

利用粗集约简挖掘生产调度决策规则

赵良辉

ZHAO Liang-hui

五邑大学 管理学院, 广东 江门 529020

School of Management, Wuyi University, Jiangmen, Guangdong 529020, China

E-mail: Xingzhe7249@163.com

ZHAO Liang-hui. Mining method for manufacture schedule with rough set reduction. Computer Engineering and Applications, 2009, 45(19): 236-238.

Abstract: Manufacture schedule is a complex decision-making procedure because of numerous factors should be considered when scheduling. To make the computer capable of scheduling, it should learn the manager's scheduling behavior, which means abstracting scheduling rules from historical scheduling cases. This paper proposes a rough set-based method of abstracting rules, and validates it with an example.

Key words: rough set; manufacture schedule; reduction of decision-table

摘要: 生产调度是一个复杂的决策过程, 其复杂性体现在决策过程中需要考虑众多因素。如果用计算机代替人进行生产调度, 计算机必须能够学习人的调度行为, 也就是从众多调度实例中提取规则。构建了一个基于粗集的调度规则提取方法, 并通过实例说明其有效性。

关键词: 粗集; 生产调度; 决策表约简

DOI: 10.3778/j.issn.1002-8331.2009.19.072 **文章编号:** 1002-8331(2009)19-0236-03 **文献标识码:** A **中图分类号:** N94

1 引言

生产调度理论的研究者经常将生产调度作为一个优化问题来研究, 其优化目标一般是一个, 或若干个目标的加权组成。但实际生产中的调度目标往往更为复杂, 通过理论研究所获得的最优调度往往并不符合生产环境的实际需要或管理者的要求。对生产调度方案评价具有影响的因素有:

(1) 时间因素

时间因素是最容易被研究的影响因素, 例如一般调度以时间越短越好, 敏捷调度以最接近交货期完工最好等等。

(2) 经济因素

经济因素中最常被研究的对象是成本, 它可以转化成时间因素被研究。但成本和时间不完全一致, 例如某高速机床可以大大缩短加工时间, 但技术工人工资、运行维护费用等如果高于普通机床很多, 其成本可能更高; 某老旧机床加工效率低下, 但管理者希望其早日折旧, 反而优先安排作业, 这体现了短期成本和长期成本的不一致。

(3) 人力或技术因素

工人的技术等级、机器的规格、精度、磨损程度都会影响调度者的决策。

(4) 管理者的心理甚至性格因素同样影响调度过程

管理者的心理和性格是调度过程中最难量化的部分, 但其对调度的影响确实存在; 例如同样一个调度方案, 在生产忙时

和闲时, 或年初和年底都可能得到不同评价; 急躁冒进的人和按部就班的人针对同一环境所做的调度方案也可能截然不同。

上述种种因素还可能互相影响互相压制, 进一步增大调度复杂性。使用计算机进行调度很难全面地考虑上述因素(尤其是管理者偏好), 因此难以得到满意的调度方案。但智能化的计算机具有学习功能, 可以从旧有调度案例中获取调度规则用于指导工作, 从而避开无法综合考虑所有影响因素的难题。

构建了一个基于粗集的决策规则提取过程。粗集理论于1982年由波兰学者 Z.Pawlak 提出, 它建立在分类机制上, 无需问题数据之外的任何先验信息, 对问题的不确定性的描述或处理较为客观; 详细内容可参考文献[1-2]。

2 决策表构造

决策规则是表征条件属性与决策属性关系的规则; 决策规则的集构成决策表; 对决策表的规则提取就是从决策表中得到简化的决策规则并将其用于指导调度系统的行为。

构造决策表首先必须确定条件属性集合。根据车间实际调度环境, 影响调度行为的最常见属性有:

(1) 完成时间 t 。定义属性值为:

1: 提前, 即作业提前于交货期完成;

0: 按期, 即作业刚好于交货期完成(处于交货期前后的一个足够小的范围内);

基金项目: 广东省自然科学基金(the Natural Science Foundation of Guangdong Province of China under Grant No.8452902001001552)。

