

基于粗糙集的属性约简方法研究综述 *

杨传健^a, 葛 浩^b, 汪志圣^c

(滁州学院 a. 计算机系; b. 电子信息工程系; c. 数学系, 安徽 滁州 239012)

摘 要: 针对经典粗糙集理论的属性约简,从三个方面对属性约简方法进行综述。最后对属性约简中存在的问题进行了分析,并提出了进一步研究的方向。

关键词: 知识发现; 粗糙集; 属性约简; 可分辨矩阵; 正区域

中图分类号: TP181 文献标志码: A 文章编号: 1001-3695(2012)01-0016-05

doi:10.3969/j.issn.1001-3695.2012.01.004

Overview of attribute reduction based on rough set

YANG Chuan-jian^a, GE Hao^b, WANG Zhi-sheng^c

(a. Dept. of Computer Science, b. Dept. of Electronic & Information Engineering, c. Dept. of Mathematics, Chuzhou University, Chuzhou Anhui 239012, China)

Abstract: For attribute reduction of classical rough set theory, this paper reviewed reduction methods in three areas. Finally, analyzed existing problems about attribute reduction and proposed the further investigation object.

Key words: knowledge discovery; rough set; attribute reduction; discernibility matrix; positive region

随着大规模数据库的广泛使用和 Internet 迅猛发展,数据库中存储的数据量越来越大,且数据还在不断动态变化。需要采用有效的工具来处理这些大量的、杂乱无章的数据,并从这些数据中提取出隐含于其中的可为人们所用的规律。粗糙集理论(rough set theory, RST)是一种处理不精确、不一致与不完全数据的数学工具,它是波兰科学家 Pawlak^[1]于 1982 年提出的,能在保持分类能力不变的前提下,通过知识约简获得知识的分类规则。与其他方法相比,粗糙集方法不需要先验知识,仅利用数据本身所提供的信息发现问题的规律。粗糙集理论具有较好的可结合性,可以与其他理论相互结合,取长补短。例如,粗糙集理论可以与模糊集理论、神经网络、遗传算法、证据理论、概念格和归纳逻辑程序设计等其他智能理论结合,实现更强大和更优良的功能。

属性约简是粗糙集理论研究的核心内容之一,它是在保持知识库分类能力不变的条件下,删除其中不相关或不重要的属性。通过属性约简可以从特征信息中提取有用的信息,简化知识处理过程。本文将对经典粗糙集理论的属性约简方法进行综述,并对今后的发展和研究进行展望。关于粗糙集理论的基本概念请参阅文献[1]。

1 属性约简方法的研究

1.1 静态信息系统属性约简

信息系统(或称决策表)的属性约简一般不是唯一的,人们期望找到所有约简或最小约简。然而 Wong 等人^[2]已经证明求解决策表的所有约简和最小约简是 NP-hard 问题。为此,通常采用启发式的方法进行属性约简。常用的启发式算法有基于信息熵的属性约简算法、基于可分辨矩阵的属性约简算法

和基于正区域的属性约简算法。大多数启发式约简算法的基本思路是以核属性为起点,然后根据属性重要性的某种测度,依次选择核属性以外未被添加到约简集中最重要的属性,将其加入到约简集中,直到满足终止条件;由此得到决策表的一个约简。

1.1.1 基于信息观的属性约简方法

苗夺谦等人^[3]从互信息量角度对决策表中属性重要性进行了定义,以此作为启发式信息,减小知识约简过程中的搜索空间,并在此基础上提出了基于互信息量的属性约简算法——MIBARK 算法。该算法多数情况下能够得到决策表的最小约简,但在决定如何扩展候选属性约简集时,必须对条件属性进行多次组合,计算量很大,算法的时间复杂度为 $O(|C||U|^2) + O(|U|^3)$,且不能保证算法的完备性。王国胤等人^[4]给出基于信息熵的 CEBARKNC 和 CEBARKNCC 算法。CEBARKNC 算法以初始条件属性集为起点,逐步删除使决策属性 D 相对条件属性 a 的条件熵 $H(D|\{a\})$ 值大的属性 a 来实现约简,其算法的时间复杂度为 $O(|C||U|^2) + O(|U|^3)$;而 CEBARKNCC 算法以核属性集为起点,逐次选择使 $H(D|B \cup \{a\})$ 最小的非核条件属性 a 添加到核属性集中,直到得到约简,算法的时间复杂度为 $O(|C|^2|U|^2)$ 。杨明^[5]针对条件信息熵约简在抗噪声干扰能力弱和约简中包含冗余属性等方面的不足,提出基于条件信息熵的近似约简概念,并给出决策表中基于条件信息熵的近似约简算法。该算法可有效增强抗噪性,且可依据实际应用的需要有效地对冗余属性进行取舍,有效增强抗噪声干扰的能力。

