

基于属性约简的恒星光谱数据分类规则挖掘系统研究^{*}

张继福^{1,2} 李银花² 郑 链¹

(北京理工大学机电工程学院 北京100081)¹

(太原重型机械学院计算机科学与技术分院 太原030024)²

摘要 知识约简与决策规则的提取是粗糙集理论研究的核心内容。本文针对新加入对象相对于原来的极小决策算法而言是全新的这一情况,提出了一种基于粗糙逻辑的增量式属性约简算法,从而避免每次从庞大的原始决策表开始约简,提高了效率。在此基础上,采用 VC++ 和 Oracle9i 为开发工具,设计与实现了基于属性约简的恒星光谱数据分类规则挖掘系统,从而为实现恒星光谱数据的自动分类提供了一种有效途径。

关键词 粗糙集,属性约简,粗糙逻辑,决策规则,光谱

Research on Classification Rules Mining System about Stellar Spectrum Data Based on Attribute Reduction

ZHANG Ji-Fu^{1,2} LI Yin-Hua² ZHENG Lian¹

(School of Electromechanical Engineering, Beijing Institute of Technology, Beijing 100081)¹

(Department of Computer Science and Technology, Taiyuan Heavy Machinery Institute, Taiyuan 030024)²

Abstract Knowledge reduction and the extraction of decision rules are very important in the rough set theory. In this paper, a dynamic algorithm of attribute reduction based on rough logic is presented. The algorithm can avoid reduction from large original decision table when new object is added, and improve the efficiency of attribute reduction. According to this algorithm, the system of classification rules about stellar spectrum data based on attribute reduction is developed through using VC++ and Oracle9i. It can afford an effective method to auto-classification of stellar spectrum data.

Keywords Rough sets, Attribute reduction, Rough logic, Decision rules, Stellar spectrum

1 引言

粗糙集是波兰学者 Z. Pawlak 教授于1982年提出的一种处理不确定、不完备、模糊数据的有效数学工具。粗糙集理论将知识定义为“区分事物的能力”,其实质是在保持分类能力不变的前提下,通过约简计算,导出问题的分类或决策规则。粗糙集理论有个最大的优点,即它无需提供问题所需处理的数据之外的任何先验信息,这一点是其他方法如:模糊集方法、概率统计方法等很难做到的。在海量数据集的信息系统中,由于属性和实例数量巨大,因此属性约简算法的效率就显得尤为重要。所以不仅要设计高效率的属性约简算法,而且也迫切需要设计高效率的增量式属性约简算法来更新、维护和管理已生成的属性约简结果,以提高属性约简的效率。目前在增量式属性约简方面,文[1]提出了信息系统属性约简的增量式算法 ASIAI 和改进算法 IASIAI。ASIAI 算法的本质是:根据新增实例构造一个动态区分矩阵,然后极小化这个动态区分矩阵。但此算法只适用于信息系统的属性约简,并不适用于决策表。文[2]分析了算法 IASIAI 的局限性,并对其做了进一步改进。与区分矩阵不同,决策逻辑是 Pawlak 提出的决策信息系统的形式化描述语言。在决策逻辑框架下,文[3]考虑了一个协调决策系统中获取极小决策算法的增量式学习算法。

目前我国正在建造一台大天区面积多目标光纤光谱望远镜(简称 LAMOST)。该望远镜计划2004年底投入运行。它是

国家“九五”计划重大工程项目,总投资达2.35亿人民币。由于 LAMOST 具有以较高效率大规模测量天体光谱的能力,可提供的研究课题将遍及天文学多个层次,从恒星、银河系、星系、星系团、活动星系核,直到宇宙大尺度结构。LAMOST 计划的主要目标是用来进行大规模光谱巡天。该望远镜装备有4000根光纤,能够同时观测4000个天体目标。预计从2004年底起,每个观测夜晚将收集2到4万条光谱的数据,预计 LAMOST 所观测到的光谱数据容量可达4TB^[5]。利用传统人工处理数据的方式将无法满足实际需求,目前急需研究新的以计算机为主的全自动分析技术。对于海量光谱数据的全自动分析技术在全世界范围内是一个全新的课题,它包括光谱的自动识别、分类和参数测定。国内外已有的光谱自动处理软件如 IRAF、Midas 都是需要人工介入的面向单光谱的半自动式处理软件。国外已有的光谱自动分析系统(如美国 SDSS、英澳的2dF)都是针对各自的望远镜观测目标设计的,它们的观测数据量级比 LAMOST 少^[5]。因此研究和开发海量光谱数据的全自动分类系统对 LAMOST 项目具有重要应用价值。

