粗糙集理论中求取最小决策规则的研究*)

潘 巍[1] 王阳生[2] 杨宏戟[3]

(首都师范大学信息工程学院 北京 100037)[1]

(中国科学院自动化研究所模式识别国家重点实验室 北京 100080)[2]

(Software Technology Research Laboratory, De Montfort University, Leicester, LE1 9BH, England)[3]

摘 要 本文探讨了粗糙集理论中最小决策规则的求取方法,提出决策依赖度的定义,尝试从最短的条件属性组合中提取尽可能多的决策规则。只有现有长度的决策规则无法完全覆盖所有样本时,才会考虑增加决策规则的长度。同时提出了3种减少计算复杂性的方案:1)引入跳跃系数λ;2) 在计算中只对具有相同决策值的样本进行等价类划分,从而避免了对含有不同决策值的等价类的无用划分;3) 设计 Remain 集合,只针对其中的样本进行等价类的划分,随着 Remain 中样本数的减少,计算量会大幅下降。此外,本文所提出的基于决策依赖度的跳跃式决策规则求取方法可以直接应用于不完备信息系统,因此具有良好的实用价值。

关键词 粗糙集理论,决策规则,条件属性,决策依赖度

A New Algorithm of Discretization of Consecutive Attributes Based on the Decision in Rough Sets

PAN Wei[1] WANG Yang-Sheng[2] YANG Hong-Ji[3]

(Institute of Information Engineering, Capital Normal University, Beijing 100037)[1]

(Institute of Automation, Chinese Science Academies, Beijing 100080)[2]

(Software Technology Research Laboratory, De Montfort University, Leicester, LE1 9BH, England)[3]

Abstract Under the base on analyzing the traditional methods of the minimal decision rules in Rough Set Theory, defines the concept of the decision dependability and proposes a novel algorithm of obtaining the "optimal" decision rules as many as possible from the shortest combinations of condition attributes. The length of decision rules is necessary to be extended only when the current decision rules can't represent all samples. At the same time, three methods are proposed to reduce the computational complexity: 1) defines the concept of bound coefficient, 2) only classify the samples with the same decision values at a time, 3) defines the Remain set. Above-mentioned methods can be used directly for incomplete information systems and have great practicability.

Keywords Rough set theory, Decision rule, Condition attribute, Decision dependability

1 引言

粗糙集理论是一种处理模糊和不精确性问题的新型数学工具,其主要思想就是在保持分类能力不变的前提下,通过知识约简找到核值,导出问题的决策或分类规则。由于粗糙集理论具备很强的定性分析能力,因此已被成功地应用于机器学习、故障诊断、决策分析、过程控制、模式识别、数据挖掘等领域,并取得了成功。

粗糙集理论的局限性主要在于:1)目前尚无关于知识约简的高效算法,同一问题在不同知识表示下的算法难度有所不同,而且粗糙集理论中的很多概念与运算的直观性较差,人们并不容易理解其本质;2)该理论是基于完备信息系统这样一个假设,而在现实生活中,由于数据的测量误差、对数据理解或知识获取的限制等原因,人们往往面临的是不完备信息系统,直接应用原有的概念或算法有时并不能得到准确的结论。为了使粗糙集的适应性更广,国内外学者纷纷展开了研究并取得了一定的成果。

2 粗糙集理论中求取最小决策规则的研究

决策规则提取是粗糙集理论中最重要的环节之一。国内外众多学者之所以热衷于研究更快、更好的属性约简算法,其最终目的也是为了决策规则的更简单、更准确,而决策规则的性能直接决定了信息系统的识别率。

2.1 常用的决策规则求取方法

2.1.1 决策树学习

决策树学习是一种逼迫离散值目标函数的方法,在这种方法中学习到的函数被表示为一棵决策树,学习到的规则可表示为多个 if-then 的规则,可读性高。

目前,很多粗糙集理论中的决策规则提取算法都是基于决策树学习的[1~5]。ID3 算法及其衍生的 C4.5[5]被公认是决策树学习中最优秀的算法,而目前基于决策树学习的算法多数是针对 ID3 和 C4.5 做进一步的改进研究。ID3 算法的缺点是当遍历决策树空间时,算法只维护单一的当前假设,而且在搜索中不进行回溯,容易收敛到局部最优而不是全局最优的答案。此外,如何确定决策树增长的深度、处理连续值的属性、选择合适的根节点、处理属性不完整的训练数据、提高计算效率等问题也是决策树学习算法中的一些实际问题。

