

# 设备管理组织方式的 决策树优化实证研究

刘希宋, 王辉坡

(哈尔滨工程大学 经济管理学院, 黑龙江 哈尔滨 150001)

**摘要:** 设备管理在实际的管理工作中处于重要的地位, 设备管理的效果直接影响企业的生产经营活动, 设备管理组织方式和设备管理的效果之间存在明显的关系, 通过对某集团 12 家分公司的设备管理组织方式进行分析, 利用决策树优化算法归纳出了设备管理组织方式优化的规则, 希望能为其它企业组织机构改造提供依据。

**关键词:** 设备管理; 组织方式; 决策树; 优化

中图分类号: F273.4

文献标识码: A

文章编号: 1001-7348(2006)09-0105-03

## 0 前言

随着我国改革开放的进程和现代管理技术的发展, 企业都对自己的设备管理机构进行了改造, 从而产生了千差万别的设备管理组织方式。笔者在对设备管理效果的研究过程中发现, 同种性质的企业, 甚至一些起点相同的企业, 在经过设备管理组织机构改造后, 其设备管理的效果差别很大, 因此, 可以认为设备管理组织方式和设备管理效果之间存在着显著的关系, 通过遵守一定准则建立的设备管理组织方式将比遵守另一种准则建立的组织方式更能有效地工作, 产生更好的效果。如何发现这些准则是本文研究的目的。数据挖掘和知识发现技术为我们的研究提供了强有力的工具, 本文通过对某集团下属的 12 个分公司的设备管理组织方式的分析, 通过决策树算法来遴选科学有效的准则, 为公司设计合理的设备管理组织方式提供借鉴。

## 1 数据挖掘与决策树算法简介

数据挖掘 (Data Mining) 是从大量数据

中开采出隐含的, 先前未知的, 对决策有潜在价值的知识与规则<sup>[1]</sup>。决策树算法在数据挖掘算法中具有重要的地位并获得了广泛的应用。决策树算法起源于概念学习系统 CLS (Concept Learning System)<sup>[2]</sup>, 与其它分类算法如贝叶斯分类 (Bayes), 神经网络 (Neural Networks), 遗传算法 (Genetic Algorithms) 相比具有分类速度快、能够提供直观的描述, 形成的分类规则易于理解、准确率高等特点。

### 1.1 基本决策树算法

决策树算法是一种归纳分类算法, 它用树状结构表示数据分类的结果, 树的非叶节点表示对数据属性的测试, 根据测试的结果选择某个分支, 每一个叶节点代表一个类。这样就形成了一个能够表示不同属性之间组合的树, 由根节点到各个叶节点的路径描述可得到各种分类规则。图 1 显示了一个简单的决策树, 它用于某项目运动员的选拔。

图 1 的决策树蕴含着 4 条规则:

规则 1: 如果该运动员的竞技水平=高, 则选中该运动员。

规则 2: 如果该运动员的竞技水平=中,

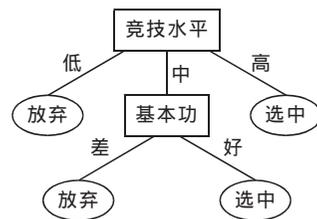


图 1 运动员选择决策树

且基本功=好, 则选中该运动员。

规则 3: 如果该运动员的竞技水平=中, 且基本功=差, 则放弃该运动员。

规则 4: 如果该运动员的竞技水平=低, 则放弃该运动员。

决策树的生成过程如下: 树以代表训练样本的单个节点开始; 如果样本都在同一类, 则该节点成为树叶, 并用该类标记。否则, 根据某一策略选择一个属性值。令该属性成为该节点的“测试”属性。对测试属性的每个已知的值, 创建一个分支, 并据此划分样本。重复 2-4 的步骤。直到给定节点的样本都属于一个类或者没有剩余属性可以用来进一步划分样本或者该分支已没有样本<sup>[2,3]</sup>。