本文以粗糙逻辑为基础,首先给出了在新实例加入论域后判断约简变化与否以及判断原极小决策算法中决策规则变化与否的判定依据,并提出了一种增量式属性约简算法。在此基础上,采用 VC++ 和 Oracle9i 为开发工具,设计与实现了基于属性约简的恒星光谱数据分类规则挖掘系统,从而为实现恒星光谱数据的自动分类提供了一种有效的途径。

^{*} 本课题得到国家“八六三”高技术研究发展计划基金项目资助(2003AA133060)。张继福 教授,博士生,研究方向:数据挖掘、人工智能及其应用。李银花 讲师,硕士,研究方向:粗糙集理论与数据挖掘。郑 链 教授,博士生导师,研究方向:模式识别与智能信息处理。

2 基于粗糙逻辑的属性约简理论

2.1 决策逻辑的语法与语义

决策表可以被看作一种特殊的逻辑模型,即决策逻辑^[4]。决策逻辑语言由原子公式组成,原子公式是一种属性-属性值对,用与、或、非等命题联结词可以组成粗糙逻辑的复合公式。设 $S=(U, A)$ 是一个模型, $P=\{a_1, a_2, \dots, a_n\} \subseteq A$ 是一个属性子集,则称形如 $(a_1, v_1) \wedge (a_2, v_2) \wedge \dots \wedge (a_n, v_n)$ 的公式为 P 基本公式。

定义1 决策规则和决策算法

任何一个蕴涵公式 $\varphi \rightarrow \psi$ 称为一个决策规则, φ, ψ 分别称为 $\varphi \rightarrow \psi$ 的前件和后件。若 $\varphi \rightarrow \psi$ 在 S 中为真,则称 $\varphi \rightarrow \psi$ 在 S 中是相容的,否则称为在 S 中是不相容的。对决策规则 $\varphi \rightarrow \psi$, 如果 φ 是一个 P 基本公式, ψ 是一个 Q 基本公式,则称 $\varphi \rightarrow \psi$ 是一个 PQ 基本决策规则,在不至引起混淆的情况下,可以简称为基本决策规则。

定义2 任何一个决策规则的有限集称为一个决策算法,任何一个基本决策规则的有限集称为一个基本决策算法。如果一个基本决策算法中的所有决策规则都是 PQ 规则,则称该算法是一个 PQ 决策算法,记作 (P, Q) 。如果模型 S 的一个 PQ 算法中的每一个决策规则都是相容的,则称该算法在 S 中是相容的,否则该算法称为是不相容的。

2.2 极小决策算法的静态生成

一个决策信息系统可以自然地看成一个模型 $S=(U, P \cup Q)$, 这里 P 表示条件属性, Q 表示决策属性,且 $P \cap Q = \emptyset$ 。由一个决策表,根据其中每一个对象在条件属性和决策属性下的描述自然地可以得到一个该决策表的 PQ 基本决策算法 (P, Q) 。文[6]给出了获得一个相容的决策算法的静态方法。这种方法可以分为以下三个步骤:第一,从整体上进行属性约简,即去掉条件属性 P 中所有多余的属性,获得一个 $P'Q$ 基本算法,这里 P' 是 P 的一个子集。第二,对每一条 $P'Q$ 基本规则进行属性约简;最后,从 $P'Q$ 基本规则集中去掉多余的规则。

3 增量式属性约简的判定依据

设 $S=(U, P \cup Q)$ 是一个决策表, M 是由 S 产生的 PQ 基本决策算法 (P, Q) 的一个极小决策算法。当一个新的实例 x 加入到决策表 S 中时,新实例可能与极小决策算法 M 产生如下四种关系。

- (1) 称 x 巩固 M , 如果存在一个决策规则 $\varphi \rightarrow \psi \in M$ 使得 $x \models_S \varphi \rightarrow \psi$ 。
- (2) 称 x 对 M 是全新的, 如果对任意的 $\varphi \rightarrow \psi \in M$ 总有 $\psi_x \neq \psi$ 。
- (3) 称 x 完全矛盾于 M , 如果 x 对 M 不是全新的且存在 $y \in U$ 使得对 y 所支持的 M 中的决策规则 $\varphi \rightarrow \psi$ 有 $\varphi_x \equiv \varphi_y$ 但 $\psi_x \neq \psi_y$ 。
- (4) 称 x 部分矛盾于 M , 如果 x 不是全新的且对任意的 $y \in U, \varphi_x \neq \varphi_y$ 。