作者简介: 赵良辉(1973-), 男, 博士, 讲师, 主要研究领域为生产调度、人工智能、电子商务。

收稿日期: 2008-09-25 **修回日期:** 2008-11-03

-1:延期,即作业落后于交货期完成。

(2)质量 q

1:质量高,超过标准;

0:合格,恰好达到标准;

-1:低于标准。

对于质量来说,高于标准往往意味着加工费用的增加,所以并非越高越好。

(3)费用 c

2:高成本;

1:经济成本;

0:低成本。

(4)机器负载均衡率 i

2:严重不均衡;

1:不均衡;

0:均衡。

机器负载均衡率可根据一定的指标,例如最大负载机器与最小机器之间负载的比值等确定。

(5)任务优先级 s

1:关键任务;

0:普通任务。

这里的关键任务指受重视的订单上的任务,与调度关键路径无关。属性集合中还可根据决策者的要求加入其他元素,如作业的复杂程度等等。

决策属性表示对条件属性组合的评价,一般只有单个属性,即决策者对条件属性组合的喜好程度,用 v 表示,取值为:

1:好;

0:合格;

-1:不好。

决策表由决策者对条件属性的不同组合做出评价(确定决策属性值)而得到。表1是一个决策表示例:

表1 决策表示例

序号	t	q	c	i	s	v
1	0	1	2	2	1	1
2	1	0	1	1	0	1
3	-1	-1	0	0	1	-1
4	0	0	1	2	0	1
5	1	1	2	2	1	0
6	-1	0	0	1	0	0
7	0	-1	0	0	0	-1
8	0	0	1	2	0	1

3 决策表约简

决策表通常包含了大量的规则,这些规则一般不能直接用于决策,须经过化简得到精炼、简单的规则再用于指导决策。决策表的化简一般分为几个步骤:

(1)属性约简:对条件属性进行相对简化,寻找表中对决策属性无影响的属性消去;

(2)消去重复的行(所有属性值与另一行完全相同的行);

(3)值约简:消去每一决策规则中的属性的冗余值。

其中条件属性约简和值约简有多种方法,简单的做法是尝试消去某一列之后查看表中是否存在互相冲突的规则,若不存在,则说明该列代表的属性可消去;同样,值约简的方法是尝试消去一行后查看表中是否存在互相冲突的规则,若存在,则被消

去列中与相冲突规则同属一行的属性须保留。这种约简方式效率较低,因此众多研究者提出了各种属性约简和值约简方法^[5-8]。本章采用另一种基于区分矩阵的方法实现属性约简和属性值约简。相关概念:

决策系统:决策系统是一个五元组:

$$S = \langle U, C, D, V, f \rangle$$

其中 U 是一个对象的有限集, C 是条件属性集合, D 是决策属性集合, V 是属性值集合, $V = \cup V_q, q=1, \dots, |C|$; f 是信息函数, $f: U \rightarrow V_q$ 。

属性约简:对决策系统 S , 条件属性集合 C 的约简 $Red(C)$ 是 C 的一个非空子集 P , 满足:

(1) $\forall a \in P, a$ 都是 P 不可省略的;

(2) $Pos_p(D) = Pos_s(D)$ 。

其中 Pos 表示属性的正域;“不可省略”表示若从 P 中去除属性 a 则 U 关于 P 的不可分辨关系将被破坏。

核: C 中所有不可省略属性的集合称为 C 的核, 记为 $Core(C)$, 且有:

$$Core(C) = \cap Red(C)$$

区分矩阵(discernibility matrix): 又称差别矩阵、分辨矩阵。设 U 的基数为 n , 属性集 $P \subseteq C$ 的区分矩阵 $M(P)$ 是一个 $n \times n$ 矩阵, 其任一元素为:

$$\delta(x, y) = \{a \in P | f(x, a) \neq f(y, a)\}, x, y \in U$$

其中 $f(x, a)$ 表示对象 x 的 a 属性的值。显然 M 是一个对称矩阵且对角线元素全为空集。定义条件属性集 C 基于决策属性 D 的区分矩阵为^[9]:

$$\delta(x, y) = \begin{cases} \{a \in P | f(x, a) \neq f(y, a)\}, & [x]_c \neq [y]_c \text{ \& } [x]_d \neq [y]_d \\ 0, & [x]_c = [y]_c \text{ or } [x]_d = [y]_d \end{cases}$$