2.1.2 概念学习[6]

对于粗糙集理论,每个决策属性值就是一个概念。给定一个决策表,根据不同的概念,系统将自动推断出该概念的一

*)国家863高技术研究发展计划项目(编号:2003AA114020)。潘 巍 博士,讲师,研究方向为模式识别、信息融合;王阳生 研究员,博士生导师,研究方向为模式识别;杨宏戟 博士生导师,英国 De Montfort 大学软件技术研究室主任,研究方向为软件工程。

般定义。概念学习使用 more-general-than 的偏序形式来搜索与训练样本相一致的假设。

1)Find-S 算法[1]。Find-S 算法中,首先将概念 c 初始化为最特殊的假设 h,然后在该假设 h 覆盖正例失败时将其一般化。在每一步中,h 只在需要覆盖新的正例时被泛化,对反例则不需要做出任何更改。Find-S 算法的重要特点是对以属性约束的合取式描述的假设空间(决策表 S),保证输出是在 S 中与正例一致的最特殊的假设。只要正确的目标概念包含在 S 中并且训练数据都是正确的,最终的假设也与所有反例一致。但 Find-S 算法没法确定它是否收敛到了正确的目标概念,而且由于它忽略所有的反例,如果训练数据中出现噪声或错误,该算法性能会受到严重破坏。此外,Find-S 算法同样面临训练中无法回溯的问题。

2)变型空间(version space)与候选消除法[6,7]。与 Find-S 算法不同,候选消除法输出的是与训练样例一致的所有假设的集合,而不仅仅是能够拟合训练样例的多个假设中的一个。候选消除法首先将变型空间初始化为包含假设空间 S 中所有的假设,然后从中去除与任一训练样例不一致的假设。变型空间随着观察到越来越多的样例而逐渐缩小,候选消除法对噪声敏感。如果训练数据中包含错误,算法肯定会从变型空间中删除正确的目标概念,因为它会删除所有与样例不一致的假设。

其它诸如等级学习(Hierarchical)、函数推理(Function induction)、增强学习等方法也可用于决策规则的求取,Witten、MacDonald[6]对此做了较为详细的介绍。

进行决策规则提取时,传统的方法是先进行条件属性的约简,然后消除冗余的属性值,得到条件属性核值表,再求取所有决策规则的属性值简化表,最后从中寻找一种决策规则的最小解。本文提出直接对决策表进行步进式决策规则求取的方法,通过决策依赖度 β 的计算和跳跃系数 λ 的选择完成最小决策规则集的选择,从而省略了计算核值表和所有简化决策规则集的过程,减少了计算复杂性。

3 基于决策依赖度的跳跃式决策规则求取方法

3.1 决策依赖度 β

设决策表 $S=\langle U,C\cup D,V,f\rangle$,$x\in U$,$B\subseteq C$,$X_i$ 是根据条件属性集 B 得到的等价类,$i=1,2,\cdots,m$,m 是条件等价类的个数,Y_j 是根据决策属性 D 得到的等价类,$j=1,2,\cdots,n$,n 是决策等价类的个数,则决策依赖度 β 的定义为

$$\beta_B(Y_j)=\begin{cases}\sum|X_i|/|Y_j|, & X_i\subseteq Y_j\\ 0, & \text{else}\end{cases}\qquad(1)$$

$\beta_B(Y_j)$ 表示根据条件属性集 B,能确定推导出属于决策类 Y_j 的对象比例。$0\leqslant\beta_B(Y_j)\leqslant 1$,$\beta_B(Y_j)=1$ 时,表示根据 B 能完全判断某个对象是否属于 Y_j。

3.2 跳跃系数 λ

设决策表 $S=\langle U,C\cup D,V,f\rangle$,根据决策规则的长度,可以把决策规则分为长度为 i 的规则,$i=1,2,\cdots,|C|$。所谓跳跃系数 λ,是指从当前长度为 i 的决策规则提取过程进入长度为 $i+\lambda$ 的决策规则提取过程。例如,当前正在处理的是长度为 1 的决策规则,$\lambda=3$ 表示跳过长度为 2 和 3 的决策规则,直接提取长度为 4 的决策规则。