在文[3]中,给出了新实例部分矛盾于已有的极小决策算法情况下(即上述第四种情况),获取新系统极小决策算法的增量式算法,比较了该算法与非增量式算法的复杂性。本文对 x 是全新的情况考虑了获取极小决策算法的增量式算法。

定理1 设 $S=(U, P \cup Q)$ 是一个决策表, R 为 P 的一个 Q 约简, M 是约简决策算法 (R, Q) 的一个极小决策算法。现有

支持 PQ 基本决策规则 $\varphi_x \rightarrow \psi_x$ 的新实例 x 加入 U , 且对 M 而言, x 是全新的。记 $U \cup \{x\} = U'$, 则 R 仍为新系统中 P 的一个 Q 约简的充分必要条件为:不存在对象 $y \in U$, 使得 $y \models_{S'} (\varphi_x/R)$, 其中 $S' = (U', P \cup Q)$ 。

证明(充分性):若 R 在新系统中不再是 P 的一个 Q 约简,这意味着新模型 $S' = (U', P \cup Q)$ 的决策算法 (R, Q) 不再是相容的,即至少存在两个决策规则矛盾。因为原模型的决策算法 (R, Q) 是相容的,故两个矛盾的决策规则必定是 x 与某个 $y \in U$ 支持的决策规则矛盾,即决策规则 $\varphi_x/R \rightarrow \psi_x$ 与 y 支持的决策规则矛盾。这表明 $y \models_{S'} (\varphi_x/R)$, 这与条件矛盾,故 R 仍为 P 的一个 Q 约简。

必要性:若 R 在新模型 $S' = (U', P \cup Q)$ 中仍为 P 的一个 Q 约简,由条件 x 对于 M 是全新的,可知对任意的 $y \in U$, 有 $\psi_x \neq \psi_y$, 故不存在对象 $y \in U$, 使得 $y \models_{S'} (\varphi_x/R)$ 。若不然,由 y 支持的决策规则与 $\varphi_x/R \rightarrow \psi_x$ 不相容。证毕。

定理2 设 $S=(U, P \cup Q)$ 是一个决策表, R 为 P 的一个 Q 约简, M 是约简决策算法 (R, Q) 的一个极小决策算法。又设当支持 PQ 基本决策规则 $\varphi_x \rightarrow \psi_x$, 且对 M 而言是全新的实例 x 加入 U 后 R 仍为 P 的一个 Q 约简。记 $U \cup \{x\} = U'$, $S' = (U', P \cup Q)$, 则 M 中的任一决策规则 $\varphi \rightarrow \psi$ 仍为新系统的极小决策规则的充分必要条件为: $\varphi \rightarrow \psi$ 的前件 φ 与 $\varphi_x \rightarrow \psi_x$ 前件 φ_x 的 $A(\varphi)$ 扩张 $\varphi_x/A(\varphi)$ 不同,即 $\varphi \neq \varphi_x/A(\varphi)$ 。

证明(充分性):设 $\varphi \rightarrow \psi \in M$ 是一极小决策规则,且满足 $\varphi \neq \varphi_x/A(\varphi)$, 则 $\varphi \rightarrow \psi$ 在新系统 $S' = (U', P \cup Q)$ 中仍为相容规则。若不然,由于 $\varphi \rightarrow \psi$ 在 $S=(U, P \cup Q)$ 中相容,故 $\varphi \rightarrow \psi$ 只与 $\varphi_x \rightarrow \psi_x$ 不相容。这表明 $\varphi_x \rightarrow \psi_x$ (或写作 $|\varphi_x| \subseteq |\varphi|$) 且 $\psi_x \neq \psi$, 即 $\varphi = \varphi_x/A(\varphi)$ 且 $\varphi \neq \psi_x$ 。这矛盾于条件 $\varphi \neq \varphi_x/A(\varphi)$, 所以 $\varphi \rightarrow \psi$ 在新系统 $S' = (U', P \cup Q)$ 中仍为相容规则。 $\varphi \rightarrow \psi$ 在 $S=(U, P \cup Q)$ 中的极小性自然蕴涵其在 $S'=(U', P \cup Q)$ 中的极小性。充分性得证。

