科中的作用有比较清楚的认识。

Knuth[29]也指出，“top-down and bottom-up were opposing methodologies: one more suitable for program exposition and the other more suitable for program creation.”一篇文章类似一个程序，Knuth 的讨论可以这样理解：自底向上的方式适用于研究阶段，而自顶向下适用于文章的写作。很可惜，文章创作过程中的很多经验和方法并不一定能在文章中体现。

粗糙集及其优点

粗糙集是 Pawlak[30,31]在 80 年代初提出的一个理论。关于粗糙集理论，在 Pawlak 等[32]是这样定义的，“Rough set theory … is a new mathematical tool to deal with vagueness and uncertainty.”关于该理论的优点，他们解释为，“One of the main advantages of rough set theory is that it does not need any preliminary or additional information about data, such as probability distribution in statistics, basic probability assignment in the Dempster-Shafer theory, or grade of membership or the value of possibility in fuzzy set theory.”

由于含糊性（vagueness）和不确定性（uncertainty）有多种解释和多面性，Pawlak 等人关于粗糙集理论的定义不很确切。由于粗糙集理论并没有给出一个处理含糊性和不确定性的普适方法，该定义有点夸大粗糙集的作用。 Pawlak 在 1982 年的文章给出了更具体的解释[30]，他指出粗糙集的主要思想是，“the exact mathematical formulation of the concept of approximative (rough) equality of sets in a given approximation space”。Pawlak[31]关于粗糙集的第一本著作是，“Rough Sets: Theoretical Aspects of Reasoning about Data”。在这本书的前言中，Pawlak 认为“The idea of the rough set consists of the approximation of a set by a pair of sets, called the lower and the upper approximation of the set”。他还指出，“The main issue we are interested in is reasoning from imprecise data, or more specifically, discovering relationships in data”。根据 Pawlak 的讨论，粗糙集主要研究的问题是集合的近似及相关的数据分析和推理方法与算法。相比而言，Pawlak 早期对粗糙集的定义及刻画更具体一点，更精确一点。值得强调的是，“精确的数学构建”体现了粗糙集理论的主要特征。粗糙集理论的重要贡献是给出了一种基于等价关系的数据分析方法，并给出了一个非常精确、严格的数学描述。对象的不可分辨性（indiscernibility）是由于它们具有相同的属性而引起，这种不可分辨性可以用等价关系表示。粗糙集理论首次形式地描述了对象不可分辨性，属性冗余性及属性约简等很多重要概念。

关于 Pawlak 等人提出的粗糙集的优点，给人一种误解，认为粗糙集理论并没有做任何假设同时也是非常客观的数据分析方法。但事实上，并非如此。比如，在很多 Pawlak 发表的文献中，都假设信息表包括所有的信息，即，封闭世界假设（closed-world assumption）。粗糙集数据分析方法也不可能是完全客观的。数据收集中属性的选择已经是非常主观了。粗糙集数据分析只提供一些数学分析工具，这同统计分析有所相同，其结果的语义并不是非常清楚，仍需要专家进行理解和解释。完全排除主观性不一定是数据分析方法的优点。更有效的数据分析模型应该有机地将客观分析方法和主观语义解释结合起来。

粗糙集近似产生的原因

作为一种数据分析方法，粗糙集主要以数据表为工具研究属性之间的依赖关系，从而获得有用的分类知识。关于粗糙集数据分析方法，很多文献解释粗糙集近似的原因是由于不精确（imprecise）和不完备（incomplete）信息产生的。这种说法值得探讨。首先，所要近似的子集是精确地，即，我们知道哪些对象属于该集合，哪些不属于。其次，信息表中关于对象的描述也是精确地，即，每一个对象在每一个属性上有一个确定值。这两者都没有什么不确定性和不完备性，这需要从其他角度解释粗糙集近似产生的原因。

粗糙集的数据分析方法是以概念（concept）的构建与分析为基础，这同形式概念分析（formal concept analysis）[33,34]有共性。在哲学、认知科学、心理学和机器学习等学科中，关于概念有多种定义和解释。概念的经典观（classical view of concepts）认为，一个概念可以用

一对概念的内涵(intension)和外延(extension)来解释和表示[35-37]。概念的内涵是概念的本质属性，基于它们可以判断一个对象是否是一个概念的实例；概念的外延是概念的实例集。

关于粗糙集中概念近似，基于 Marek 和 Truszczynski[38]的工作，我们可以有下面两种解释[39]。

不可定义概念和近似：粗糙集理论用决策逻辑语言(decision logic language) [31]来描述概念。由于语言的限制，一些对象不能通过有限属性集上的值进行区分。因此，该语言有可能不能定义某些对象集。一个对象子集可以表示一个概念的外延。如果不能找到语言中的公式来表示这个对象集，那么，这个概念就是一个在该逻辑语言下的不可定义概念。粗糙集理论研究用可定义概念去逼近不可定义概念[40]。

复杂概念和近似：在很多时候，我们知道一个概念的准确外延，并且这个概念可以准确地定义。但是，这个描述可能过于复杂而失去实用价值：理解和操作复杂概念可能会非常困难。这就需要我们去用一些简单描述的概念去逼近复杂概念。

在粗糙集数据分析中，一个信息表中不同的属性集给出不同的可定义概念。不可定义概念的产生是因为我们所选的属性不够多，这使得从属性值上看等价的对象有些属于一个概念，另一些不属于该概念。通过增加属性，我们可以将不可定义概念转换为可定义概念。粗糙集近似产生于属性不够多，或者说，属性集本身是不完备的，而不是对象的属性值是不精确或含糊的，也不是这个概念的外延是不精确或含糊的。通过这个观察，可以看到在数据分析中，含糊性、不确定性和不完备性有不同的解释。一个非常重要的语义问题就是准确地、严格地定义这些具有多重性的概念。

属性约简是粗糙集分析的一个重要问题之一。属性约简的目的与用简单概念近似复杂概念是一致的。简单概念只需要少量属性就能定义，复杂概念需要多个属性定义。近似属性约简就是实现用简单概念近似复杂概念。

信息表与概念表示

在粗糙集理论中，概念的内涵和外延可以通过信息表精确地定义。信息表可以定义为以下四元组：

$$S = (U, At, \{V_a \mid a \in At\}, \{I_a \mid a \in At\}),$$

其中，

U 是一个有限、非空的对象集合，

At 是一个有限、非空的属性集合，

V_a 是属性 $a \in At$ 的一个非空属性值集合，

$I_a : U \to V_a$ 是一个信息函数，将 U 中的一个对象映射到 V_a 上的一个值。

信息表给出了一种用一个有限的属性集合来表示一个有限的对象集合的精确方法。在一个信息表中，属性集的任何子集定义一个 U 上的等价关系：$P \subseteq At,$

$$xE_p y \Leftrightarrow \forall a \in P(I_a(x) = I_a(y)).$$

两个对象 x 和 y 等价，当且仅当它们在 P 中的所有属性上具有相同的值。这个等价关系给出了 U 的一个划分(partition)，记为 $U / E_p = \{[x]_p \mid x \in U\}$，其中

$$[x]_p = [x]_{E_p} = \{y \mid xE_p y\}$$

是包含 x 的等价类。 当 P 在讨论中很清楚时，我们也常常省略下标 P ，将 E_P 和 $[x]_P$ 分别记为 E 和 $[x]$。

一个 U 上的等价关系给出了 U 的一个划分；相反的，一个 U 的划分唯一地定义了一个 U 上的等价关系。所有 U 上的等价关系和所有 U 的划分是一一对应的。因此，在很多讨论中等价关系和划分可以交替使用。划分之间的偏序关系可以通过等价关系之间的包含关系来定义：

$$U / E \le U / E' \Leftrightarrow E \subseteq E'.$$

如果 $U / E \le U / E'$ ，每一个 U / E' 中的块必定是某些 U/E 中块的并集。因此， U/E 称为 U/E' 的细化， U/E' 称为 U/E 的粗化。对于不同属性集的子集，可以建立以下关系： $P, P' \subseteq At, x \in U,$