其中 $[x]_c$ 表示对象 x 关于属性集 C 的等价类;也可理解为 x 关于 C 的属性值集。该区分矩阵将不协调对象(各条件属性值完全相同而决策属性值不同的对象)亦归入无区分对象行列, 对应矩阵中元素皆为0。由区分矩阵可方便地得到核:

$$Core(C) = \{a | a \in C \wedge \exists x, y (\delta(x, y) = \{a\})\}$$

差别函数:差别函数是一个布尔函数,用 Δ 表示,定义为:

$$\Delta = \prod_{(x, y) \in U \times U} \delta(x, y), (\sum \delta(x, y) = \bigvee a_i, a_i \in \delta(x, y))$$

对差别函数根据布尔代数中的分配律和吸收律进行化简即得属性约简的析取式。以表1为例,对应区分矩阵为(上三角部分略):

对象	1	2	3	4	5	6	7	8
1								
2	0							
3	$tqci$	$tqci s$						
4	0	0	$tqci s$					
5	t	$qcis$	$tqci$	$tqcs$				
6	$tqci s$	tc	qis	tci	0			
7	$qcis$	$tqci$	0	qci	$tqci s$	tqi		
8	0	0	$tqci s$	0	$tqcs$	tci	qci	

对应差别函数为(略去重复项):

$$\Delta = (t \vee q \vee c \vee i)(t \vee q \vee c \vee i \vee s)(t \vee q \vee c \vee i \vee s)(t \vee q \vee c \vee s)(t \vee c)(q \vee i \vee s)(t \vee c \vee i)(q \vee c \vee i)(t \vee q \vee i) = t(q \vee i \vee s)(q \vee c \vee i) = tq \vee ti$$

得该决策表的约简 $\{t, q\}$ 和 $\{t, i\}$,两者都可作为决策表的约简;核是 $\{t\}$ 。取约简 $\{t, q\}$ 并去掉重复行(第8行)后的决策表如表2所示。得到属性约简后可根据区分矩阵继续求值约简,方法为:

表2 约简后的决策表

序号	t	q	v
1	0	1	1
2	1	0	1
3	-1	-1	-1
4	0	0	1
5	1	1	0
6	-1	0	0
7	0	-1	-1

(1)将原区分矩阵各元素中不属于约简的属性删除(并删除重复对象)获得新区分矩阵。对属性约简后的决策表直接计算区分矩阵可得同样结果;

(2)将新区分矩阵每行视为一个单独的决策表求属性约简,每个约简在行中对应的值组成一个决策;

(3)所有行约简获得的决策剔除重复项后组成最终决策集。

此方法借鉴了文献[4],但步骤更简练且意义明了。以表2对应的决策表为例,其对应区分矩阵为(需要对整行约简,故上三角元素不能省略):

对象	1	2	3	4	5	6	7
1			tq		t	tq	q
2			tq		q	t	tq
3	tq	tq		tq	tq	q	
4			tq		tq	t	q
5	t	q	tq	tq			tq
6	tq	t	q	t			tq
7	q	tq		q	tq	tq	

对各行求约简得:

第1行:约简为 $\{tq\}$,对应决策 $t_1q_1 \rightarrow v_1$ (下标表示属性的取值);

第2行:约简为 $\{tq\}$,对应决策 $t_1q_0 \rightarrow v_1$;

第3行:约简为 $\{q\}$,对应决策 $q_{-1} \rightarrow v_{-1}$;

第4行:约简为 $\{tq\}$,对应决策 $t_0q_0 \rightarrow v_1$;

