

DOI:10.16652/j.issn.1004-373x.2018.01.038

# 基于决策树的工程项目管理优化研究

蒲天添

(云南财经大学, 云南 昆明 650000)

**摘要:** 为了使企业能够更好地适应社会的需求,对市场的变化做出快速地反应。针对工程项目的特点,运用改进的ID3决策树算法对工程项目管理进行优化。首先,构建工程项目管理的优化属性,聘请相关专家对属性与工程项目风险进行评分,然后,对评分进行离散化处理,最后利用ID3决策树算法对工程项目信息进行充分挖掘,生成工程项目管理优化决策树,找出其中影响工程项目优化管理的分类规则,分析工程项目的实施情况,给出个性化提示,从而达到对企业工程项目的有效控制和优化。

**关键词:** 电力系统; 工程项目管理优化; 信息熵; ID3决策树; 数据挖掘; 离散化处理

**中图分类号:** TN911-34

**文献标识码:** A

**文章编号:** 1004-373X(2018)01-0169-04

## Study on engineering project management optimization based on decision tree

PU Tiantian

(Yunnan University of Finance & Economics, Kunming 650000, China)

**Abstract:** In order to let the enterprises adapt to the society requirement better, and respond to the changes in the market rapidly, an improved ID3 decision tree algorithm is used to optimize the engineering project management according to the characteristics of engineering project. The optimization attribute of the engineering project management was constructed. The attribute and engineering project risk are graded by relevant experts. And then the scores are discredited. The ID3 decision tree algorithm is adopted to mine the engineering project information fully, generate the decision tree of the engineering project management optimization, find out the classification rules influencing the engineering project management optimization, analyze the implementation of the engineering project, and give the personalized reminder. The ID3 decision tree algorithm can control and optimize the enterprise's engineering project effectively.

**Keywords:** electric power system; engineering project management optimization; information entropy; ID3 decision tree; data mining; discretization

## 0 引言

随着中国经济的快速发展,为了使企业能够更好地满足社会的需求,适应市场的快速变化,据相关研究<sup>[1-2]</sup>,我国有90%以上的国有大中型企业都实施了工程项目管理模式,这些企业因为采用工程项目管理模式,能够对市场的变化做出快速地反应,运行状况良好,都取得了不错的成果。但是,由于我国的管理技术落后于西方发达国家,我国推行工程项目管理起步较晚,与西方发达国家的工程项目管理运行效果进行比较,可以发现我国企业推行的工程项目管理与发达国家之间还存在一定的差距<sup>[3]</sup>。在配套改革、制度创新和机制转变等方面还需要不断发展与完善<sup>[4]</sup>。由于企业在管理方式、竞争程度等方面存在着一定的差异,因此,不同企业的工程项目管理效果有着明显的不同,因此,为了更好地适应

社会、市场的需求,不同的企业应根据自身的资源优势、行业竞争状况,对工程项目进行优化管理<sup>[5]</sup>。

本文将决策树算法引入企业工程项目管理优化过程中,决策树算法主要用来对研究数据进行预测与分类,是目前一种应用最为广泛的数据挖掘算法<sup>[6]</sup>。研究中,为了加快工程项目管理优化的速度,对传统的ID3决策树算法进行改进,将改进的ID3决策树算法运用于工程项目管理中,通过对企业工程项目的实施信息进行充分挖掘,找到对工程项目实施效果有较大影响的那些分类规则,用于分析企业工程项目的实施情况,对企业的工程项目实施方式给出针对性建议,从而使企业的工程项目管理模式得到优化。

## 1 ID3决策树算法

### 1.1 决策树算法

决策树算法主要用于对数据的分类与预测,是目前

收稿日期:2017-05-05

修回日期:2017-06-22

一种应用最为广泛的数据挖掘算法之一。它从根节点开始遍历,在搜索的过程中利用归纳算法选择其中一个分支,直到选择某个叶子节点,从而构造成决策树<sup>[7]</sup>。在决策树中,为了实现数据的分类,按照 IF-THEN 的分类规则,从根节点到叶子节点将决策树转化成多个分类规则。决策树算法的优点是生成的模式简单、鲁棒性强,同时分类的精度高、速度快。决策树算法有多种,其中应用最广泛的是 ID3 决策树算法<sup>[8]</sup>。