（i） $$E_P = \bigcap_{a \in P} E_{\{a\}}$$

$$E_{P \cup P'} = E_P \cap E_{P'}$$

（ii） $$[x]_P = \bigcap_{a \in P} [x]_{\{a\}}$$

$$[x]_{P \cup P'} = [x]_P \cap [x]_{P'}$$

（iii） $$P \subseteq P' \Rightarrow E_{P'} \subseteq E_P$$

$$P \subseteq P' \Rightarrow U / E_{p'} \le U / E_P$$

性质(i)和(ii)说明，由属性集子集定义的等价关系或划分可以通过单点属性集定义的等价关系或划分来构造。性质(iii)说明，对于属性集合之间的包含关系，细化-粗化关系≤是单调的。

每一个等价类 $[x]$ 可以看作是一个概念的外延。要定义概念的内涵，需要引入基于信息表的逻辑语言。在一个信息表中，决策逻辑语言 L 可以递归地定义为：原子公式 $a = v$ ，其中 $a \in At, v \in V_a$ ；如果 ϕ 和 ψ 是公式，那么 $\phi \wedge \psi$ 是一个公式。这个决策语言是Pawlak决策逻辑语言(decision logic language)[31]的一个子语言。通过将逻辑连接符限定为∧，我们仅考虑可以通过一组原子公式合取定义的概念。

一个公式的含义(meaning)可以通过对象对公式的可满足性(satisfiability)来定义。如果一个对象 x 满足公式 ϕ ，那么记为 $x \vDash \phi$ 。可满足性定义为：

$$x \vDash (a = v) \text{ iff } I_a(x) = v,$$
$$x \vDash \phi \wedge \psi \text{ iff } x \vDash \phi \text{ 与 } x \vDash \psi,$$

即，对象 x 满足公式 ϕ ，当且仅当其满足 ϕ 中的所有原子公式。对于一个公式 ϕ ，集合 $m(\phi)$ 称为公式 ϕ 的含义，可定义为：

$$m(\phi)=\{x\in U \mid x\vDash\phi\}.$$

通过以上定义，公式的含义具有以下两个性质：

$$m(a=v)=\{x\in U \mid I_a(x)=v\},$$
$$m(\phi\wedge\psi)=m(\phi)\bigcap m(\psi).$$

基于逻辑语言，一个概念可以精确地表示为一个二元组$(\phi,m(\phi))$，其中ϕ是概念的内涵，由L中的逻辑公式定义，$m(\phi)$是概念的外延，由满足该逻辑公式的对象集合来构造。

给定一个公式，可以构造唯一的U的子集作为它的含义。反之，对于任意一个U的子集X，不一定能找到一个定义该子集的公式。如果存在一个公式ϕ，使得$m(\phi)=X$，那么子集$X\subseteq U$称为一个合取可定义集合；否则，X是不可定义的。对于U的一个子集，也许可以找到多个公式来表示。即，U的某些子集有多种表示。

合取可定义集合和基于属性集的子集定义的等价关系的等价类之间有如下关系：

(1) $$m(a=I_a(x))=[x]_{\{a\}},$$

(2) $$m(\bigwedge_{a\in P} a=I_a(x))=[x]_P=\bigcap_{a\in P}[x]_{\{a\}},$$

即，通过使用决策逻辑语言，可以构造概念$(a=I_a(x),[x]_{\{a\}})$和$(\wedge_{a\in P} a=I_a(x),[x]_P)$。合取可定义概念是粗糙集理论进行数据分析及规则学习的基础。

Pawlak 上、下近似及正、负、边界域

在一个信息表中，一个属性子集定义一个对象集上的等价关系，记为E，其等价类是基本的可定义子集。通过等价类，我们可以描述或近似U的任何一个子集。设子集$C\subseteq U$表示一个概念所包含的对象集，即，该概念的外延，它不一定可以准确地用E的等价类来描述，即，C不一定是一组等价类的并集。因此，我们用一对上近似和下近似来刻画C：

$$\underline{apr}(C)=\{x\in U \mid [x]\subseteq C\},$$
$$\overline{apr}(C)=\{x\in U \mid [x]\bigcap C\neq\varnothing\}.$$

这是基于元素的定义，一个对象x属于下近似$\underline{apr}(C)$当且仅当所有和x等价的元素都属于C，x属于上近似$\overline{apr}(C)$当且仅当有一个和x等价的元素属于C。上、下近似还有另一个等价的定义：

$$\underline{apr}(C)=\bigcup\{[x]\in U/E \mid [x]\subseteq C\},$$
$$\overline{apr}(C)=\bigcup\{[x]\in U/E \mid [x]\bigcap C\neq\varnothing\}.$$

这是基于粒的定义，下近似$\underline{apr}(C)$是所有包含于集合C中的等价类的并集；上近似$\overline{apr}(C)$是所有跟集合C有非空交集的等价类的并集。这也就是说，上下近似可以用等价类准确地描述。

给定任何一个子集$C\subseteq U$，基于它的上下近似，我们可以得到U的一个划分：

$$\begin{aligned}
\mathrm{POS}(C) &= \underline{apr}(C) \\
&= \{x \in U \mid [x] \subseteq C\}, \\
\mathrm{NEG}(C) &= U - \overline{apr}(C) \\
&= \{x \in U \mid [x] \bigcap C = \emptyset\}, \\
\mathrm{BND}(C) &= \overline{apr}(C) - \underline{apr}(C) \\
&= \{x \in U \mid [x] \bigcap C \neq \emptyset \wedge \neg([x] \subseteq C)\}.
\end{aligned}$$

这三个子集分别称为 C 的正域 $\mathrm{POS}(C)$，负域 $\mathrm{NEG}(C)$ 和边界域 $\mathrm{BND}(C)$。如果 $x \in \mathrm{POS}(C)$，则 x 一定属于集合 C；如果 $x \in \mathrm{NEG}(C)$，则 x 一定不属于集合 C，即属于 C 的补集 C^c；如果 $x \in \mathrm{BND}(C)$，则不能判断 x 一定属于或者不属于集合 C。

上、下近似从定性的角度考虑了两种情况，即，必然性和可能性。所有等价类 $[x]$ 中的对象都由逻辑表达式 $\wedge_{a \in P} a = I_a(x)$ 描述。下近似解释为：如果任何一个对象有该描述，那么它必然属于 C；上近似解释为：如果一个对象有该描述，那么它可能属于 C。这种基于必然性和可能性的解释在粗糙集理论和模态逻辑之间建立了一个很紧密的关系，这种关系使得我们能够用模态逻辑的结果和方法研究粗糙集，其结果是基于一般二元关系的拓广粗糙集[14]。

正、负及边界域更准确地体现粗糙集的基本原理，即，粗糙集近似给出了一个三枝分类决策：

如果 $[x] \subseteq C$，则 $x \in \mathrm{POS}(C)$，
如果 $[x] \bigcap C = \emptyset$，则 $x \in \mathrm{NEG}(C)$，
如果 $[x] \bigcap C \neq \emptyset \wedge \neg([x] \subseteq C)$，则 $x \in \mathrm{BND}(C)$。

三枝分类更显式地表明了粗糙集同其他二枝分类的区别。一般二枝分类强行地将对象分为正或负两个域，这种分类是以分类错误为代价的。三枝分类的一个主要特点是，关于正、负域的分类是百分之百正确的，没有任何不确定性。所有的不确定对象都分到了边界域。

Pawlak 粗糙集的另一种表示

Pawlak 粗糙集可以看做是一种定性的近似，下近似由集合包含定义而上近似由集合相交非空定义。该定义没有任何不确定性，这种优点也同时成为它的局限性。为了获得定量粗糙集，我们首先考虑 Pawlak 粗糙集的另一种表示。

Wong 和 Ziarko[41]于 1987 年将概率近似空间引入到粗糙集的研究中。令 $Pr(C \mid [x])$ 表示任何一个实体在属于 $[x]$ 的条件下属于 C 的条件概率。那么，我们可以获得下面的等价条件，

$$\begin{aligned}
&Pr(C \mid [x]) = 1 \Leftrightarrow [x] \subseteq C, \\
&Pr(C \mid [x]) = 0 \Leftrightarrow [x] \bigcap C = \emptyset, \\
&0 < Pr(C \mid [x]) < 1 \Leftrightarrow [x] \bigcap C \neq \emptyset \wedge \neg([x] \subseteq C).
\end{aligned}$$