定理 1 设决策表 $S=\langle U,C\cup D,V,f\rangle$,$B\subseteq C$,如果 $\beta_B(Y_j)=0$,则 $\forall E\subset B$,$\beta_E(Y_j)=0$。

证明:集合论中有性质:$\Phi\subset\Gamma$,若 $\Phi\not\subset H$,则 $\Gamma\not\subset H$。$E\subset B$ 表明 B 是比 E 更细的条件划分,即 B 的任何一个等价类 $[x_i]_B$ 都可以看作是 E 的某一等价类 $[z_k]_E$ 的子集,且 $\cup[x_i]_B=\cup[z_k]_E=U$。$[x_i]_B$ 表示包含元素 $x_i\in U$ 的 B 的一个等价类。$\beta_B(Y_j)=0$ 表示 B 的任何一个等价类都不是 Y_j 的子集,这表示 E 的任何一个等价类也都不是 Y_j 的子集,故 $\beta_E(Y_j)=0$。

跳跃系数 λ 可以加速决策规则的提取。在求取决策规则时,可以先从寻找较长的决策规则入手。例如,如果 $\beta\{a,b,c\}(Y_j)=0$,$a,b,c\in C$,根据定理 1 可直接得到 $\beta_Z(Y_j)=0$,$z\in\{\{a\},\{b\},\{c\},\{ab\},\{ac\},\{bc\}\}$;如果 $\beta\{a,b,c\}(Y_j)\neq 0$,则再返回计算较短长度的决策规则,这样可以大大减少计算的复杂度。当然,跳跃系数 λ 应根据实际情况而定,如果决策表中条件属性个数较多,可以让跳跃的步伐大一些,反之则要小一些。λ 可根据需要在程序中随时调整,λ=1 表示根据长度顺序依次进行决策规则的提取。

3.3 决策规则提取算法

在介绍具体算法之前,先对要用到的主要变量进行说明:

决策表 $S=\langle U,C\cup D,V,f\rangle$,$U$ 是样本集合,C 是条件属性集,D 是决策属性集。

Decision(集合):用于存放决策的等价类;

Remain(集合):用于存放尚无法进行明确分类的样本,形式与 Decision 相同;

ConditionCombination(集合):用于存放当前要计算的条件属性组合的等价类;

Belta[属性组合][决策种类数]:决策依赖度;

CurrentRuleLength:当前正在计算的决策规则长度;

m:决策种类数。

最优决策规则求取算法步骤如下:

1)对决策表 S 进行属性约简,并得到约简后的决策表,不妨仍记作 S,读入数据,Remain 初始化为全体样本,即此时还没有任何样本可以明确分类;λ=1,CurrentRuleLength=1,BeltaY[j]=0,$j=1,2,\cdots,m$,计算 Decision 的等价类。

2)生成长度为 CurrentRuleLength 的属性组合,Belta[i][j]初始化为 0,针对 Remain 中的样本,计算各属性组合 Condition Combination (CurrentRuleLength) 的等价类。计算时采用异决策比较方式,如果两个样本的属性值相同,但决策值不同,则不再对该属性值进行计算并放弃依据它划分的等价类。对其作好标记后,转入下一个样本的属性值提取。将提取的属性值与做过标记的所有属性值相比,如果属性值是新值(没有做过标记),则对其进行等价类的划分,否则再提取下一个样本。完成对所有的样本的等价类划分后,对所提取的等价类计算 Belta[i][j]值,$i=1,2,\cdots$,等价类数,$j=1,2,\cdots,m$。由于在计算中提取的都是具有相同决策值的等价类,因此 Belta[i][j]值均不为 0。

3)for $j=1$ to m

for $i=1$ to 当前属性组合总数,将最大的 Belta[i][j]所对应的 Condition Combination[i]中的样本从 Remain 中相应的 j 类中排除,如果排除的个数不为 0,则提取对应的分类规则。特别地,如果 maxBelta[i][j]=1,表明当前的属性组合能完全确定第 j 类决策属性。如果第 j 类经排除后为空,表明针对第 j 类的决策规则提取已完成,在以后的计算中不再予以考虑,直接转入对下一个决策类的计算。如果第 j 类不为空,再从大到小依次考虑其它 Belta[i][j]! =0 的情况,如果它所对应的 ConditionCombination[i]中的样本从 Re-

main 中相应的 j 类中排除的个数不为 0，则提取对应的分类规则。

endfor

endfor

4)检查 Remain 中各决策等价类的剩余样本情况，如果已有等价类为空，则 $m=m-1$。CurrentRuleLength= CurrentRuleLength+1，生成相应的属性组合，返回步骤 2)，仅针对 Remain 中剩余的样本进行等价类的划分。如果 Remain 为空，表明所有样本已有明确分类，决策规则提取完毕，算法结束。

