

<<Reduct理论>>

图书基本信息

书名：<<Reduct理论>>

13位ISBN编号：9787302219576

10位ISBN编号：7302219575

出版时间：2010-4

出版时间：清华大学出版社

作者：韩素青，赵岷 著

页数：362

版权说明：本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问：<http://www.tushu007.com>

## &lt;&lt;Reduct理论&gt;&gt;

## 前言

Pawlak研究Rough Sets的原始动机是针对数据集合（决策表），找到一种介于严格统计学和随意经验之间的能够更好地描述知识不确定性的严谨方法。

然而，由于基于Rough Sets描述知识不确定性的度量Roughness只与给定的决策表相关，因此，Roughness能否真实描述决策表相对应的问题世界，取决于给定的决策表对于问题世界的概括程度。在以泛化为目标的统计机器学习中，如果样本集是由对问题世界的有限次观测构成的，那么，该样本集同样需要满足一个表示“概括程度”的条件：样本集合与问题世界同分布，这是讨论泛化问题的基础。

Rough Sets中的Roughness与统计学的概率定义在形式上有些相似，但是，两者的本质有所不同，这一点，我们在本书的第1章中作了说明。

这种本质上的不同导致我们无法从Roughness进一步推广并且定义类似于统计学中分布这样的概念，这意味着，Rough Sets不具备从有限观测（给定数据集合）推测问题世界的理论基础。

另外，由于Roughness来源于给定的数据集合，因此，这个不确定度量取决于且仅仅取决于给定的数据集合。

这一点与模糊集等依赖主观经验的不确定描述方法也不相同。

这些原因，使得Roughness是否是问题世界的真实描述，不得不依赖于给定数据集合对问题世界的概括程度。

更进一步，考虑到Roughness所依据的基础——等价关系对数据集合的限制——符号集合，那么把从给定决策表获得的Roughness作为描述问题世界知识不确定性的度量，其可信度就更存在疑问了。

这是本书为什么没有像大多数有关论著那样以Rough Sets为题的原因所在。

## <<Reduct理论>>

### 内容概要

本书系统介绍了基于用户需求的Reduct理论。

主要内容包括Reduct理论、Reduct典型算法、用户需求描述、基于用户需求的Reduct理论、Reduct与特征选择、数据描述的“规则+例外”模型以及基于边缘区域的例外分析等。