1.1.2 基于可分辨矩阵的属性约简方法

由于基于差别矩阵的属性约简方法简洁、易于理解,得到

收稿日期: 2011-07-25; 修回日期: 2011-08-28 基金项目: 安徽省高等学校省级自然科学研究项目(KJ2011Z276);安徽省高等学校省级优秀青年人才基金项目(2011SQRL123);滁州学院科学研究项目(2010kj014B,2011kj003Z)

作者简介: 杨传健(1978-),女,安徽来安人,副教授,硕士,CCF 会员,主要研究方向为数据挖掘、粗糙集(tocjy474@126.com);葛浩(1976-),男,安徽滁州人,副教授,硕士,CCF 会员,主要研究方向为粒计算、粗糙集;汪志圣(1978-),男,安徽六安人,讲师,主要研究方向为数据挖掘。

许多学者的关注。下面从不同角度来对该方法进行综述。

1) 以属性重要性为启发式信息

Hu 等人^[6]根据 Skowron 可分辨矩阵^[7]提出一种基于属性重要性的启发式属性约简算法。该算法首先创建 Skowron 可分辨矩阵,由可分辨矩阵获得核属性,以核属性为起点,通过对正区域计算获得属性重要性,选择具有最大重要性的属性加入约简集中。该算法的时间复杂度和空间复杂度分别为 $O(|C|^2|U|^2)$ 和 $O(|C||U|^2)$ 。同样, Jelonek 等人^[8]也提出了基于属性重要性的启发式算法。该算法是从核属性开始的,以属性增益作为重要性,将具有最大属性增益的属性加入约简集中,直到属性集变成协调集为止。由于必须计算不同属性子集的近似精度,因此该算法的时间复杂度为 $O(|C|^3|U|^2)$,算法的效率不高。为此,叶东毅^[9]对 Jelonek 的算法进行改进,利用单属性的近似精度决定如何扩展候选属性,并利用改进的可分辨矩阵进行修正,得到一个新的属性约简算法,提高了计算速度。该算法的时间复杂度为 $O(|C|^2|U|^2)$ 。由于没有考虑到决策表中不一致情况,因此在处理不一致决策表时,Hu 的算法不能保证获得正确的核属性,从而不能确保获得正确的属性约简。张文修等人^[10,11]给出改进的可分辨矩阵建立方法,通过将一致对象和不一致对象区分开后再建立可分辨矩阵,很好地处理了不一致问题,然后利用逻辑运算对分辨函数简化,形成最小析取范式,获得约简。该算法的时间和空间复杂度与 Hu 的算法是相同的,因而有必要进一步改进。

2) 以属性频率为启发式信息

在可分辨矩阵中,一个属性出现越频繁,它潜在的辨识能力就越大;可分辨矩阵中元素项的长度越短,则它对区分这一项所对应的两个对象的作用就越大,该项中属性的重要性也就越大。为此,胡可云^[12]给出了属性频率函数: $f(a) = f(a) + |a|/|c|$,设计以属性频率函数为启发式信息的属性约简 (HORAFA 算法)。HORAFA 算法进行约简处理时不依赖于原有的表,有较好的独立性。但是该算法并不一定能找到决策表的最优约简或较优约简,算法效率也不特别理想。胡或等人^[13]对 HORAFA 算法提出改进,给出 HORAFA-A 算法。HORAFA 算法和 HORAFA-A 算法均没有考虑到决策表中存在的不一致的情况。为了避免决策表中存在不一致对象而造成的求解错误,蔡卫东等人^[14]改进了可分辨矩阵,给出简化可分辨矩阵,并证明了基于该简化可分辨矩阵的属性约简与基于修正的可分辨矩阵的属性约简^[11]是等价的,并以简化的可分辨矩阵中属性频率为启发式信息,设计了基于简化可分辨矩阵的属性约简算法,但该算法并不完备。葛浩等人^[15]通过添加一个属性列来区分不一致对象,将不一致决策表转换为一致决策表,在此基础上创建可分辨矩阵,同时修正可分辨矩阵,并以属性频率为启发式信息,分别给出基于修正可分辨矩阵的属性约简算法和基于差别集的属性约简算法。