4 实验

实验采用 Lenses 数据集，用 ID3 算法提取的决策规则数为 9 条，其中长度为 1 的规则 1 条、长度为 2 的规则 2 条、长度为 4 的规则 6 条。表 1 是采用本文方法所提取的决策规则。与用 ID3 算法提取的决策规则相比，虽然都是 9 条规则，但本文所提取的决策规则最长为 3，显然更为简洁，而表 1 中所列的决策规则也确实是该数据集的最小决策规则。

表 1 Lenses 数据集的决策规则

规则	条件属性				决策属性
	a	b	c	d	
1	—	—	—	1	3
2	2	2	2	—	3
3	3	2	2	—	3
4	3	1	1	—	3
5	1	—	1	2	2
6	2	—	1	2	2
7	—	2	1	2	2
8	1	—	2	2	1
9	—	1	2	2	1

分析本文提出的决策规则提取方法，我们总是试图从最短的条件属性组合中提取尽可能多的决策规则，只有现有长度的决策规则无法完全覆盖所有样本时，才会考虑增加决策规则的长度。

由于在计算过程中，需要计算的条件属性的组合数是条件属性的幂集，因此，为减少计算复杂性，本文在程序中加入 3 个环节：

1)引入跳跃系数 λ，先计算较长的条件属性组合的决策依赖度 β。如果 β 为 0，则表明该属性组合的所有子集均不必再计算；如果 β 不为 0，再跳回其较短的子集进行计算。如此跳跃式计算，在属性较多的情况下可以大大减少计算复杂性。以实验一为例，如果起始时 $\lambda=2$，ab、ac、bc 的 β 值均为 0，而 ad、bd、cd 的 β 值不为 0，由此可知，只有 d 的 β 值可能不为 0，跳回去只计算 d，d 的 β 值不为 0，由此得到基于 d 的决策规则。

2)第二个减少计算量的方法是在计算中只对具有相同决策值的样本进行等价类划分，这样处理的好处是避免了对含有不同决策值的等价类的划分。

3)第三个减少计算量的方法是设计 Remain 集合，只针对 Remain 中的样本进行等价类的划分。实验一经过第一轮计算后，Remain 中的样本减为 12 个。以样本 2 为例，在新的属性组合(例如 ab)，只要发现样本 2 具有属性值相同，但决策不同的情况，即放弃对样本 2 的分类，转入下一个样本 4。通过与第二个方法的结合，随着 Remain 中样本数的减少，计算量会大幅下降。

5 基于不完备决策表的决策规则求取方法与实验

对不完备决策表直接进行决策规则求取的方法不是很多[8,9]，而这些方法的思路也主要是基于概率统计或信息熵的方法，或多或少都带着推测的成分。以文[8]为例，通过对决策表中出现的所有的属性组合进行统计，并对大于某一阈值 T_1 的属性组合计算悲观可信度(也可以是乐观度)。若悲观可信度大于预定阈值 T_2，则该属性组合就成为一条可能决策强规则。这种方法有个明显的缺陷，就是如果决策表中某一属性组合的出现次数少于设定阈值 T_1，该属性组合就会被直接删除，这是不太合理的，可能会把确实为真但出现次数较少的属性组合误删掉。

本文在基于决策依赖度的跳跃式最优决策规则求取方法的基础上提出基于不完备决策表的决策规则直接求取方法，就是充分利用决策表中的每一个已知信息而不是人为地添加信息，因为添加的内容很可能与实际不符，反而会造成识别的不准确。与基于完备决策表的计算流程基本相似，都是要划分各条件属性组合的等价类并计算决策依赖度，只是在划分等价类时，只考虑已知属性值，忽略缺失值。

