

# 基于粗糙集和 SVM 的工程项目投标风险研究

李顺国<sup>1</sup>, 卢新元<sup>2</sup>

LI Shun-guo<sup>1</sup>, LU Xin-yuan<sup>2</sup>

1. 武汉理工大学 土木工程与建筑学院, 武汉 430070

2. 华中师范大学 信息管理系, 武汉 430074

1. School of Civil Engineering and Architecture, Wuhan University of Technology, Wuhan 430070, China

2. Department of Information Management, Huazhong Normal University, Wuhan 430074, China

**LI Shun-guo, LU Xin-yuan. Research of constructional engineering project bidding risk analysis based on rough sets and support vector machine. Computer Engineering and Applications, 2008, 44(17): 224-227.**

**Abstract:** Aimed at the uncertainty and risk analysis in project bidding, the knowledge expression system of bidding risk decision analysis is described. The reduction of attributes of the bidding risk decision data with rough set theory, and multi-classification of data with Least Squares SVM are discussed.

**Key words:** constructional engineering project; bidding risk; rough set; Support Vector Machine

**摘要:** 针对建筑工程项目投标过程中的不确定性和风险分析问题, 运用 Rough 集理论和方法对投标风险决策表的属性进行约简, 再采用支持向量机对决策数据进行风险分类, 帮助决策者迅速对工程项目的投标风险进行评估和预测。

**关键词:** 建筑工程项目; 投标风险; Rough 集; 支持向量机

**DOI:** 10.3778/j.issn.1002-8331.2008.17.067 **文章编号:** 1002-8331(2008)17-0224-04 **文献标识码:** A **中图分类号:** TP18

## 1 引言

现代建筑工程项目大多规模大、技术难度高、建设周期长、参加单位多、环境复杂, 在招标投标过程中, 将涉及到多种类型风险, 决策者需要通过对现有的不确定性风险信息进行有效分析, 产生合理的决策方案, 从而更科学、高效地做到招投标风险的规避。目前已有学者运用模糊理论、灰色系统理论<sup>[1]</sup>、粗糙集 (Rough 集)<sup>[2]</sup>等对投标风险进行了定量分析。由于企业工程项目投标风险数据缺乏记录, 而可辨识的风险因素较多, 在投标风险定量研究中出现了样本数据少、样本数据维数大的问题。针对这一问题, 提出基于 Rough 集和支持向量机 (SVM) 的分析方法, 以进一步提高决策精度。

论文将从 Rough 集和 SVM 理论的基本体系出发, 讨论如何用 Rough 集和 SVM 理论和方法, 建立投标风险的决策模型, 用于考察同类工程不同投标方案 (公司) 的风险, 提高投标风险决策的科学性。

## 2 投标风险特性描述

构成工程项目风险的因素大致可以分为系统内部因素和外部因素两类。内部因素指投标企业在标书中所提及的有关技术和商务标内容如工程造价、工期、技术方案、质量标准及施工机械的配套能力等; 外部因素是指建筑工程项目所在地的相关政策、法律环境及相关的经济、技术环境。结合成功者以