这样，就得到了 Pawlak 三个域的另一种表示：

$$\begin{aligned}
\mathrm{POS}(C) &= \{x \in U \mid Pr(C \mid [x]) \geq 1\},\\
\mathrm{NEG}(C) &= \{x \in U \mid Pr(C \mid [x]) \leq 0\},\\
\mathrm{BND}(C) &= \{x \in U \mid 0 < Pr(C \mid [x]) < 1\}.
\end{aligned}$$

显然，定性粗糙集三个域仅用了概率的两个极端值，即 0 和 1。这种表示为定量粗糙集给出了一个很好的启示。如果我们将 0 和 1 用其他的值来表示，那么就可以获得一种定量粗糙集。

条件概率 $Pr(C \mid [x])$ 的定义及估计是一个非常重要的问题。但在定义概率粗糙集时，我们并不需要知道 $Pr(C \mid [x])$ 是怎样获得的。在这个阶段的研究中，我们假定条件概率已经给出，研究问题的焦点是怎样定义概率粗糙集。这样做的好处是可以不受到条件概率获取方法的影响。关于条件概率的定义及估计，可以作为另一个研究课题，在后面的平凡贝叶斯粗糙集中会谈到。

0.5-概率粗糙集

在 1985 年的科技报告中，Wong 和 Ziarko[42]提出了 0.5-概率粗糙集模型，该模型随后在 Pawlak，Wong 和 Ziarko[43]的文章中有更近一步的介绍。这个模型的主要理论依据是多数规则（majority rule）。它用一个 0.5 概率阈值来定义概率正、负和边界域：

$$\begin{aligned}
\mathrm{POS}_{0.5}(C) &= \{x \in U \mid Pr(C \mid [x]) > 0.5\};\\
\mathrm{BND}_{0.5}(C) &= \{x \in U \mid Pr(C \mid [x]) = 0.5\};\\
\mathrm{NEG}_{0.5}(C) &= \{x \in U \mid Pr(C \mid [x]) < 0.5\}.
\end{aligned}$$

阈值 0.5 定量地刻画了多数规则，当等价类 $[x]$ 中超过一半的元素属于 C 时，我们可以将 x 放到 C 的正域中；当超过一半的元素不属于 C 时，我们可以将 x 放到 C 的负域中；当刚好一半的元素属于 C 时，我们可以将 x 放到 C 的边界域中。

在 0.5-概率模型中，正、负域引入了决策错误。对正域来说，引入的错误分类率为 $1 - Pr(C \mid [x]) < 0.5$；对负域来说，引入的错误分类率为 $Pr(C \mid [x]) < 0.5$。从很大程度上讲，0.5-概率模型体现了四舍五入的进位思想。但用 0.5 定义边界域显得有些不自然，因为可以很简单地将概率为 0.5 的对象放到正或负域里。

决策粗糙集主要结果

在 1990 年，Yao, Wong 和 Lingras[1]提出了一个更一般性的概率粗糙集模型，称为决策粗糙集模型。该模型用一对概率阈值来定义概率正、负和边界域。设 $0 \leq \beta < \alpha \leq 1$, 决策粗糙集模型的一个基本结果是 (α, β) -概率正、负和边界域：

$$\begin{aligned}
\mathrm{POS}_{(\alpha,\beta)}(C) &= \{x \in U \mid Pr(C \mid [x]) \geq \alpha\},\\
\mathrm{BND}_{(\alpha,\beta)}(C) &= \{x \in U \mid \beta < Pr(C \mid [x]) < \alpha\},\\
\mathrm{NEG}_{(\alpha,\beta)}(C) &= \{x \in U \mid Pr(C \mid [x]) \leq \beta\}.
\end{aligned}$$

同 Pawlak 三个域相比，α 和 β 分别取代了概率极值 1 和 0。同 0.5-概率模型相比，边界域显得更实际、更有意义。

概率正、负域包含错误分类。正域的错误分类率是$1-Pr(C\,|\,[x]) \le 1-\alpha$，负域的错误分类率是$Pr(C\,|\,[x]) \le \beta$。这为$\alpha$和$\beta$给出了一种基于错误分类率的解释。该解释有其直观易懂的优点。如果决策粗糙集仅停留在这种解释上，那么它的意义并不是很大。我们在实际应用中还是没有一种指导思想和一套有效地方法来解释和获得这两个阈值。也就是说，概率粗糙集仅仅是对经典粗糙集的一个简单拓广。

决策粗糙集真正的贡献在于它不仅给出了概率正、负和边界域这个结果，更重要的是，它给出了基于贝叶斯决策论的一个语义模型，同时也给出了一个实际、有效的解释和计算阈值的方法。也就是说，决策粗糙集是一个有坚实理论基础同时又实用的一个模型。

决策粗糙集研究的三个问题

从形式上看，(α,β)-正、负和边界域给出了一个定量的概率粗糙集模型。对于该模型来讲，这还远远不够，我们需要探讨和解释它所用到的基本概念和基本量。最少有以下三个问题需要解决：

- 阈值α,β的解释与计算[1,2]
- 概率正、负及边界域的解释与应用[8-10]
- 条件概率$Pr(C\,|\,[x])$的估计[11]

它们代表我在决策粗糙集中的三个不同阶段的研究。

阈值的解释与计算

关于贝叶斯决策论，可参见 Duda 和 Hart[44]书中第二章的介绍。

决策粗糙集模型可以认为是贝叶斯决策理论的一个简单应用，其描述如下。对于一个子集$C \subseteq U$，可以构造一个含有两个状态的集合$\Omega=\{C, C^c\}$；对应于粗糙集中的三个域，我们可以构造一个决策动作集$A=\{a_P, a_B, a_N\}$，其中，a_P, a_B和a_N分别代表对一个对象分类的动作，即，选择$x \in \mathrm{POS}(C)$，$x \in \mathrm{BND}(C)$或$x \in \mathrm{NEG}(C)$。不同的决策会引导不同的分类错误，也将产生不同的后果。这可以表示为一个由3×2的矩阵定义的损失函数：

	C（Positive instance）	C^c（Negative instance）
a_P（选择POS域）	$\lambda_{PP}=\lambda(a_P\,|\,C)$	$\lambda_{PN}=\lambda(a_P\,|\,C^c)$
a_B（选择BND域）	$\lambda_{BP}=\lambda(a_B\,|\,C)$	$\lambda_{BN}=\lambda(a_B\,|\,C^c)$
a_N（选择NEG域）	$\lambda_{NP}=\lambda(a_N\,|\,C)$	$\lambda_{NN}=\lambda(a_N\,|\,C^c)$

其中，λ_{PP}，λ_{BP}和λ_{NP}分别表示当一个对象属于集合C时，采用动作a_P，a_B和a_N的损失；λ_{PN}，λ_{BN}和λ_{NN}分别表示当一个对象不属于集合C时，采用这些动作的损失。

损失函数的意义决定于具体的应用。通常，它可以用其他更直观的概念定义和解释，比

如，钱、时间、人力等资源的度量，或者是不同后果的危险程度的度量。一方面，在建立一个数学模型时，不需要考虑一个具体的解释，因为这会限制模型的一般性。另一方面，非常需要建立这种抽象概念和应用中的具体概念的关系，因为这才能保证在实际应用中对抽象概念赋予有意义的解释。

在早期的决策粗糙集文献中，我们用a_1, a_2, a_3表示动作，用$\lambda_{11}, \lambda_{21}, \lambda_{31}$等表示损失函数。从形式上讲，这种表示没有错，但却增加了文章阅读的认知负担。在读一个公式时，因为数字 1 和正域没有显式的联系，因此需要将a_1翻译成所对应的选择正域的动作。在新的符号里，下标P是正域(Positive region)的首字母，因此，a_P明确地表示了选择正域的动作。同样的解释也适应于损失函数的表示。这套新的符号系统是由 Harbert 建议的。这个小例子给出了一个很有用的写作启示：一篇文章的可读性和有效性不仅由它的内容决定，在某些程度上也由它的形式决定。

对于$[x]$中的对象，采用不同的动作所产生的损失表示如下：

$$\begin{aligned}
R(a_P \mid [x]) &= \lambda_{PP} Pr(C \mid [x]) + \lambda_{PN} Pr(C^c \mid [x]),\\
R(a_B \mid [x]) &= \lambda_{BP} Pr(C \mid [x]) + \lambda_{BN} Pr(C^c \mid [x]),\\
R(a_N \mid [x]) &= \lambda_{NP} Pr(C \mid [x]) + \lambda_{NN} Pr(C^c \mid [x]).
\end{aligned}$$

贝叶斯决策论给出了以下最小风险决策规则：

(P) 如果$R(a_P \mid [x]) \le R(a_B \mid [x])$且$R(a_P \mid [x]) \le R(a_N \mid [x])$，则选择$x \in \mathrm{POS}(C)$，

(B) 如果$R(a_B \mid [x]) \le R(a_P \mid [x])$且$R(a_B \mid [x]) \le R(a_N \mid [x])$，则选择$x \in \mathrm{BND}(C)$，

(N) 如果$R(a_N \mid [x]) \le R(a_P \mid [x])$且$R(a_N \mid [x]) \le R(a_B \mid [x])$，则选择$x \in \mathrm{NEG}(C)$.

