

粗糙集在销售决策系统中的应用

邓松^{1,2}, 王映龙^{1,2}, 何火娇^{1,2}, 罗东平¹, 袁威¹

(1. 江西农业大学 软件学院, 江西 南昌 330045;

2. 江西农业大学 江西省高等学校农业信息技术重点实验室, 江西 南昌 330045)

摘要:在销售决策支持系统中,存在着大量的信息和很多不确定的因素,这使得做出科学合理的决策变得很困难。粗糙集理论是处理不确定性知识与不完整数据的有效工具,因此可以根据粗糙集理论通过分析推理找出销售数据中存在的有用的知识。依据粗糙集理论实现了一种对销售决策表知识简化的方法,采用粗糙集理论处理大量销售信息,从中提取有用规则,通过分析和推理产生最小决策规则。通过实例分析,验证了粗糙集理论与销售决策支持系统相结合方法的可行性。该方法有效地解决了智能销售决策支持系统中决策规则的获取与理解等问题。

关键词:决策支持系统;粗糙集;知识发现

中图分类号:TP311

文献标识码:A

文章编号:1673-629X(2011)09-0160-03

Application of Rough Set in DSS for Sales

DENG Song^{1,2}, WANG Ying-long^{1,2}, HE Huo-jiao^{1,2}, LUO Dong-ping¹, YUAN Wei¹

(1. Department of Software, Jiangxi Agriculture University, Nanchang 330045, China;

2. Key Laboratory of Agricultural Information Technology of Jiangxi Province, Jiangxi Agriculture University, Nanchang 330045, China)

Abstract: There is lots of sales information and uncertain factors in DSS of sales, so it is difficult to make a scientific and rational decision. Rough Set theory is an important tool dealing with uncertainty and incomplete information. Therefore, according to analysing and reasoning Rough Set theory, find useful knowledge in sales information. Present a method that forms the reduction of decision representation using Rough Set. In this paper, use Rough Set theory to deal with useful rules which extract from a large amounts of sales information, analyzing and deducing in order to get the minimal no redundant subsets of attributes and minimized rules. Through the result analysis of a case, the experiment results verify the feasibility of the Rough Set theory and DSS combining method. This combining method is an effective solution for getting and understanding of decision rules for sale IDSS.

Key words: decision support system; rough set; knowledge discovery

0 引言

决策支持系统(DSS)是在管理信息系统和运筹学的基础上发展起来的。起初DSS被广泛地用于支持单个决策者的管理活动,所以当时DSS普遍采用模型辅助决策的方法。由于人工智能技术的快速发展,研究者开始采用推理技术来模拟专家解决问题,利用神经网络等新技术模拟大脑学习推理。当人工智能技术和DSS相结合,就形成了智能决策支持系统^[1]。知识推理机制是智能决策支持系统的核心,通过已获取的

数据经过分析、推理,从而产生合理的决策规则,最终形成有用的知识^[2]。决策支持系统中,知识表达系统的条件属性以及结果属性通常可以先从数据库中抽取获得,再利用不确定数据处理方法求出最简的决策规则。粗糙集主要是用来处理不确定和不精确的信息。粗糙集采用基于知识的分类观点,在知识库中研究如何在维持分类能力不变的情况下,通过知识简约以决定如何做出适当的决策。目前该理论已经在人工智能、决策支持系统、知识与数据发现、模式识别与分类等方面取得了较为成功的应用^[3,4]。

因此,在文中把粗糙集理论与销售决策支持技术相结合获取销售领域的相应知识,也可以解决在销售领域的知识发现中的不易于解释的缺陷。

1 粗糙集基本概念

粗糙集理论是由波兰华沙理工大学 Z. Pawlak 教授

收稿日期:2011-02-26;修回日期:2011-05-30

基金项目:江西农业大学自然科学基金项目(09003335);江西省教育厅科技项目(赣教技字 GJJ08517)

