

基于优势关系粗糙集的自主式学习模型

邓维斌^{1,2)}, 王国胤²⁾, 胡峰²⁾

¹⁾(西南交通大学信息科学与技术学院 成都 610031)

²⁾(计算智能重庆市重点实验室(重庆邮电大学) 重庆 400065)

摘要 为了增强对不一致有序信息系统的处理能力, 变精度优势关系粗糙集通过引入变精度阈值增加了对不一致信息的适应性, 其分类性能受变精度阈值大小的影响. 然而, 变精度阈值往往依赖于领域先验知识或通过反复尝试确定, 极大地影响了算法的实用性. 针对在不确定下如何进行信息系统知识获取的自主控制难题, 在分析了变精度优势关系粗糙集所存在问题的基础上, 首先定义了优势关系信息系统中决策表的整体确定性, 最大整体确定性, 整体不确定性, 最小整体不确定性等度量准则, 进而提出了对各决策类集的最大确定性进行度量的准则和算法. 在此基础上, 提出了将各决策类集的最大确定性作为该决策类集的变精度阈值进行知识获取的自主式学习模型. 该模型不仅避免了知识获取过程中对先验知识的依赖, 也增强了对处理不一致信息系统的适应性. 通过与现有算法的仿真实验对比分析, 发现该自主式学习方法对处理具有较高不一致性的有序信息系统具有比较突出的优势.

关键词 优势关系粗糙集; 自主式学习; 不确定性度量; 知识获取; 机器学习

中图法分类号 TP181

Self-learning Model Based on Dominance-based Rough Set Approach

Deng Wei-Bin^{1), 2)}, Wang Guo-Yin²⁾, Hu Feng²⁾

¹⁾(School of Information Science and Technology, Southwest Jiaotong University, Chengdu 610031)

²⁾(Chongqing Key Laboratory of Computational Intelligence (Chongqing University of Posts and Telecommunications), Chongqing 400065)

Abstract In order to improve the processing ability for inconsistent preference ordered information systems, the dominance-based rough set approach (DRSA) has been extended to variable consistency dominance-based rough set approach (VC-DRSA). VC-DRSA can tolerate some inconsistency through set a threshold value of consistency level and its classification performance is affected by the threshold value. The threshold value, however, is usually set according to prior domain knowledge or by a trial-and-error procedure which restricts the applications of the algorithm to a large extent. To address learning uncertain knowledge automatically deriving by the monotonic information systems themselves, the inadequacies of VC-DRSA are analyzed. The integral certainty measure, maximum integral certainty measure, integral uncertainty measure and minimum integral uncertainty measure of ordered information system are defined. Furthermore, the maximum certainty of every class union is measured and a corresponding computing algorithm is proposed. Based on them, a self-learning model based on DRSA is proposed which using the max uncertainty coefficients of every class union as the consistency threshold value respectively. In this model, knowledge can be learned automatically without depending on prior domain knowledge. Besides, it strengthens the adaptability for dealing with inconsistent information systems. Through

本课题得到国家自然科学基金 (61073146, 61309014)、中国-波兰政府间科技合作项目(34-5)、重庆市自然科学基金(cstc2012jjA40032, cstc2013jcyjA40063)、重庆市/信息产业部计算机网络与通信技术重点实验室开放基金(CY-CNCL-2010-05)资助. 邓维斌, 男, 1978年生, 博士研究生, 副教授, 主要研究领域为智能信息处理和不确定性决策. E-mail: dengwb@cqupt.edu.cn. 王国胤(通讯作者), 男, 1970年生, 博士, 教授, 博士生导师, 主要研究领域为Rough集理论、神经网络、机器学习、数据挖掘等. 胡峰, 男, 1978年生, 博士, 教授, 主要研究领域为智能信息处理和数据挖掘.

comparing with the existing algorithms, the efficiency of this self-learning method is illustrated. Especially, we found that the method has advantages for dealing with high inconsistency preference-ordered information systems.

Key words dominance-based rough set approach; self-learning; uncertainty measure; knowledge acquisition; machine learning

1 引言

机器学习是当前人工智能研究中较为重要的领域, 机器学习的实质是知识表示形式的转化, 在其中起决定性作用的是数据所具有的本质特性^[1]. 由于实际信息系统中不一致性的存在, 在传统的机器学习(如概率论, 模糊集等)研究中, 人们往往都借助于部分领域先验知识. 如何在不确定性条件下摆脱学习过程中对先验知识的依赖, 进行自主地学习, 是人工智能知识获取研究中的一个难题^[2].

由于粗糙集(Rough Set, 简称RS)能有效处理不精确, 不一致及不完备信息, 能够较好地摆脱学习过程中对先验知识的依赖, 自Pawlak^[3]于1982年提出以来, 从众多机器学习理论中脱颖而出. 近年来在机器学习, 医疗诊断, 市场决策及信息安全等众多领域得到了广泛应用^[4-6], 很多学者在属性约简和不完备信息系统处理等方面开展了大量的研究^[7-9]. 针对不一致性的处理, Ziarko^[10]提出了变精度粗糙集(VPRS)^[10], Yao^[11]等提出了三枝决策粗糙集^[11-13], 而这些模型仍然有部分先验知识的引入. 为了实现数据驱动的自主式学习, 王国胤等学者对基于粗糙理论的自主式模型和方法进行了有益的探索, 构建了面向领域, 数据驱动的数据挖掘(Domain-oriented Data-driven Data Mining, 简称3DM)基础理论^[2]. 并通过对基于等价关系决策表局部一致性, 整体一致性及其关系进行了深入研究, 将其成果应用到粗糙集的知识获取^[14], 决策树自动预修剪^[15], 概念格学习^[16], 客户分类^[17]等过程中.

为了处理具有连续属性和优势关系的信息系统, Greco^[18,19]提出了优势关系粗糙集(Dominance-based Rough Set Approach, 简称DRSA), 用优势关系代替了经典粗糙集的不可分辨关系, 很多学者对优势关系下不一致, 不完备信息等系统也进行了大量研究^[20-22]. 为了增强对不一致信息的适应性, VC-DRSA^[23], VP-DRSA^[21]和

ISVP-DRSA^[24]等变精度模型先后被提出. 在这些变精度模型中, 对象是否进入下近似取决于一致性阈值 β 的大小, 因此设置一个合适的 β 值对模型的分类性能至关重要, 虽然Blaszczynski^[25], Hu^[26,27]等学者对优势关系下对象和属性的一致性进行了研究, 但由于在优势关系下对决策表和决策类集的不一致性度量变得更为复杂, 目前还没有有效的方法. 为了避免不一致信息的影响, 王珏等^[28]提出了“规则+例外”的思想, Deng^[29]提出了将优势关系决策表进行一致性转化, 并将阈值 β 设置为1, 虽然简化了对最优阈值的反复尝试, 但这种方法可能会导致部分有效信息的丢失, 所获取的规则也可能会缺乏对噪声数据的适应性.

为了实现优势关系下的自主式学习, 充分体现粗糙集理论“让数据自己说话”的突出优点, 本文在分析了现有变精度优势关系粗糙集分类性能不足的基础上, 通过度量优势关系下决策表和决策类集的不确定性, 提出了一种新的局部变精度优势关系粗糙集模型, 以各决策类集的最大确定性为该决策类集的变精度阈值控制规则获取. 实现了一种在不确定条件下, 完全由数据自主控制的机器学习方法, 建立了一种不确定条件下的自主式知识获取模型.

2 基本理论

2.1 优势关系粗糙集

为了叙述方便, 先介绍优势关系粗糙集理论的一些基本概念.