当两个动作有同样的风险，需要引入决胜规则(tie-breaking rule)，这样每一个对象只划到唯一的域中。

因为$Pr(C \mid [x]) + Pr(C^c \mid [x]) = 1$，我们可以只使用概率$Pr(C \mid [x])$和损失函数$\lambda$来简化这些规则。首先，考虑一个满足下述条件的损失函数：

$$\text{(c0).} \qquad \lambda_{PP} \le \lambda_{BP} < \lambda_{NP}, \qquad \lambda_{NN} \le \lambda_{BN} < \lambda_{PN}.$$

就是说，将属于C的对象x分到正域$\mathrm{POS}(C)$的损失小于或者等于将其分到边界域$\mathrm{BND}(C)$的损失，并且这个两个损失都严格的小于将其分到负域$\mathrm{NEG}(C)$的损失。相反顺序用来描述对一个不属于C对象分类。在条件(c0)下，我们可以简化决策规则$(\mathrm{P})-(\mathrm{N})$。规则中的六个条件可以表示为：

$$R(a_P\,|[x]) \le R(a_B\,|[x]) \Leftrightarrow Pr(C\,|[x]) \ge \frac{(\lambda_{PN}-\lambda_{BN})}{(\lambda_{PN}-\lambda_{BN})+(\lambda_{BP}-\lambda_{PP})} = \alpha,$$

$$R(a_P\,|[x]) \le R(a_N\,|[x]) \Leftrightarrow Pr(C\,|[x]) \ge \frac{(\lambda_{PN}-\lambda_{NN})}{(\lambda_{PN}-\lambda_{NN})+(\lambda_{NP}-\lambda_{PP})} = \gamma,$$

$$R(a_B\,|[x]) \le R(a_P\,|[x]) \Leftrightarrow Pr(C\,|[x]) \le \frac{(\lambda_{PN}-\lambda_{BN})}{(\lambda_{PN}-\lambda_{BN})+(\lambda_{BP}-\lambda_{PP})} = \alpha,$$

$$R(a_B\,|[x]) \le R(a_N\,|[x]) \Leftrightarrow Pr(C\,|[x]) \ge \frac{(\lambda_{BN}-\lambda_{NN})}{(\lambda_{BN}-\lambda_{NN})+(\lambda_{NP}-\lambda_{BP})} = \beta,$$

$$R(a_N\,|[x]) \le R(a_P\,|[x]) \Leftrightarrow Pr(C\,|[x]) \le \frac{(\lambda_{PN}-\lambda_{NN})}{(\lambda_{PN}-\lambda_{NN})+(\lambda_{NP}-\lambda_{PP})} = \gamma,$$

$$R(a_N\,|[x]) \le R(a_B\,|[x]) \Leftrightarrow Pr(C\,|[x]) \le \frac{(\lambda_{BN}-\lambda_{NN})}{(\lambda_{BN}-\lambda_{NN})+(\lambda_{NP}-\lambda_{BP})} = \beta,$$

其中所用到的三个参数可以如下计算：

$$\alpha = \frac{(\lambda_{PN}-\lambda_{BN})}{(\lambda_{PN}-\lambda_{BN})+(\lambda_{BP}-\lambda_{PP})},$$

$$\beta = \frac{(\lambda_{BN}-\lambda_{NN})}{(\lambda_{BN}-\lambda_{NN})+(\lambda_{NP}-\lambda_{BP})},$$

$$\gamma = \frac{(\lambda_{PN}-\lambda_{NN})}{(\lambda_{PN}-\lambda_{NN})+(\lambda_{NP}-\lambda_{PP})}.$$

决策规则 (P) – (N) 可以简明地表示为：

(P). 如果 $Pr(C\,|[x]) \ge \alpha$ 且 $Pr(C\,|[x]) \ge \gamma$, 则选择 $x \in \mathrm{POS}(C)$;

(B). 如果 $Pr(C\,|[x]) \le \alpha$ 且 $Pr(C\,|[x]) \ge \beta$, 则选择 $x \in \mathrm{BND}(C)$;

(N). 如果 $Pr(C\,|[x]) \le \beta$ 且 $Pr(C\,|[x]) \le \gamma$, 则选择 $x \in \mathrm{NEG}(C)$.

规则 (B) 表明限定 $\alpha > \beta$ 是合理的，这样边界域不会为空。令 $\alpha > \beta$，则损失函数应满足如下条件：

$$\text{(c1).}\qquad \frac{\lambda_{NP}-\lambda_{BP}}{\lambda_{BN}-\lambda_{NN}} > \frac{\lambda_{BP}-\lambda_{PP}}{\lambda_{PN}-\lambda_{BN}}.$$

条件 (c0) 和 (c1) 合起来隐含 $1 \ge \alpha > \gamma > \beta \ge 0$。因此，可以获得以下简化规则：

(P). 如果 $Pr(C\,|[x]) \ge \alpha$, 则选择 $x \in \mathrm{POS}(C)$;

(B). 如果 $\beta < Pr(C\,|[x]) < \alpha$, 则选择 $x \in \mathrm{BND}(C)$;

(N). 如果 $Pr(C\,|[x]) \le \beta$, 则选择 $x \in \mathrm{NEG}(C)$.

这给出了解释和计算阈值的理论和方法。

三枝决策

规则学习及分类是粗糙集的一个主要应用，其思想是用合取可定义概念来构造规则。三枝决策是基于粗糙集三个域和统计学中的假设验证提出的，它更精确地反映了粗糙集的近似原理，并可以用来解释实际应用中很多决策现象。

三枝决策是实际生活中常用的策略之一。比如，医生通常采用三枝决策，即，根据初步诊断对病人实施治疗、不治疗或进一步观察。这三类措施可能有错，因而导致不同的风险。杂志编辑也常采用三枝决策处理稿件，即，根据审稿人意见决定接收、拒绝或进一步审查。投资管理者也采用投资、不投资或进一步观察的三枝决策。这样的例子还有很多。从很大程度上讲，粗糙集的正、负和边界域为三枝决策提供了理论基础。具体的说，正域所对应的规则，简称正规则，表示接收；负域对应的规则，简称负规则，表示拒绝；边界域对应的规则，简称边界规则，对应不做决定或者推迟决定。不论是接收还是拒绝都可能带有错误，即，接收错误或拒绝错误，这同统计学中对一个假设的接收和拒绝相似。

设 $\mathrm{Des}(x) = \wedge_{a \in P}(a = I_a(x))$ 是合取可定义概念 $[x]_P \subseteq U$ 的描述，即，内涵。从粗糙集的正、负和边界域可推导出如下三枝决策规则：

$$
\begin{aligned}
&\mathrm{Des}(x) \to \text{接收 } x \in C, && [x]_P \subseteq \mathrm{POS}(C),\\
&\mathrm{Des}(x) \to \text{既不接收也不拒绝 } x \in C, && [x]_P \subseteq \mathrm{BND}(C),\\
&\mathrm{Des}(x) \to \text{拒绝 } x \in C, && [x]_P \subseteq \mathrm{NEG}(C).
\end{aligned}
$$

