

基于 SVM 的区域土地资源生态安全评价研究

张志勇, 刘希玉

ZHANG Zhi-yong, LIU Xi-yu

山东师范大学 管理与经济学院, 济南 250014

Department of Management and Economics, Shandong Normal University, Jinan 250014, China

E-mail: zyzhang_2006@sina.com.cn

ZHANG Zhi-yong, LIU Xi-yu. Research on land resource eco-security assessment based on Support Vector Machines. Computer Engineering and Applications, 2009, 45(10): 245-248.

Abstract: Support vector machine is a new machine learning technique developed on statistical learning theory. Because of its excellent learning performance, this technology has become the new research hotspot in the field of international machine learning. In this paper, the principle of support vector machine is introduced, and an integrated assessment model for regional land resource eco-security is established based on it. Test result indicates that it is feasible to apply support vector machine to assess the level of regional land resource eco-security.

Key words: eco-security; land resource; support vector machine; multi-classification

摘要: 支持向量机是在统计学习理论上发展起来的一种新的机器学习方法, 由于其出色的学习性能, 该技术已成为国际机器学习领域新的研究热点。介绍了支持向量机算法的原理, 建立了基于支持向量机的区域土地资源生态安全评价的综合评判模型。实验结果表明, 采用支持向量机对区域土地资源进行生态安全评价是可行的。

关键词: 生态安全; 土地资源; 支持向量机; 多分类

DOI: 10.3778/j.issn.1002-8331.2009.10.073 文章编号: 1002-8331(2009)10-0245-04 文献标识码: A 中图分类号: TP181

1 引言

土地资源是一种战略性资源, 对于区域生态环境和区域可持续发展具有重要意义。土地资源生态安全评价是区域可持续发展测度的一个重要内容, 也是土地资源合理配置的基础, 但目前土地资源生态安全评价研究尚属起步阶段, 评价方法也比较单一, 较多是采用综合指数法^[1]。

支持向量机(Support Vector Machine, SVM)是由 Vapnik 于 1995 年首先提出的一类新型机器学习方法, 它采用的是结构风险最小化原则, 整个求解过程转化为一个凸二次规划问题, 解是全局最优的和唯一的。SVM 能较好地解决小样本、非线性、高维数和局部极小点等实际问题, 在很大程度上解决了模型选择与过学习问题、非线性和维数灾难问题以及局部极小点等问题, 因此, 正成为继模式识别和神经网络研究之后机器学习领域新的研究热点^[2-3]。

将 SVM 用于建立区域土地资源生态安全评估模型, 应用于宁南黄土丘陵区土地资源生态安全评价, 作为土地资源生态安全评价方法的一种尝试与探索, 取得了较好的效果。

2 支持向量机

2.1 基本思想

SVM 的核心技术是构建最优超平面^[4]。其基本思想可用图 1 说明。图中实心点和空心点分别代表两类样本, H 为分类超平面, H_1 和 H_2 分别为过两类样本中距离分类超平面 H 最近的样本且平行与 H 的超平面, 它们之间的距离叫分类间隔。最优分类超平面不但能将两类正确分开, 而且使分类间隔最大, 如果 H 满足最优超平面的条件, 则 H_1, H_2 上的训练样本点就称做支持向量。

2.2 线性支持向量分类机

(1) 线性可分情况

设训练样本为 $(x_i, y_i), i=1, 2, \dots, l, x_i \in R^n, y_i \in \{-1, +1\}$ 。设两类数据可被超平面 $w \cdot x + b = 0$ 正确划分, 通过对向量系数 w 进行归一化, 所有训练样本被正确划分时, 应满足以下约束条件:

$$y_i(w \cdot x_i + b) \geq 1, i=1, 2, \dots, l \quad (1)$$

设 x_1 和 x_2 为在边界超平面 H_1 和 H_2 上的支持向量, 连接它们的向量在划分超平面法方向上的投影即是点 x_1 与最优超

基金项目: 山东省教育厅科技计划项目资助课题(No. J05G01)。

作者简介: 张志勇(1977-), 男, 硕士研究生, 主要研究领域为数据挖掘、计算智能; 刘希玉(1964-), 男, 教授, 博士生导师, 主要研究领域为数据挖掘、计算智能、计算机辅助设计等。