## 1.2 ID3 决策树算法的构造

ID3 决策树算法是将信息增益用作选取分支节点的测试属性,首先将信息增益最高的节点作为决策树的根节点,然后找到次高节点,作为决策树的分支,依次类推,最后递归生成一棵简单的决策树<sup>[9]</sup>。

### 1.2.1 设定信息熵

假设训练集  $S$  中包含了  $S_i$  (其中  $i=1,2,\dots,m$ ) 个数据样本,分成  $m$  个类  $c_1, c_2, \dots, c_i$  (其中  $i=1,2,\dots,m$ )。

$$I(S_1, S_2, \dots, S_m) = -\sum_{i=1}^m P_i \log_2 P_i \quad (1)$$

式中:因为信息采用二进制编码,所以使用的函数为以 2 为底的对数;  $P_i$  表示训练集  $S$  中某随机数据样本属于类  $c_i$  的概率,  $P_i$  使用  $S_i/S$  进行估算。

### 1.2.2 选择根节点

设属性  $A$  为树的根节点,  $A$  中包含  $v$  个值  $\{a_1, a_2, \dots, a_v\}$ 。训练集  $S$  被分为子集  $\{S_1, S_2, \dots, S_v\}$ , 设其中  $A$  的值为  $a_j$ , 某子集为  $S_{j \circ s_{ij}}$  是  $S_j$  中的样本属于  $C_i$  的个数,则:

$$E(A) = \sum_{j=1}^v \frac{s_{1j} + s_{2j} + \dots + s_{mj}}{S} I(s_{1j}, s_{2j}, \dots, s_{mj}) \quad (2)$$

式中:  $\frac{s_{1j} + s_{2j} + \dots + s_{mj}}{S}$  是子集  $S_j$  的权,利用式(2)推导出集合  $S$  的信息熵为:

$$I(s_{1j}, s_{2j}, \dots, s_{mj}) = -\sum_{i=1}^m P_{ij} \log_2 P_{ij} \quad (3)$$

式中:  $P_{ij}$  使用  $s_{ij}/|S_j|$  估计,表示子集  $S_j$  中某随机数据样本属于类  $C_i$  的概率。

分支  $A$  的信息增益为:

$$\text{Gain}(A) = I(S_1, S_2, \dots, S_m) - E(A) \quad (4)$$

信息增益用来描述样本子集的纯度,信息增益越大,表明划分后的样本子集越纯,越有利于进行集合的优化与分类。

### 1.2.3 ID3 决策树算法性能分析

ID3 决策树算法从空树开始,利用爬山策略遍历假设空间,对大规模的数据集进行处理,其优缺点如下<sup>[10]</sup>:

优点: ID3 决策树算法以爬山策略对搜索空间进行遍历,以信息增益作为测试属性的技术,为了能够获得

被测数据的最大类别信息,对每个非叶子节点都进行测试,如果类别信息属性是非离散性的,为了降低训练样本的敏感性,需要对类别信息进行离散化处理。

缺点: 由于决策结算法是从根节点开始,对整个决策树进行遍历,如果再增加训练样例,则需要重新构造新的决策树,通用性很弱。在选择分裂节点时,需要对每个属性的信息熵进行计算来判断其分裂属性,计算量较大,如果属性较多,则增大了决策树的产生成本,同时,运行的效率亦会受到很大的影响。

## 1.3 ID3 决策树算法的改进

为了提高 ID3 决策树算法的运行效率,利用泰勒公式和麦克劳林公式对传统的 ID3 决策树算法进行改进。当  $|x - x_0|$  很小时,高等数学中的微分理论知识见式(5):

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + o(x - x_0) \quad (5)$$

麦克劳林公式如下:

$$f(x) = f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \dots + \frac{f^{(n)}(0)}{n!}x^n + R(x) \quad (6)$$

其中,  $R(x) = \frac{f^{(n+1)}(x)}{(n+1)!}(x - x_0)^{n+1}$

麦克劳林公式的近似公式见式(7):

$$f(x) \approx f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \dots + \frac{f^{(n)}(0)}{n!}x^n \quad (7)$$