就正规则而言，由于 $[x]_P \subseteq \mathrm{POS}(C)$，它的置信度 (confidence) $\delta = Pr(C \mid [x]_P) = 1$，同时错误率为 $1-\delta = 0$。对负规则而言，由于它的置信度 $\delta' = Pr(C^c \mid [x]_P) = 1$，错误率为 $1-\delta' = 0$。对于边界规则，它的置信度和错误率介于 0 和 1 之间。因为正、负规则不产生错误，所以基于 Pawlak 粗糙集的三枝决策可以理解为定性决策规则。

在决策粗糙集模型中，正域、边界域和负域由一对阈值 (α, β) 决定，其中 $0 \le \beta < \alpha \le 1$。因此，它们能够容忍一定程度的错误，所对应的定量三枝决策规则是：

$$
\begin{aligned}
&\mathrm{Des}(x) \to \text{接收 } x \in C, && [x]_P \subseteq \mathrm{POS}_{(\alpha,\beta)}(C),\\
&\mathrm{Des}(x) \to \text{既不接收也不拒绝 } x \in C, && [x]_P \subseteq \mathrm{BND}_{(\alpha,\beta)}(C),\\
&\mathrm{Des}(x) \to \text{拒绝 } x \in C, && [x]_P \subseteq \mathrm{NEG}_{(\alpha,\beta)}(C).
\end{aligned}
$$

根据决策粗糙集中正域的定义，正规则的置信度满足 $\delta = Pr(C \mid [x]_P) \ge \alpha$，其错误率满足 $1-\delta \le 1-\alpha$。负规则的置信度满足 $\delta' = Pr(C^c \mid [x]_P) = 1 - Pr(C \mid [x]_P) \ge 1-\beta$，错误率满足 $1-\delta' \le \beta$。当阈值 $\alpha = 1$ 且 $\beta = 0$ 时，经典粗糙集模型可以看作是决策粗糙集模型的一个特例。

定性和定量三枝决策的主要区别是定性决策不允许有错误，而定量决策容忍一定的错误。在实际应用中，很难达到没有错误的决策，因而定量的接收和拒绝显得更有意义。用另一句话说，当支持证据很强时，就可以采取接收策略，当反证很强时，就可以采取拒绝策略，当证据不足时，可以延迟决策，并寻求新的证据。决策粗糙集中的阈值 (α, β) 反映了所需证据的强弱。这样，决策粗糙集可以和统计学中假设验证相联系，也为实际中的三枝决策提

供了一个理论模型。

平凡贝叶斯粗糙集

Yao 和 Zhou[11]提出平凡贝叶斯粗糙集模型并给出估计条件概率$Pr(C\,|\,[x])$的一个简单实用方法。

由于条件概率$Pr(C\,|\,[x])$很难从可观察到的数值中估计，通常用贝叶斯公式进行如下变换：

$$Pr(C\,|\,[x]) = \frac{Pr([x]\,|\,C)Pr(C)}{Pr([x])},$$

其中，$Pr(C\,|\,[x])$称为后验概率(a posteriori probability)，$Pr(C)$称为先验概率(a priori probability)，$Pr([x]\,|\,C)$称为似然(likelihood)函数。在实际应用中，也常用到赔率(odds)变换：

$$\begin{aligned} O(Pr(C\,|\,[x])) &= \frac{Pr(C\,|\,[x])}{1-Pr(C\,|\,[x])} \\ &= \frac{Pr(C\,|\,[x])}{Pr(C^c\,|\,[x])} \\ &= \frac{Pr([x]\,|\,C)}{Pr([x]\,|\,C^c)}\frac{Pr(C)}{Pr(C^c)} \\ &= \frac{Pr([x]\,|\,C)}{Pr([x]\,|\,C^c)}O(Pr(C)), \end{aligned}$$

其中，$O(Pr(C\,|\,[x]))$为后验赔率，$O(Pr(C))$为先验赔率，$Pr([x]\,|\,C)/Pr([x]\,|\,C^c)$为似然比(likelihood ratio)。它的优点是不需要估计等价类$[x]$的概率$Pr([x])$。如果在赔率进一步应用对数函数，那么就得到logit 变换：

$$\begin{aligned} \text{logit}(Pr(C\,|\,[x])) &= \log(O(P(C\,|\,[x]))) \\ &= \log\frac{Pr([x]\,|\,C)}{Pr([x]\,|\,C^c)} + \log\frac{Pr(C)}{Pr(C^c)} \end{aligned}$$

通过这些变换，条件概率$Pr(C\,|\,[x])$的估计转换为似然比$Pr([x]\,|\,C)/Pr([x]\,|\,C^c)$的估计。

假设$[x]$由一个信息表的属性子集$P\subseteq At$定义，即$[x]=[x]_P=\bigcap_{a\in P}[x]_{\{a\}}$。为了估计$Pr([x]\,|\,C)$和$Pr([x]\,|\,C^c)$，平凡概率粗糙集做了下面的概率独立假设：

$$\begin{aligned} Pr([x]\,|\,C) &= Pr([x]_P\,|\,C) \\ &= Pr(\bigcap_{a\in P}[x]_{\{a\}}\,|\,C) \\ &= \prod_{a\in P} Pr([x]_{\{a\}}\,|\,C)\ , \end{aligned}$$

$$\begin{aligned} Pr([x] \mid C^c) &= Pr([x]_P \mid C^c) \\ &= Pr(\bigcap_{a\in P}[x]_{\{a\}} \mid C^c) \\ &= \prod_{a\in P} Pr([x]_{\{a\}} \mid C^c) \ . \end{aligned}$$

将它们带入logit公式，得到

$$\mathrm{logit}(Pr(C \mid [x]_P)) = \sum_{a\in P}\log\frac{Pr([x]_{\{a\}} \mid C)}{Pr([x]_{\{a\}} \mid C^c)} + \log\frac{Pr(C)}{Pr(C^c)} \ .$$

基于信息表中的数据，概率$Pr([x]_{\{a\}} \mid C)$和$Pr([x]_{\{a\}} \mid C^c)$可以如下估计：

$$\begin{aligned} Pr([x]_{\{a\}} \mid C) &= \frac{|[x]_{\{a\}} \cap C|}{|C|} \ , \\ Pr([x]_{\{a\}} \mid C^c) &= \frac{|[x]_{\{a\}} \cap C^c|}{|C^c|} \ , \end{aligned}$$

其中$|\cdot|$表示集合的势。通常，$[x]_{\{a\}}$的对象个数会多一点，相对而言这些估计可能更可靠一点。

就阈值来说，可以建立以下的关系：

$$\begin{aligned} Pr(C \mid [x]) \ge \alpha &\Leftrightarrow \frac{Pr(C \mid [x])}{Pr(C^c \mid [x])} \ge \frac{\alpha}{1-\alpha} \\ &\Leftrightarrow \frac{Pr([x] \mid C)}{Pr([x] \mid C^c)}\frac{Pr(C)}{Pr(C^c)} \ge \frac{\alpha}{1-\alpha} \\ &\Leftrightarrow \log\frac{Pr([x] \mid C)}{Pr([x] \mid C^c)} + \log\frac{Pr(C)}{Pr(C^c)} \ge \log\frac{\alpha}{1-\alpha} \end{aligned}$$

在概率独立假设下有，

$$Pr(C \mid [x]) \ge \alpha \Leftrightarrow \sum_{a\in P}\log\frac{Pr([x]_{\{a\}} \mid C)}{Pr([x]_{\{a\}} \mid C^c)} \ge \log\frac{Pr(C^c)}{Pr(C)} + \log\frac{\alpha}{1-\alpha}$$

关于β，可以建立类似的关系。令

$$\begin{aligned} \alpha' &= \log\frac{Pr(C^c)}{Pr(C)} + \log\frac{\alpha}{1-\alpha} \\ \beta' &= \log\frac{Pr(C^c)}{Pr(C)} + \log\frac{\beta}{1-\beta} \end{aligned}$$

则可以获得以下概率正、负和边界域的定义：

$$\mathrm{POS}_{(\alpha,\beta)} = \{x \in U \mid \sum_{a\in P}\log\frac{Pr([x]_{\{a\}} \mid C)}{Pr([x]_{\{a\}} \mid C^c)} \ge \alpha'\}$$

$$\mathrm{NEG}_{(\alpha,\beta)} = \{x \in U \mid \sum_{a\in P}\log\frac{Pr([x]_{\{a\}} \mid C)}{Pr([x]_{\{a\}} \mid C^c)} \ge \beta'\}$$

$$\mathrm{BND}_{(\alpha,\beta)} = \{\, x \in U \mid \beta' < \sum_{a\in P}\log\frac{Pr([x]_{\{a\}} \mid C)}{Pr([x]_{\{a\}} \mid C^c)} < \alpha'\}$$