收稿日期: 2008-04-10 修回日期: 2008-07-10

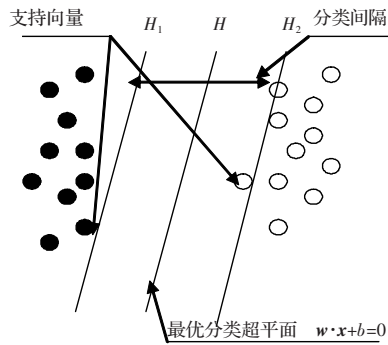


图1 线性可分情况下的最优分类超平面

平面间最大间隔的2倍。因此求最优超平面问题即是下面的约束最优化问题:

$$\begin{cases} \min \varphi(\mathbf{w}) = \frac{1}{2} \|\mathbf{w}\|^2 = \frac{1}{2} \mathbf{w} \cdot \mathbf{w} \\ \text{s.t. } y_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1 \geq 0 \quad i=1, 2, \dots, l \end{cases} \quad (2)$$

引入拉格朗日(Lagrange)函数求解:

$$L(\mathbf{w}, b, \boldsymbol{\alpha}) = \frac{1}{2} (\mathbf{w}, \mathbf{w}) - \sum_{i=1}^l \alpha_i (y_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1) \quad (3)$$

上式中 $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_l)^T$ 是 Lagrange 乘子矢量, $\alpha_i \geq 0$ 。由式(3)可得:

$$\frac{\partial L(\mathbf{w}, b, \boldsymbol{\alpha})}{\partial \mathbf{w}} = 0, \Rightarrow \mathbf{w} = \sum_{i=1}^l \alpha_i y_i \mathbf{x}_i \quad (4)$$

$$\frac{\partial L(\mathbf{w}, b, \boldsymbol{\alpha})}{\partial b} = 0, \Rightarrow \sum_{i=1}^l \alpha_i y_i = 0, \alpha_i \geq 0 \quad (5)$$

将上两式代入式(3),即知原始优化问题的对偶问题可表示为

$$\begin{cases} \max_{\boldsymbol{\alpha}} \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j (\mathbf{x}_i \cdot \mathbf{x}_j) \\ \text{s.t. } \sum_{i=1}^l \alpha_i y_i = 0, \alpha_i \geq 0 \end{cases} \quad (6)$$

设参数对 (\mathbf{w}^*, b^*) 是原始问题式(2)的解,又设 $\boldsymbol{\alpha}^* = (\alpha_1^*, \dots, \alpha_l^*)^T$ 是对偶问题的解,由式(4)可得

$$\mathbf{w}^* = \sum_{i=1}^l \alpha_i^* y_i \mathbf{x}_i \quad (7)$$

由 KKT 条件,故

$$\alpha_i^* (y_i((\mathbf{w}^* \cdot \mathbf{x}_i) + b^*) - 1) = 0, i=1, \dots, l$$

对 $\alpha_i^* > 0$, 上式意味着

$$y_j((\mathbf{w}^* \cdot \mathbf{x}_j) + b^*) - 1 = 0$$

将式(7)代入上式,得到

$$b^* = y_j - \sum_{i=1}^l y_i \alpha_i^* (\mathbf{x}_i \cdot \mathbf{x}_j) \quad (8)$$

用求出的参数得到最优分类面方程为:

$$g(\mathbf{x}) = \mathbf{w}^* \cdot \mathbf{x} + b^* = \sum_{i=1}^l \alpha_i^* y_i (\mathbf{x}_i \cdot \mathbf{x}) + b^* = 0 \quad (9)$$

基于最优分类面的分类规则可用以下指示函数表示:

$$f(\mathbf{x}) = \text{Sgn}(g(\mathbf{x})) \quad (10)$$

(2) 线性不可分情形

实际情况中,考虑到噪声或其它干扰因素造成的线性不可分情况,为了减少泛化出错的风险,就需要允许有不满足约束条件(1)的样本点存在,通过引入松弛变量 $\xi_i \geq 0, i=1, \dots, l$, 在目标函数中引入错误惩罚分量后,原始问题式(2)改为以下优

化问题:

$$\begin{cases} \min_{\mathbf{w}, b, \xi} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l \xi_i \\ \text{s.t. } y_i((\mathbf{w} \cdot \mathbf{x}_i) + b) \geq 1 - \xi_i, i=1, 2, \dots, l \\ \xi_i \geq 0, i=1, \dots, l \end{cases} \quad (11)$$