往的经验及当前实际<sup>[3]</sup>, 一般工程项目投标风险体系归纳如图 1 所示。

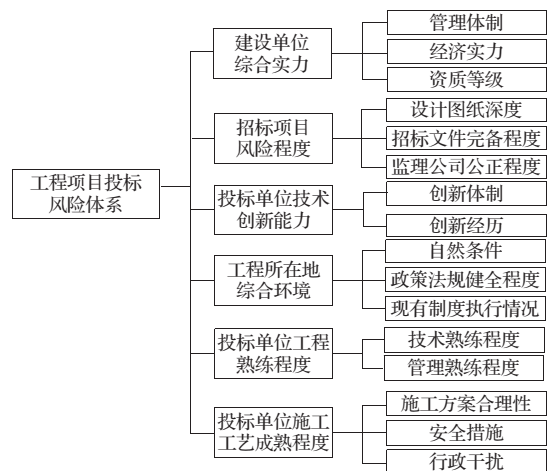


图 1 工程项目投标风险体系

体系中的六类指标均为定性指标, 需要根据实际情况, 由专家评价得到量化值。

## 3 粗糙集的基本原理

在粗糙集理论中, 知识的表达是通过信息系统 (决策表)  $S = \{U, R, V, f\}$  来实现的。如表 1 所示。

表1 基于决策表的投标风险知识表达系统

$U$	条件属性 1	条件属性 2	...	条件属性 $n$	决策属性
对象 1	$Obj1-atr1-value$	$Obj1-atr2-value$	...	$Obj1-atrin-value$	$Obj1-decision-value$
对象 2	$Obj2-atr1-value$	$Obj2-atr2-value$	...	$Obj2-atrin-value$	$Obj2-decision-value$
...	...	...	...	...	...
对象 $m$	$Objm-atr1-value$	$Objm-atr2-value$	...	$Objm-atrin-value$	$Objm-decision-value$

(1)论域 $U=\{x_1, x_2, \dots, x_m\}$ 为有限数量的对象集。

(2) $R$  为有限数量的投标风险属性集,由条件属性集合  $C$  和决策属性集合  $D$  组成,  $C \cup D = R$ , 并且  $C \cap D = \emptyset$ , 不失一般性, 设  $D = \{d\}$ 。

(3) $V = \bigcup_{a \in R} V_a$ ,  $V_a$  为属性  $a \in R$  的值域;  $f: U \rightarrow V_a$  为一单射,

使论域  $U$  中的任一对象取属性  $a$  在  $V_a$  中有唯一确定值。

在决策表中,不同的条件属性具有不同的重要程度,一些属性提供了丰富的信息,对决策起着至关重要的作用,而其他的一些属性对决策的影响力相对较弱。为此,对属性进行重要性估计必不可少,而属性依赖度可以作为度量标准来分析每一个属性的重要性。

决策属性的重要程度  $SGF(C, D, c_i)$  的计算方法可参见文献[2],在此不予赘述。 $SGF(C, D, c_i)$  越大,表明当从条件属性中去掉属性  $c_i$  以后再对论域中对象进行分类时,分类  $P(ID)$  的正域所受影响越大,则条件属性  $1 \leq i \leq s, 1 \leq j \leq t$  对投标决策越重要。

设阈值  $\theta$  为最小决策属性重要度,即  $SGF(C, D, c_i) < \theta$  时,认为即使去掉属性  $c_i$ ,也不会对投标决策产生显著影响。因此,可忽略样本数据中属性  $c_i$  对应的风险因素值,从而在保证正确的决策能力的同时达到对样本数据进行约简的目的。

## 4 SVM 多值分类模型

(1)SVM 基本理论

SVM 是从线性可分情况下的最优分类超平面发展而来的。对两类分类问题,设训练样本集为  $(x_i, y_i) (i=1, 2, \dots, n)$ , 其中  $n$  为训练样本的个数,  $x \in R^d$  为训练样本,  $y_i \in \{1, -1\}$  是输入样本  $x_i$  的类标记(期望输出)。SVM 算法的出发点是寻找最优分类超平面。最优分类超平面不但能将所有样本正确分开(训练错分率为 0),而且能够使两类间的边际(margin)最大,边际定义为训练数据集到该分类超平面的最小距离之和。最优分类超平面意味着对测试数据平均分类误差最小。

若样本集线性可分,  $d$  维空间中线性判别函数  $g(x) = \omega \cdot x + b$ , 分类面方程为  $\omega \cdot x + b = 0$ 。将判别函数归一化,使离分类面最近样本的  $|g(x)| = 1$ 。若分类面对所有样本都能够正确分类,则满足:

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

分类超平面  $H_0: \omega \cdot x_1 + b = 0$  为最优当且仅当  $(\omega, b)$  是下面优化问题的解:

$$\begin{cases} \min: \frac{\|w\|^2}{2} \\ \text{subject to: } y_i (\omega \cdot x_i + b) \geq 1 \end{cases}$$