其中数据描述的“规则+例外”模型源自认知科学，不仅与数据挖掘密切相关，而且与用户需求密切相关。

本书适合从事机器学习、数据挖掘、人工智能、信息处理研究和应用的科技人员学习参考。

## &lt;&lt;Reduct理论&gt;&gt;

## 书籍目录

第1章 概述 1.1 Rough的含义 1.2 Reduct 1.3 Reduct 计算 1.4 用户需求描述 1.5 次属性定理 1.6 基于用户需求的最优Reduct计算 1.7 规则+例外 1.8 符号机器学习 1.9 特征选择 1.10 小结第2章 Reduct理论与计算 2.1 引言 2.1.1 初等范畴与基本范畴 2.1.2 集合的近似 2.1.3 信息系统的知识表示 2.1.4 信息系统的属性约简 2.1.5 信息系统的范畴约简 2.1.6 决策表的知识表示 2.1.7 决策表的属性约简 2.1.8 决策表的范畴约简 2.1.9 决策表约简 2.2 差别矩阵原理 2.2.1 信息系统的差别矩阵 2.2.2 决策表的差别矩阵 2.3 Reduct计算 2.3.1 基于属性独立性的约简算法 2.3.2 基于正区域的约简算法 2.3.3 基于互信息的约简算法 (MIBARK算法) 2.3.4 基于差别矩阵原理的约简算法 2.3.5 基于先验知识的约简算法 2.4 小结第3章 用户需求描述 3.1 属性的用户偏好 3.1.1 属性的定量评价描述 3.1.2 属性的定性评价描述 3.2 属性定量评价与定性评价之间的关系 3.3 属性子集的用户偏好 3.3.1 基本性质 3.3.2 属性子集的定量评价 3.3.3 属性子集的定性评价 3.4 Reduct的用户偏好 3.5 小结第4章 基于差别矩阵的属性序Reduct算法 4.1 属性序 4.2 属性序Reduct算法及性质 4.2.1 基本概念 4.2.2 属性序Reduct算法 4.2.3 算法解的完备性及唯一性 4.3 基于自由属性的属性序Reduct算法 4.3.1 基本概念 4.3.2 基于自由属性的Reduct算法 4.3.3 算法解的完备性 4.4 基于差别矩阵初等运算的属性序Reduct算法 4.4.1 差别矩阵的初等运算 4.4.2 基于初等运算的Reduct算法 4.4.3 基于初等运算的属性序Reduct算法 4.4.4 基于条件偏好关系的属性序Reduct算法 4.5 小结第5章 基于属性?值树的属性序Reduct算法 5.1 基本属性?值树及生成算法 5.1.1 初等范畴和基本范畴 5.1.2 树结构 5.1.3 基本属性?值树 5.1.4 基本属性?值树的生成算法 5.2 完全属性?值树 5.3 正区域的属性?值树表示 5.3.1 属性?值树表示下正区域的定义与性质 5.3.2 属性?值树表示下正区域 5.4 封闭属性?值树 5.4.1 死子树与活子树 5.4.2 封闭属性?值树表示 5.5 Core属性的属性?值树表示 5.5.1 属性?值树表示下Core属性的定义与性质 5.5.2 属性?值树表示下Core的计算 5.6 Reduct的属性?值树表示 5.6.1 Reduct的计算方法 5.6.2 Reduct算法的完备性 5.7 属性值?Core与属性值?Reduct的属性?值树表示 5.7.1 属性值?Core的属性?值树表示 5.7.2 属性值?Reduct的属性?值树表示 5.8 属性序Reduct算法与属性?值树Reduct算法的等价性 5.9 关于树结构的讨论 5.10 小结第6章 属性序空间与Reduct空间之间的关系 6.1 满足用户偏好最优Reduct的计算复杂度 6.2 属性序偶与属性序Reduct算法的形式化描述 6.2.1 基本概念 6.2.2 属性序Reduct算法的形式化描述 6.2.3 属性序偶的性质 6.3 邻近属性序偶Reduct的基本判定 6.3.1 差别元素聚合命题 6.3.2 等价类分解命题 6.3.3 邻近属性序偶基本判定定理 6.4 邻近属性序偶Reduct判定规则 6.4.1 无条件判别规则 6.4.2 子区间判别规则 6.4.3 单向与双向判别规则 6.5 小结第7章 次属性原理及属性?值树次属性算法 7.1 次属性 7.1.1 基本概念 7.1.2 次属性原理 7.1.3 次属性定理 7.2 属性?值树次属性算法及算法的完备性 7.2.1 差别矩阵与属性?值树表示 7.2.2 属性?值树次属性算法 7.2.3 属性?值树次属性算法的完备性 7.3 小结第8章 任意属性序偶Reduct判定 8.1 属性序之间的关系及属性移动基本规则 8.2 次属性变化规律 8.3 任意属性序偶Reduct是否相同的判定问题 8.3.1 任意属性序偶Reduct基本判定 8.3.2 任意属性序偶Reduct判定 8.4 属性范序与属性序偶Reduct判定 8.4.1 基本概念 8.4.2 基于属性范序的属性序偶Reduct判定 8.5 小结第9章 基于用户偏好最优Reduct计算 9.1 满足用户偏好的最优Reduct 9.2 次属性定理与最优Reduct计算 9.2.1 最优Reduct的定量描述 9.2.2 次属性定理与搜索策略 9.2.3 最优Reduct逼近算法 9.2.4 算法复杂性分析 9.3 小结第10章 特征选择与Reduct计算 10.1 特征选择概述 10.1.1 最优特征子集的搜索问题 10.1.2 特征和特征子集评价问题 10.1.3 特征子集的产生方式 10.1.4 特征选择和学习算法之间的关系 10.1.5 特征选择和特定应用之间的关系 10.2 Reduct与特征选择之间的关系 10.2.1 基本概念 10.2.2 Reduct的搜索与评价问题 10.2.3 Reduct的产生方式以及与学习算法之间的关系 10.2.4 基于删除策略的Reduct计算 10.2.5 基于添加+删除搜索策略的Reduct计算 10.2.6 基于添加策略的Reduct计算 10.3 小结第11章 数据描述的“规则+例外”模型 11.1 认知心理学关于概念的研究 11.1.1 概念结构的假说 11.1.2 概念形成 11.2 规则归纳 11.2.1 基本搜索策略 11.2.2 样例与规则相结合的方法 11.2.3 常用归纳算法 11.2.4 规则归纳小结 11.3 粒度与粒计算 11.3.1 粒度 11.3.2 粒计算 11.4 例外分析 11.4.1 例外与“Outlier” 11.4.2 例外分析的应用 11.4.3 基于建模的例外分析方法 11.4.4 基于模式的例外分析方法 11.4.5 关于例外分析的讨论 11.5 规则+例外模型 11.5.1 脊椎动物世界——一个例子 11.5.2 “规则+例外”模型研究 11.6 正区域和边缘区域扩展研究 11.6.1 正区域