收稿日期: 2005-09-19

作者简介: 刘希宋(1936-), 湖南新化人, 哈尔滨工程大学经济管理学院教授、博士生导师, 研究方向为工业经济管理、现代管理理论与方法和企业现代化管理; 王辉坡(1979-), 男, 河北栾城人, 哈尔滨工程大学博士研究生。

1.2 划分属性的选择方法<sup>[3-6]</sup>

决策树一般使用信息增益 (information gain) 作为选择分裂属性的标准。信息增益是指分裂前的“熵”值减去分裂后各子类“熵”值的“加权和”。熵是衡量系统混乱程度的统计量,为了使得分类后的分支更加趋于有序,要选择能够使得熵减少最多的属性作为分裂属性,即选择信息增益最大者。这种选择分裂属性的方法可以使样本分类所需的信息量达到最小,并能使对样本分类的随机性和不纯度达到最小。

信息论中对信息和熵分别作了如下定义<sup>[7-9]</sup>:

$$I_i = -\log_2 p_i$$

$$E = -\sum_{i=1}^n p_i \log_2 p_i$$

在决策树选择分裂属性时,其具体步骤如下:

(1) 计算分裂前的样本的“熵”值。假设集合 S 中有 s 个记录,有  $A_1, A_2, \dots, A_m, C$  共 n+1 个属性,按照分类属性 C 的 m 个不同值  $\{a_1, a_2, \dots, a_m\}$ , 将 s 个记录分成 m 个不同的类  $c_1, c_2, \dots, c_m$ , 每类的记录数分别为  $s_1, s_2, \dots, s_m$ , 那么分裂前集合 S 的总熵为:

$$E(C) = -\sum_{i=1}^n p_i \log_2 p_i \quad (1)$$

式中  $p_i$  为集合 S 中任意记录属于类  $c_i$  的概率,用  $s_i/s$  来估计。

(2) 计算分裂后各子类“熵”值的“加权和”及信息增益。

设属性  $A_1$  有 v 个不同的属性值,它将集合 S 分为 v 个子集,  $\{S_1, S_2, \dots, S_v\}$ 。其中,  $S_j$  中的记录数为  $d_j$ , 按照分类属性 C 将  $d_j$  分为  $\{d_{1j}, d_{2j}, \dots, d_{mj}\}$ , 则子集的熵为:

$$E(S_j) = -\sum_{i=1}^n p_{ij} \log_2 p_{ij} \quad (2)$$

式中  $p_{ij}$  为集合  $S_j$  中任意记录属于类  $c_i$  的概率,用  $d_{ij}/d_j$  来估计。用上式可以计算出  $E(S_1), E(S_2), \dots, E(S_v)$ 。

求  $E(S_1), E(S_2), \dots, E(S_v)$  的“加权和”,用  $d_j/s$  作为  $E(S_j)$  的权重。计算公式为:

$$E(A_1) = \sum_{j=1}^v \frac{d_j}{s} E(S_j) \quad (3)$$

计算  $A_1$  作为分裂属性时的信息增益,计算公式如下:

$$\text{Gain}(A_1) = E(A) - E(A_1) \quad (4)$$

重复步骤 ~ , 计算出  $\text{Gain}(A_2),$

$\text{Gain}(A_3), \text{Gain}(A_4), \dots, \text{Gain}(A_n)$ 。

(3) 选择信息增益最大的作为分裂属性。

## 2 设备管理组织方式实证研究的指标选择

随着现代管理在企业的逐步推进,管理理念不断创新,设备管理的组织方式在理论上应该达到如下目的:首先,先进的设备管理组织方式应该尽量降低企业的设备运行成本,企业的设备运行成本在企业中特别是在制造业企业中占有很重要的地位,先进的管理组织方式可以有效地控制设备运行中的物料消耗和能源消耗,从而降低设备运行成本;其次,先进的设备管理组织方式可以对外部环境做出快速的反应,保证企业生产任务调整的完成;最后,先进的设备管理组织方式有利于设备技术水平的不断提升、有利于本行业先进制造技术在本企业的推广应用。所以,本文以上述目的为标准,从以下几个角度来考察企业设备管理的组织方式。