其中 $C > 0$ 是对越界样本点施加的惩罚参数,从而在算法的复杂度和样本的错分率之间进行折衷^[3]。与原始问题式(2)的解决方法一样,以上优化问题转变成形式和式(6)一样的对偶问题,只是约束条件变成

$$0 \leq \alpha_i \leq C, i=1, \dots, l \quad (12)$$

2.3 非线性支持向量分类机

非线性支持向量机的基本思想是^[4],通过事先选定的内积核函数,将输入空间中的数据非线性映射到高维特征空间,从而将输入空间中的非线性问题转化为某个高维特征空间中的线性可分问题,随后在高维特征空间构造一个最优分类超平面。通过引入核函数,将高维特征问题转化到输入空间进行计算。核函数定义如下:假设有非线性映射 $\phi(\mathbf{x}): \mathbf{x} \in R^n \rightarrow H$, 将输入空间中的样本映射到高维特征空间 H 中, H 中的内积可以定义为 $k(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j)$, $k(\mathbf{x}_i, \mathbf{x}_j)$ 称为核函数,它的选择需要满足 Mercer 条件。

采用不同的内积核函数将导致不同的支持向量机算法,目前采用较多的3类核函数包括:

(1) 多项式内积函数:

$$k(\mathbf{x}, \mathbf{y}) = [(\mathbf{x} \cdot \mathbf{y}) + 1]^l$$

(2) 径向基函数:

$$k(\mathbf{x}, \mathbf{y}) = \exp\left[-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{\sigma^2}\right]$$

(3) S 型内积函数:

$$k(\mathbf{x}, \mathbf{y}) = \tanh(v(\mathbf{x}, \mathbf{y}) + c)$$

此时,最优分类面问题的对偶问题如下:

$$\begin{cases} \max \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j k(\mathbf{x}_i, \mathbf{x}_j) \\ \text{s.t. } \sum_{i=1}^l \alpha_i y_i = 0 \\ 0 \leq \alpha_i \leq C, i=1, \dots, l \end{cases} \quad (13)$$

2.4 多类分类问题

目前多分类方法有很多,在应用方面各有利弊,本文采用一类对余类方法^[2]。它的思想是对于 $M (M \geq 2)$ 类分类问题,把类1作为一类,其余类视为一类,用支持向量机构造决策函数 $f^1 = \text{Sgn}(g^1(\mathbf{x}))$, 把类1分出来。类似的构造能把其余各个类分出来的决策函数 $f^2 = \text{Sgn}(g^2(\mathbf{x})), \dots, f^m = \text{Sgn}(g^m(\mathbf{x}))$, 按这种方法,共需要训练出 M 个 SVM 分类器。如果这些决策函数都能准确反映实际情况,那么对任意的输入 \mathbf{x} , 函数 f^1, \dots, f^m 中有且只有一个为正值。由此容易判定它的归属。然而构造的决策函数总是有误差的,可能对某些输入来说,或者没有 $f^i(\mathbf{x}), (i=1, \dots, m)$ 取正值,或者有多个 $f^i(\mathbf{x})$ 取正值。本文处理这种情况的方法是:判定 \mathbf{x} 属于 $g^i(\mathbf{x}), i=1, \dots, m$ 中取值最大者所对应的类^[2]。此方法的另一缺点是,这些两类问题是不对称的。在任一两类问题中,负类样本点显然要多于正类样本点,本文采用一种对样本加权的方法^[2]来处理这种不对称,这样最优优化问题就变为

$$\begin{cases} \min_{w,b,\xi} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \lambda_i \xi_i \\ \text{s.t. } y_i((w \cdot x_i) + b) \geq 1 - \xi_i, i=1, 2, \dots, l \\ \xi_i \geq 0, i=1, \dots, l \end{cases} \quad (14)$$

上式中, $\lambda_i \geq 1$, 对于某类较少的样本点, 取 $\lambda_i > 1$, 这样对样本点较少的正类就采用了较大的惩罚, 提高了分类的可靠度。