其中α'和β'可由α，β，$Pr(C)$和$Pr(C^c)$来导出。

变精度粗糙集

在 1993 年，Ziarko[45]发表了一篇变精度粗糙集(Variable precision rough set model)的文章。该文章给出了另一个定量粗糙集模型。虽然得出的结果是决策粗糙集的特例，但在早期的文章中，Ziarko 并没有采用概率的描述和解释。变精度粗糙集的概率解释在他后期的文章中才出现[46]。变精度粗糙集的出发点是一个相对错误分类度(relative degree of misclassification)的函数。设X和Y是两个集合，集合X关于集合Y的错误分类度定义为：

$$mc(X,Y)=\begin{cases}1-\dfrac{|X\cap Y|}{|X|}, & |X|>0,\\ 0, & |X|=0,\end{cases}$$

显然，

$$X\subseteq Y\Leftrightarrow mc(X,Y)=0.$$

将这个关系一般化，可定义一个z-优势关系(z- majority relation)：$0\le z<0.5$

$$X\subseteq_z Y\ \Leftrightarrow mc(X,Y)\le z.$$

z-优势关系可以用Y包含X的程度解释，称为包含度(inclusion degree)[47]。基于z-优势关系，z-正、负和边界域定义为：

$$\begin{aligned}\mathrm{ZPOS}_z(C)&=\{x\in U\mid[x]\subseteq_z C\}\\&=\{x\in U\mid mc([x],C)\le z\}\end{aligned}$$

$$\begin{aligned}\mathrm{ZNEG}_z(C)&=(\mathrm{ZPOS}_z(C^c))^c\\&=(\{x\in U\mid mc([x],C^c)\le z\})^c\\&=\{x\in U\mid mc([x],C)\ge 1-z\}\end{aligned}$$

$$\begin{aligned}\mathrm{ZBND}_z(C)&=(\mathrm{ZPOS}_z(C)\cup\mathrm{ZNEG}_z(C))^c\\&=\{x\in U\mid z<mc([x],C)<1-z\}\end{aligned}$$

基于一个阈值z的变精度粗糙集称为对称变精度粗糙集。通过使用一对阈值(l,u)，$0\le l<u\le 1$，可以引入非对称变精度粗糙集，定义如下[46]：

$$\mathrm{ZPOS}_{(l,u)}(C)=\{x\in U\mid mc([x],C)\le l\}$$
$$\mathrm{ZBND}_{(l,u)}(C)=\{x\in U\mid l<mc([x],C)<u\}$$
$$\mathrm{ZNEG}_{(l,u)}(C)=\{x\in U\mid mc([x],C)\ge u\}$$

对称变精度粗糙集是非对称变精度粗糙集的一个特例，其中$l=z$，$u=1-z$和$z<0.5$。

关于变精度粗糙集模型，需要几个说明。一，为了讨论方便，我在前面的描述中用了基于元素的定义，而 Ziarko 用了基于粒(即等价类)的定义；同时，为了讨论的一致性也用了不同的符号系统，比如，z 对应于文献[45]中所用的 β，mc 对应于 c。虽然描述不同，但所定义的正、负和边界域是相同的。二，在 Ziarko 的讨论中，参数 z 的引入显得有点突然。这里，需要引入一些中间联系。基于 z-优势关系的定义，对于等价类 $[x]\neq\emptyset$，我们有

$$\begin{aligned}
&0\le mc([x],C)\le 1,\\
&[x]\subseteq C \Leftrightarrow mc([x],C)=0,\\
&[x]\bigcap C\neq\emptyset \wedge \neg([x]\subseteq C) \Leftrightarrow 0< mc([x],C)<1,\\
&[x]\bigcap C=\emptyset \Leftrightarrow mc([x],C)=1.
\end{aligned}$$

因此，Pawlak 三个域可以等价地定义为：

$$\begin{aligned}
\mathrm{POS}(C)&=\{x\in U\mid mc([x],C)\le 0\},\\
\mathrm{BND}(C)&=\{x\in U\mid 0< mc([x],C)<1\},\\
\mathrm{NEG}(C)&=\{x\in U\mid mc([x],C)\ge 1\}.
\end{aligned}$$

从形式上来讲，它们暗示了变精度模型的拓广定义。三，相对错误分类度 $mc([x],C)$ 与 C 在 $[x]$ 下的相对频率或 C 在 $[x]$ 下的条件概率有如下关系：

$$Pr(C\mid[x])=\frac{|C\bigcap[x]|}{|[x]|}=1-mc([x],C).$$

通过该关系，可以在决策粗糙集和变精度粗糙集之间建立一种对应关系。令 $l=1-\alpha$ 和 $u=1-\beta$，

$$\begin{aligned}
\mathrm{POS}_{(\alpha,\beta)}(C)&=\{x\in U\mid Pr(C\mid[x])\ge\alpha\}\\
&=\{x\in U\mid 1-Pr(C\mid[x])\le 1-\alpha\}\\
&=\{x\in U\mid mc([x],C)\le l\}\\
&=\mathrm{ZPOS}_{(l,u)}(C)\\
\mathrm{BND}_{(\alpha,\beta)}(C)&=\{x\in U\mid \beta<Pr(C\mid[x])<\alpha\}\\
&=\{x\in U\mid 1-\alpha<1-Pr(C\mid[x])<1-\beta\}\\
&=\{x\in U\mid l<mc([x],C)<u\}\\
&=\mathrm{ZBND}_{(l,u)}(C)\\
\mathrm{NEG}_{(\alpha,\beta)}(C)&=\{x\in U\mid Pr(C\mid[x])\le\beta\}\\
&=\{x\in U\mid 1-Pr(C\mid[x])\ge 1-\beta\}\\
&=\{x\in U\mid mc([x],C)\ge u\}\\
&=\mathrm{ZNEG}_{(l,u)}(C)
\end{aligned}$$

从这个特别角度来看，可以说变精度粗糙集是决策粗糙集的一个特例。另一方面，由于 $mc([x],C)$ 可有其他的解释，变精度模型实际上给出了另一类定量、非概率粗糙集模型。张文修,梁怡和徐萍[47]在《基于包含度的不确定推理》中，用包含度解释 $mc([x],C)$，并给

出了很多其他的包含度函数，这为粗糙集拓广提供了另一种非概率方法。

研究体会

决策粗糙集和变精度粗糙集既有联系又有区别。关于这两个模型的研究和所发表的文献，有以下几个体会：

一、认真阅读原始文献。从严格意义上讲，Ziarko 最早的文章并没有明确地给出了决策粗糙集中所提出的概率正、负和边界域。在后期的文章中，他将变精度的结果用概率的形式重新给出，给读者一种错觉。随后的很多作者都是用概率的形式来阐述变精度思想，而没有认识到概率粗糙集是在决策粗糙集中显式、明确提出来的。这说明非原始文献，特别是非直接引用，并不完全能给出原始文件中的信息和结果，而这些结果对于后面的研究是非常重要的。

二、坚持自己独特的学术观点。变精度粗糙集模型的概率描述丢失了其基于包含度的解释。本来，包含度是变精度模型的主要贡献之一，但在基于概率的描述中却丢失了。这也许是一大损失。一个更优的方案既保持它本身的特性，又建立和其他理论的联系。

三、客观评价研究工作。任何一个理论都有其优点和缺点，客观评价一个理论是非常难的一件事，但却非常重要。在关于决策粗糙集的讨论中，很多作者认为获得损失函数是很困难的也不实际的，但却没有用同样的准则去讨论变精度粗糙集所需要的阈值是否更难解释或获得。