## &lt;&lt;Reduct理论&gt;&gt;

11.6.2 认知正区域与认知边缘区域 11.7 文本粒度与文本粒子 11.7.1 文本粒度 11.7.2 文本粒子 11.8  
小结第12章 边缘区域与例外分析 12.1 边缘区域 (BR) 的结构研究 12.1.1 例子 12.1.2 BR的结构  
12.1.3 “活的”与“死的” CPOS——关于边缘区域的进一步讨论 12.2 基于BR的差别矩阵研究 12.2.1  
BR的差别矩阵 12.2.2 合并问题 12.2.3 CPOS的死活问题 12.3 基于CPR的Reduct计算 12.4 Core属性与  
例外鉴别 12.4.1 Core属性 12.4.2 Core属性的性质 12.4.3 差别矩阵中Core的分布 12.5 基于差别矩阵  
的例外鉴别 12.5.1 从PRAS中鉴别例外 12.5.2 从正区域的PR中鉴别例外 12.5.3 例子和讨论 12.6 基  
于概念结构的例外鉴别 12.6.1 例——基于原型的方法 12.6.2 例——基于异类之间相似度的方法 12.7  
小结参考文献算法索引

## 章节摘录

插图：Rough在Rough Sets中的含义可以解释为一种Rough度量，用于表示知识的不确定程度。

目前存在多种关于Rough度量的描述方法，这里仅就Pawlak给出的原始描述方法进行讨论，其他描述方法大同小异。

首先考虑Pawlak对于矛盾对象的描述：如果决策表中两个对象关于条件属性集中所有属性的值完全相同，但关于决策属性集中有些属性的值不同，则称这两个对象是矛盾的。

例如，对于两个记录在案的病例，如果病人的症状都是发烧、咳嗽，但医生的最后诊断分别为感冒和肺炎，那么这两个病例是一对有矛盾的对象。

其中，两个病例表示两个对象，“发烧”和“咳嗽”是两个条件属性，“疾病”是决策属性，“感冒”和“肺炎”是对象关于决策属性的属性值。

显然，这个例子中的两个对象关于条件属性“发烧”和“咳嗽”的属性值完全相同，即这两个对象都以“是”作为它们关于条件属性“发烧”和“咳嗽”的属性值，但是，它们关于决策属性的属性值不同，因此，这两个病例是矛盾的对象。

Pawlak对于矛盾对象的刻画采用了严格的数学形式。

在给出矛盾对象的定义之前，首先引入了上、下近似两个概念尽管用“矛盾对象集合”（即，用表示知识的Rough程度）描述知识的不确定性，具有直观的优点，但是，对于与“知识（规则）”相关的正区域的描述，则不得不基于“矛盾对象集合”来间接定义。

我们猜测，作为数学家的Pawlak大概不喜欢这种使用次要概念间接描述更重要概念的方式，也不喜欢在定义阶段就将两个概念相互联系，由此，他引进了上下近似的概念，从而，可以分别独立地定义正区域和边缘区域。

，然后基于这两个概念，决策表的论域被划分为两个部分，分别称为正区域（positive region）和边缘区域（boundary region），并且正区域和边缘区域满足：（1）正区域和边缘区域的交集为空集；（2）正区域和边缘区域的并集为论域。

其中只有边缘区域与矛盾对象的描述有关。

为了便于读者理解Pawlak关于矛盾对象的描述思想，在这一章，我们采用直观的方法对两个区域进行刻画，精确的定义参见本书第2章。

<<Reduct理论>>

编辑推荐

《Reduct理论》：中国计算机学会学术著作丛书

<<Reduct理论>>

版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问:<http://www.tushu007.com>