3 区域土地资源生态安全评价模型

3.1 土地资源生态安全评价指标体系

构建能够反映区域特点的土地资源生态安全评价指标体系, 建立科学的土地资源生态安全评价模型, 对于了解该区域土地资源生态安全状态及其演变趋势十分重要。

本文参考文献[1], 根据 20 世纪 80 年代末, 经济合作发展组织(OECD)与联合国环境规划署(UNEP)共同提出的环境指标的 P-S-R 概念模型, 即压力(Pressure)-状态(State)-响应(Response)模型, 建立区域土地资源生态安全评价指标体系及评价等级标准, 见表 1。其中压力指标反映人类活动给资源环境造成的负荷; 状态指标表征自然资源与生态系统的状况; 响应指标表征人类面临资源环境问题所采取的对策与措施^[1]。

3.2 基于 SVM 模型的实现

如表 1 所示, 区域土地资源生态安全有 5 个安全等级, 分别为安全、较安全、临界安全、较不安全和不安全。在每级标准范围内, 各随机生成 20 个样本, 5 级共生成 100 个随机样本, 目标值分别设置为 1, 2, 3, 4, 5。设待评价的指标体系中的数据为 x_{ij} , 其中 j 表示评价指标 $j=1, \dots, 19$, i 表示训练样本 $i=1, \dots, 100$, 为了消除不同指标的量纲及数值大小带来的影响, 各指标数值用下述公式进行归一化。

(1)当指标为正向指标时

$$x_{ij} = \frac{r_{ij} - \min_i r_{ij}}{\max_i r_{ij} - \min_i r_{ij}} \quad (2)$$

当指标为逆向指标时

$$x_{ij} = \frac{\max_i r_{ij} - r_{ij}}{\max_i r_{ij} - \min_i r_{ij}}$$

其中 r_{ij} 代表第 i 个样本的第 j 个指标的实际取值, r_{ij} 为它的归一化后的值。在 100 个样本中, 选择 80 个作为训练样本, 其余 20 个为测试样本。因为有 5 个安全等级, 所以这是一个 5 分类问题, 结合上面的讨论, 此实验需要建立 5 个 SVM 两分类器。在对上述 3 类核函数的多次实验中, 径向基核函数结果稍优于其它两类, 故本实验采用径向基核函数, 参数取值为 $\sigma=2$, 惩罚参数 C 为 150, 5 个 SVM 两分类器中的所有正类训练样本的权为 $\lambda=1.5$, 所有负类训练样本的权为 $\lambda=1$ 。以文献[1]中彭阳县 2004 年的评价指标数据作为待测样本。具体数据见表 2。

3.3 结果与分析

实验在 MATLAB7.0 中完成。把待测样本值分别输入训练好的 5 个 SVM 两分类器中, 得到 5 个决策函数值, 待测样本属于决策函数值最大的分类器所分出的类。上文中 5 个安全等级作为 5 个类, 分别用 1, 2, 3, 4, 5 代表。把表 2 数据作为待测样本, 得到的评价结果如表 3 所示。

如表 3 所示, 目前研究区土地生态安全水平处于临界安全级别, 系统仍不稳定, 从表 2 单指标的安全级别来看, 目前大部分指标处于临界安全水平, 有待于进一步改善, 而地表起伏度和自然灾害受灾面积比分别处于不安全和较不安全级别, 可见该区域土地生态系统本底环境非常脆弱。表中同时列出了文献[1]中物元模型对研究区 2004 年的评价结果, 比较发现, 两种模型的评价结果是完全一致的, 这说明采用 SVM 对土地资源的