作者简介:邓松(1982-),男,江西南昌人,博士研究生,讲师,主要研究领域为Web数据挖掘、决策支持;王映龙,博士,教授,主要研究领域为数据挖掘。

等在1982年提出的^[5],粗糙集可以用于研究不完整数据、不精确知识的表达、学习、归纳等。粗糙集理论不仅为信息科学和认知科学提供了新的研究方法,并为智能决策提供非常有效的处理手段。

下面介绍粗糙集的一些基本的概念定义。

定义1 信息表与信息系统。

粗糙集把客观世界或对象世界抽象成了一个信息系统,该信息系统可以通过以下四元组 S 表示:

$$S = \langle U, A, V, f \rangle \quad (1)$$

其中, U 是事例或者对象的有限集合,表示为: $U = \{u_1, u_2, \dots, u_n\}$; A 是所有属性的有限集合, $A = \{a_1, a_2, \dots, a_m\}$; V 是属性值域集合, $V = \{v_1, v_2, \dots, v_m\}$, v_k 是属性 a_i 的值域; f 为信息函数, $f: U \times A \rightarrow V, f(x_j, a_k) \in v_k$ 指定 U 中每个对象 x 的属性值。

定义2 不可分辨关系。

等价关系也称为不可分辨关系,是粗糙集理论的基本概念,它们在信息系统中的定义如下:

$$R(B) = \{(x_1, x_2) \mid f(x_1, b) = f(x_2, b), \forall b \in B\} \quad (2)$$

B 是 A 的子集,对于任意的 $x_1, x_2 \in X_i$, 有 $x_1, x_2 \in R$; 对于任意的 $x_1 \in X_i, x_2 \in X_j, i \neq j$, 都有 $(x_1, x_2) \notin R$ 。

定义3 近似空间^[6,7]。

所谓的近似空间实际上是一个二元组 $\langle U, R(B) \rangle$, 其中 B 是 A 的属性子集合。任意集合 A , 当集合 X 能够表示成基本等价类的并集时,集合 X 则可以被称为是可以精确定义的,否则就只能通过逼近的方式来刻画。粗糙集中存在两个精确集合即上近似集和下近似集,粗糙集通过近似集刻画不精确性。

设 $X \subseteq U$ 是一个集合, R 是一个定义在 U 上的等价关系,则 X 的 R 下近似集定义如下:

$$R_-(X) = \cup \{Y \in U/R \mid Y \subseteq X\} \quad (3)$$

X 的 R 上近似集定义如下:

$$R_+(X) = \cup \{Y \in U/R \mid Y \cap X \neq \emptyset\} \quad (4)$$

定义4 精确度与准确度。

对于某个等价关系 R 下面的集合 X , 下近似集与上近似集之比反映了集合 X 对关系 R 的精确度,

$$\alpha_R(X) = \frac{R_-(X)}{R_+(X)} \text{ 用于计算集合 } X \text{ 对关系 } R \text{ 的精确度。}$$

粗糙度则可以定义为:

$$\rho_R(X) = 1 - \alpha_R(X), 0 < \alpha_R(X), \rho_R(X) < 1 \quad (5)$$

定义5 属性依赖度。

属性依赖度也称为近似质量^[8]。属性集合 D 对 B 的依赖度定义如下:

$$\gamma(B, D) = \frac{|\text{pos}_B(D)|}{|U|} \quad (6)$$

$\text{pos}_B(D)$ 是根据属性集合 B 划分的正区域,如果 B 为条件属性, D 是全部的决策属性,则 $\gamma(B, D)$ 刻画了条件属性 B 描述决策属性 D 的能力,同时也表示用 B 对 U 划分后, $\forall x \in U$ 能划分到决策类的概率。