定义 1. 决策表信息系统^[2]. 一个决策表信息系统(简称决策表) $S=\langle U, R, V, f \rangle$, 其中, U 是对象的集合, 也称为论域, $R=C \cup D$ 是属性集合, 子集 C 和 D 分别称为条件属性集和决策属性集, $D \neq \emptyset$, $V=\bigcup_{r \in R} V_r$ 是属性值的集合, V_r 表示属性 $r \in R$ 的属性值范围, 即属性 r 的值域, $f: U \times R \rightarrow V$ 是一个信息函数, 它指定 U 中每一个对象 x 的属性值. 在优势关系粗糙集中, 将有序的属性称为准则.

定义 2. 优势关系^[9]. 令 \succeq_q 为论域 U 上的弱

偏序关系. $y \succeq_q z$ 是指在准则 q 上 y 至少和 z 一样好. 如果对 $\forall q \in P$ 都有 $y \succeq_q z$, 则 y, z 在准则 $P \subseteq C$ 上的优势关系表示为 $y D_P z$. 显然, 优势关系具有自反性和传递性.

定义 3. 上并集和下并集^[19]. 假设决策属性 d 将论域 U 分成有限的类集 $\mathcal{X} = \{X_i, t \in T\}$, $T = \{1, \dots, n\}$, $\forall y \in U$ 属于一个且仅属于一个 X_i , 且假定这种分类是有序的, 即对所有的 $i, j \in T$, 如果 $j > i$, 则 X_j 中的任一对象优于(或劣于) X_i 中的对象. 则上并集和下并集表示为:

$$X_i^{\geq} = \bigcup_{t \geq i} X_t \quad (i=2,3,\dots,n), \quad (1)$$

$$X_i^{\leq} = \bigcup_{t \leq i} X_t \quad (i=1,2,\dots,n-1). \quad (2)$$

定义 4. 优势集和劣势集^[19]. 给定 $P \subseteq C$, $\forall y, z \in U$, y 关于 P 的优势集和劣势集分别表示为:

$$D_P^+(y) = \{z \in U : z D_P y\}, \quad (3)$$

$$D_P^-(y) = \{z \in U : y D_P z\}. \quad (4)$$

定义 5. 上近似、下近似和边界域^[19]. 为了简化描述, 在不致混淆的情况下, 用符号 X 表示类集 X_i^{\geq} 或 X_i^{\leq} , 用符号 $E_P(y)$ 表示优势集 $D_P^+(y)$ 或劣势集 $D_P^-(y)$, $\forall y \in U$. 则 X 关于 P 的下近似 $\underline{P}(X)$ 和上近似 $\overline{P}(X)$ 分别表示为:

$$\underline{P}(X) = \{y \in U : E_P(y) \subseteq X\}, \quad (5)$$

$$\overline{P}(X) = \bigcup_{y \in X} E_P(y). \quad (6)$$

则 X 关于 P 的边界域表示为:

$$Bn_P(X) = \overline{P}(X) - \underline{P}(X). \quad (7)$$

定义 6. 不一致对象. 如果 $y \in X_i^{\geq}$, 但 $D_P^+(y) \cap X_{i-1}^{\leq} \neq \emptyset$ 或 $y \notin X_i^{\leq}$, 但 $D_P^-(y) \cap X_{i+1}^{\geq} \neq \emptyset$, 则 y 称为优势关系信息系统中关于 P 的不一致对象.

定义 7. 近似分类质量^[19]. 设 $P \subseteq C$, 决策类集 \mathcal{X} 的近似分类质量定义为:

$$\gamma_P(\mathcal{X}) = \frac{|U - (\bigcup_{t \in T} Bn_P(X))|}{|U|}, \quad (8)$$

其中 $|X|$ 表示集合 X 的基.

粗糙集求取下近似的目的是为了决策规则的提取, 以便对其他样本进行预测. 下面对决策规则

及相关性质进行简单介绍.

定义 8. 决策规则^[19,30]. 决策规则 r 的一般表现形式为

$$\text{if } \Phi \text{ then } \Psi, \quad (9)$$

其中, Φ 是规则的条件部分, Ψ 是规则的决策部分. 条件部分 Φ 是条件值对的合取. 如果决策规则 r 的决策部分有多个决策属性, 那么 Ψ 将是多个基本决策的析取. 本文仅讨论决策集属性个数为 1 的情况.

在优势关系粗糙集中, 确定性规则都是从类集 X 的下近似 $\underline{P}(X)$ 中导出. 所导出的规则的一般表现形式为:

$$\text{if } c_{i_1}(y) \geq t_{i_1} \wedge \dots \wedge c_{i_p}(y) \geq t_{i_p} \wedge c_{i_{p+1}}(y) = t_{i_{p+1}} \wedge \dots \wedge c_{i_z}(y) = t_{i_z} \text{ then } y \in X_i^{\geq}, \quad (10)$$

$$\text{if } c_{i_1}(y) \leq t_{i_1} \wedge \dots \wedge c_{i_p}(y) \leq t_{i_p} \wedge c_{i_{p+1}}(y) = t_{i_{p+1}} \wedge \dots \wedge c_{i_z}(y) = t_{i_z} \text{ then } y \in X_i^{\leq}, \quad (11)$$

其中, c_{i_1}, \dots, c_{i_p} 表示具有优势关系的条件属性, $c_{i_{p+1}}, \dots, c_{i_z}$ 表示没有优势关系的条件属性, t_{i_j} 表示属性 c_{i_j} 的一个值, $i_j \in \{i_1, \dots, i_z\} \subseteq \{1, \dots, |C|\}$.

从类集 X 导出的某条决策规则用 r_i 表示, r_i 的条件部分和决策部分分别用 $\Phi(r_i)$ 和 $\Psi(r_i)$ 表示. 此外, 分别用 $\{\Phi(r_i)\}$ 和 $\{\Psi(r_i)\}$ 分别表示满足条件部分和决策部分的对象集合.

定义 9. 决策规则的最小性^[30,31]. 如果从类集 X 导出的某条决策规则 r_i 只能覆盖 X 中的对象, 则称 r_i 为区分规则. 如果 r_i 去掉其条件部分的任一部分后不再是区分规则, 称 r_i 为最小规则. 换言之, 如果 r_i 是最小规则, 则不存在从类集 Y 中导出的规则 r_j , 使得 $\{\Phi(r_i)\} \subseteq \{\Phi(r_j)\}$ 且 $X \supseteq Y$.

定义 10. 决策规则集的完备性和非冗余性^[30-32]. 假设从类集 X 导出的决策规则集是 R , 如对 $\forall y \in X$

都至少能被 R 中的一条决策规则覆盖, 则称从 X 导出的决策规则集 R 是完备的. 如果去掉 R 中任一决策规则 r_i , 都影响了 R 的完备性, 则称决策规则集 R 是非冗余的.

根据决策规则导出的算法不同, 一个决策表可以导出多个不同的具有完备性和非冗余性的决策规则集^[30-32]. 有些学者试图提出一种能提取最少规则数量, 每条规则都是最小规则的最小决策规则集提取算法, 但这已被证明是一个 NP-hard 难题^[33-34]. 为了尽量提取规则数较少的决策规则集, 一般采用启发式搜索方法. 针对经典粗糙集 Grzymala 等提出了 LM2 算法^[33-35], 针对优势关系粗糙集, Greco 等提出了 DomLEM 算法^[30].