表 1 黄土丘陵区土地资源生态安全评价指标及其评价等级标准

项目层	因素层	指标层	评价标准				
			1级安全	2级较安全	3级临界安全	4级较不安全	5级不安全
土地资源压力	人口压力	C1 人口密度/(人·km ²)	[0, 80]	[80, 150]	[150, 350]	[350, 500]	[500, 800]
		C2 人口增长率(%)	[0, 5]	[5, 10]	[10, 15]	[15, 20]	[20, 25]
		C3 人均耕地/(hm ² ·人 ⁻¹)	[0.25, 0.7]	[0.1, 0.25]	[0.07, 0.1]	[0.05, 0.07]	[0.03, 0.05]
	环境压力	C4 单位面积耕地化肥负荷/(kg·hm ⁻²)	[1, 255]	[255, 450]	[450, 825]	[825, 1 200]	[1 200, 1 500]
		C5 单位面积耕地农药负荷/(kg·hm ⁻²)	[0, 150]	[150, 300]	[300, 450]	[450, 600]	[600, 750]
		C6 地表起伏度	[0, 0.2]	[0.2, 0.3]	[0.3, 0.4]	[0.4, 0.5]	[0.5, 0.7]
		C7 自然灾害受灾面积比例/(%)	[0, 5]	[5, 8]	[8, 10]	[10, 20]	[20, 30]
社会经济压力	C8 经济密度/(万元·km ²)	[35, 50]	[25, 35]	[15, 25]	[3, 15]	[1, 3]	
	C9 城镇化率/(%)	[15, 30]	[10, 15]	[7, 10]	[3, 7]	[1, 3]	
土地资源状态	土地质量	C10 耕地粮食单产/(kg·hm ⁻²)	[2 250, 4 500]	[1 800, 2 250]	[900, 1 800]	[450, 900]	[150, 450]
		C11 土壤侵蚀模数/(t·km ⁻² ·a ⁻¹)	[1 000, 2 000]	[2 000, 3 000]	[3 000, 6 400]	[6 400, 7 400]	[7 400, 10 000]
	土地利用结构	C12 水土流失面积比例/(%)	[5, 10]	[10, 20]	[20, 50]	[50, 90]	[90, 100]
		C13 森林覆盖率/(%)	[40, 60]	[30, 40]	[15, 30]	[5, 15]	[0, 5]
人文环境影响	调整压力	C14 25°以上坡耕地面积比/(%)	[0, 2]	[2, 5]	[5, 10]	[10, 15]	[15, 30]
		C15 有效灌溉面积比/(%)	[15, 50]	[8, 15]	[6, 8]	[4, 6]	[0, 4]
		C16 环保治理投资占 GDP 比例/(%)	[3, 5]	[1.5, 3]	[1, 1.5]	[0.1, 1]	[0, 0.1]
	改善状态	C17 人均 GDP/元	[3 500, 5 000]	[2 500, 3 500]	[1 500, 2 500]	[300, 1 500]	[100, 300]
		C18 第三产业占国民生产总值比重/(%)	[45, 50]	[35, 45]	[25, 35]	[15, 25]	[1, 15]
		C19 农民人均纯收入/元	[2 500, 3 000]	[2 000, 2 500]	[1 000, 2 000]	[250, 1 000]	[100, 250]

表 2 2004 年彭阳县土地资源生态安全评价指标值

C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19
100	13.63	0.27	633.19	131.87	0.54	10.26	22.44	7.95	1 705.79	3 600	27.8	16.6	1.78	7.02	1.2	2 241.98	27.87	1 518.82

表3 单指标及综合指标的各个决策函数值及评价结果

指标	决策函数值					安全等级	
	决策函数 1	决策函数 2	决策函数 3	决策函数 4	决策函数 5	SVM	物元模型
C1	-0.358 4	-0.224 2	-0.330 0	-1.134 9	-5.981 9	2	2
C2	-4.374 4	-0.339 6	-0.160 1	-0.246 4	-3.470 5	3	3
C3	0.048 2	-0.026 5	-0.250 0	-0.237 6	-6.340 6	1	1
C4	-4.429 4	-0.260 7	-0.055 0	-0.366 3	-4.797 2	3	3
C5	0.604 2	-0.234 1	-0.721 6	-1.316 2	-8.740 7	1	1
C6	-7.372 0	-0.729 8	-0.369 4	-0.208 6	0.618 6	5	5
C7	-2.976 7	-0.248 0	-0.226 4	-0.212 8	-2.951 4	4	4
C8	-3.171 5	-0.244 3	-0.129 5	-0.317 0	-5.463 1	3	3
C9	-3.305 6	-0.404 2	-0.216 8	-0.232 3	-2.816 4	3	3
C10	-2.518 2	-0.148 7	-0.144 0	-0.366 7	-6.135 3	3	3
C11	-3.247 3	-0.194 5	0.006 2	-0.622 0	-7.993 1	3	3
C12	-3.094 4	-0.214 8	-0.091 8	-0.901 2	-6.379 2	3	3
C13	-5.270 2	-0.586 5	-0.184 5	-0.203 4	-2.843 3	3	3
C14	0.155 4	-0.230 7	-0.317 0	-0.958 1	-6.027 0	1	1
C15	-1.791 2	-0.252 7	-0.229 9	-0.231 0	-1.240 9	3	3
C16	-3.582 4	-0.417 1	-0.197 0	-0.233 8	-3.059 4	3	3
C17	-3.543 9	-0.221 4	-0.139 4	-0.328 9	-5.448 0	3	3
C18	-4.771 8	-0.308 2	-0.169 2	-0.262 4	-3.435 5	3	3
C19	-4.680 3	-0.293 1	-0.014 5	-0.295 4	-4.913 4	3	3
综合	-2.358 1	0.400 8	0.799 1	-2.196 3	-2.937 0	3	3