式中:若  $f(x) = \ln(1+x)$ , 且当  $x \rightarrow 0$ , 存在  $\ln(1+x) \approx x$ 。

## 1.4 改进后属性信息熵的运算公式推导

由等式:

$$I(p_i, n_i) = -\frac{p_i}{p_i + n_i} \log_2 \frac{p_i}{p_i + n_i} - \frac{n_i}{p_i + n_i} \log_2 \frac{n_i}{p_i + n_i} \quad (8)$$

$$E(A) = \sum_{i=1}^n \frac{p_i + n_i}{p + n} I(p_i, n_i) \quad (9)$$

整理得:

$$E(A) = \sum_{i=1}^n \frac{1}{(p+n) \ln 2} \left( -p_i \ln \frac{p_i}{p_i + n_i} - n_i \ln \frac{n_i}{p_i + n_i} \right) \quad (10)$$

因为  $\frac{1}{(p+n) \ln 2}$  是常量, 设  $G(A) =$

$\sum_{i=1}^n \left( -p_i \ln \frac{p_i}{p_i + n_i} - n_i \ln \frac{n_i}{p_i + n_i} \right)$ , 存在以下关系式成立:

$$\ln(1+x) \approx x \quad (11)$$

$$\ln \frac{p_i}{p_i + n_i} = \ln \left( 1 - \frac{n_i}{p_i + n_i} \right) \approx -\frac{n_i}{p_i + n_i} \quad (12)$$

同理:

$$\ln \frac{n_i}{p_i + n_i} \approx -\frac{p_i}{p_i + n_i} \quad (13)$$

则:

$$G(A) = \sum_{i=1}^n \left( p_i \frac{n_i}{p_i + n_i} + n_i \frac{p_i}{p_i + n_i} \right) = \sum_{i=1}^n \frac{2p_i n_i}{p_i + n_i} \quad (14)$$

改进后的属性信息熵公式为:

$$H(A)=\left(\sum_{i=1}^n\frac{2p_in_i}{p_i+n_i}\right)N\tag{15}$$

式中  $N$  表示属性值的个数。

2 改进ID3决策树算法的工程项目管理优化算法

- 1) 选择企业的某个工程项目,构建工程项目管理优化的相关属性,聘请相关专家对属性值进行打分,收集专家的打分结果,构建工程项目管理优化的训练集;
- 2) 利用粗糙集理论,对训练集进行离散化处理;
- 3) 利用改进ID3决策树算法,选择信息增益最高的节点属性作为根节点;
- 4) 然后找到次高节点,作为决策树的分支,依次类推,最后递归生成一棵决策树;
- 5) ID3决策树算法中每循环执行一次,生成的决策树都会被更新一次。

3 仿真试验

3.1 工程项目管理的三要素分析

工程项目管理以最低成本均衡资源,控制工程质量为目标,根据相关研究成果,可以将工程项目管理属性归纳为计划、技术、费用和进度4个基本属性。在进行工程项目管理时,计划是为了确保项目的顺利实施与完成,事先需要做好周密的计划安排,预测工程项目在实施时的各种风险;技术是确保项目在设计 and 实施过程中的技术保障;费用是保证项目 in 实施过程中的财务预算;进度是指项目设计、建造和安装等所估算和分配的作业时间。从经济学角度来看,工程项目管理中计划、技术、费用和进度的相互制约关系产生了多目标协同问题,即多目标优化。

3.2 基于ID3决策树算法的工程项目管理优化

在Java平台的Weka软件进行试验仿真,以某国有企业工程项目管理优化为目标属性,选择工程项目管理的计划、技术、费用和进度4个属性对工程项目管理进行优化,利用ID3决策树算法提取有用的分类规则,指导企业工程项目优化运行。

3.3 数据准备

分析该国有企业的某个工程管理项目的运行情况,聘请相关专家对这4个指标分别根据样本实际情况给出评分,并按一定规则得到综合评分,最后,根据各属性的综合评分情况对该工程项目的风险进行综合评分。表1为收集的10个专家对该工程项目的评分。

3.4 传统ID3决策树算法的信息增益

为了减少运行的误差,利用粗糙集算法对该工程项目的专家评分表进行离散化处理,然后利用式(1)~式(3)