必要性:必要性是显然的。因为由条件 x 对于 M 是全新的可知 $\psi_x \neq \psi$, 若 $\varphi = \varphi_x/A(\varphi)$, 则有 $\varphi \rightarrow \psi$ 与 $\varphi_x \rightarrow \psi_x$ 不相容,这与“ $\varphi \rightarrow \psi$ 仍为新系统的极小决策规则”矛盾,故有 $\varphi \neq \varphi_x/A(\varphi)$ 。证毕。

由定理1及定理2,我们立即有下面的推论。

推论 设 $S=(U, P \cup Q)$ 是一个决策表, R 为 P 的一个 Q 约简, M 是约简决策算法 (R, Q) 的一个极小决策算法。又设支持 PQ 基本决策规则 $\varphi_x \rightarrow \psi_x$ 的新实例 x 对 M 而言是全新的。记 $U \cup \{x\} = U'$, $S' = (U', P \cup Q)$ 。如果不存在对象 $y \in U$, 使得 $y \models_{S'} (\varphi_x/R)$, 且对 M 中的任一决策规则 $\varphi \rightarrow \psi$ 有 $\varphi \neq \varphi_x/A(\varphi)$, 则 $M \cup \{r\}$ 乃为新系统的极小决策算法,这里决策规则 r 是 $\varphi_x/R \rightarrow \psi_x$ 在 $M \cup \{\varphi_x/R \rightarrow \psi_x\}$ 中的极小化。

定理3 设 $S=(U, P \cup Q)$ 是一个决策表, R 为 P 的一个 Q 约简, M 是约简决策算法 (R, Q) 的一个极小决策算法。又设当支持 PQ 基本决策规则 $\varphi_x \rightarrow \psi_x$, 且对 M 而言是全新的实例 x 加入 U 后 P 的 Q 约简 R 发生了变化,记 $U \cup \{x\} = U'$, $S' = (U', P \cup Q)$ 可以证明新系统的约简 R' 一定包含 R , 所以在考虑新系统的约简时只需逐步添加 $P \setminus R$ 中的属性即可。

证明:首先证明 R 中的属性均是必要的。若 R 中的某一属性 a 在加入 X 后是冗余的,即算法 $(P - \{a\}, Q)$ 是相容的,这样对原系统 U 来讲,算法 $(P - \{a\}, Q)$ 也相容,故 a 冗余,这与已知条件 R 为 P 的一个 Q 约简相矛盾,因此,如果 R 发生了变化,则 R 中的属性均为必要的,我们只需考虑添加 $P \setminus R$ 中的属性即可。

4 增量式属性约简算法

依据上节给出的判定依据,可得如下增量式属性约简算法:

输入:决策表 $S=(U, P \cup Q)$, P 的一个 Q 约简 R , 约简决策算法 (R, Q) 的一个极小决策算法为 M 。支持 PQ 基本决策规则 $\varphi \rightarrow \psi$, 且对 M 而言是全新的实例 x 。

输出:决策表 S 的约简及其极小决策算法。

步骤1:首先利用定理1判断当全新实例 x 加入 U 后 P 的 Q 约简 R 是否发生了变化。

步骤2:如果 R 发生了变化,则根据定理3先求出新的 R 然后再根据非增量式算法求极小决策算法。否则转步骤3。

步骤3:根据定理2判断原极小决策算法 M 中的任一决策规则 $\varphi \rightarrow \psi$ 是否仍为新系统的极小决策规则。如果 M 中的任一决策规则 $\varphi \rightarrow \psi$ 仍为新系统的极小决策规则,根据推论即可很快求得新系统的极小决策算法。否则转步骤4。

步骤4:将 M 中发生变化的那些决策规则 $\varphi \rightarrow \psi$ 所对应的实例 x 与新实例记为集合 A , 求出集合 A 中所有对象的约简,与 M 中那些没有发生变化的决策规则 $\varphi \rightarrow \psi$ 组合在一起便可求得新系统的极小决策算法。