生态安全等级进行评价是完全可行的,并且因为在评价过程中,SVM模型不需要人为的对指标数据分配权重,所以评价结果更具客观性。

4 结束语

SVM是一种基于小样本学习理论的通用学习算法,具有严格的理论基础,能较好地解决小样本、非线性、高维数和局部极小点等实际问题。本文提出了一种有效的、基于SVM的区域土地资源生态安全评估模型,案例研究证明了利用该模型进行区域土地资源生态安全评价是可行的,可作为土地资源生态安全评价研究的新方法。目前,SVM在其它领域的应用及参数选择方法的研讨正逐渐成为大家关注的课题。

(上接 244 页)

(3)如果如果在网时长大于等于 7,通话时长变化率大于等于-0.5 的用户,那么其流失概率是 2.8%;

(4)如果在网时长大于等于 7,通话时长变化率小于-0.5,并且年龄小于等于 30 岁的用户,那么其流失概率为 0%;

(5)如果在网时长大于等于 7,通话时长变化率小于-0.5,并且年龄大于 30 岁,主叫通话变化率在-0.5 和 0.5 之间的用户,那么其流失概率为 10%;

(6)如果在网时长大于等于 7,通话时长变化率小于-0.5,并且年龄大于 30 岁,主叫通话变化率小于-0.5,长途占比等于 0 的用户,那么其流失概率为 20%;

(7)如果在网时长大于等于 7,通话时长变化率小于-0.5,并且年龄大于 30 岁,主叫通话变化率小于-0.5,长途占比大于 0 的用户,那么其流失概率为 0,即该类用户 100%不会流失。

5 结束语

本文运用新的信息增益度计算算法对 ID3 决策树算法进

参考文献:

- [1] 张虹波,刘黎明,张军连,等.区域土地资源生态安全评价的物元模型构建及应用[J].浙江大学学报:农业与生命科学版,2007,33(2): 222-229.
- [2] 邓乃扬,田英杰.数据挖掘中的新方法——支持向量机[M].北京:科学出版社,2004.
- [3] 梅建新,段汕,潘继斌,等.支持向量机在小样本识别中的应用[J].武汉大学学报:理学版,2002,48(6):732-736.
- [4] 李柞泳,汪嘉杨,熊建秋,等.可持续发展评价模型与应用[M].北京:科学出版社,2007.
- [5] 徐红敏.基于支持向量机理论的水环境质量预测与评价方法研究[D].长春:吉林大学,2007.

行了改进,并将该算法应用于某移动通信公司的客户流失预测(二元决策)。结果表明:改进算法不仅能更快速、更高效地建立模型,而且可以提供更高的检测率和更低的误检率;通过熵函数的递推性质可以将该算法推广到多元决策情况。

参考文献:

- [1] Quinlan J R. Induction of decision tree[J]. Machine Learning, 1986, 1(1): 81-106.
- [2] 段云峰.数据仓库及其在电信领域中的应用[M].北京:电子工业出版社,2003.
- [3] 叶进,程泽凯,林士敏.基于贝叶斯网络的电信客户流失预测分析[J].计算机工程与应用,2005,41(14):212-214.
- [4] 夏火松.数据仓库与数据挖掘技术[M].北京:科学出版社,2004.
- [5] Han J W, Kamber M.数据挖掘概念与技术[M].北京:机械工业出版社,2001.
- [6] 朱浩刚.基于数据挖掘的移动通讯业客户流失管理[J].计算机工程与应用,2004,40(1):215-219.