3) 属性删除的方法

可分辨矩阵在处理大数据集时将占用很大的存储空间,为此 Wang 等人^[16,17]给出了基于差别集的属性约简算法。由于不需要存储可分辨矩阵,因而大大减少了算法所占用的空间。Wang 等人^[17]通过对差别集中的属性实施反向删除,每次选择一个删除属性,当出现单属性元素,则将其加入约简集,如此反复,直到差别集为空。这种属性约简方法是完备的,同时避免了二次约简,但该算法却不能保证获得最小约简。蒙祖强等人^[18]在差别集的基础上提出一种启发式属性约简算法。该算法在每次迭代过程中从差别集中删除属性频率最小的属性,这

样确保了该算法在很大程度上能找到决策表的最小属性约简。Yao 等人^[19]通过对可分辨矩阵进行吸收操作,获得简化的可分辨矩阵,在简化的可分辨矩阵基础上提出删除单个属性和属性集两种删除属性的方法进行属性约简。由于删除属性集的方法一次性可以删除多个属性,减少了迭代的次数,大大加快了属性约简的速度,但该方法并不能保证最大可能性获得最小属性约简。

4) 基于属性序的方法

多数情况下用户对属性的评价是根据属性自身的语义给出的,在一些特定情况下,用户对属性有偏好,WANG 等人^[16,20]将用户对条件属性的偏好表示为属性之间的一个排列,即属性序,并提出以一个给定的属性序进行约简。文献^[16]给出的属性约简算法 (RA-Order) 是具有多项式时间复杂度的属性约简完备算法,该算法建立在预先定义条件属性集的一个全序关系上,这样可以充分利用已有的专家经验,并且可以保证相同的全序关系下求得的约简是相同的。如果事先没有预定义一种条件的属性序,则该算法就很难实现。胡峰等人^[21]在给定的属性序下,将分治法的思想融入到决策表的正区域计算和属性约简过程中,提出一种新的属性约简算法,该算法得到的属性约简是给定的属性序下的唯一约简。官礼和等人^[22]在给定的属性序下。提出了一种基于可分辨矩阵元素集的增量式 Pawlak 约简算法,该算法根据新增对象的不同情况快速更新可分辨矩阵元素集,并依据元素的变化,在给定的属性序下更新 Pawlak 约简,该方法进一步丰富了属性序的约简理论和方法。

5) 二进制可分辨矩阵的方法

二进制可分辨矩阵相对一般的可分辨矩阵可以减少至少一半的存储空间,其属性约简方法与基于可分辨矩阵的方法本质是相同的。

Felix 等人^[23]于 1999 年提出了一种基于二进制的可分辨矩阵,该可分辨矩阵元素由 0 和 1 组成,根据该矩阵可以判别出决策表是否是一致性的。支天云等人^[24]对二进制可分辨矩阵进行分析并进行约简变换,给出基于二进制可分辨矩阵的属性约简算法。但该算法存在一些不足:获得的约简中不能指明哪个约简是最优约简;求得的约简与正区域算法求得的约简不一致;算法时间复杂度为 $O(|C|^4 + |U|^2)$,并不很理想。徐章艳等人^[25]提出了简化二进制可分辨矩阵的定义,并证明了该定义与基于正区域的属性约简的定义是等价的,在此基础上设计了一个快速属性约简算法。杨萍等人^[26]从区分度和区分率两个不同层次计算属性的重要性,区分度越强,属性的重要性越大;区分率越强,属性的重要性越大,通过选择重要性大的属性,进行属性约简,以获得最小约简,但其空间复杂度仍为 $O(|C|^2|U|^2)$ 。蒙祖强等人^[27]利用吸收操作来构造简化二进制可分辨矩阵,大大降低了二进制可分辨矩阵的空间占用,然后设计属性约简算法,其时间复杂度逼近此类算法的时间复杂度下界 $O(|C|^2|U/C|)$ 。

1.1.3 基于正区域的属性约简方法

基于正区域的属性约简方法因不需要建立可分辨矩阵,其空间和时间开销相对较小,故许多学者也对这种方法做了大量研究。

等价类划分是该类约简算法的关键步骤。赵军等人^[28,29]采用了快速排序法划分等价类 U/C ,其时间复杂度为 $O(|C||U|\log|U|)$ 。刘少辉等人^[29]还进一步研究了正区域的渐增式计算方法,由此设计了完备的属性约简算法,该算法时间复杂