2.2 变精度优势关系粗糙集

在 DRSA 中, 由于部分噪声数据的存在, 会导致其他一致对象变成不一致对象, 在实际应用中定义 5 对下近似的求解过于严格, 如果系统中存在少量极端不一致对象, 就会导致包含在下近似中的对象非常少. 为了增强对噪声数据的适应性, 学者们分别提出了 VC-DRSA, VP-DRSA 和 ISVP-DRSA 的变精度模型. 其核心思想是设置一个变精度阈值 β , 扩充下近似中所包含的对象, 以增强对部分不一致信息的适应性.

在所有的变精度 DRSA 模型中, 均需对对象的一致性进行度量, 将其表示为 $\Theta_x : U \rightarrow [0, \infty)$. 对象的一致性 Θ_x 既可以是一个 gain 类型(值越大, 对象的一致性越高), 也可以是一个 cost 类型(值越小, 对象的一致性越高).

定义 11. 变精度 DRSA 的下近似^[23,25,30]. 给定 $X \subseteq U$, $\forall y \in U$, 对象一致性度量准则 Θ_x , 和变精度阈值 β , 类集 X 的下近似定义为:

$$\underline{P}^\beta(X) = \{y \in X : \Theta_x(y) \geq \beta\}, \quad (12)$$

其中, ∞ 在 gain 类型的一致性度量中表示大于等于 (\geq), 在 cost 类型的一致性度量中表示小于等于 (\leq). 通常, 变精度阈值 $\beta \in [0, 1]$, 其作用有两个, 一是控制对象是否进入下近似, 二是在规则提取中作为控制规则是否生成的标准, 在 VC-DRSA, VP-DRSA 和 ISVP-DRSA 等几种变精度模型中, 人为地对一个决策系统中所有类集的变精度值设置得相等.

对基于变精度粗糙集的决策规则提取问题, 文献[31]在对 DomLEM 算法改进的基础上, 提出了

VC-DomLEM 算法, 使用迭代方法不断从训练集中提取规则, 最终从学习数据集中提取出一个规则数相对较少, 完备且非冗余的决策规则集. 目前 VC-DomLEM 算法已被集成到波兰科学院 Slowinski 院士团队所开发的基于优势关系粗糙集的智能决策分析系统 jMAF¹ 中.

3 基于优势关系粗糙集的自主式学习模型

首先, 分析一下经典变精度优势关系粗糙集所存在的问题.

表 1 学生成绩评价数据

U	数学	文学	综合评价
S_1	Excellent	Very Good	Yes
S_2	Excellent	Medium	Yes
S_3	Very Good	Very Good	No
S_4	Very Good	Good	Yes
S_5	Very Good	Bad	Yes
S_6	Very Good	Utterly Bad	No
S_7	Good	Excellent	Yes
S_8	Medium	Excellent	Yes
S_9	Medium	Bad	Yes
S_{10}	Bad	Medium	No
S_{11}	Bad	Very Bad	No
S_{12}	very Bad	Medium	Yes
S_{13}	Very Bad	Very Bad	No
S_{14}	Very Bad	Utterly Bad	Yes
S_{15}	Utterly Bad	Bad	Yes
S_{16}	Utterly Bad	Very Bad	Yes
S_{17}	Utterly Bad	Utterly Bad	No

通过定义 11 可知, 阈值 β 的设置对变精度 DRSA 的分类性能影响较大, 但设置一个合适的阈值并非易事, 目前主要是依赖领域先验知识或不断尝试的方法来取得这个值. 在 gain 类型的一致性度量准则中通常设为 0.9, 0.95 等, 在 cost 类型的一致性度量准则中通常设为 0.05, 0.025 等^[23,36]. 为什么设置这些值, 并不能给出一个合理的解释, 且对所有有序决策信息系统都这样设置 β 值是否能达到最佳分类性能也值得商榷. 下面通过一个实例来分析变精度阈值 β 设置对系统分类性能的影响.

例 1. 表 1 是一个包含 17 个学生成绩评价的决策表, $U = \{S_1, S_2, \dots, S_{17}\}$, $C = \{\text{数学 (Math)}, \text{文学 (Lit)}\}$, $d = \{\text{综合评价 (PS)}\}$, $V = \{\text{Utterly Bad, Very Bad, Bad, Medium, Good, Very Good, Excellent, Yes}\}$.

¹ <http://idss.cs.put.poznan.pl/site/software.html>

No}, 其中决策值 Yes 表示学生通过了综合考核, No 表示学生未通过综合考核. 显然这是一个在优势关系下不一致的决策表, 如 S_3 在 Math 和 Lit 成绩上均优于 S_9 , 但决策属性上却比 S_9 劣. 根据定义 6, $S_3, S_4, S_5, S_6, S_9, S_{10}, S_{11}, S_{12}, S_{13}, S_{14}, S_{15}, S_{16}$ 等 12 个对象均是不一致对象, 决策表的近似分类质量仅为 5/17. 可知这是一个具有高度不一致性的决策表. 下面通过实验来分析阈值 β 的设置对分类性能的影响. 在实验中先用这 17 条记录作为学习数据, 用 jMAF 系统进行规则提取, 并对原学习数据集进行分类, 一致性阈值 β 从 0.05 开始, 以 0.05 的步长递增到 1.0, 一致性阈值 β 在不同取值下的分类效果如图 1 所示.

从图 1 可看出, 当 $\beta > 0.9$ 时分类效果最差, 介于 0.75 到 0.8 之间时分类性能最好. 通过分析发现, 根据 VC-DomLEM 算法^[36], 随着 β 值的增加, 具有较高支持度(覆盖度较低)的规则会替换具有较低支持度(覆盖度较高)的规则.

这种通过不断尝试对数据集设置一个分类精度较高的 β 阈值^[36,37]的方法需反复进行多次学习和测试, 具有较高的时间复杂度, 很难满足对海量数据集或适时性要求较高的智能数据分析要求.

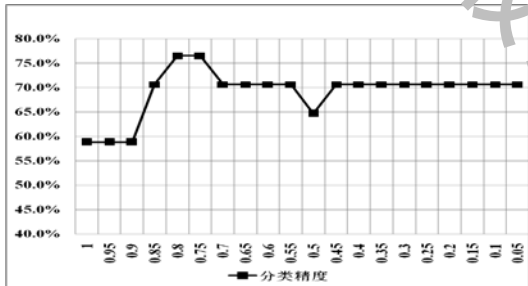


图 1 用 jMAF 对表 1 在不同变精度阈值 β 下的分类效果

如何根据不同数据集自身所隐含的特征确定一个合适的变精度阈值是一个难题, 也是自主式学习中需要解决的核心问题. 为了实现自主式学习, 受文献[2]和[14]的启发, 首先对优势关系下决策表和决策类集的一致性进行度量.

3.1 优势关系下的一致性度量

定义 12. 相对一致对象. 假设 $y \in U$ 是不一致对象, 如果将其他不一致对象从决策系统中删除后, y 变成了一致对象, 则 y 称为相对一致对象.

定理 1. 决策信息系统中的不一致对象均是相对一致对象.

证明. 假设决策信息系统 S 中包含 y_1, y_2, \dots, y_n 等共 n 个对象, 对其中的某个对象 y_i , 如果有其他 k 个对象与 y_i 相冲突, 则将这 k 个对象删除后, y_i 变

成了一致对象. 极端情况下, $y_1, y_2, \dots, y_{i-1}, y_{i+1}, \dots, y_n$ 等 $n-1$ 个对象均与 y_i 冲突, 则将这 $n-1$ 个对象均删除后, 决策信息系统 S 中仅剩 y_i 一个对象, 由于 y_i 与其自身不冲突, 则 y_i 变成了一致对象. 证毕.