第5行:约简为 $\{tq\}$,对应决策 $t_1q_1 \rightarrow v_0$;

第6行:约简为 $\{tq\}$,对应决策 $t_{-1}q_0 \rightarrow v_0$;

第7行:约简为 $\{q\}$,对应决策 $q_{-1} \rightarrow v_{-1}$ 。

对上述约简综合可得简化的决策规则:

$$t_0(q_0 \vee q_1) \vee t_1q_0 \rightarrow v_1$$

$$t_1q_1 \vee t_{-1}q_0 \rightarrow v_0$$

$$q_{-1} \rightarrow v_{-1}$$

所得决策规则用于计算机调度过程的指导。为了更好的做出满意的调度,计算机还需学习决策者针对某种调度状态所采取的行为,例如将作业改派其他机器、延迟某些作业等;这些行为可以直接记录在数据库中以备随时调用,或采用与本章所述类似的规则提取方式获得。

4 结论

提出的约简方法将属性约简和值约简都通过区分矩阵操作进行,步骤简单明了,非常利于计算机实现。将其应用于生产调度决策规则的提取获得了较好效果。该方法的缺点在于当决策表比较庞大时,对应区分矩阵占用资源较大,对系统运行效率有消极影响。

参考文献:

- [1] 张文修.粗集理论与方法[M].北京:科学出版社,2001.
- [2] Ziarko W, Cercone N, Hu X. Rule discovery from databases with decision matrices[C]//9th Int Symposium on Foundation of Intelligent System, 1996:653-662.
- [3] 胡寿松,何亚群.粗糙决策理论与应用[M].北京:北京航空航天大学出版社,2006.
- [4] 顾军华,周艳聪,宋洁,等.一种新的求解属性值约简算法[J].南开大学学报:自然科学版,2003,36(4):38-42.
- [5] 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.
- [6] 叶东毅,黄翠微,赵斌.基于逼近精度的一个粗糙集属性约简算法[J].福州大学学报,2000,28(1):7-10.
- [7] 石峰,娄臻亮,张永清.一种改进的粗糙集属性约简启发式算法[J].上海交通大学学报,2002,36(4):478-481.
- [8] 夏克文,沈钧毅,李昌彪.样本信息处理中一种属性约简方法的研究[J].西安交通大学学报,2005,39(6):558-602.

(上接206页)

5 结论

以上是通过主成分分析和因子分析对我国区域经济发展进行比较分析的一个应用实例,采用了20世纪80年代的历史数据进行了比较分析,从结果中可以看出完全符合当时我国区域经济的发展状况。总之,应用主成分分析的方法即可以在运算结果的基础上,对各省区经济发展战略模式和经济发展总体水平进行综合性的比较、分析和评价,为制定决策提供科学的依据。该方法一直在区域宏观经济分析系统中使用,也得到了有关方面的一致好评,但由于实际数据的保密性,就不便将结果显示出来了。实践证明该方法不仅实用,而且也有着很好的发展前景,随着其研究和应用的进一步深入,目前已在数据库和数据挖掘等领域得到广泛的应用,并取得了很好的社会效益。

参考文献:

- [1] 何晓群.现代统计分析方法与应用[M].北京:中国人民大学出版社,1998.
- [2] Lu Cong-de, Zhang Chun-mei, Zhang Tai-yi, et al. Kernel based symmetrical principal component analysis for face[J]. Classification, Neurocomputing, 2006.
- [3] Boivin J, Ng S. Are more data always better for factor analysis?[J]. Journal of Econometrics, 2006.
- [4] 朱建平.数据挖掘的统计方法及实践[M].北京:中国统计出版社,2005.
- [5] 丁士晟.多元分析方法及其应用[M].吉林:吉林出版社,1987.
- [6] 高惠璇.应用多元统计分析[M].北京:北京大学出版社,2005.
- [7] 向东进.实用多元统计分析[M].武汉:中国地质大学出版社,2005.
- [8] 张润楚.多元统计分析[M].北京:科学出版社,2006.