度为 $O(|C|^2|U|\log|U|)$ 。徐章艳等人^[30]对等价类划分算法进行改进,采用链式基数排序算法划分等价类 U/C ,其时间复杂度为 $O(|C||U|)$;并以快速缩小搜索空间为目的,提出了一种改良的近似质量计算方法,使整个约简算法时间复杂度降低为 $\max\{O(|C||U|), O(|C|^2|U/C|)\}$,但由于条件过于宽松,该方法并不能保证求得的约简是完备的。葛浩等人^[31]分析了徐章艳算法的局限性,提出了分布计数的基数排序方法划分等价类 U/C ,其时间复杂度为 $O(|C||U|)$;并给出了改进的属性约简算法,该算法的效率优于刘少辉的方法,并且解决了徐章艳算法中存在的不足,但该算法并不适用于不一致决策表。刘勇等人^[32]从决策系统的不一致情况出发,给出了不一致度的概念及其性质,并证明了不一致记录数与正区域的等价关系;然后提出了基于 hash 的正区域计算方法,其时间复杂度降为 $O(|U|)$;同时给出 hash 快速属性约简算法,其时间复杂度为 $O(|C|^2|U/C|)$ 。蒋瑜等人^[33]利用桶排序思想设计了等价类划分算法,其时间复杂度也为 $O(|C||U|)$,并给出一种无须求解正区域便能判断正区域是否变化的方法;最后提出一种快速属性约简算法,该算法的时间复杂度也为 $O(|C|^2|U/C|)$ 。从上述提出的约简算法中可以发现,正区域约简相对于可分辨矩阵的约简方法,在时间和空间性能上均有较大的优势。

1.2 动态信息系统属性约简

现实世界是发展变化的,信息系统中的对象在不断动态变化,已得到的属性约简将可能不再有效,这就需要对属性约简进行动态修改。刘宗田^[34]提出了当数据库记录增加后,可在已有最小属性约简簇集的基础上求解新的最小属性约简簇集的方法,但该增量式方法是针对非决策表的。

Wang 等人^[16]给出了一种基于 Skowron 分辨矩阵的属性约简的增量模型,但不能保证得到一个 Pawlak 约简,也不能处理不一致决策表。Hu 等人^[35]给出了基于正区域的属性约简增量式更新算法,该算法提高了属性约简的更新效率,可以处理不一致决策表,并得到一个完备的 Pawlak 约简,但该算法不能得到最小约简。胡峰等人^[36]给出一种改进的分辨矩阵增量式属性约简算法,该算法利用布尔逻辑运算可得到动态变化决策表的所有属性约简,但对于新增对象引起的不一致情况还需改进。王元珍等人^[37]分析了对象集的增加与属性约简的关系,给出了增量式属性约简算法,该算法的时间复杂度为 $O(K|A||U|^2 + K^2|A|^3|U|)$,算法效率并不理想。杨明^[38]提出一种基于可分辨矩阵的增量式属性约简算法,该算法可通过快速更新可分辨矩阵,在动态求解核的基础上,利用原有的属性约简 R 对属性约简进行增量式更新,算法的时间复杂度为 $O(|R||U_{pos}|(|U_{pos}| + |U_{neg}|))$ 。官礼和等人^[39]提出了基于分辨矩阵元素集的属性约简集增量式更新算法,该算法依据分辨矩阵元素集中增加和减少的元素有效地更新原属性约简集,快速得到新的最小属性约简,算法的时间复杂度为 $O(|C||U_1|(|U_1| + |U_2|))$,空间复杂度为 $O(|C||U_1|(|U_1|/2 + |U_2|))$ 。由于采用可分辨元素集的形式来存储可分辨矩阵,有效地降低了算法的空间开销,也提高了约简算法的处理效率。冯少荣等人^[40]提出一种不存储差别矩阵的高效属性约简算法,用于处理对象动态增加情况下属性约简的更新问题。但该方法只是动态更新了核属性,然后在核属性的基础上采用正区域的方法来进行属性约简,并非是完整意义上的增量式属性约简。谭旭^[41]给出了基于条件熵的分辨矩阵定义,提出了相应的基于条件熵的属性约简算法,并改进分辨矩阵,给出了一个增量式

条件属性约简算法,该约简为信息观点下的属性约简。

上述增量式约简研究均是针对对象添加情况下的增量式属性约简,而在实际环境下删除和修改对象的情况是普遍存在的,因此这些方面有深入研究的意义。

1.3 海量信息系统属性约简

随着信息技术的发展,信息系统中的数据正以惊人的速度递增,形成大规模或者超大规模数据集。在现实中,大规模数据集中包含大量对象、大量属性,并且决策类数量庞大。采用传统的方法处理大规模数据集时,将会出现算法复杂度上升、分类准确率降低,以及所需的存储空间受限等问题。如何实现海量数据集进行知识获取,将面临着较大的挑战。