定理 1 说明, 任何不一致的信息系统均可在删除不一致信息后转变成一致信息系统, 转变成一致信息系统之后的记录集合不为空. 此外, 根据删除不一致信息的方法不同, 转变成的一致信息系统也不同.

定义 13. 决策表的整体确定性和整体不确定性. 如果将不一致对象集 $O_u = \{y_1, y_2, \dots, y_k\}$ 删除后, 决策表 S 变成了一致性决策表, 则决策表 S 相对于不一致性对象集 O_u 的整体确定性定义为:

$$\mu_c = \frac{|U - O_u|}{|U|}. \quad (13)$$

与此相对应, 决策表 S 相对于不一致性对象集 O_u 的整体不确定性定义为:

$$\mu_{uc} = 1 - \mu_c = \frac{|O_u|}{|U|}. \quad (14)$$

根据定义 12 和定理 1, 将决策表 S 变成一致性决策表的方法有多种, 删除不同的不一致对象集 O_u , 将得到不同的决策表整体确定性和整体不确定性. 这与经典粗糙集对决策表的一致性和不一致性度量具有较大区别.

定义 14. 决策表的最大整体确定性和最小整体不确定性. 给定决策表 S , 在删除任意的不一致对象集 O_{u_k} 后均可变成一致性的决策表, 则决策表的最大整体确定性定为:

$$\mu_{mc} = \frac{|U - \min\{O_{u_k}\}|}{|U|}, \quad (15)$$

称 $O_{\min} = \min\{O_{u_k}\}$ 为最小不确定性对象集, 则决策表的最小整体不确定性为:

$$\mu_{muc} = 1 - \mu_{mc} = \frac{|O_{\min}|}{|U|}. \quad (16)$$

定理 2. 由决策表 S 得到的完备且非冗余的决策规则集 R , 在决策表能充分反映领域样本的情况下, 用 R 对样本数据进行测试的最大可能正确率 η 等于决策表的最大整体确定性 μ_{mc} , 即

$$\eta = \mu_{mc} = \frac{|U - O_{\min}|}{|U|}.$$

证明. 当删除了最小不一致对象集 O_{\min} 后, 决策表 S 转变成了一致决策表 S' , 由于决策规则集 R 具有完备性和非冗余性的特点, 除不一致对象集

O_{min} 中的对象外, 其余对象均可被正确分类, 则最大识别正确率为 $\frac{|U - O_{min}|}{|U|}$. 证毕.

定理 2 说明, 在决策表 S 中除不一致对象集 O_{u_k} 中的对象外, 其余对象均进入下近似有利于分类性能的改善. 如果不一致对象集 O_{u_k} 中所包含的对象越少, 最大可能分类正确率就越大, 当 $O_{u_k} = O_{min}$ 时, 最大可能分类正确率最高.

定义 15 决策类集的最大确定性和最小不确定性. 给定决策表 S 和最小不一致对象集 O_{min} , 类集 X_i 的最大确定性表示为:

$$\kappa_{X_i} = \frac{|X_i'|}{|X_i' \cup O_{X_i}'|}, \quad (17)$$

其中, $X_i' = X_i - O_{min}$, O_{X_i}' 表示 O_{min} 中所有影响 X_i' 中对象一致性的对象集合. 决策类集的最小不确定性表示为:

$$\kappa'_{X_i} = 1 - \kappa_{X_i}. \quad (18)$$

决策表 S 的整体最小不确定性反映了决策表 S 的最小冲突情况, 决策类集 X_i 的最小不确定性反映了决策类集 X_i 的最小冲突情况.

下面设计一个算法来计算决策类集的最大确定性.

算法 1. 计算决策类集 X_i 的最大确定性算法.

输入: 决策类集 X_i 和最小不一致对象集 O_{min}

输出: 决策类集 X_i 的最大确定性 κ_{X_i}

1. BEGIN
2. $X_i' := X_i - O_{min}$;
3. $O_{X_i}' := \emptyset$;
4. FOR EACH $y \in O_{min}$ DO
5. {FOR EACH $x \in X_i'$ DO
6. {IF ($y \in D_p^-(x) \cap X_{i+1}^{\geq}$ OR
7. $y \in D_p^+(x) \cap X_{i-1}^{\leq}$) THEN
8. $O_{X_i}' := O_{X_i}' \cup y$;}}
9. $\kappa_{X_i} := \frac{|X_i'|}{|X_i' \cup O_{X_i}'|}$;
10. RETURN κ_{X_i}
11. END

定义 16. 规则可信度^[14]. 对于决策表 $S=(U, R,$

$V, f), R=C \cup D$ 是属性集合, 子集 C 和 D 分别为条件属性集和决策属性集, 决策规则 r 的可信度定义为 $cr = |\{\Phi(r)\}| / |\{\Psi(r)\}|$.

定理 3. 在变精度优势关系粗糙集中, 用决策类集 X_i 的最大确定性 κ_{X_i} 作为该决策类集可信度阈值提取的决策规则集 R , 属于 R 的规则为 r_1, r_2, \dots, r_n 对应的可信度为 cr_1, cr_2, \dots, cr_n , 则规则集 R 的最小可信度 $\min\{cr_1, cr_2, \dots, cr_n\} \geq \kappa_{X_i}$.

证明: 如果将决策类集 X_i 的最大确定性 κ_{X_i} 作为该决策类集的可信度阈值, 即令 $\beta_i = \kappa_{X_i}$. 根据变精度优势关系粗糙集规则提取思想, 对拟提取的任意决策规则 r_i , 其对应的可信度均需满足 $cr_i \geq \beta_i = \kappa_{X_i}$. 则有 $\min\{cr_1, cr_2, \dots, cr_n\} \geq \kappa_{X_i}$. 证毕.

定理 3 表明用决策类集的最大确定性作为可信度阈值提取决策规则集后, 其规则集的可信度反映了该决策类集的最大确定性.

为了度量决策表的最大整体确定性, 最小整体不确定性和决策类集的最大确定性, 均需找到决策表的最小不一致对象集 O_{min} , 文献[37]给出了一种优势关系下对象全局不一致性的度量准则

$$\alpha(y) = \max \left\{ \begin{array}{l} \alpha^-(y) = \frac{|D_p^-(y) \cap X_{i+1}^{\geq}|}{|D_p^+(y) \cap X_i^{\leq}|}, \\ \alpha^+(y) = \frac{|D_p^+(y) \cap X_{i-1}^{\leq}|}{|D_p^-(y) \cap X_i^{\geq}|} \end{array} \right\}, \quad (19)$$

并根据迭代思想提出了一种将不一致决策表转变成一致决策表的算法 TIPSStoC, 在算法的每次运行中, 总是删除不一致性最大的对象, 直到决策表变成一致的. 当决策表变成一致后, 被删除的所有对象构成了不一致对象集 O_u , 具体过程如算法 2.

算法 2. TIPSStoC 算法.