2 粗糙集理论在销售决策系统中知识发现的研究

在销售决策支持系统中,存在着大量的信息和很多不确定的因素,这使得做出科学合理的决策变得很困难。粗糙集理论可以允许决策对象中不完整与不明确属性的存在,因此可以根据粗糙集理论通过分析推理找出销售数据中存在的有用的知识。利用粗糙集理论发现知识流程图如图1所示,具体可以分为如下几个步骤:

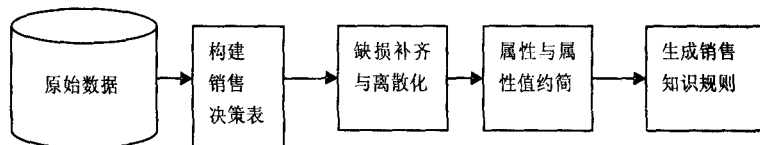


图1 基于粗糙集理论的知识发现流程图

(1) 收集数据: 根据历史记录中的原始数据, 可以从问题及期望得到的规则形式出发来确定决策表属性与决策属性集合。

(2) 数据的预处理: 因为原始数据并不一定适合直接用于知识的获取, 所以数据的预处理是建立任何知识系统需要解决的问题。对于原始的数据中的缺失信息, 需要进行填充, 也就是要进行数据补齐操作。文中采用基于 Rough 集理论不完备数据分析方法, 主要的思想是使得补齐后的完整信息表所产生的规则具有尽可能高的支持度和集中度。由于粗糙方法只能处理离散化的对象, 因此需要对数据进行离散化。文中采用等频率划分算法来对数据进行离散化。此种算法是根据用户给定参数 N 把 M 个对象分成 N 段, 每段中含有 M/N 个对象。

(3) 属性与属性值的约简^[9,10]: 在数据离散化之后, 还需要对决策表中的属性进行约简和属性值进行约简, 所谓属性的约简就是删除其中不相关和不重要的属性, 从而简化原有的信息。 $x \in U, g_x$ 表示决策规则, $g_x: \text{des}([x]_c) \Rightarrow \text{des}([x]_d), d_x(a) = a(x), a \in C \cap D$ 。

决策表中的每一个条件属性 a_i 被删除之后, 仍有 $\text{POS}_{(P \cup \{a_i\})}(Q) = \text{POS}_P(Q)$, 则说明该属性是不必要的。

(4) 生成规则集合: 对于每一条决策规则查看是否有 $[x]_{C \cup \{d\}} \subset [x]_D$, 若是, 则 r 为规则中可以省略的属性值。最后, 就是将简化后的结果用 IF-THEN 语句

或者 XML 语言表示出来^[11],形成容易理解的规则形式。

3 粗糙集理论在销售决策系统知识发现中的应用

文中用一个实例来说明粗糙集理论应用于销售决策支持系统知识发现的方法。文中数据来自江西某 IT 公司,首先从问题及期望得到的规则形式出发来确定决策表属性与决策属性集合。接着对数据库中属性值进行离散化,采用等频率划分算法来对数据进行离散化。年龄分四档(0-22岁),(23-35岁),(36-50岁),(50岁以上),分别用1,2,3,4四个数字表示。收入分为三个档次,2500以下,2500-4500,4500以上,分别用1,2,3代表。地区分省城,地市,乡镇;学历分为高中中专或以下学历,本专科学历,硕士及以上学历,分别用A,B,C和1,2,3表示。在离散化后,获取年龄,收入,地区,学历,购买电脑价格区间5个字段值完全一致,且重复记录在100条以上的销售记录。文中以销售记录为研究对象,销售记录 $U = \{C1, C2, C3, C4, C5, C6, C7, C8, C9\}$ 。由粗糙集理论的相关知识,通过数据离散化,可以获取如表1所示的知识表达系统。

表1 知识表达系统

客户	年龄 A	收入 B	地区 C	学历 D	P
C1	1	1	B	1	1
C2	1	1	A	2	2
C3	2	2	C	1	3
C4	2	2	C	3	3
C5	3	2	B	1	3
C6	4	3	A	3	4
C7	3	2	B	2	3
C8	3	3	B	2	4
C9	4	2	B	1	2