(1) 设备管理组织的深度( $A_1$ )。设备管理组织的深度是指设备管理机构的级别,不同的级别拥有不同的权限,在实际工作过程中,超过自身权限时要报上一级管理部门批准。上一级的设备管理部门,负责监督下一级管理部门的工作。设备管理组织方式的分级别设置可以实现良好的监督和制约,防止徇私舞弊的产生,同时过多的级别将大大降低设备管理的灵活性和对行业新技术的适应能力。

(2) 设备管理组织的广度( $A_2$ )。设备管理组织的广度是指同一级的设备管理部门的数量,现代的设备管理也是基于流程作业的管理,每一设备从提出需求、投资到安装调试乃至到最后的报废,都需要有专门的人员负责。按照作业流程设计的管理部门必须做到简洁而不简单,这样才能使得设备管理工作顺利完成又不会由于各部门之间相互推诿导致效率的降低。

(3) 设备维修体系的社会化程度( $A_3$ )。近年来供应链技术被引入到了设备管理当中,从而使得设备维修的社会化得到了相应的发展,设备维修的社会化有利于设备维修工作的专业化,完善的社会化设备维修体系可以降低企业设备运行的成本。本文中,设备维修体系的社会化程度用企业社会化的维修费用占全部维修费用的比例来衡量。

(4) 技术管理人员与经济管理人员的比例( $A_4$ )。设备管理工作分为技术管理和经济管理,这是一个事物的两个方面,技术管理是从实物的角度来进行设备的维护、保养,运行状态的保持等工作;经济管理是指从资产的角度来对设备进行账面管理、合理提取补偿资金,保持企业持续生产能力,这一指标就是为了寻找这两者之间合理的配比关系。

(5) 设备管理人员占企业全部管理人员的比例( $A_5$ )。设备管理是企业管理中一个重要的组成部分,由于设备管理工作的复杂性,一个企业要想做好设备管理工作必须配备一定的人员,人员数量和企业规模的大小有关,本文用设备管理人员占全部管理人员的比例来判断人员配备是否有冗余。

(6) 设备管理的综合效果(C)。设备管理的效果可以根据不同的目的设计不同的指标,现在较常用的指标包括正面指标和负面指标,前者如设备完好率、设备综合利用率;后者如设备事故率、设备故障停机率以及设备维修费用率等。本文采取综合考虑的方法,首先把负指标经过一定的数学处理转化为正指标,然后采取加权合成的技术,从而得到综合效果指标。

## 3 设备管理组织方式优化实证

## 3.1 数据的获取

本文对某集团下属的 12 家分公司的设备管理组织方式进行了实证考察,得到各属性的值如表 1 所示。

表 1 设备管理组织方式的原始数据

分公司	$A_1$	$A_2$	$A_3$	$A_4$	$A_5$	C
$F_1$	4	7	0.2	0.8	0.08	0.62
$F_2$	4	3	0.7	0.4	0.04	0.57
$F_3$	3	6	0.6	0.4	0.07	0.89
$F_4$	4	6	0.4	0.7	0.05	0.84
$F_5$	6	5	0.5	0.5	0.05	0.41
$F_6$	5	4	0.5	0.6	0.06	0.64
$F_7$	4	7	0.3	0.4	0.07	0.59
$F_8$	3	5	0.6	0.5	0.05	0.88
$F_9$	6	6	0.7	0.3	0.08	0.57
$F_{10}$	4	5	0.7	0.2	0.09	0.65
$F_{11}$	3	7	0.5	0.4	0.07	0.71
$F_{12}$	5	4	0.6	0.5	0.08	0.69