输入: 决策表 $S=(U, R, V, f)$ 其中 $R=C \cup D, U$ 是决策表中所有对象的集合, 子集 C 和 D 分别为条件属性集和决策属性集

输出: 一致对象集 $O_c \subseteq U$, 不一致对象集 $O_u \subseteq U$

1. BEGIN
2. $O_u := \emptyset$;
3. $[O_1, \alpha_1] := findInconsistentObjects(U)$;
4. WHILE($O_1 \neq \emptyset$)
5. $\{y_{max} := findMaxInconsistentObject(O_1, \alpha_1)$;
6. $O_u := O_u \cup \{y_{max}\}$;
7. $[O_1, \alpha_1] := findInconsistentObjects(O_1 \setminus \{y_{max}\})$;
8. RETURN [$U \setminus O_u, O_u$]
9. END

其中, 函数 $findInconsistentObjects(U)$ 的功能是查找所有不一致性系数 $\alpha > 0$ 的对象, 函数

$findMaxInconsistentObject(O_1, \alpha_1)$ 的功能是查找不一致对象集 O_1 中不一致性系数最大的不一致对象。

在不一致对象查找过程中,对象全局不一致性度量准则起着至关重要的作用,文献[37]也定义了另外两种度量准则 $\varepsilon(y)$ 和 $\mu(y)$, 但通过大量的对比实验发现,在决策表的一致性转化过程中, $\alpha(y)$ 准则所删除的不一致对象最少,则本文用 α -TIPStoC 所删除的不一致对象构成决策表的最小不一致对象集 O_{min} . 通过算法 2 得到了决策表的最小不一致对象集 O_{min} 后,就可通过算法 1 计算每个决策类集的最大确定性,下面通过一个实例来说明算法 1 的工作过程.

例 2. 对如表 1 所示的决策表,用 α -TIPStoC 删除的不一致记录为 S_3, S_4, S_{16} 和 S_{10} , 在迭代过程中所对应的全局不一致系数 α 为 1.0, 4.0, 3.0 和 2.0, 则有:

$$O_{min}=\{S_3, S_{14}, S_{16}, S_{10}\},$$

$$X_{Yes}^{\geq}=\{S_1, S_2, S_4, S_5, S_7, S_8, S_9, S_{12}, S_{15}\},$$

$$X_{No}^{\leq}=\{S_6, S_{11}, S_{13}, S_{17}\},$$

$$\eta = \mu_{mc} = \frac{|U - O_{min}|}{|U|} = 0.7647.$$

由于 S_3 影响了 S_4, S_5 的一致性, S_{10} 影响了 S_{12}, S_{15} 的一致性, 则 $\kappa_{X_{Yes}^{\geq}} = 9/(2+9)=0.818$.

由于 S_{14} 影响了 S_6, S_{11}, S_{13} 的一致性, S_4 影响了 S_{11}, S_{13} 的一致性, 则 $\kappa_{X_{No}^{\leq}} = 4/(2+4)=0.667$.

3.2 优势关系下的自主式学习算法

在经典变精度 DRSA 中,对象是否进入下近似

表 2 用 SL-DRSA 算法对表 1 所提取的规则集

规则	规则描述	支持对象	覆盖对象	可信度
r_1	If Math>=UtterlyBad & Lit>=Bad then PS>=Yes	$S_1, S_2, S_4, S_5, S_7, S_8, S_9, S_{12}, S_{15}$	$S_1, S_2, S_3, S_4, S_5, S_7, S_8, S_9, S_{10}, S_{12}, S_{15}$	0.818
r_2	Lit<= VeryBad then PS<=No	$S_6, S_{11}, S_{13}, S_{17}$	$S_6, S_{11}, S_{13}, S_{14}, S_{16}, S_{17}$	0.667

例 3. 对如表 1 所示的决策表, X_{Yes}^{\geq} 的下近似 $\underline{P}(X_{Yes}^{\geq})=\{S_1, S_2, S_4, S_5, S_7, S_8, S_9, S_{12}, S_{15}\}$, X_{No}^{\leq} 的下近似 $\underline{P}(X_{No}^{\leq})=\{S_6, S_{11}, S_{13}, S_{17}\}$, 分别用 $\kappa_{X_{Yes}^{\geq}}=0.818$ 和 $\kappa_{X_{No}^{\leq}}=0.667$ 作为可信度阈值控制 $\underline{P}(X_{Yes}^{\geq})$ 和 $\underline{P}(X_{No}^{\leq})$ 的规则提取,得到的规则如表 2 所示. 用表 2 所示两条规则对原数据进行分类,除 S_3, S_{10}, S_{14} 和 S_{16} 外,其余对象均可被正确分类,分

和规则提取时都存在对先验知识的依赖. 为了实现知识获取过程中的自主式控制,可以从以下两方面考虑: 其一是让尽可能多的对象进入下近似,可让除了最小不一致对象集 O_{min} 中的对象外,其余对象均进入下近似,其公式如式(20)所示; 其二是根据各决策类集的确性的不同,分别用各决策类集的最大确定性作为对应的可信度阈值控制规则获取. 优势关系下的自主式学习过程如算法 3 所示.

$$\underline{P}(X)=\{y \in X : y \notin O_{min}\}. \quad (20)$$

算法 3. 优势关系下的自主式学习算法 (SL-DRSA)

输入: 决策表 $S=(U, R, V, f)$, 其中 $R=C \cup D$, U 是决策表中所有对象的集合,子集 C 和 D 分别为条件属性集和决策属性集

输出: 规则集 R

Step1 根据式(19)分别计算各对象的全局不一致性 α ;

Step2 根据对象的全局不一致性,用算法 2 将决策表 S 中的对象分成一致性对象集 O_c 和不一致对象集 O_{min} , 其中 $O_c \cup O_{min}=S, O_c \cap O_{min}=\emptyset$;

Step3 根据式(20)计算各决策类集 X_i 的下近似 $\underline{P}(X_i)$;

Step4 根据算法 1 计算各决策类集 X_i 的最大确定性 κ_{X_i} ;

Step5 调用 VC-DomLEM 算法,并用各决策类集 X_i 的最大确定性 κ_{X_i} 分别作为各决策类集的可信度阈值 β_i 控制规则获取,得出规则集 R .

类正确率达到最高,为 $13/17=76.47\%$.

4 仿真实验

拟通过与文献[31]和[36]中的 ε -VC-DomLEM 算法和文献[29]和[37]的 α -TIPStoC 算法的对比实验来验证本文提出的自主式学习算法的效果,共选择了 16 个数据集进行测试,数据集的具体信息如表 3 所示. 其中 ERA, ESL, LEV 和 SWD 来源于文献[38],

denbosch 来源于文献[39], windsor 来源于文献[23], GansuMobile 来源于中国联通某分公司的客户消费数据, 其余数据集来源于 UCI (<http://www.ics.uci.edu/~mllearn/MLRepository.html>) 等公共数据集. 整个实验分以下几步进行.

第 1 步 对数据集的预处理

(1)由于 cpu, housing 和 windsor 三个数据集的决策属性是连续值, 为了与文献[37]等实验结果进行对比, 采用与其相同的离散化方法, 即将决策属

$$\lambda(\chi) = \frac{|X_1 \cap POS(X_1^{\leq})|}{|U|} + \frac{|\bigcup_{i=2}^{k-1} (X_i \cap (POS(X_i^{\geq}) \cup POS(X_i^{\leq})))|}{|U|} + \frac{|X_k \cap POS(X_k^{\geq})|}{|U|} \quad (21)$$

三个不同的度量值均与数据集的一致性呈正相关, 从表 3 可以看出, balance, cpu 和 housing 等三个数据集是完全一致的, bank-g, Australian, breast-w, car, denbosch 和 fame 六个数据集具有较高的一致性, SWD, LEV 和 ERA 等三个数据集的不一致性很高.

对于完全一致的数据集, 近似分类质量 γ , λ 度量准则和最大整体确定性 μ_{mc} 均为 1; 对一致性较高的几个数据集三个度量值比较接近, 而对于具有高度不一致性的数据集, 不同度量准则的差距较大. 对于三个度量准则, 均是值越大, 表明数据集越易于学习.