5.1 实验

为方便比较，对文[8]提供的不完备数据集进行决策规则的直接求取，在划分等价类时忽略缺失值。决策规则如表 2 所示，文[8]的决策强规则如表 3 所示。经验证，本文提取的决策规则不能覆盖样本 6，于是在不与决策规则冲突的条件下，对样本 6 进行缺失值的属性组合枚举，结果显示各组合均能获得明确分类，没有拒识。文[8]所获取的决策强规则数虽然也是 7 条，但与现有样本冲突的有 4 条规则(1,2,6,7)，此外强规则 5 是强规则 1 的冗余。

表 2 本文方法提取的决策规则

规则	舒适程度	价格	里程	速度	百公里耗油	外观	决策
1						好	好
2						差	差
3			5000				好
4	差		2000				差
5	好			高			好
6	差				100		差
7		低		高			好

表 3 文[8]中的决策强规则

规则	舒适程度	价格	里程	速度	百公里耗油	外观	决策
1	好						好
2		低					好
3			5000				好
4	差		2000				差
5	好			高			好
6		低			80		好
7				高	80		好

本文的对不完备决策表进行直接决策规则提取的方法是行之有效的，能挖掘出现有决策表中最丰富的数据内在联系，而且计算流程与基于完备决策表的方法基本一致，因此具有很好的实用价值。

(下转第 191 页)

确识别率。可见，选用了新的柔性可信度函数后，对提取出的规则集而言，最大识别率要优于使用阶梯函数的效果。

5.3 lenses 数据集

lenses 数据集是完备、无冲突、无噪声的，其中有 24 个样本，每个样本包含 4 个条件属性和 1 个决策属性。决策属性有 3 个决策类，且这 3 个决策类是不等量划分的，分别占样本总数的 16.67%，20.83%和 62.5%。在可信度阈值 $Tr=0.95$ 和陡度因子 $a=25$ 的条件下，得到 9 条规则，使得样本的识别率为 100%，与标准结果相同。这说明，本文的规则度量函数可以很好地处理决策类不等量划分的情况。

结论 本文使用免疫算法作为分类规则提取的工具，结合小生境技术，提出了柔性可信度的概念和新的规则重要性度量方法，将其作为免疫算法中抗体亲和力的度量公式进行了仿真实验。实验结果表明，无论是处理无噪声的数据，还是有噪声或决策类不均衡的数据，本文的方法都可以取得较好的效果。

参 考 文 献

1 Khoo L P, Zhai L Y. A prototype genetic algorithm-enhanced rough set-based rule induction system. Computers in Industry, 2001, 46(1): 95～106
2 He M, et al. Genetic algorithm based multi-knowledge extraction method for rough set. Mini-Micro System, 2005, 26(4): 651～654
3 Freitas A A. On rule interestingness measure. Knowledge-Based System, 1999, 12(4): 309～315
4 黄席樾, 等. 现代智能算法理论及应用. 北京: 科学出版社, 2005
5 Ghosh A, Nath B. Multi-objective rule mining using genetic algorithms. Information Sciences, 2004, 163(1-3): 123～133
6 Wu Z J, et al. A Chromosome-based Evaluation Model for Computer Defense Immune Systems. In: Proceedings of 2003 IEEE Congress on Evolutionary Computation, 2003. 1363～1369
7 Zheng J G, Wang L, Jiao L C. An immune algorithm for generalized rule induction. In: Proceedings of CIE International Conference of Radar Proceedings, 2001. 1023～1026
8 Goldberg D E, Richardson J J. Genetic algorithms with sharing for multimodal function optimization. Genetic algorithms and their application. In: Proceedings of the Second International Conference of Genetic Algorithms, 1987. 41～49

（上接第 184 页）

由表 3 所示集值决策信息系统在关系 $R_A^{\cap}$ 和 $R_A^{\cup}$ 下的广义决策辨识矩阵，可以得到基于 $R_A^{\cap}$ 的广义决策辨识矩阵中的非空元素是基于 $R_A^{\cup}$ 的广义决策辨识矩阵中的相应元素的子集。但是由关系 $R_A^{\cap}$ 得到的最优广义决策规则与由关系 $R_A^{\cup}$ 得到的最优广义决策规则并无关联。