1.3.1 高效算法的研究

为了提高属性约简算法处理效率,许多学者在算法优化方面做了大量的研究。基于正区域的求核和属性约简方法因不需要建立可分辨矩阵,因而其空间和时间开销相对较小,算法优化的效果比较明显。叶东毅曾给出对 Jelonek 约简算法^[8]的改进算法,其时间复杂度最坏情况下为 $O(|C|^2|U|^2)$ 。而基于正区域求解约简过程中,等价类划分是约简算法的关键步骤,因此,刘少辉采用了快速排序法对属性集排序后划分等价类,其时间复杂度为 $O(|C||U|\log|U|)$,由此设计的属性约简算法时间复杂度为 $O(|C|^2|U|\log|U|)$ 。徐章艳等人^[30]采用链式基数排序算法划分等价类,其时间复杂度为 $O(|C||U|)$,使整个约简算法时间复杂度降低为 $\max\{O(|C||U|), O(|C|^2|U/C|)\}$ 。刘勇提出了基于 hash 的正区域计算方法,并设计 hash 属性约简算法,其时间复杂度降为 $O(|C|^2|U/C|)$ 。蒋瑜利用桶排序思想设计等价类划分算法,并提出一种时间复杂度为 $O(|C|^2|U/C|)$ 的属性约简算法。在不借用已有的技术和算法情况下,时间复杂度为 $O(|C|^2|U/C|)$ 的约简算法是目前比较快速的。

1.3.2 数据集分解方法的研究

对于海量数据,仅仅靠优化约简算法还是不够的。Kusiak^[42]指出降低大型数据集分析和挖掘过程中计算复杂度的方法之一是减少每次处理的数据量,可以用分解的方法来实现这一目的;通过分解可以减少计算时间、增强数据挖掘过程的透明度、适用于并行数据挖掘算法并能增强决策的有效性。目前国内外学者在决策表分解研究方面做了许多工作,提出了一些决策表的分解思想和方法,包括利用函数分解思想进行决策表分解、基于属性核的分解、基于属性聚类分解、基于属性重要性的分解、基于决策模板的分解、基于决策一致性的分解等。

Blaz 等人^[43]将函数分解的思想应用于决策表的分解,将不精确的单一决策目标逐步分层求精为子目标和更小的决策目标,建立层次决策模型。该方法可将大型属性集分解为具有层次关系的易于分析的属性子集,提高了数据分析效率;但是该分解方法仅适用于一致决策表,容噪能力较差。樊群等人^[44]将原始决策表分解为两个子系统,一个子系统的条件属性由原决策表的核属性组成,另一个子系统的条件属性由核之外的属性以及第一个子系统的决策属性组成。但该方法只能在属性集上将原决策表分为两个子系统,不能适应更加细致的分解。赵卫东等人^[45]提出了基于属性重要性的决策表属性集分解方法,但该分解方法可能会导致属性重要性较大的条件属性所形成的子决策表对决策属性具有较差的正区域分类能力,不利于知识获取。Hu 等人^[46]结合分治法思想,将原决策表分成多个独立的子决策表,对它们分别求解,在得到子决策表

的属性约简后,给出了一种决策表合并约简算法,这样就可以将大数据的决策表分解为多个较小的子问题,提高了算法效率。吴子特等人^[47]提出一种可伸缩的快速属性约简算法,该方法通过将决策表纵向分割成不同的属性列表存储于磁盘,仅在需要时才将其调入内存。由于常驻内存只有一张决策类表,有利于处理大数据集,为海量信息系统的约简提供一条较为可行的途径。

针对不一致决策表存在的不一致问题,刘清^[48]提出将初始决策表分解为一致性决策表和不一致性决策表两个部分,分解后再对一致性决策表和不一致性决策表分别计算条件属性的约简。但该方法会造成由于子系统获取的规则可能与原系统规则出现矛盾。杨明等人^[49]针对决策表存在一致和不一致的情况,提出一种基于水平分布的多决策表全局属性核求解方法。该方法通过传送压缩的局部决策表,可有效减少网络通信量,达到提高全局属性核求解的效率,为分布式环境下的多决策表属性核求解提供了一条有效途径,同时为分布式属性约简奠定了基础。