第 3 步 分类性能的交叉验证

为了增加实验结果的客观性, 测试采用 10 重交叉验证 5 次循环的方法(记为 5×10-f cv)^[31,36,37], 即独立运行 10 重交叉验证 5 次. 在每次 10 重交叉过程中, 随机将数据平均分成 10 份, 先用其中的 9 份作学习数据获取规则, 对余下的 1 份数据进行测试, 如此轮循 10 遍. 对文献[31]和[36]的 ε -VC-DomLEM 算法, 取不同的变精度参数 β 运行多次 5×10-f cv, 取其中最好的分类效果. 对文献[29]和[37]的 α -TIPStoC 算法和本文所提出的自主式学习算法, 由于不需要变精度参数的设置, 只需运行 1 次 5×10-f cv.

第 4 步 测试结果

一般分类器的性能通过分类正确率(PCC)来体现, 而对于有序分类, 还常用对象的实际类别与预测类之间的平均绝对误差(MAE)来测度, 公式为:

$$MAE = \frac{\sum_{i=1}^k |f(y_i, d) - f'(y_i, d)|}{|U_T|}, \quad (22)$$

其中, U_T 表示测试数据集, 共有 k 个测试对象, $f(y_i, d)$ 表示对象 y_i 在决策属性 d 上的实际值, $f'(y_i, d)$ 表示对象 y_i 的分类预测值.

性分成 4 段, 每段的频数相等.

(2)数据集中如果存在缺失值, 直接将具有缺失数据的记录删除.

第 2 步 数据集的一致性度量

一致性越好的数据集, 可学习性就越强. 为了度量数据集的一致性, 分别计算了各个数据集的近似分类质量 γ , λ ^[40] 度量准则(式 21)和最大整体确定性 μ_{mc} , 具体数值如表 3 所示.

MAE 越小分类性能越好. 对两个测量准则计算出 5×10-f cv 的平均值和标准偏差, 具体结果如表 4 所示.

表 3 实验数据集特征及相应的 γ , λ 和 μ_{mc} 值

序号	数据集	对象数	条件属性数	分类数	γ	λ	μ_{mc}
1	balance	625	4	3	1.000	1.000	1.000
2	cpu	209	6	4	1.000	1.000	1.000
3	housing	506	13	4	1.000	1.000	1.000
4	bank-g	1411	16	2	0.980	0.980	0.997
5	australian	690	14	2	0.977	0.977	0.990
6	breast-w	699	9	2	0.976	0.976	0.991
7	car	1296	6	4	0.972	0.986	0.986
8	denbosch	119	8	2	0.899	0.899	0.950
9	fame	1328	10	5	0.894	0.983	0.974
10	GansuMobile	4661	6	3	0.606	0.793	0.920
11	windsor	546	10	4	0.348	0.696	0.764
12	breast-cancer	286	7	2	0.238	0.238	0.822
13	ESL	488	4	9	0.182	0.850	0.795
14	SWD	1000	10	4	0.018	0.487	0.640
15	LEV	1000	4	5	0.007	0.412	0.665
16	ERA	1000	4	9	0.000	0.113	0.336

第 5 步 结果分析

(1)用不同分类方法在所有数据集上测试性能的平均排序来度量整体分类性能的好坏^[31,36], 平均排序值越小, 其分类性能越好. 可以看出, 不管采用 PCC 度量标准还是采用 MAE 度量标准, SL-DRSA 方法的都比其他两种方法好.

(2)为了检验不同方法在分类性能上是否具有显著性差异, 取 0.05 的显著性水平对三种方法的实验结果进行相关样本 Wilcoxon 非参数检验^[24,37], 发现 SL-DRSA 与 ε -VC-DomLEM, α -TIPStoC 在 PCC 和 MAE 度量准则上的显著性概率均小于 0.05, 说明本文提出的方法显著优于其他两种方法. 而 ε -VC-DomLEM 与 α -TIPStoC 方法在 PCC 和 MAE 两种度量标准下比较的显著性概率均大于 0.7, 说明这两种方法在分类性能上无显著性差异.

表4 实验结果—PCC和MAE表

序号	数据集	PCC			MAE		
		ϵ -VC-DomLEM	α -TIPStoC	SL-DRSA	ϵ -VC-DomLEM	α -TIPStoC	SL-DRSA
1	balance	85.6(2) \pm 0.441	85.6(2) \pm 0.441	85.6(2) \pm 0.441	0.170(2) \pm 0.006	0.170(2) \pm 0.006	0.170(2) \pm 0.006
2	cpu	91.3(2) \pm 1.070	91.3(2) \pm 1.070	91.3(2) \pm 1.070	0.087(2) \pm 0.011	0.087(2) \pm 0.011	0.087(2) \pm 0.011
3	housing	69.8(2) \pm 1.590	69.8(2) \pm 1.590	69.8(2) \pm 1.590	0.343(2) \pm 0.017	0.343(2) \pm 0.017	0.343(2) \pm 0.017
4	bank-g	95.3(1) \pm 0.226	95.0(3) \pm 0.359	95.2(2) \pm 0.287	0.047 (2) \pm 0.002	0.050(3) \pm 0.004	0.046 (1) \pm 0.003
5	australian	75.1(2) \pm 0.502	75.3(1) \pm 0.973	75.0(3) \pm 0.891	0.249(3) \pm 0.005	0.247(1) \pm 0.01	0.248(2) \pm 0.04
6	breast-w	95.8(3) \pm 0.249	96.1(1) \pm 0.210	95.9(2) \pm 0.230	0.042(3) \pm 0.002	0.039(1) \pm 0.002	0.041(2) \pm 0.003
7	car	97.2 (1) \pm 0.076	96.7(3) \pm 0.113	97.1 (2) \pm 0.108	0.033 (1) \pm 0.001	0.040(3) \pm 0.001	0.036(2) \pm 0.002
8	denbosch	86.6 (2) \pm 1.590	84.5(3) \pm 1.260	86.8 (1) \pm 1.190	0.134 (2) \pm 0.016	0.155(3) \pm 0.013	0.133 (1) \pm 0.005
9	fame	67.4 (1) \pm 0.402	65.8(3) \pm 0.577	66.3(2) \pm 0.384	0.343 (1) \pm 0.004	0.359(3) \pm 0.007	0.352(2) \pm 0.008
10	GansuMobil	87.7(2) \pm 0.124	88.0(2) \pm 0.142	88.1(1) \pm 0.147	0.126(3) \pm 0.001	0.123 (1) \pm 0.002	0.125(2) \pm 0.002
11	windsor	51.3(2) \pm 1.862	51.4(3) \pm 1.400	55.6 (1) \pm 0.958	0.541(2.5) \pm 0.007	0.541(2.5) \pm 0.017	0.489 (1) \pm 0.012
12	breast-cancer	74.3(2) \pm 0.713	74.3(3) \pm 0.720	75.7 (1) \pm 0.751	0.251(2) \pm 0.007	0.257(3) \pm 0.007	0.217 (1) \pm 0.004
13	ESL	63.2(3) \pm 0.824	63.4(2) \pm 1.150	69.6 (1) \pm 0.965	0.409(3) \pm 0.008	0.341 (2) \pm 0.011	0.340 (1) \pm 0.001
14	SWD	56.5(2) \pm 0.222	56.1(3) \pm 1.020	57.2 (1) \pm 0.442	0.460(2) \pm 0.003	0.465(3) \pm 0.01	0.445 (1) \pm 0.009
15	LEV	54.7(3) \pm 0.450	62.6 (1) \pm 0.813	62.3 (2) \pm 0.941	0.503(3) \pm 0.005	0.405 (2) \pm 0.007	0.403 (1) \pm 0.002
16	ERA	25.0(3) \pm 0.723	26.8(1) \pm 0.922	27.8 (1) \pm 0.685	1.380(3) \pm 0.016	1.330 (2) \pm 0.018	1.320 (1) \pm 0.016
平均排序		2.12	2.25	1.63	2.28	2.22	1.53

注: 结果表示为 $a(b) \pm c$ 形式, 其中 a 表示 PCC 或 MAE 的 5×10 -f cv 测试结果的平均值, b 表示平均值在三种方法中的排序, c 表示多次测试结果的标准差.