5 恒星光谱数据分类规则挖掘系统

现代天文学的基本概念之一是光谱。对恒星来说,星光经过光谱仪的分光,可拍摄到恒星的光谱,就像百货公司商品上的条码一样。恒星的光谱可以告诉我们有关恒星的资讯。如恒星表面的温度等,进而我们可以确定出恒星的大概颜色。20世纪初,美国哈佛大学对50万颗恒星进行了光谱研究,并对恒星光谱中谱线(如图1所示)出现情况进行了分类,结果发现它们跟颜色有关系,即蓝色的“O”型,蓝白色的“B”型,白色的“A”型,黄白色的“F”型,黄色的“G”型,橙色的“K”型,红色的“M”型等主要类型。实际上这只是一个恒星表面温度序列。对于相同表面温度的恒星,可依其光度进一步分类,形成恒星的光度级。恒星的光度级分类有:超巨星(I)、亮巨星(II)、巨星(III)、亚巨星(IV)和主序星(V)等五类。

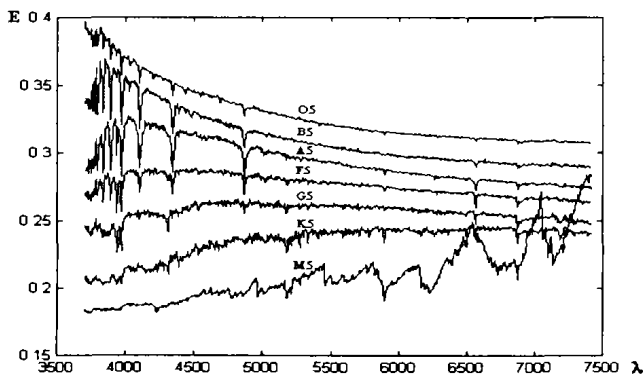


图1 按光谱型分类的标准恒星光谱

综上所述,恒星光谱从大的方面可以分为七类:O B A F G K M。为了通过某颗恒星的观测值来确定该恒星的类别以了解它,需要对恒星的观测值进行预处理,以便形成决策表。

5.1 恒星光谱数据决策表

由于粗集理论方法分析的决策是有限维的离散化数据表,因此对光谱数据进行预处理时首先要提取条件属性。在本问题中以波长作为条件属性,自然地以该波长处所对应的流量为条件属性值。只是由于流量的取值是连续的,因此存在一

个将数据离散化的问题。为了兼顾最终得出的极小决策算法的范化性与预测的准确性,在对属性值进行离散化时采用了粗集理论中的局部离散化方法中的一种——等间隔划分法。利用这种方法将整个属性值的域分为十级:这样经过预处理之后,得到了粗糙集的模式——决策表。

5.2 系统功能模块及体系结构

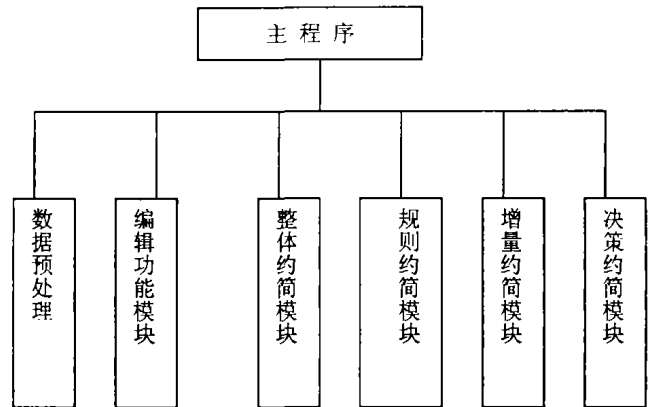


图2 功能模块

图2为恒星光谱数据分类规则挖掘系统的功能模块,图3为该系统的软件体系结构。在功能模块图1中,数据预处理模块是将原始的恒星光谱数据进行离散化,得到决策表,(具体的离散化方法在5.3节中介绍);编辑功能模块是进行插入和删除操作。整体约简模块是将所有条件属性中冗余的属性进行约简,它是规则约简、决策约简的基础;规则约简模块是在整体约简的基础上,针对每条规则进行约简;决策约简是在规则约简的基础上,去掉多余的规则,得到极小决策算法。增量约简模块是在原极小决策算法的基础上,以第3节为理论依据得到新决策表的极小决策算法;在上述模块中,整体约简模块、规则约简模块和决策约简模块正是静态属性约简算法的三个步骤。