2 有待进一步研究的内容

粗糙集理论从提出至今近三十年,取得了令人瞩目的研究成果,在决策分析、机器学习、模式识别等计算机领域得到成功的应用,并且逐渐被人们所重视。然而属性约简作为粗糙集理论的核心内容之一,仍有许多挑战性问题需要解决和继续研究。

a) 高效约简算法研究。高效的约简算法是海量数据知识发现的基础,目前尚缺乏一种非常有效的方法,研究海量数据情况下属性约简算法的数据结构和内存管理问题,寻求高效的约简算法仍然是今后主要的研究方向之一。

b) 动态数据的研究。现实中的数据是不断更新的,如数据的添加、删除和修改等。完善大型数据库的动态知识约简方法,仍然是目前研究的重点方面。

c) 分布式知识约简方法研究。在分布式数据环境下,不完全的或不确定的知识表示和多 agent 之间知识转换化问题也是粗糙集理论研究的一个热点问题。如何利用并行算法在分布式环境下进行知识约简是粗糙集理论研究的一个重要方向。

d) 不确定数据的知识约简研究。在现实世界中,数据数量庞大且数据质量较低,使得许多数据存在不确定性。常见的不确定性大致可分为随机性、模糊性、不完全性、不稳定性和不一致性等。在对这些不确定性数据进行挖掘中,研究有效方法处理各种不确定知识,仍是目前粗糙集知识约简研究的重要内容之一。

3 结束语

粗糙集理论提供了一整套的方法和技术,国内外许多学者进行了许多研究。本文着重介绍了经典粗糙集理论的属性约简方法,涉及到基于信息熵的属性约简方法、基于可分辨矩阵的属性约简方法、基于正区域的属性约简方法以及海量信息系统的属性约简方法。最后指出了粗糙集理论知识约简有待进一步解决的问题和研究的方向。

参考文献:

[1] PAWLAK Z. Rough sets[J]. International Journal of Parallel and Programming, 1982, 11(15): 341-356.
[2] WONG S K M, ZIARKO W. Optimal decision rules in decision table

[J]. Bulletin of Polish Academy of Sciences, 1985, 33(11-12): 693-696.
[3] 苗夺谦, 胡桂荣. 知识约简的一种启发式算法[J]. 计算机研究与发展, 1999, 36(6): 681-684.
[4] 王国胤, 于洪, 杨大春. 基于条件信息熵的决策表约简[J]. 计算机学报, 2002, 25(7): 759-766.
[5] 杨明. 决策表中基于条件信息熵的近似约简[J]. 电子学报, 2007, 35(11): 2156-2160.
[6] HU Xiao-hua, CERCONE N. Learning in relational databases: a rough set approach[J]. Computational Intelligence, 1995, 11(2): 323-337.
[7] SKOWRON A, RAUSZER C. The discernibility matrices and functions in information systems[C]//Proc of Intelligent Decision Support-Handbook of Applications and Advances of the Rough Sets theory. Dordrecht: Kluwer Academic Publisher, 1991: 331-362.
[8] JELONEK J, KRAWIEC K, SLOWINSKI R. Rough set reduction of attributes and their domains for neural networks[J]. International Journal of Computational Intelligence, 1995, 11(2): 339-347.
[9] 叶东毅. Jelonek 属性约简算法的一个改进[J]. 电子学报, 2000, 28(12): 81-82.
[10] 张文修, 吴伟志, 梁吉业, 等. 粗糙集理论与方法[M]. 北京: 科学出版社, 2001.
[11] 刘文军, 谷云东, 冯艳宾, 等. 基于可辨别矩阵和逻辑运算的属性约简算法[J]. 模式识别与人工智能, 2004, 17(1): 119-123.
[12] 胡可云. 基于概念格和粗糙集的数据挖掘方法研究[D]. 北京: 清华大学, 2001.
[13] 胡斌, 李智玲, 李春伟. 一种基于区分矩阵的属性约简算法[J]. 计算机应用, 2006, 26(6): 80-82.
[14] 蔡卫东, 李凡, 徐章艳, 等. 基于修正差别矩阵的高效属性约简算法[J]. 华中科技大学学报: 自然科学版, 2007, 35(9): 110-113.
[15] 葛涛, 李龙澍, 杨传健. 新的可分辨矩阵及其约简方法[J]. 控制与决策, 2010, 25(12): 1891-1895.
[16] WANG Jue, WANG Ju. Reduction algorithms based on discernibility matrix: the ordered attributes method[J]. Journal of Computer Science and Technology, 2001, 16(6): 489-504.
[17] WANG Bing, CHEN Shan-ben. A complete algorithm for attribute reduction[C]//Proc of Robotic Welding, Intelligence and Automation. Berlin: Springer-Verlag, 2004: 345-352.
[18] 蒙祖强, 史忠植. 一种新的启发式知识约简算法[J]. 小型微型计算机系统, 2009, 30(7): 1249-1255.
[19] YAO Yi-yu, ZHAO Yan. Discernibility matrix simplification for constructing attribute reducts[J]. Information Sciences, 2009, 179(7): 867-882.
[20] HAN Su-qing, WANG Jue. Reduct and attribute order[J]. Journal of Computer Science and Technology, 2004, 19(4): 429-449.
[21] 胡峰, 王国胤. 属性序下的快速约简算法[J]. 计算机学报, 2007, 30(8): 1429-1435.
[22] 官礼和, 王国胤, 于洪. 属性序下的增量式 Pawlak 约简算法[J]. 西南交通大学学报, 2011, 46(3): 461-468.
[23] FELIX R, USHIO T. Rough sets-based machine learning using a binary discernibility matrix[C]//Proc of the 2nd International Conference on Intelligent Processing and Manufacturing of Materials. 1999: 299-305.
[24] 支天云, 苗夺谦. 二进制可辨矩阵的变换及高效属性约简算法的构造[J]. 计算机科学, 2002, 29(2): 140-142.
[25] 徐章艳, 杨炳儒, 宋威. 基于简化的二进制差别矩阵的快速属性约简算法[J]. 计算机科学, 2006, 33(4): 155-158.
[26] 杨萍, 李济生, 黄永宣. 一种基于二进制区分矩阵的属性约简算法