(3) 为了检验不同方法在所有数据集上分类性能优、劣和相当程度, 对几种方法的实验结果分别进行比较, 结果如表 5 所示. 其中 $w-t-l$ 分别表示当前行所在方法比当前列所在方法性能优、相当和劣的数据集个数, 如第一行第一列中的“6-3-7”表示 α -TIPStoC 分类性能比 ϵ -VC-DomLEM 方法优的数据集有 6 个, 性能相当的有 3 个, 性能不足的有 7 个. 从表 5 也可看出, 本文提出的自主式学习方法 SL-DRSA 分类性能也优于其他两种方法, 而其他两种方法之间无明显差异, 这与采用 Wilcoxon 非参数检验的结果是一致的.

表5 几种方法的测试结果比较

	PCC		MAE	
	ϵ -VC-DomLEM	α -TIPStoC	ϵ -VC-DomLEM	α -TIPStoC
α -TIPStoC	6-3-7		6-4-6	
SL-DRSA	9-3-4	10-3-3	11-3-2	10-3-3

(4) 结合表 3 分析, 发现在三个完全一致的数据集 (balance, cpu 和 housing) 上, 由于不存在不一致数据, 三种方法分类性能完全一样; 在一致性较高的六个数据集 (bank-g, Australian, breast-w, car, denbosch 和 fame) 上, 三种方法性能差异不明显; 而在不一致性较高的几个数据集 (windsor,

breast-cancer, ESL, SWD, LEV 和 ERA) 上, 自主式学习方法的性能表现得比较明显.

5 结论及展望

针对优势关系下不一致信息系统的智能决策问题, 提出了一种基于优势关系的自主式学习算法. 该方法用每个决策类集的最大确定性作为该类集的可信度阈值进行规则获取, 避免了传统方法对变精度阈值选取的随意性和不断尝试所带来的计算复杂性, 也避免了不一致信息一致性转化方法对不一致信息缺乏容忍度的缺陷. 通过大量的实验表明, 该方法可实现根据数据集本身特性自主确定各决策类的变精度阈值, 且表现出了较好的分类性能, 特别是在不一致性较高的数据集中体现得比较明显.

为了度量决策表和决策类集的不确定性, 需确定决策表的最小不一致对象集. 文中采用了优势关系下对象全局不一致性的 α 度量准则, 是否还有更好的度量方法, 以及如何对大数据集进行属性约简将是下一步研究的主要内容.

参 考 文 献

- [1] Shi Zhong-Zhi. Knowledge Discover. Second Edition. Beijing: Tsinghua University Press, 2011 (in Chinese)
(史忠植. 知识发现(第2版). 北京: 清华大学出版社, 2011)
- [2] Wang Guo-Yin, Wang Yan. 3DM: domain-oriented data-driven data mining. *Fundamenta Informaticae*, 2009, 90(4): 395-426
- [3] Pawlak Z. Rough set. *International Journal of Computer and Information Sciences*, 1982, 11(5): 341-356
- [4] Zhang Wen-Xin, Qiu Guo-Fang. Uncertainty Decision Based on Rough Set. Beijing: Tsinghua University Press, 2005 (in Chinese)
(张文信, 仇国芳. 基于粗糙集的不确定性决策. 北京: 清华大学出版社, 2005)
- [5] Chen Hong-Mei, Li Tian-Rui, Ren Da, Lin Jian-Hui, Hu Cheng-Xiang. A rough-set based incremental approach for updating approximations under dynamic maintenance environment. *IEEE Transactions on Knowledge and Data Engineering*, 2013, 25(2): 174-284
- [6] Wang Guo-Yin, Yao Yi-Yu, Yu Hong. A survey on rough set theory and applications. *Chinese Journal of Computers*, 2009, 32(7): 1229-1246 (in Chinese)
(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, 32(7): 1229-1246)
- [7] Kryszkiewicz M. Comparative study of alternative types of knowledge reduction in inconsistent systems. *International Journal of Intelligent Systems*, 2001, 16(1): 105-120
- [8] Qian Yu-Hua, Liang Ji-Ye, Pedrycz W, Dang Chuang-Yin. An efficient accelerator for attribute reduction from incomplete data in rough set framework. *Pattern Recognition*, 2011, 44(8): 1658-1670
- [9] Min Fan, He Hua-Ping, Qian Yu-Hua, Zhu William. Test-cost-sensitive attribute reduction. *Information Sciences*. 2011, 181(25): 4928-4942
- [10] Ziako W. Variable precision rough set model. *Journal of Computer and System Science*, 1993, 46(1): 39-59
- [11] Yao Yi-Yu, Wong S K M. A decision theoretic framework for approximating concepts. *International Journal of Man-Machine Studies*, 1992, 37(6):793-809
- [12] Yao Yi-Yu. Three-way decisions with probabilistic rough sets. *Information Sciences*, 2010, 180(3): 341-353
- [13] Yao Yi-Yu. Decision-theoretic rough set model//Proceedings of 2nd International Conference on Rough Set and Knowledge Technology. Toronto, Canada, 2007: 1-12
- [14] Wang Guo-Yin, He Xiao. A self-learning model under uncertain condition. *Journal of Software*, 2003, 14(6): 1096-1102 (in Chinese)
(王国胤, 何晓. 一种不确定条件下的自主式知识学习模型. 软件学报, 2003, 14(6): 1096-1102)
- [15] Yin De-Sheng, Wang Guo-Yin, Wu Yu. A self-learning algorithm for decision tree pre-pruning//Proceedings of the 3rd International Conference on Machine Learning and Cybernetics. Shanghai, China, 2004: 2140-2145
- [16] Wang Yan, Wang Guo-Yin, Deng Wei-Bin. Concept lattice based data-driven uncertain knowledge acquisition. *Pattern Recognition and Artificial Intelligence*, 2007, 20(5): 636-642 (in Chinese)
(王燕, 王国胤, 邓维斌. 基于概念格的数据驱动不确定性知识获取. 模式识别与人工智能, 2007, 20(5): 636-642)
- [17] Deng Wei-Bin, Wang Yan. 3DM based customer classification algorithm of B2C. *Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition)*, 2009, 21(4): 568-572 (in Chinese)
(邓维斌, 王燕. 基于3DM的B2C客户分类算法. 重庆邮电大学学报(自然科学版), 2009, 21(4): 568-572)
- [18] Greco S, Matarazzo B, Slowinski R. Rough approximation of a preference relation by dominance relations. *European Journal of Operational Research*, 1999, 117(1): 63-68
- [19] Greco S, Matarazzo B, Slowinski R. Rough sets theory for multi-criteria decision analysis. *European Journal of Operational Research*, 2001, 129(1): 1-47
- [20] Yang Xi-Bei, Yang Jing-Yu, Wu Chen, Yu Dong-Jun. Dominance-based rough set approach and knowledge reductions in incomplete ordered information system. *Information Sciences*, 2008, 178(4): 1219-1234
- [21] Inuiguchi M, Yoshioka Y, Kusunoki Y. Variable-precision dominance-based rough set approach and attribute reduction. *International Journal of Approximate Reasoning*, 2009, 50(8): 119-134
- [22] Xie Wei-Hua, Zhang Xiao-Yan, Zhong Jian-Min, Zhang Wen-Xiu. Attribute reduction in ordered information systems based on evidence theory. *Knowledge and Information Systems*, 2010, 25(1): 169-184
- [23] Greco S, Matarazzo B, Slowinski R, Stefanowski J. Variable consistency model of dominance-based rough sets approach // Proceedings of the 2nd International Conference on Rough Sets and Current Trends in Computing. Banff, Canada, 2000: 170-181
- [24] Deng Wei-Bin, Wang, Guo-Yin, Hu Feng. An improved variable precision model of dominance-based rough set approach// Proceedings of the 13th International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing. Moscow, Russia, 2004: 65-67
- [25] Blaszczynski J, Greco S, Slowinski R, Szlag M. Monotonic variable consistency rough set approaches. *International Journal of Approximate Reasoning*, 2009, 50(7): 979-999
- [26] Hu Qing-Hua, Guo Mao-Zu, Yu Da-Ren, Liu Jin-Fu. Information entropy for ordinal classification. *Science China Information Sciences*, 2010, 53(6): 1188-1200
- [27] Hu Qing-Hua, Pan Wei-Wei, Zhang Lei, Zhang David, Song Yan-Ping, Guo Mao-Zu, Yu Da-Ren. Feature Selection for Monotonic Classification. *IEEE Transactions on Fuzzy Systems*, 2012, 20(1):