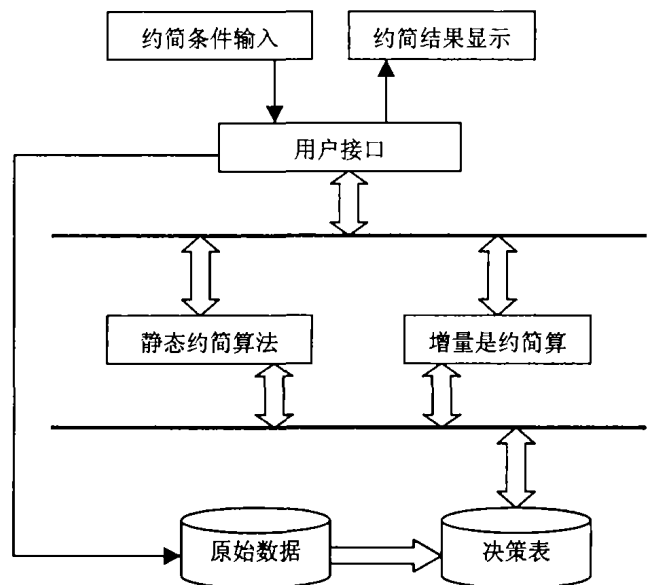


图3 软件体系结构

5.3 运行结果

以恒星光谱数据为数据集,经过以下预处理,1)选定70个波长为条件属性,这70个波长分别为3510, 3580, 3650, ... 8330,即每间隔70取一个波长,并将它们分别记为 $A_1, A_2, A_3, \dots, A_{70}$ 。恒星的所属类别作为决策属性。2)将流量按等间

(下转第130页)

基础上的查询重构、查询条件的重写规则上进一步优化现在的实现算法,在不过多损失系统总体性能的基础上,达到更高的查询效率(查全率、查准率)。

参考文献

- 1 Yu H-J, Yang Jiong, Han Jia-Wei. Research track: Classifying large date sets Using SVMs with hierarchical clusters. In: Proc. 9th ACM SIGKDD, Aug. 2003
- 2 Kleinberg J. Authoritative sources in a hyperlinked environment. In: Proc. of the 9th ACM SIAM New Orleans: ACM Press, 1997. 668~677
- 3 王晓宇,熊方,凌波,周傲英. 一种基于相似度分析的主题提取和发现算法. 软件学报, 2003, 14 (9): 1578~1585
- 4 biteboul S A, Segoufin L, Vianu V. Representing and querying XML with incomplete information. In: Proc. of PODS, 2001. 40~

- 50
- 5 Fernandez M, Suci D, Tan W C. SilkRoute: Trading between relations and XML. In: Proc. of 9th Intl. Conf. on World Wide Web, 2000
- 6 Popa L, Deutsch A, Sahuguet A, tannen V. A chase too far? In: Proc. of SIGMOD, 2000
- 7 Calvanese D, Giacomo G D, Lenzerini M, Vardi M. View-Based Query Processing for Regular Path Queries with Inverse In: proc. of PODS, 2000. 58~66
- 8 Pottinger R, Levy A R. A scalable algorithm for answering queries using views. In: Proc. of the 26th Intl. Conf. on VLDB, 2000. 484~485
- 9 Ullman J D. Information Integration Using Logical Views. In: Proc. of ICOT'97, 1997. 19~40
- 10 Raza-Yates R, Ribeiro-Neto B. Modern Information Retrieval. ACM, 1999

(上接第120页)

隔划分法分成十个区间: [0, 20], 0; [21, 40], 1; [41, 60], 2; [61, 80], 3; [81, 100], 4; [101, 120], 5; [121, 140], 6; [141, 160], 7; [161, 180], 8; [181, 201], 9. 这样便可以得到决策表。

Id	A18	A23	A27	A29	A35	A42	A43	A52	A58	A59	A66	E
39	6	-1	5	-1	-1	-1	-1	-1	-1	-1	-1	-1 A
40	7	-1	4	-1	-1	-1	-1	-1	-1	-1	-1	-1 A
43	-1	-1	-1	-1	-1	-1	-1	-1	2	1	-1	-1 A
46	7	-1	-1	-1	-1	-1	-1	-1	2	1	-1	-1 A
50	-1	6	-1	-1	-1	-1	3	-1	-1	-1	-1	-1 A
41	-1	5	-1	-1	-1	-1	-1	-1	1	1	-1	1 B
42	-1	5	-1	-1	-1	-1	-1	-1	1	1	-1	-1 B
45	7	-1	-1	5	3	3	3	2	1	1	1	-1 B
47	7	-1	-1	-1	-1	-1	-1	1	1	1	1	1 B
48	-1	-1	-1	-1	-1	2	-1	2	1	1	1	1 B
6	2	-1	-1	-1	-1	-1	-1	5	6	6	6	-1 C
10	2	-1	-1	-1	-1	-1	-1	-1	-1	5	-1	-1 C
26	4	-1	-1	-1	-1	-1	4	4	4	4	-1	-1 F