- [J]. 信息与控制, 2009, 38(1): 70-74.
- [27] 蒙祖强, 史忠植. 一种新的基于简化二进制可辨矩阵的相对约简算法[J]. 控制与决策, 2008, 23(9): 976-980.
- [28] 赵军, 王国胤, 吴中福, 等. 一种高效的属性核计算方法[J]. 小型微型计算机系统, 2003, 24(11): 1590-1593.
- [29] 刘少辉, 盛秋骥, 吴斌, 等. Rough 集高效算法的研究[J]. 计算机学报, 2003, 26(5): 524-529.
- [30] 徐章艳, 刘作鹏, 杨炳儒, 等. 一个复杂度为 $\max(O(|C||U|), O(|C|^2|U/C|))$ 的快速属性约简算法[J]. 计算机学报, 2006, 29(3): 391-399.
- [31] 葛浩, 李龙澍, 杨传健. 改进的快速属性约简算法[J]. 小型微型计算机系统, 2009, 30(2): 308-312.
- [32] 刘勇, 熊蓉, 褚健. Hash 快速属性约简算法[J]. 计算机学报, 2009, 32(8): 1493-1499.
- [33] 蒋瑜, 刘胤田, 李超. 基于 Bucket Sort 的快速属性约简算法[J]. 控制与决策, 2011, 26(2): 207-212.
- [34] 刘宗田. 属性最小约简的增量式算法[J]. 电子学报, 1999, 27(11): 96-98.
- [35] HU Feng, WANG Guo-yin, HUANG Hai, *et al.* Incremental attribute reduction based on elementary sets[C]//Proc of the 10th International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing. Heidelberg Regina: Physica-Verlag, 2005: 185-193.
- [36] 胡峰, 代劲, 王国胤. 一种决策表增量属性约简算法[J]. 控制与决策, 2007, 22(3): 268-271.
- [37] 王元珍, 裴小兵. 增量式的高效属性约简算法[J]. 小型微型计算机系统, 2005, 26(11): 193-195.
- [38] 杨明. 一种基于改进差别矩阵的属性约简增量式更新算法[J]. 计算机学报, 2007, 30(5): 815-822.
- [39] 官礼和, 王国胤. 决策表属性约简集的增量式更新算法[J]. 计算机科学与探索, 2010, 4(5): 436-444.
- [40] 冯少荣, 张东站. 一种高效的增量式属性约简算法[J]. 控制与决策, 2011, 26(4): 495-500.
- [41] 谭旭. 改进分辩矩阵下的增量式条件属性约简算法[J]. 系统工程理论与实践, 2010, 30(9): 1684-1694.
- [42] KUSIAK A. Decomposition in data mining: an industrial case study [J]. IEEE Trans on Electronics Packaging Manufacturing, 2000, 23(4): 345-353.
- [43] BLAZ Z, MARKO B, IVAN B, *et al.* A dataset decomposition approach to data mining and machine discovery [C]// Proc of the 3rd International Conference on Knowledge Discovery and Data Mining. [S. l.]: AAAI Press, 1997: 299-303.
- [44] 樊群, 赵卫东, 达庆利. 一种基于粗集的实例分解归纳学习方法[J]. 管理工程学报, 2001, 15(2): 79-81.
- [45] 赵卫东, 李旗号. 复杂决策表的特征提取方法研究[J]. 小型微型计算机系统, 2002, 10(23): 1241-1244.
- [46] HU Feng, FAN Xing-hua, YANG S X, *et al.* A novel reduction algorithm based decomposition and merging strategy [C]//Proc of International Conference on Intelligent Computing. 2006: 790-796.
- [47] 吴子特, 叶东毅. 一种可伸缩的快速属性约简算法[J]. 模式识别与人工智能, 2009, 22(2): 234-239.
- [48] 刘清. Rough 集及 Rough 推理[M]. 北京: 科学出版社, 2001.
- [49] 杨明, 吴永芬. 一种基于水平分布的多决策表全局属性核求解算法[J]. 控制与决策, 2008, 23(2): 127-132.