69-81

- [28] Wang Jue, Yao Yi-Yu, Wang Fei-Yue. 'Rule + Exception' learning based on reduct. Chinese Journal of Computers, 2005, 28(11): 1778-1789 (in Chinese)
(王珏, 姚一豫, 王飞跃. 基于 Reduct 的“规则+例外”学习. 计算机学报, 2005, 28(11): 1778-1789)
- [29] Deng Wei-Bin, Wang Guo-Yin, Yang Shuang-Xia, Hu Feng. A new method for inconsistent multi-criteria classification// Proceedings of the 6th International Conference on Rough Sets and Knowledge Technology, Banff, Canada, 2011: 600-609
- [30] Greco S, Matarazzo B, Slowinski R, Stefanowski J. An algorithm for induction of decision rules consistent with dominance principle// Proceedings of the 2th Rough Sets and Current Trends in Computing, Banff, Canada, 2000: 30-31
- [31] Blaszczynski J, Greco S, Slowinski R, Szelag M. Sequential covering rule induction algorithm for variable consistency rough set approaches. Information Sciences, 2011, 181(6): 98-100
- [32] Pei Xiao-Bing, Wu Tao, Lu Yong-Zhong. Calculating method for a minimal set of decision rules. Transactions on Intelligent Systems, 2007, 2(6): 65-67
(裴小兵, 吴涛, 陆永忠. 最小化决策规则集的计算方法. 智能系统学报, 2007, 2(6): 65-67)
- [33] Grzymala-Busse J W. A new version of the rule induction system LERS. Fundamenta Informaticae, 1997, 31(1): 27-39
- [34] Grzymala-Busse J W. LERS – A system for learning from examples based on rough sets// Slowinski R. Intelligent Decision Support. Handbook of Applications and Advances of the Rough Sets Theory. Dordrecht: Kluwer Academic Publishers, 1992: 3-18
- [35] Grzymala-Busse J W, Stefanowski J. Three discretization methods for rule induction. International Journal of Intelligent Systems, 2001, 16(1): 29-38
- [36] Blaszczynski J, Deng Wei-Bin, Hu Feng, Slowinski R, Szelag M, Wang Guo-Yin. On different ways of handling inconsistencies in ordinal classification with monotonicity constraints//Proceedings of the 14th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems. Catania, Italy, 2012: 300-309
- [37] Deng Wei-Bin, Wang Guo-Yin, Hu Feng, Blaszczynski J, Slowinski R, Szelag M. A novel method for elimination of inconsistencies in ordinal classification with monotonicity constraints. Fundamenta Informaticae, 2013, 126(4): 377-395
- [38] Ben-David A. Monotonicity maintenance in information-theoretic machine learning algorithms. Machine Learning, 1995, 19(1): 29-43
- [39] Daniels H, Kamp B. Applications of MLP networks to bond rating and house pricing. Neural Computation and Applications, 1999, 8(3): 226-234
- [40] Blaszczynski J, Slowinski R, Szelag M. Learnability in rough set approaches// Proceedings of the 7th International Conference on Rough Sets and Current Trends in Computing. Warsaw, Poland, 2010: 402-411



Deng Wei-Bin, born in 1978, Ph.D. candidate. His main research interests include intelligent information processing and uncertainty decision.

Wang Guo-Yin, born in 1970, professor and Ph.D. supervisor. His

research interests include intelligent information processing, data mining, rough set and neural network, et al.

Hu Feng, born in 1973, Ph.D. and professor. His research interests include intelligent information processing and data mining.

Background

Data with monotonicity constraints involves a monotonic relationship between the description of an object and the class to which it is assigned. Dominance-based rough set approach (DRSA) has been proposed by Greco, Matarazzo and Slowinski to deal with information system with monotonicity constraints. DRSA generalizes the classical rough set approach (CRSA) by substituting the indiscernible relation, used in CRSA, by a dominance relation. It can be used not only for classification, but also for choice, ranking, and sorting. Nowadays, many researchers have interests on DRSA and the fundamental theory, including some basic definitions, rule extraction and classification algorithm has been constructed. It has being a new research hotspot in decision support, machine learning and artificial intelligence et al. Owing to the regular attributes have been substituted by ordinal criteria, it is more difficult to deal with inconsistent objects. Some extended models, e.g. VC-DRSA, VP-DRSA and ISVP-DRSA, have been proposed, but all of them depending on domain prior knowledge or through a trial-and-error process to set the threshold variable consistency level.

In this paper, we consider learning uncertain knowledge automatically driving only by the data sets for multi-criteria classification. The primary contributions of this paper are as follows: Firstly, the certainty and uncertainty, e.g. integral certainty, integral uncertainty, and maximum integral certainty of the information system are measured. Secondly, we define the maximum certainty measure for the class union and propose a corresponding algorithm. What's more, a self-learning model, which can learn knowledge without depending on prior knowledge, based on variable consistency DRSA is proposed. This method strengthens the adaptability for dealing with inconsistent preference-ordered information systems.

This work has been partially supported by the National Natural Science Foundation of China (No. 61073146, 61309014), Inter-governmental Science and Technology Cooperation of China and Poland (No. 34-5), Natural Science Foundation Project of CQCSTC (NO. cstc2012jjA40032, cstc2013jcyjA40063) and Chongqing Key Lab of Computer Network and Communication Technology (No. CY-CNCL-2010-05).

Our research group has been working on pattern recognition, machine learning techniques, uncertain decision, et al. We have accumulated several related work. For example, we proposed Domain-oriented Data-driven Data mining (3DM) theory published in *Fundamenta Informaticae* and some applications have been published in *Pattern Recognition* and international conferences. We developed a particle swarm optimization model and its applications have been published in *International Journal of Computational Intelligence Systems*. Generalized approximations of rough sets have been defined by us published in *Information Sciences*. And the uncertainty measures of hierarchical quotient space structure have been studied in *Mathematical Problems in Engineering*. It is worth mentioning that we have been studying on ordinal classification, uncertainty decision, et al. with Roman Slowinski's team, a famous Polish study team in rough set and dominance-based rough set approach, financial supported by Inter-governmental Science and Technology Cooperation of China and Poland since 2010.