计算出该工程项目的项目风险、计划、技术、费用和进度的期望信息。

表1 电力系统工程项目管理指标评分

Tab. 1 Indicators grading of electric power system engineering project management

项目	计划	技术	费用	进度	风险程度
1	0.7	0.6	0.8	0.6	0.7
2	0.9	0.7	0.8	0.7	0.8
3	0.8	0.6	0.7	0.6	0.7
4	0.8	0.7	0.5	0.5	0.7
5	0.9	0.6	0.8	0.4	0.7
6	0.8	0.5	0.4	0.5	0.6
7	0.9	0.6	0.5	0.5	0.7
8	0.9	0.7	0.7	0.6	0.7
9	0.8	0.4	0.8	0.7	0.7
10	0.8	0.7	0.5	0.8	0.7

项目风险的期望信息:Info(风险程度)=0.998 8

计划的期望信息:Info(计划)=0.998 7

技术的期望信息:Info(技术)=0.994 5

费用的期望信息:Info(费用)=0.987 9

进度的期望信息:Info(进度)=0.991 9

利用式(4)计算得到4个属性的信息增益:

Gain(计划)=0.998 8-0.998 7=0.000 1

Gain(技术)=0.998 8-0.994 5=0.004 3

Gain(费用)=0.998 8-0.987 9=0.010 9

Gain(进度)=0.998 8-0.991 9=0.006 9

由于进度在属性中具有最高信息增益,它被选作测试属性。决策树创建过程为:创建一个节点,用进度指标作为标记,并对每个属性值引出一个分枝,对各个分枝进行如上所述递归运算,最终得到的决策树如图1所示。

从生成的决策树来看,为了实现该项目的有效管理,首先必须要保证该项目的进度,然后在确保计划的完善,最后,要充分确保项目实施技术的支撑能力,这样就可以确保该项目的顺利实施与完成。

3.5 改进ID3决策树算法后的信息增益

为了提高工程项目管理优化效率,对ID3决策树算法进行改进,利用式(15)进行计算:

项目风险的期望信息:Info(风险程度)=0.999 7

计划的期望信息:Info(计划)=0.675 3

技术的期望信息:Info(技术)=0.453 7

费用的期望信息:Info(费用)=0.653 4

进度的期望信息:Info(进度)=0.562 4

得到各属性的属性信息熵,计算出各指标的信息增益如下:

Gain'(计划)= 0.999 7-0.675 3=0.324 4

Gain'(技术)=0.999 7-0.453 7=0.546



$$\text{Gain' (费用)} = 0.999\ 7 - 0.653\ 4 = 0.346\ 3$$

$$\text{Gain' (进度)} = 0.999\ 7 - 0.562\ 4 = 0.437\ 3$$

Gain' (进度) 相对不大, 即应把进度指标作为测试属性创建决策树, 然后对各个分支进行递归运算, 构建的决策树与图1相同。为了比较决策树ID3决策树算法改进前后性能的变化, 针对不同个数的工程项目管理, 处理相同数据所需的耗时对比见图2。由图2可见, 改进后的ID3决策树算法的速度要远远高于传统的ID3决策树算法, 工作效率大大增强。