表1中的P表示电脑价格,用区间表示,文中价格区间分为以下四段:0-3000,3000-4500,4500-6500,

6500以上。接着利用以下近似区分矩阵找出所有的非必要属性。

表3 所有决策规则的核值表

客户	年龄(a)	收入(b)	学历(d)	P
C1	1	-	1	1
C2	1	-	2	2
C3	2	-	-	3
C4	-	-	-	3
C5	3	-	1	3
C6	-	3	-	4
C7	-	2	2	3
C8	-	-	-	4
C9	4	-	1	2

从表2可以看出,特征集合 $\{a, b, d\}$ 可以作为决策系统识别的特征集合,也就是模式分类的核。在表1中除去地区属性后,可以获取到所有决策规则的核值表如表3。根据表3,得到所有的决策规则如下:(1) $a1d1 \rightarrow p1$; (2) $a1d2 \vee a4d1 \rightarrow p2$; (3) $a2 \vee a3d1 \vee b2d2 \rightarrow p3$; (4) $b3 \rightarrow p4$ 。

以上决策规则用 IF-THEN 语句可以表示为:

IF 年龄在22岁以下且为高中中专或以下学历者 THEN 倾向购买3000元以下的低价电脑

IF 年龄在22岁以下且为本专科学历者,或者年龄在50岁以上且学历为高中中专及以下者 THEN 倾向购买价格为3000元到4500元价位的电脑

IF 年龄在23至35岁之间者,或者年龄在36至50之间学历为高中中专及以下者,或收入在2500元到4500元且学历为本专科者 THEN 倾向于购买价格为4500-6500元价位的电脑

IF 收入在4500以上者,倾向于购买价格在6500以上的电脑

4 结束语

粗糙集是解决和处理不确定性的有效工具,具有能够消除冗余属性、运算量较小、精度高等优点^[12]。

表2 近似区分矩阵

	C1	C2	C3	C4	C5	C6	C7	C8	C9
C1		cd	abc	abcd	ab	abcd	abd	abd	ab
C2			abcd	abcd	abcd	abd	abc	abc	abcd
C3				d	ac	abcd	acd	abcd	ac
C4					acd	abc	acd	abcd	acd
C5						abcd	d	bd	a
C6							abcd	acd	bcd
C7								b	ad
C8									abd
C9									

的情况,使用单个任务调用该函数时运行正确,在对程序进行分析后,在函数的入口处加信号量,在函数结束后释放信号量来保证函数被调用的完整性与原子性,对修改后的程序再进行测试,同样 8 个任务运行正确,示例代码如下:

```
int DMA_TRANS(const channel,
const length,
const src_addr,
const dst_addr)
{
semTake( Dma_Lock[ channel], WAIT_FOREVER); /*
DMA 通道加锁,防止多任务访问出错 */
将源地址、目的地址和 DMA 长度写入相应的寄存器;
启动 DMA;
查询 DMA 是否完成;
semGive( Dma_Lock[ channel]);
return 0;
}
```

4 结束语

在较为复杂的多任务程序设计时,进行合理的任务划分、任务优先级的设置、任务间通信机制的选择对整个系统的稳定性、可靠性有至关重要的作用,文中结合 FC 通信系统实例,探讨了多任务程序设计的要点,在实际应用中,程序经过长时间验证运行稳定、可靠,对基于 VxWorks 的嵌入式实时系统的程序设计具有借鉴作用。

(上接第 162 页)

将粗集理论应用于决策支持系统中的知识发现中,可以有效地获取简约的知识规则。文中利用粗集理论对销售数据进行了分析、推理,得到了可能的简化决策规则,有效地解决了智能销售决策支持系统中决策规则的获取与理解等问题。

参考文献:

- [1] 陈文伟. 决策支持系统及其开发[M]. 北京:清华大学出版社, 2002:15-20.
- [2] 刘振岩,王 勇,陈立平,等. 基于 SVM 的农业智能决策 Web 服务的研究与实现[J]. 计算机技术与发展,2010,20(6):213-216.
- [3] 张文修,吴伟志. 粗糙集理论介绍和研究综述[J]. 模糊系统与数学, 2000,14(4):1-12.
- [4] 王 珏,苗夺谦,周育健. 关于 Rough Sets 理论与应用的综述[J]. 模式识别与人工智能,1996(9):337-344.
- [5] Pawlak A, Rough S. Theoretical Aspects of Reasoning about Data[M]. [s. l.]:Kluwer Academic Pub, 1992.

参考文献:

- [1] 彭 华,沈为群. 一种基于 VxWorks 的飞行仿真实时管理系统[J]. 系统仿真学报,2003,15(7):966-968.
- [2] 邝 坚. Tornado / Vxworks 入门与提高[M]. 北京:科学出版社,2004.
- [3] 周启平,张 杨. VxWorks 下设备驱动程序及 BSP 开发指南[M]. 北京:中国电力出版社,2004.
- [4] 王才善,陈光武. 基于 VxWorks 的实时多任务程序设计机制分析[J]. 铁路计算机应用,2009,18(5):12-15.
- [5] 谢拴勤,杨 阳,李宇翔. 基于 VxWorks 的先进飞机电气负载管理中心软件设计[J]. 计算机测量与控制,2009,17(11):2302-2305.
- [6] Wind River Inc. VxWorks Programmer's Guide 5. 5 [EB/OL]. [2002]. <http://www.windriver.com>.
- [7] 周 谷,宋李彬. 基于 VxWorks 的机载雷达模拟吊舱软件设计[J]. 现代雷达,2009,31(9):84-87.
- [8] 李 俊. VxWorks 下多任务间通信方式的比较与分析[J]. 科技资讯,2009(25):24-25.
- [9] 黄 琪,姜 涛. 星载计算机中实时多任务软件的设计[J]. 通信对抗,2009(1):58-61.
- [10] ANSI. Fibre Channel Framing and Signaling-2 (FC-FS-2), Rev 0. 01 [M]. US:ANSI,2003.
- [11] ANSI. Fibre Channel Avionics Environment-Anonymous Subscriber Messaging (FC-AE-ASM), Rev 1. 2 [M]. US:ANSI, 2006.
- [12] Freescale Semiconductor Inc. MPC8641D Integrated Host Processor Family Reference Manual, Rev. 2 [EB/OL]. [2008]. <http://www.freescale.com/suppot>.
- [6] Chess B, McGraw G. Static analysis for security[J]. IEEE Security & Privacy, 2004(6): 32-35.
- [7] 刘 文,祝 峰,余 堃,等. 粗糙集决策表边界属性分析[J]. 计算机工程与应用,2010,46(28):33-35.
- [8] 王国胤. Rough 集理论与知识获取[M]. 西安:西安交通大学出版社,2001.
- [9] 张祥德,张 巍,刘玉蓉. 数据挖掘分类问题的贪婪粗糙集约简算法[J]. 东北大学学报(自然科学版),2001,22(5):580-583.
- [10] He Aijing, Zhu Yaoyao, Mazlack L J. Data discovery using rough set based reductive partitioning: some experiments [C]//IFSA World Congress and 20th NAFIPS International Conference. [s. l.]:[s. n.], 2001:203-208.
- [11] 李雪峰,刘 鲁. 基于 XML 决策支持系统决策模型共享机制的研究[J]. 计算机集成制造系统,2004,10(8):903-907.
- [12] 杨宝华,胡学钢. 一种基于 Rough 集的数据归约算法的实现[J]. 佳木斯大学学报(自然科学版),2003, 21(3):277-279.