用 Lagrange 乘子法解上述 QP 问题,等价于解对偶问题:

$$\begin{cases} \max: L_D = \sum_{i=1}^n \alpha_i - \frac{1}{2} \|w\|^2 - \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j x_i \cdot x_j \\ \text{subject to: } \sum_{i=1}^n \alpha_i y_i = 0, y_i \geq 0 \end{cases} \quad (2)$$

对每个训练样本  $x_i$ , 都有一个 Lagrange 乘子  $\alpha_i$ 。解  $\alpha_i > 0$  所对应的  $x_i$  称为支持向量,它满足式(1)中的等式约束。支持向量距最优超平面最近,通常只是全体样本中的很少一部分,是对分割两类非常重要的样本点。

若  $\alpha_i$  为最优解,则:

$$\omega = \sum_{i=1}^n \alpha_i^* y_i x_i \quad (3)$$

所以得到的分类决策函数为:

$$g(x) = \text{sgn}(w^* \cdot x + b^*) = \text{sgn}\left(\sum_{i \in SV} \alpha_i^* y_i x_i \cdot x + b^*\right) \quad (4)$$

其中  $x$  为待分类样本,  $SV$  为支持向量集,  $b^*$  为分类阈值,可用任一支持向量求得。

(2)利用最小二乘 SVM 构造多元分类器

以 Suykens 在文献[4]中提出了最小二乘 SVM(LS-SVM),而文献[5]以 LV-SVM 为基础,根据多元分类的特点采用以类别为划分标准的新样本集形式:  $(x_i, y_{ij})$ 。其中:  $i=1, \dots, n; j=1, \dots, L$  ( $L$  是样本集所包含类别总数);  $y_{ij}$  表示第  $i$  个样本属于第  $j$  类。根据新样本集的表示,目标函数可以描述为:

$$\text{Min} \Phi(\omega_j, \xi_i, j) = \frac{1}{2} \sum_{j=1}^L \omega_j \cdot \omega_j + \frac{1}{2} C \sum_{i=1}^n \sum_{j=1}^L \xi_{i,j}^2 \quad (5)$$

其约束函数则为:

$$y_i [\omega_i \cdot \phi(x_i) + b_i] + \xi_{i,i} = 1 \quad (6)$$

式中:  $i=1, \dots, n; j=1, \dots, L$ 。

构造拉格朗日函数为:

$$\begin{aligned} L^{(m)}(\omega_j, b_j, \xi_{i,j}; a_{i,j}) = \\ \Phi^{(m)}(\omega_j, b_j, \xi_{i,j}) - \sum_{i,j} a_{i,j} \{y_i [\omega_j \cdot \phi(x_i) + b_j] - 1 + \xi_{i,j}\} \end{aligned} \quad (7)$$

对式(7)中的  $\omega_j, b_j, \xi_{i,j}, a_{i,j}$  分别求导,与二元分类时的情况相类似,式(9)也可以转化为矩阵形式求解,可得:

$$\begin{bmatrix} 0 & Y_M \\ Y_M^T & \Gamma_M \end{bmatrix} \times \begin{bmatrix} b_M \\ a_M \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad (8)$$

其中:  $Y_M$  和  $\Gamma_M$  都是对角阵,  $Y_M = \{[y_1^{(1)}, \dots, y_n^{(1)T}], \dots, [y_1^{(L)}, \dots, y_n^{(L)T}]\}^T$ ; 而  $\Gamma_M = [\Gamma^{(1)}, \dots, \Gamma^{(1)T}]$ ;  $a_M = [a_{(1,1)}, \dots, a_{(n,1)}, \dots, a_{(1,L)}, a_{(n,L)}]$ ;  $b_M = [b_1, \dots, b_L]$ 。

$$\Gamma_{m,n}^{(j)} = y_m^{(j)} y_n^{(j)} K_j(x_m, x_n) + C^{-1} I \quad (9)$$

由式(8)、(9)可得出多元分类最小 SVM 的分类决策函数为:

$$F(x) = \text{sgn}\left[\sum_{i=1}^n a_{i,j} y_i^{(j)} K_j(x_i, x_j) + b_j\right] \quad (10)$$