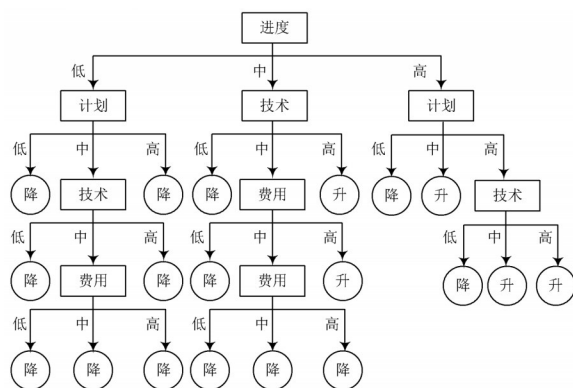


图1 最终决策树

Fig. 1 Final decision tree

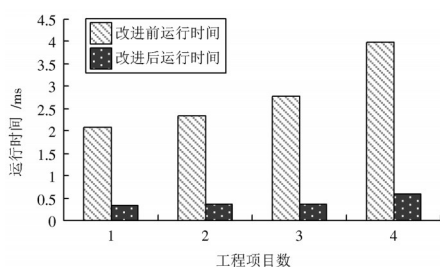


图2 ID3决策树算法改进前后运行时间对比

Fig. 2 Comparison of running time of ID3 decision tree algorithms

## 4 结论

为了企业能够更好地适应社会、市场的需求, 传统的管理方法已经很难适应环境的变化, 因此, 对企业工程项目进行针对性的管理成为一种必然趋势。本文主要通过ID3决策树算法构造企业工程项目管理决策树, 得到分类规则, 对项目的管理信息进行充分挖掘, 实现工程项目管理的个性化提示和针对化管理。

## 参考文献

- [1] 梁晨. 基于神经网络的机械设备故障检测技术[J]. 计算机与数字工程, 2015(9): 1685-1688.  
LIANG Chen. Equipment failure prediction technology based on system network [J]. Computer and digital engineering, 2015

(9): 1685-1688.

- [2] 刘菁, 赖茂宇, 孙晓丹. 基于物元分析法的施工项目管理绩效评价模型[J]. 工程管理学报, 2015(1): 124-128.  
LIU Jing, LAI Jiyu, SUN Xiaodan. The performance evaluation of construction project management model based on matter-element analysis method [J]. Journal of engineering management, 2015(1): 124-128.
- [3] 段彦斌. 电力能信工程项目管理组织优化研究[D]. 北京: 华北电力大学, 2014.  
DUAN Yanbin. The power to channel project management organization optimization research [D]. Beijing: North China Electric Power University, 2014.
- [4] 郭鹏, 文晓阁. 基于BP人工神经网络的BOT项目风险评估研究[J]. 科技管理研究, 2015, 35(21): 210-214.  
GUO Peng, WEN Xiaoge. Research of BOT project risk assessment based on BP neural network [J]. Science and technology management research, 2015, 35(21): 210-214.
- [5] 林红, 孙雅娟. 基于优化决策树的化工企业风险监测算法[J]. 计算机仿真, 2013, 30(8): 368-371.  
LIN Hong, SUN Yajuan. Chemical enterprise production steady quantitative measurement based on optimization decision tree [J]. Computer simulation, 2013, 30(8): 368-371.
- [6] 樊妍妍. ID3决策树算法对教学辅助系统的优化[J]. 齐齐哈尔大学学报(自然科学版), 2016, 32(1): 85-89.  
FAN Yanyan. Application of ID3 decision tree mining in teaching assistant system [J]. Journal of Qiqihar University (natural science edition), 2016, 32(1): 85-89.
- [7] 杨洋. 决策树ID3算法及其改进[J]. 软件导刊, 2016, 15(8): 46-48.  
YANG Yang. Decision tree ID3 algorithm and its improvement [J]. Software guide, 2016, 15(8): 46-48.
- [8] 王小巍, 蒋玉明. 决策树ID3算法的分析与改进[J]. 计算机工程与设计, 2011, 32(9): 3070-3072.  
WANG Xiaowei, JIANG Yuming. Analysis and improvement of ID3 decision tree algorithm [J]. Computer engineering and design, 2011, 32(9): 3070-3072.
- [9] 贾志刚, 贺蓉, 李仁发, 等. 一种基于决策树分类算法的家庭能量动态调度系统[J]. 计算机应用研究, 2016, 33(9): 2619-2624.  
JIA Zhigang, HE Rong, LI Renfa, et al. Dynamic scheduling system of home energy based on decision tree classification algorithm [J]. Application research of computers, 2016, 33(9): 2619-2624.
- [10] 温瑞英, 王红勇. 基于岭回归: BP神经网络的管制工作负荷预测方法[J]. 交通运输系统工程与信息, 2015, 15(1): 123-129.  
WEN Ruiying, WANG Hongyong. A forecasting method of controller's workload based on ridge regression-BP neural network [J]. Journal of transportation systems engineering and information technology, 2015, 15(1): 123-129.

作者简介: 蒲天添(1981—), 女, 湖南洪江人, 博士, 讲师, 工程师。从事建筑工程项目、国际工程项目管理研究。