结论 通常我们所处理的信息表中属性的值是单一的并且是确定的，但这只是一种理想状态。在实际情况中，我们所得到的数据往往既不单一也不确定，可能缺失部分数据，可能是数据集的一个子集。本文用集值信息系统来处理数据不完全的信息系统。集值信息系统是处理不完备信息系统的一种方法。本文在集值信息系统上定义了两种关系：相容关系和优势关系。同时，给出了集值信息系统在这两种关系下的上下近似的定义及其性质。在关系 $R_A^{\cup}$ 下关于属性集 B 的下近似和上近似优于在关系 $R_A^{\cap}$ 下关于属性集 B 的下近似和上近似。讨论了集值决策信息系统基于两种不同关系的广义决策约简，给出了广义决策约简的判定定理及辨识矩阵，进而从集值决策信息系统中提取了最优广义决策规则。

参 考 文 献

1 Pawlak Z. Rough Sets: Theoretical Aspects of Reasoning About Data. Kluwer Academic Publishers, Boston, 1991
2 Stefanowski J, Tsoukias A. On the Extension of Rough sets under Incomplete Information. In: Zhong N, et al. eds. RSFDGrC 1999, LNAI 1711, 1999. 73～82
3 Kryszkiewicz M. Rule in incomplete information systems. Information Science, 1999, 113: 271～292
4 Grzymala-Busse J W. Incomplete Data and Generalization of Indiscernibility Relation, Definability, and Approximations. In: Slezak, et al. eds. RSFDGrC 2005, LNAI 3641, 2005. 244～253
5 Slowinski R, Vanderpooten D. A Generalized Definition of Rough Approximations Based on Similarity. IEEE Transactions on Knowledge and Data Engineering, 2000, 12(2): 331～336
6 张文修, 梁怡, 吴伟志. 信息系统与知识发现. 北京: 科学出版社, 2003
7 王国胤. Rough 集理论与知识获取. 西安: 西安交通大学出版社, 2001

（上接第 187 页）

小结 本文探讨了粗糙集理论中最小决策规则的求取方法，提出决策依赖度的定义，尝试从最短的条件属性组合中提取尽可能多的决策规则。只有现有长度的决策规则无法完全覆盖所有样本时，才会考虑增加决策规则的长度。由于条件属性的所有组合是条件属性数的幂集，本文提出了 3 种减少计算复杂性的方案：1）引入跳跃系数 λ，先计算较长的条件属性组合的决策依赖度 β。如果 β 为 0，则表明该属性组合的所有子集均不必再计算；如果 β 不为 0，再回溯到其较短的子集进行计算。如此跳跃式计算，在属性较多的情况下可以大大减少计算复杂性。而回溯功能则可以避免决策树学习方法经常会遇到的收敛到局部最优答案的情况。2）在计算中只对具有相同决策值的样本进行等价类划分，这样处理的好处是避免了对含有不同决策值的等价类的无用划分。3）设计 Remain 集合，只针对 Remain 中的样本进行等价类的划分。随着 Remain 中样本数的减少，计算量会大幅下降。实验表明，本文所提出的基于决策依赖度的跳跃式决策规则求取方法可以直接应用于不完备信息系统，因此具有良好的实用价值。

参 考 文 献

1 曾华军, 张银奎译. 机器学习. 北京: 机械工业出版社, 2003
2 文硕频, 乔胜勇, 陈彩云, 等. 基于决策树的不完全决策表的数据及规则提取. 计算机应用, 2003, 23(11): 17～22
3 孙长嵩, 董西国, 张健沛. 一个基于粗糙集和决策树的最简分类规则集生成算法. 哈尔滨工程大学学报, 2002, 23(5): 87～91
4 Yang Jing, Wang Hao, Hu Xue-Gang, et al. A new classification algorithm based on rough set and entropy. International Conference on Machine Learning and Cybernetics, 2003, 1: 364～367
5 Quinlan J R. C4. 5: Programs for machine learning. San Mateo, CA: Morgan Kaufmann
6 Witten I H, MacDonald B A. Using concept learning for knowledge acquisition. International Journal of Man-Machine Studies, 1988, 27: 349～370
7 Mitchell T M, Utgoff P E, Banerji R. Learning by experimentation: acquiring and modifying problem-solving heuristics. Machine Learning, 1: 163～190
8 王东锴, 梁梁. 不完全决策表中的决策规则发现. 计算机科学, 2001, 28(5. 专刊): 83～85
9 文硕频, 乔胜勇, 陈彩云, 等. 基于决策树的不完全决策表的数据及规则提取. 计算机应用, 2003, 23(11): 17～22