由于表 1 的原始数据呈现连续化的趋势,使得决策树过于庞大、计算复杂,使得有用的规则被大量的信息所淹没,不利于决策树方法的使用,因此需要对上述原始数据进行离散化处理,处理过程可按表 2 进行。

表2 原始数据离散化准则

属性	0	1	2
C	<0.60	0.60~0.80	>0.80
A <sub>1</sub>	<4	4~5	>5
A <sub>2</sub>	<5	5~6	>6
A <sub>3</sub>	<0.4	0.4~0.7	>0.7
A <sub>4</sub>	<0.4	0.4~0.7	>0.7
A <sub>5</sub>	<0.05	0.05~0.08	>0.08

3.2 生成设备管理组织方式优化的决策树

(1) 建立决策树。

利用公式(1)计算分裂前样本的熵值, E(C)=1.5546。

利用公式(2)、(3)计算 E(A<sub>1</sub>), E(A<sub>2</sub>), ..., E(A<sub>5</sub>), 结果如下:

E(A<sub>1</sub>)=1.0339

E(A<sub>2</sub>)=1.1887

E(A<sub>3</sub>)=1.4758

E(A<sub>4</sub>)=1.3554

E(A<sub>5</sub>)=1.3091

利用公式(4)计算 Gain(A<sub>1</sub>), Gain(A<sub>2</sub>), ..., Gain(A<sub>5</sub>), 结果如下:

Gain(A<sub>1</sub>)=0.5207

Gain(A<sub>2</sub>)=0.3659

Gain(A<sub>3</sub>)=0.0788

Gain(A<sub>4</sub>)=0.1992

Gain(A<sub>5</sub>)=0.2455

选取能够使熵值减少最多的属性作为分裂属性, 即选择, 按照以上步骤循环操作, 即可形成决策树, 见图2。

4 结论

通过上面的计算我们得到了图2所示的决策树, 该决策树蕴含着9条规则, 它们就是通过数据挖掘而得到的知识, 如表3所示。

通过表3我们可以发现在设备组织方式设计时, 应该尽量向规则2和规则6靠近; 此外, 该表还蕴含着如下信息:

表3 设备管理组织方式优化规则

规则	A <sub>1</sub>	A <sub>2</sub>	A <sub>3</sub>	A <sub>4</sub>	A <sub>5</sub>	C
1	<4	>6	—	—	—	0.6~0.8
2	<4	5~6	—	—	—	>0.80
3	4~5	<5	—	—	<0.05	<0.6
4	4~5	<5	—	—	0.05~0.08	0.6~0.8
5	4~5	5~6	—	<0.4	—	0.6~0.8
6	4~5	5~6	—	0.4~0.7	—	>0.8
7	4~5	>6	—	0.4~0.7	—	<0.6
8	4~5	>6	—	>0.7	—	0.6~0.8
9	>5	—	—	—	—	<0.6