四、重视一个理论的解释和指导功能。一个好的理论不但能解释自然现象，同时也给出正确使用该理论的条件。用阈值处理问题是人们常用的一种方式。比如，当一个函数的值大于一个阈值时，我们做一种选择；小于时，做另一种选择。决策粗糙集和变精度粗糙集都使用了阈值。变精度粗糙集并没有提供一个对阈值满意的解释，更没有提供估计阈值的方法，这使得后期的很多工作只能靠猜或试验来决定。相比而言，决策粗糙集为阈值提供了更实际的解释并给出了估计方法。用另一句话说，决策粗糙集不但解释了为什么用阈值，即，求最小决策风险，又给出了一套计算方法。从这点来说，决策粗糙集不但从时间上早于变精度粗糙集，在理论和应用上也有其优势。

研究与交流

研究的乐趣既在于发现一个新的结果，也在于结交朋友。以文会友，读一篇优美的文章是一种享受；学术交流，分享朋友优秀的研究成果令人高兴；共同合作，一起探索新的研究领域会有加倍的快乐。在决策粗糙集研究中，我得到很多良师益友的帮助与建议。特别要指出的是，过去几年中，大家一起在国际会议中组织了关于决策粗糙集的专题讨论会，推动了决策粗糙集研究。

我的决策粗糙集研究既受益于与朋友的交流，也受益于他们研究成果的启发。利用这个机会，我非常高兴地将他们关于决策粗糙集的优秀成果介绍给大家：在具有重要影响的《粗

糙集理论与方法》一书中，张文修教授、吴伟志教授、梁吉业教授和李德玉教授首先将决策粗糙集的介绍给国内读者[48]；姚静涛教授和 Joseph Harbert 博士将决策粗糙集和博弈粗糙集结合起来[49-54]; 周献中教授、李华雄博士及其研究小组给出了决策粗糙集的一个多视角解释，讨论了决策粗糙集属性约简问题及应用[55-59]; 李天瑞教授、刘盾博士及其研究小组讨论了多类决策粗糙集和应用[60-64]; 苗夺谦教授、Pawan Lingras 教授及其研究小组将决策粗糙集同聚类分析结合起来[65-68]; 杨晓平教授等将决策粗糙集与多代理(multi-agent)结合起来[69,70]; 商琳教授和贾修一等讨论了决策粗糙集中的属性约减问题[71-73]; 王国胤教授和于洪博士等讨论了决策粗糙集在聚类中的应用，也讨论了决策粗糙集的研究现状及新趋势[74-77]; Yuefeng Li 教授等将决策粗糙集用到信息检索和信息过滤中[78,79]; Salvatore Greco 教授和 Roman Slowinski 教授等将决策粗糙集和基于优势的粗糙集结合起来[80]; 周冰给出了多类决策粗糙集的一种新描述[81]。

结束语

每写一篇文章，对自己来说都是一个新的挑战。怎样和读者交流，并系统、完整地表达自己的研究成果和思想？怎样客观、不偏不倚地介绍和评论现有的工作？怎样精确地引用和借鉴其他研究者的成果并描述其对自己研究所起的作用？这些问题常常闪现在脑海中，因此，对每个符号、每句话都仔细推敲，尽量避免歧义或错误。写文章如履薄冰，一方面，不想自己的文章没有人读，成为垃圾；另一方面，更不想研究缺乏创新或文章晦涩难懂而浪费读者宝贵的时间。这篇文章是对自己关于决策粗糙集研究的整理和小结，为进一步研究做点准备。希望对决策粗糙集有兴趣的读者能从中找出有用的东西。

参考文献

[1] Yao Y Y, Wong S K M, Lingras P. A decision-theoretic rough set model\\ The 5th International Symposium on Methodologies for Intelligent Systems, 1990.

[2] Yao Y Y, Wong S K M. A decision theoretic framework for approximating concepts. International Journal of Man-machine Studies, 1992, 37(6): 793-809.

[3] Yao Y Y. Probabilistic rough set approximations. International Journal of Approximate Reasoning, 2008, 49(2): 255-271.

[4] Yao Y Y. Probabilistic approaches to rough sets. Expert Systems, 2003, 20(5): 287-297.

[5] Yao Y Y. Decision-theoretic rough set models. Rough Sets and Knowledge Technology// The 2nd International Conference on Rough Sets and Knowledge Technology, RSKT'07, 2007.

[6] Yao Y Y. Two semantic issues in a probabilistic rough set model. Fundamenta Informaticae, 2011, 108(3-4):249-265.

[7] Yao Y Y, Zhao Y. Attribute reduction in decision-theoretic rough set models. Information

Sciences, 2008, 178(17): 3356-3373.

[8] Yao Y Y. Three-way decision: an interpretation of rules in rough set theory\\ The 4th International Conference on Rough Sets and Knowledge Technology, RSKT'09, 2009.

[9] Yao Y Y. Three-way decisions with probabilistic rough sets. Information Sciences, 2010, 180(3): 341-353.

[10] Yao Y Y. The superiority of three-way decisions in probabilistic rough set models. Information Sciences, 2011, 181(6): 1080-1096.

[11] Yao Y Y, Zhou B. Naive Bayesian Rough Sets\\ The 5th International Conference on Rough Sets and Knowledge Technology, RSKT'10, 2010.

[12] Feynman R P. The Character of Physical Law. Cambridge: the MIT Press, 1967.

[13] Yao Y Y. On generalizing rough set theory// Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing, Proceedings of the 9th International Conference, RSFDGrC 2003, 2003.

[14] Yao Y Y, Lin T Y. Generalization of rough sets using modal logic. Intelligent Automation and Soft Computing, An International Journal, 1996, 2(2): 103-120.

[15] Yao Y Y. Relational interpretations of neighborhood operators and rough set approximation operators. Information Sciences, 1998, 111(1-4): 239-259.

[16] Yao Y Y. On generalizing Pawlak approximation operators// Rough Sets and Current Trends in Computing, Proceedings of the 1st International Conference, RSCTC'98, 1998.

[17] Bateson G. Mind and Nature: A Necessary Unity. New York: E. P. Dutton, 1979.

[18] Hempel C G. Philosophy of Natural Science. Englewood Cliffs, N.J. : Prentice-Hall, 1966.

[19] Glaser B G, Strauss A L. The Discovery of Grounded Theory: Strategies for Qualitative Research. Chicago: Aldine, 1967.

[20] Yao Y Y. Granular computing: basic issues and possible solutions// The 5th Joint Conference on Information Sciences, JCIS 2000, 2000.

[21] Yao Y Y. Granular computing: past, present and future// The 2008 IEEE International Conference on Granular Computing, GrC 2008, 2008.

[22] Yao Y Y. The art of granular computing// Lecture Notes in Artificial Intelligence, LNAI 4585, 2007, 101-112.

[23] Yao Y Y. The Rise of granular computing. Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition), 2008, 20(3): 299-308.

[24] Yao Y Y. A unified framework of granular computing// Handbook of Granular Computing. New Jersey: Wiley, 2008, 401-410.

[25] Yao Y Y. Artificial intelligence perspectives on granular computing// Pedrycz W and Chen S M (Eds.). Granular Computing and Intelligent Systems Design with Information Granules of Higher Order and Higher Type. Intelligent Systems Reference Library, Springer, Berlin, 2011, 13:17-34.

[26] 姚一豫. 粒计算三元论// 张燕平, 罗斌, 姚一豫, 苗夺谦, 王国胤, 张铃, 张钹. 商空间与粒计算-结构化问题求解理论与方法. 北京: 科学出版社, 2010, 115-143.

[27] 姚一豫. 粒计算的艺术// 苗夺谦, 王国胤, 刘清, 林早阳, 姚一豫. 粒计算: 过去、现在与展望. 北京: 科学出版社, 2007, 1-20.

[28] Lindsay P H, Norman D A. An Introduction to Psychology. 2nd edition. New York: Academic Press, 1977.

[29] Knuth D E. Literate programming. The Computer Journal, 1984, 27(2): 97-111.

[30] Pawlak Z. Rough sets. International Journal of Computer and Information Sciences, 1982, 11(5): 341-356.

[31] Pawlak Z. Rough Sets, Theoretical Aspects of Reasoning about Data. Dordrecht: Kluwer Academic Publishers, 1991.

[32] Pawlak Z, Grzymala-Busse J W, Slowinski R, Ziarko Z. Rough sets. Communications of the ACM, 1995, 38: 88-95.

[33] Wille R. Restructuring lattice theory: an approach based on hierarchies of concepts// Ordered Sets, Rival I (Ed.), Reidel, Dordrecht-Boston, 1982, 445-470.