## 5 投标风险分析方法的实现

投标风险分析系统的学习样本是由属性集(风险条件属性集和风险类别属性集)组成的,其决策是根据风险类别制定的,因而系统主要的工作是正确识别风险的类别。系统首先利用 Rough 集对训练集样本进行约简,分辨出属性重要度小于阈值

$\theta$  的样本属性,将对应的风险条件属性从属性集中删去,从而获得了由最少属性构成的训练集样本。然后将样本输入 SVM 进行分类。重复上述过程直到能全部正确划分所有训练集样本为止。算法如下:

**步骤 1** 输入训练样本数据,输出决策表  $S=\{U,R,V,f\}$ ,构建知识系统  $K$ 。

input 训练样本数据集  $V$

论域  $U, u=1, U[0]=u$

决策表  $S(S[i]$  表示  $S$  的第  $i$  条记录)

foreach(record in  $V$ )

{ $S[i]$  条件属性  $c_i$ =风险因素  $v_i$  专家评估值

$S[i]$  决策属性  $d$ =风险类别  $d$

$u++$ ,  $U[i]=u$

$S[i]=L(S[i],LT)/L$  为离散化和归一化函数,  $LT$  为离散和归一规则表

}

output  $S$

**步骤 2** 输入决策表  $S$ , 计算属性重要度, 输出可删去条件属性的索引值。

input  $S$

属性重要度  $AI=SGF(S)/SGF$  为属性重要度计算函数

初始化待删去条件属性数组  $DEL$  为 1

foreach(record in  $AI$ )

if( $AI[i]<\theta$ ) $DEL[i]=0$

output  $DEL$

**步骤 3** 输入可删去条件属性的索引值, 删去训练集中对应的风险因素值, 输出约简后的训练集  $V'$

input  $DEL$

for each(record in  $V$ )

if( $DEL[i]$ )  $v_i=v_{i+1}$

$V'=Retrieval(V)/Retrieval$  是冗余函数, 即出现相同的记录只保留一条

output  $V'$

**步骤 4** 将  $V'$  输入 LS-SVM, 输出模型  $M$

$M=LS-SVM(V')$

**步骤 5** 输入测试数据  $V_t$ , 输出分类结果向量  $Decision$ , 计算正确率

$Decision=LS-SVM(V_t, M)$

$Accuracy=Decision/D//D$  是类别向量

## 6 实例分析

本文的数据均来自于第十五冶金建筑某公司, 风险因素 a-f 分别对应工程项目投标风险体系的六类指标。取训练样本数据和测试样本数据各 200 条, 限于篇幅, 文中只列出前 10 条训练样本数据, 如表 2 所示。

建立基于粗糙集的投标风险决策分析的知识表达系统  $S=\{U,R,V,f\}$ , 其中, 论域  $U$  设为投标方式集合; 条件属性的集合  $C=(a,b,c,d,e,f)$ ; 决策属性的集合  $D=\{1,2,3\}$ , “1”表示一类投标风险, “2”表示二类投标风险, “3”表示三类投标风险。对样本数据进行离散化和归一化后得到决策表如表 3 所示。

接下来, 系统计算出条件属性的重要性如表所示, 计算方法见文献[2]。

表 2 工程项目投标风险决策表

投标方式	风险因素						风险类别
	a	b	c	d	e	f	
1	0.85	0.18	8	1	3	1	1
2	0.62	0.33	11	1	2	1	1
3	0.57	1.68	10	1	3	1	1
4	2.65	1.85	4	1	7	1	2
5	0.54	1.16	5	1	4	1	2
6	2.68	0.46	4	1	7	1	2
7	0.47	0.94	5	0	4	1	3
8	0.59	0.36	6	0	3	0	3
9	0.88	0.75	3	0	5	0	3
10	0.36	1.44	7	0	6	0	3