图4 分类结果1

该系统主要有两项功能:静态属性约简与增量式属性约简。其中在静态属性约简模块中又分三块:整体约简模块、规则约简模块、求极小决策算法模块。在增量式属性约简模块中也分了三块:添加模块、删除模块、求极小决策算法模块。具体形式及运行结果如图4和图5所示。

Id	A18	A23	A27	A29	A35	A42	A43	A52	A58	A59	A66	E
39	6	-1	5	-1	-1	-1	-1	-1	-1	-1	-1	-1 A
42	-1	-1	-1	-1	-1	-1	-1	-1	2	1	-1	-1 A
45	7	-1	-1	-1	-1	-1	-1	-1	2	1	-1	-1 A
49	-1	6	-1	-1	-1	-1	3	-1	-1	-1	-1	-1 A
40	-1	5	-1	-1	-1	-1	-1	-1	1	1	1	1 B
41	-1	5	-1	-1	-1	-1	-1	-1	1	1	-1	-1 B
44	7	-1	-1	5	3	3	3	2	1	1	1	-1 B
46	7	-1	-1	-1	-1	-1	-1	1	1	1	1	1 B
47	-1	-1	-1	-1	-1	-1	-1	2	1	1	1	1 B
5	2	-1	-1	-1	-1	-1	-1	-1	5	6	6	-1 C
9	2	-1	-1	-1	-1	-1	-1	-1	-1	5	-1	-1 C
25	4	-1	-1	-1	-1	-1	4	4	4	4	-1	-1 F

图5 分类结果2

上述运行结果表明:经过整体约简后70个条件属性中仅有11个必要属性 A18, A23, A27, A29, A35, A42, A43, A52, A58, A59, A66, 经过规则约简与决策约简后得到极小决策算法。上述结果中的每一行是一条分类规则,如第一行 A18₆A23₋₁A27₅A29₋₁A35₋₁A42₋₁A43₋₁A52₋₁A58₋₁A59₋₁A66₋₁EA 表示:属性 A18、A27 分别取 6 和 5 即在波长为 4700、5300 处流量值分别处于区间 [141, 160] 与 [121, 140] 时该恒星所属类别为 A。属性值为 -1 表示该属性值可以约简掉。对于其他行的结果意义相同。

结束语 本文提出的基于粗糙逻辑的增量式属性约简算法,其特点是在原有的极小决策算法的基础上,并依据第3节中给出的理论求得新系统的约简与极小决策算法,从而实现了原极小决策算法的更新与维护。以此为基础,采用 VC++ 和 Oracle9i 为开发工具,设计与实现了基于属性约简的恒星光谱数据分类规则挖掘系统,从而为恒星光谱数据的自动分类实现提供了一种有效的途径。但由于人类对天体认识还很有限,以及天体光谱的多样性和信噪影响,要想真正解决天体光谱的识别和分类问题,必须有足够的能以准确的分类的天文光谱专家知识。如何将天文光谱专家知识进行规则化并融入相应的分类算法中,以尽可能地实现计算机自动较准确地进行分类,是天体光谱数据挖掘分类中的一个重要研究问题。

参考文献

- 1 刘宗田. 属性最小约简的增量式算法. 电子学报, 1999, 27(11): 96~98
- 2 陈云化, 叶东毅. 基于粗糙集理论的一个增量算法. 计算机科学, 2002, 29(9): 53~55
- 3 Bang W-C, Bien Z. New incremental inductive learning algorithm in the framework of rough set theory. International Journal of Fuzzy Systems, 1999, 1 (1): 25~36
- 4 Pawlak Z. Rough sets: Theoretical Aspects of Reasoning about Data. Kluwer Academic Publishers, Dordrecht, 1991
- 5 覃冬梅. 天体光谱信号的自动识别方法研究: [博士论文]. 中国科学院自动化研究所, 2003
- 6 Bonikowski Z. Algebraic Structures of Rough Sets. In: Ziarko W, ed. Rough Set, Fuzzy Sets and Knowledge Discovery. Springer-Verlay, 1994. 242~247