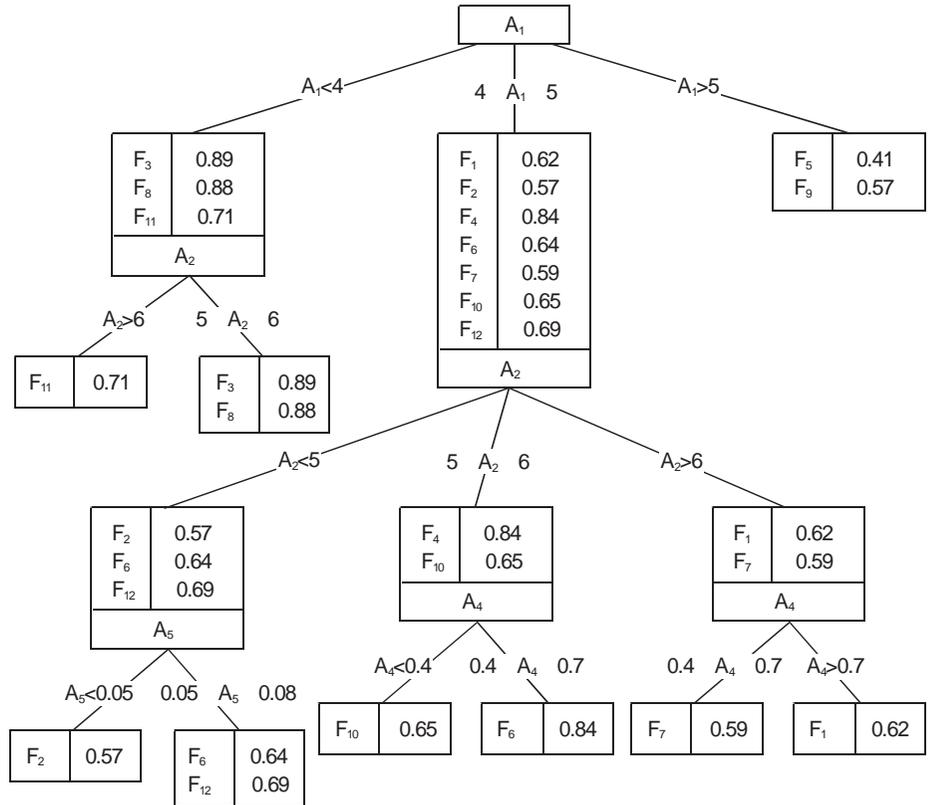


图2 设备管理组织方式优化的决策树

(1) 组织方式的深度应尽量减少, 在4层以下是效率最高, 在4~5层时, 经过适当的调整也可以达到较好的管理效果。

(2) 组织方式的广度应该在5~6之间, 这样可以满足部门分工的需要, 又不至于造成机构臃肿。

(3) 从理论上讲, 设备维修社会化程度的提高可以降低企业的设备维修成本, 但是由于我国的特殊国情, 这方面的效果并不明显, 一方面是由于我国专业化的维修体系还不健全; 另一方面, 大部分企业维修社会化的合作单位本来就是从本企业分出去的, 即企业在实行主业和副业分离时, 往往为了安抚副业部门的人心, 而对副业部门给予很多优惠措施。

(4) 应适当提高技术管理人员的比重, 现代大部分企业设备管理的技术管理和经济管理是分开的, 适当地加强技术管理保证设备运行的可靠性和经济性可以提高设备管理的效果。

(5) 设备管理人员占管理人员总数的比例应适当控制, 过高会造成人员冗余, 过少又会导致

人手不足。

设备管理是企业的重要组成部分, 希望本文通过决策树优化算法得出的结论能够为企业的决策者提供一些决策参考。

参考文献:

- [1] 高洪深. 决策支持系统(DSS)理论. 方法. 案例[M]. 北京: 清华大学出版社, 南宁: 广西科学技术出版社, 2000.
- [2] 田苗苗. 数据挖掘之决策树方法概述[J]. 长春大学学报, 2004, (6).
- [3] 张德明, 曹秀英, 刘希宋. 基于判定树归纳法的公司治理组合机制实证研究[J]. 哈尔滨工程大学学报, 2004, (4).
- [4] 贺宪民, 孟虹, 王忆勤, 郎庆波, 范思昌. 基于熵的决策树理论及其在中医证型研究中的应用[J]. 数理统计与管理, 2004, (9).
- [5] 梁静国, 张祥林. 管理决策仿真[M]. 哈尔滨: 哈尔滨工程大学出版社, 2003.
- [6] 胡小刚. 数据挖掘中决策树分类算法的研究[D]. 华中师范大学硕士学位论文, 2002.18-20.
- [7] C.E.Shannon: A Mathematical Theory of Communication. B.S.T.J.vol.27, July 1948
- [8] 藤田宏一. 基础信息论[M]. 魏鸿骏, 陈尚勤译. 北京: 国防工业出版社, 1982.

(责任编辑: 汪智勇)