表 3 预处理后的工程项目投标风险决策表

投标方式	风险因素						风险类别
	a	b	c	d	e	f	
1	2	1	2	2	1	1	1
2	1	1	2	2	1	1	1
3	1	3	2	2	2	1	1
4	2	2	2	2	1	1	1
5	3	3	2	2	1	1	1
6	2	3	1	2	2	1	2
7	1	2	1	2	1	1	2
8	1	3	1	2	2	1	2
9	2	2	1	2	1	1	2
10	2	1	1	2	2	1	2

表 4 投标风险决策表条件属性重要性表

属性	a	b	c	d	e	f
SGF	0.231	0.077	0.269	0.269	0.154	0

根据专家建议, 取最小决策属性重要度  $\theta=0.1$ , 则认为属性 b 和属性 f 不会对决策产生显著影响。从训练样本中去掉 b 和 f 对应的数据维, 将约简后的样本数据如表 5 所示。

表 5 预处理后的工程项目投标风险决策表

$U$	条件属性				决策属性
	a	c	d	e	
1	2	2	2	1	1
2	1	2	2	1	1
3	1	2	2	2	1
4	2	2	2	1	1
5	3	2	2	1	1
6	2	1	2	2	2
7	1	1	2	1	2
8	1	1	2	2	2
9	2	1	2	1	2
10	2	1	2	2	2

将表 4 输入 LS-SVM 进行训练得到模型, 然后输入测试数据, 结果如表 5 所示。

由表可见, 只有第 6 条记录的决策结果出现了错误。

同时将此方法与直接采用 LS-SVM 的方法进行了对比试验, 试验结果如表 7 所示。

## 7 总结与展望

由表 7 可见, 基于 Rough 集和 SVM 的工程项目投标风险分析方法较仅基于 SVM 方法来说, 准确率有了显著提高。由于 Rough 集无法直接对连续的数据进行处理而必须进行离散化

表 6 工程项目投标风险决策测试结果表

投标方式	风险因素						风险类别	人工分类
	a	b	c	d	e	f		
1	0.9	0.23	9	1	4	1	1	1
2	0.67	0.38	12	1	3	1	1	1
3	0.62	1.73	11	1	4	1	1	1
4	2.70	1.90	5	1	8	1	2	2
5	0.59	1.21	6	1	5	1	2	2
6	0.29	1.87	4	1	9	1	2	3
7	0.52	0.99	6	0	5	1	3	3
8	0.64	0.41	7	0	4	0	3	3
9	0.93	0.8	4	0	6	0	3	3
10	0.41	1.49	8	0	7	0	3	3

表 7 风险分类试验结果对比表

方法	LS-SVM/%	Rough Set & LS-SVM/%
准确率	90.00	97.50
回归率	96.52	87.75

和归一化。而离散过程中分割点集的选取不当,可能增加离散数据的不确定性。同时 Rough 集对数据的变化较为敏感,微小的数据改动就可能产生不同的约简决策表<sup>[6]</sup>,得到不同的属性重要度,导致决策精度的降低;而采用仅基于 SVM 的方法,虽然避免数据改动带来的不确定性,但是无法得到风险因素对决

(上接 205 页)

- [45] Kelly F, Steinberg R. A combinatorial auctions with multiple winners for universal service[J]. Management Science, 2000, 46(4): 586-596.
- [46] Jones J L, Koehler G J. Combinatorial auctions using rule-based bids[J]. Decision Support Systems, 2002, 34: 59-74.
- [47] 黄宛宁, 龚建伟, 王鹏辉. 基于改进遗传算法的多机器人任务分配方法[J]. 计算机仿真, 2006, 23(11): 164-167.
- [48] Dias M B. TraderBots: a new paradigm for robust and efficient multirobot coordination in dynamic environment[D]. Robotics Institute, Carnegie Mellon University, 2004.
- [49] Stentz A, Dias M B, Zlot R, et al. Market-based approaches for coordination of multi-robot teams at different granularities of interaction [C]//Proceedings of Conference on Robotics and Remote

(上接 223 页)