[34] Ganter B, Wille R. Formal Concept Analysis, Mathematical Foundations. Berlin: Springer, 1999.

[35] Mechelen I, Hampton J, Michalski R S, Theuns P (Eds.). Categories and Concepts, Theoretical Views and Inductive Data Analysis. New York: Academic Press, 1993.

[36] Sowa J F. Conceptual Structures: Information Processing in Mind and Machine. Massachusetts: Addison-Wesley, 1984.

[37] 张文修. 信息论与或然逻辑. 西安: 西安交通大学出版社, 1992.

[38] Marek V W, Truszczynski M. Contributions to the theory of rough sets. Fundamenta Informaticae, 1999, 39: 389-409.

[39] Yao Y Y. Interval sets and interval-set algebras// Proceedings of the 8th IEEE International Conference on Cognitive Informatics, ICCI’09, 2009: 307-314.

[40] Yao Y Y. A note on definability and approximations. LNCS Transactions on Rough Sets VII, LNCS 4400, 2007, 274-282.

[41] Wong S K M, Ziarko W. Comparison of the probabilistic approximate classification and the fuzzy set model. Fuzzy Sets and Systems, 1987, 21: 357-362.

[42] Wong S K M, Ziarko W. A Probabilistic Model of Approximate Classification and Decision Rules with Uncertainty in Inductive Learning. Technical Report CS-85-23, Department of Computer Science, University of Regina, 1985.

[43] Pawlak Z, Wong S K M, Ziarko W. Rough sets: probabilistic versus deterministic approach. International Journal of Man-Machine Studies, 1988, 29: 81-95.

[44] Duda R O, Hart P E. Pattern Classification and Scene Analysis. New York: Wiley, 1973.

[45] Ziarko W. Variable precision rough set model. Journal of Computer and System Sciences, 1993, 46: 39–59.

[46] Ziarko W. Set approximation quality measures in the variable precision rough set model// Proceedings of the 2nd International Conference on Hybrid Intelligent Systems, HIS'02, 2002.

[47] 张文修, 梁怡, 徐萍. 基于包含度的不确定推理. 北京: 清华大学出版社, 2007.

[48] 张文修, 吴伟志, 梁吉业, 李德玉. 粗糙集理论与方法. 北京: 科学出版社, 2001.

[49] Yao J T, Herbert J P. A game-theoretic perspective on rough set analysis. Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition), 2008, 20(3): 291-298.

[50] Herbert J P, Yao J T. Game-theoretic risk analysis in decision-theoretic rough sets// The 3rd International Conference on Rough Sets and Knowledge Technology, RSKT'08, 2008.

[51] Herbert J, Yao J T. Learning optimal parameters in decision-theoretic rough sets// The 4th International Conference on Rough Sets and Knowledge Technology, RSKT'09, 2009.

[52] Herbert J P, Yao J T. Criteria for choosing a rough set model. Computers and Mathematics

with Applications, 2009, 57(6): 908-918.

[53] Herbert J P, Yao J T. Analysis of data-driven parameters in game-theoretic rough sets// The 6th International Conference on Rough Sets and Knowledge Technology, RSKT'11, 2011.

[54] Herbert J P, Yao J T. Game-theoretic rough sets. Fundamenta Informaticae, 2011, 108(3-4): 267-286.

[55] Zhou X Z, Li H X. A multi-view decision model based on decision-theoretic rough set// The 4th International Conference on Rough Sets and Knowledge Technology, RSKT'09, 2009.

[56] Li H X, Zhou X Z. Risk decision making based on decision-theoretic rough set: a multi-view decision model. International Journal of Computational Intelligence Systems, 2011, 4(1): 1-11.

[57] Li H X, Zhou X Z, Zhao J B, Liu D. Attribute reduction in decision-theoretic rough set model: a further investigation// The 6th International Conference on Rough Sets and Knowledge Technology, RSKT'11, 2011.

[58] Liu D, Li H X, Zhou X Z. Two decades’ research on decision-theoretic rough sets// Proceedings of the 9th IEEE International Conference on Cognitive Informatics, ICCI’10, 2010.

[59] 李华雄, 刘盾, 周献中. 决策粗糙集模型研究综述. 重庆邮电大学学报(自然科学版), 2010, 22(5): 624-630.

[60] Liu D, Li T R, Hu P, Li H X. Multiple-category classification with decision-theoretic rough sets// The 5th International Conference on Rough Sets and Knowledge Technology, RSKT'10, 2010.

[61] Liu D, Li T R, Ruan Da. Probabilistic model criteria with decision-theoretic rough sets. Information Sciences, 2011, 181: 3709-3722.

[62] Liu D, Yao Y Y, Li T R. Three-way investment decisions with decision-theoretic rough sets. International Journal of Computational Intelligence Systems, International Journal of Computational Intelligence Systems，2011, 4(1): 66-74.

[63] Liu D, Li T R and Liang D. A new discriminant analysis approach under decision-theoretic rough sets// The 6th International Conference on Rough Sets and Knowledge Technology, RSKT'11, 2011.

[64] 刘盾, 姚一豫, 李天瑞. 三枝决策粗糙集. 计算机科学, 2011, 38(1): 245-250.

[65] Li W, Miao D Q, Wang W L, Zhang N. Hierarchical rough decision theoretic framework for text classification// Proceedings of the 9th IEEE International Conference on Cognitive Informatics, ICCI’10, 2010..

[66] Lingras P, Chen M, Miao D Q. Rough multi-category decision theoretic framework// The 3rd International Conference on Rough Sets and Knowledge Technology, RSKT'08, 2008.

[67] Lingras P, Chen M, Miao D Q. Semi-supervised rough cost/benefit decisions. Fundamenta Informaticae, 2009, 94 (2): 233-244.

[68] Lingras P, Chen M, Miao D Q. Rough cluster quality index based on decision theory. IEEE Transactions on Knowledge and Data Engineering, 2009, 21 (7): 1014-1026.

[69] Yang X P, Yao J T. A multi-agent decision-theoretic rough set model// The 5th International Conference on Rough Sets and Knowledge Technology, RSKT'10, 2010.

[70] Yang X P, Song H G, Li T J. Decision making in incomplete information system based on decision-theoretic rough sets// The 6th International Conference on Rough Sets and Knowledge Technology, RSKT'11, 2011.

[71] Jia X Y, Li W W, Shang L, Chen J J. An optimization viewpoint on decision-theoretic rough set model// The 6th International Conference on Rough Sets and Knowledge Technology, RSKT'11, 2011.

[72] 贾修一, 商琳, 陈家骏. 基于三值决策的属性约简// 中国人工智能进展, 2009, 北京: 北京邮电大学出版社, 2009, 193-198.

[73] 贾修一, 商琳, 陈家骏. 决策风险最小化属性约简. 计算机科学与探索, 2011, 5(2): 155-160

[74] Yu H, Chu S S, Yang D. Autonomous knowledge-oriented clustering using decision-theoretic rough set theory// The 5th International Conference on Rough Sets and Knowledge Technology, RSKT'10, 2010.

[75] Yu H, Liu Z G, Wang G Y. Automatically determining the number of clusters using decision-theoretic rough set// The 6th International Conference on Rough Sets and Knowledge Technology, RSKT'11, 2011.

[76] 王国胤, 于洪, 姚一豫. 决策粗糙集研究综述. 手稿, 2011.

[77] 王国胤, 姚一豫, 于洪. 粗糙集理论与应用研究综述. 计算机学报, 2009, 1229-1247.

[78] Li Y F, Zhang C Q, Swanb J R. Rough set based model in information retrieval and filtering// Proceeding of the 5th International Conference on Information Systems Analysis and Synthesis, ISAS’99, 1999.

[79] Li Y F, Zhang C Q, Swanb J R. An information filtering model on the web and its application in job agent. Knowledge-Based Systems, 2000, 13: 285-296.

[80] Greco S, Slowinski R, Yao Y Y. Bayesian decision theory for dominance-based rough set approach// The 2nd International Conference on Rough Sets and Knowledge Technology, RSKT'07, 2007.

[81] Zhou B. A new formulation of multi-category decision-theoretic rough sets// The 6th International Conference on Rough Sets and Knowledge Technology, RSKT'11, 2011.