(上接第15页)

- [35] SOMASUNDARA A, RAMAMOORTHY A, SRIVASTAVA M. Mobile element scheduling with dynamic deadlines[J]. IEEE Trans on Mobile Computing, 2007, 6(4): 395-410.
- [36] JEA D, SOMASUNDRA A, SRIVASTAVA M. Multiple controlled mobile elements (data mules) for data collection in sensor networks [C]//Proc of IEEE DCSS. 2005: 244-257.
- [37] GU Yao-yao, BOZDAG D, EKICI E. Mobile element based differentiated message delivery in wireless sensor networks [C]// Proc of IEEE International Symposium on World of Wireless, Mobile and Multimedia Networks. 2006.
- [38] XING Guo-liang, WANG Tian, XIE Zhi-hui, *et al.* Rendezvous planning in mobility-assisted wireless sensor networks [C]//Proc of the 28th IEEE Real-time Systems Symposium. 2007: 311-320.
- [39] JAIN S, FALL K, PATRA R. Routing in delay tolerant network [C]// Proc of ACM SIGCOMM. [S. l.]: ACM Press, 2004: 145-158.
- [40] HANDOREAN R, GILL C, ROMAN G C. Accommodating transient connectivity in Ad hoc and mobile settings [C]//Proc of Pervasive. 2004: 305-322.
- [41] MERUGU S, AMMAR M H, ZEGURA E W. Routing in space and time in networks with predictable mobility, GIT-CC-04-7 [R]. [S. l.]: Georgia Institute of Technology, 2004.
- [42] WANG Yong, JAIN S, MARTONOSI M, *et al.* Erasure-coding based routing for opportunistic networks [C]// Proc of ACM SIGCOMM Workshop on Delay Tolerant Networking. [S. l.]: ACM Press, 2005: 229-236.
- [43] ZHU Hong-zi, ZHU Yan-min, LI Ming-lu, *et al.* HERO: online real-time vehicle tracking in Shanghai [C]//Proc of the 27th IEEE Conference on Computer Communication. 2008: 942-950.
- [44] SKORDYLIS A, TRIGONI N. Delay-bounded routing in vehicular Ad hoc networks [C]//Proc of the 9th ACM International Symposium on Mobile Ad hoc Networking and Computing. New York: ACM Press, 2008: 341-350.
- [45] BALASUBRAMANIAN A, LEVINE B N, VENKATARAMANI A. Enhancing interactive Web applications in hybrid networks [C]// Proc of the 14th ACM International Conference on Mobile Computing and Networking. New York: ACM Press, 2008: 70-80.
- [46] VAHDAT A, BECKER D. Epidemic routing for partially connected Ad hoc networks, CS-2006-06 [R]. Durham: Duke University, 2000.
- [47] TCHAKOUNTIO F, RAMANATHAN R. Tracking highly mobile endpoints [C]//Proc of the 4th ACM International Workshop on Wireless Mobile Multimedia. New York: ACM Press, 2001: 83-94.
- [48] ZHANG Xiao-lan, KUROSE J, LEVINE B N, *et al.* Study of a bus-based disruption-tolerant network: mobility modeling and impact on routing [C]//Proc of the 13th Annual ACM International Conference on Mobile Computing and Networking. [S. l.]: ACM Press, 2007: 195-206.
- [49] ACER U G, KALYANARAMAN S, ABOUZEID A A. Weak state routing for large scale dynamic networks [C]//Proc of the 13th ACM International Conference on Mobile Computing and Networking. New York: ACM Press, 2007: 290-301.
- [50] BANERJEE N, CORNER M D, TOWSLEY D, *et al.* Relays, base stations, and meshes: enhancing mobile networks with infrastructure [C]//Proc of the 14th ACM International Conference on Mobile Computing and Networking. New York: ACM Press, 2008: 81-91.