表 1 是采用直接 Zernike 矩多目标识别、基于圆投影的 Zernike 矩多目标识别方法<sup>[2]</sup>和基于弧半径投影的图像识别方法的对 3 种不同大小模板的识别时间的比较。从表 1 数据可以看出基于弧半径投影的图像识别方法在识别速度上有了很大的提升。

表 1 三种方法识别速度比较

方法	模板大小		
	43×43	51×51	81×81
Zernike 矩	35.3	46.76	67.54
基于圆投影 Zernike 矩	3.64	4.53	5.59
基于弧半径投影	2.17	2.34	3.38

## 5 总结

本文针对工业 CT 图像识别,提出了一种基于弧半径投影的多目标图像识别方法,该方法通过求取多个工件所在的扇环带区域,然后利用 Zernike 矩进行精确识别。该方法可以一次识

别多个工件对象,略过无目标区域的识别,减少了识别时间,且具有较高的识别率。通过实验结果表明,本文的方法可有效地降低识别耗时。

## 参考文献:

- [1] 齐锡晶, 刘莉, 王树伟. 施工企业投标风险评价[J]. 沈阳建筑工程学院学报, 1999, 15(4): 378-382.
- [2] 刘清. Rough 集及 Rough 推理[M]. 北京: 科学出版社, 2001.
- [3] Vapnik V. Statistical Learning Theory[M]. New York: John Wiley, 1998.
- [4] Suykens J A K, Vandewalle J. Least squares support vector machine classifiers[J]. Neural Processing Letters, 1999, 9(3): 293-300.
- [5] Suyken J A K, Lusas L, Van D P, et al. Least squares support vector machine classifiers: a large scale algorithm[C]//In Proceedings of the European Conference on Circuit Theory and Design (ECCTD 99). Italy: Stresa, 1999.
- [6] Banerjee M, Chakraborty M K. A category for rough sets[J]. Foundations of Computing and Decision Sciences, 1993, 18(3-4): 167-180.
- [7] Crammer K, Singer Y. Ultraconservative online algorithms for multi-class problems[J]. Journal of Machine Learning Research, 2003, 3: 951-991.
- Systems, Gainesville, FL, United State: American Nuclear Society, LaGrange Park, United States, 2004: 145-151.
- [50] Zlot R, Stentz A. Complex task allocation for multiple robots[C]//Proceedings of the 2005 IEEE International Conference on Robotics and Automation, 2005: 1515-1522.
- [51] Rothkopf M H, Pekec A, Harstad R M. Computationally manageable combinatorial auctions[J]. Management Science, 1998, 44(8).
- [52] 宋梅萍, 顾国昌, 张汝波. 多移动机器人协作任务的分布式控制系统[J]. 机器人, 2003, 25(5): 456-460.
- [53] 王进戈, 王强. 基于模糊评判的机器人足球比赛策略研究[J]. 哈尔滨工业大学学报, 2005, 37(7): 943-946.
- [54] 原魁, 李园, 房立新. 多移动机器人系统研究发展近况[J]. 自动化学报, 2007, 33(8): 785-794.

## 参考文献:

- [1] 郭志平. 工业 CT 技术发展概况[J]. 无损检测, 1996, 18(1): 27-30.
- [2] Choi M S. A novel two stage template matching method for rotation and illumination invariance[J]. Pattern Recognition, 2002, 35: 119-129.
- [3] Khoanzad A, Hong Y H. Invariant image recognition by zernike moments[J]. IEEE Trans Pattern Anal Machine Intell, 1990, 12(5): 487-497.
- [4] 沈霄凤. Zernike 矩与应用[J]. 华东师范大学学报, 1997, 1: 104-107.
- [5] 金敏, 徐守时, 汪行. 不变矩在模式识别中的应用研究[J]. 计算机工程与应用, 2004, 40(25): 65-67.
- [6] 姚军, 蒋晓瑜, 黄应清. 目标识别中 Hu 矩、Zernike 矩和小波矩的比较[J]. 装甲兵工程学院学报, 2006(3): 9-12.
- [7] 何斌, 马天予, 王运坚. Visual C++ 数字图像处理[M]. 2 版. 北京: 人民邮电出版社, 2002.