

# 基于属性约简—光滑支持向量机的中小企业信息化评价研究

戴 强

(太极计算机股份有限公司, 北京 100083)

✉daiqiang@taiji.com.cn



**摘 要:** 本文以企业信息化的软硬件保障水平、信息化组织水平、信息技术应用与盈利水平、信息化能力水平四个角度建立中小企业信息化评价体系, 应用粗集的属性约简理论对指标进行约简, 同时构建了一个二阶分段光滑函数支持向量机模型, 利用光滑支持向量机对企业的信息化水平进行了评价, 最后, 通过具体实例进行实证分析, 以证明方法的有效性。

**关键词:** 企业信息化; 光滑支持向量机; 属性约简

**中图分类号:** TP301 **文献标识码:** A

## Research on Information Evaluation of Small and Medium Sized Enterprises based on Attribute Reduction Smooth Support Vector Machine

DAI Qiang

(Taiji Computer Corporation Limited, Beijing 100083, China)

✉daiqiang@taiji.com.cn

**Abstract:** In this research, the information evaluation system for SMEs is established from the four perspectives of enterprise informationization: software and hardware security level, information organization level, information technology application and profit level, and the informationization level. Through applying the attribute reduction theory of rough set to the index and meanwhile constructing the second-order piecewise smooth function support vector machine model, the informationization level of the enterprises is evaluated with the smooth support vector machine. Finally, the method is proved to be effective through the empirical analysis of the real cases.

**Keywords:** enterprise informationization; smooth support vector machine; attribute reduction

### 1 引言(Introduction)

随着大数据、边缘计算、物联网、区块链等技术的快速发展, 企业的信息化数据呈几何趋势增长, 根据互联网数据中心预测, 2020年全球企业信息化的数据量将达到35ZB<sup>[1]</sup>。2016年大数据产业峰会4月27日到28日举行<sup>[2]</sup>, 在工业和信息化部指导下, 中国信息通信研究院和数据中心联盟面向企业和政府部门征集大数据在政府及相关部门应用的优秀技术成果和案例, 形成政府大数据应用领域的优秀案例汇编。总之, 在大数据时代背景下, 中小企业信息化水平影响着中小企业的健康发展, 然而企业的信息化建设需要投入大量的前期成本, 信息化投入的成本能否给中小企业带来效益还是未知<sup>[3]</sup>, 因此对中小企业信息化水平进行有效地评估显得尤为重要。

支持向量机(Support Vector Machine, SVM)是人工智能算法一个较新的研究成果<sup>[4-7]</sup>, 其应用的领域越来越广泛, SVM的优点是在处理小样本数据时克服了神经网络的过拟合、易入局部极小的问题<sup>[8-10]</sup>。然而SVM在数据输入训练时, 其分类和预测的光滑性略显不足, 同时其不能智能的确定输入的训练数据的冗余性, 这样就增加了计算运行的复杂度, 而粗糙集的属性约简可以有效地降低输入数据中的噪声<sup>[11]</sup>, 降低输入数据的冗余性。因此, 构造一个二阶分段光滑函数, 增强支持向量机的非凸分类和预测性能, 将光滑支持向量机智能算法与粗糙集属性约简联合是互补的, 将属性约简后的训练指标数据输入到光滑支持向量机中进行训练学习, 然后对测试样本进行预测或评价, 这样在消除冗余信息减少光滑支持向量机的学习与预测复杂度的同时, 提高了光滑支持向

量机的精度。因此，本文应用属性约简—光滑支持向量机算法对中小企业信息化水平评价进行了研究。

## 2 属性约简—光滑支持向量机模型构建 (Attribute reduction-construction of smooth support vector machine model)

### 2.1 模型构建描述

本文的属性约简—光滑支持向量机联合评价模型主要是应用在中小企业信息化水平评价上，因此构建模型的基本思路是：第一步应用粗糙集的属性约简剔除光滑支持向量机输入的指标数据进行有效地去噪处理，将约简得到的核心属性指标数据分为训练样本和预测样本送进光滑支持向量机进行学习与预测，进而进行有效地评价。

### 2.2 属性约简

属性约简是粗糙集理论的主要核心思想，其在处理不确定或者不精确问题上有较强的优势。假设有一个知识表达系统可以表示为  $S = \langle U, C, D, V, f \rangle$ <sup>[12]</sup>，在表达系统中，决策表用  $T = (U, A = CUD)$  表示，而属性约简主要是针对整个属性集  $A$  进行的，而约简的只是条件属性集。属性约简主要分为两步，分别是属性集约简和属性值约简<sup>[13-15]</sup>。

第一步：属性集约简

假如  $P \subseteq C$ ，存在  $P$  关于  $D$  独立的，同时  $Pos_P(D) = Pos_C(D)$ ，那么  $P$  可以成为是  $C$  的  $D$  约简。在  $C$  中的所有  $D$  约简后的交  $\cap RED_C(D)$  被称为  $C$  的核，记为  $Core_D(C)$ 。

第二步：属性值约简

假设一致性决策表为  $T = (U, CUD)$ ，同时  $P \subseteq C$  是  $C$  的  $D$  约简。那么值约简是针对  $P$  来说的，也可以理解为属性值约简是针对决策规则来约简的。决策表中的条件属性值可以被约去，但是必须保证约简掉该属性值后，仍然保持该条规则的一致性<sup>[15]</sup>。

### 2.3 构建光滑支持向量机

标准的SVM算法如下<sup>[16]</sup>

$$\begin{aligned} \min_{w,b,C} \|w\|^2 / 2 + Ce^T y \\ \text{s.t. } D \cdot (Aw + eb) \geq e, y \geq 0 \end{aligned} \quad (1)$$

2005年，台湾大学的Lee等人<sup>[17]</sup>首次引入光滑的概念，将支持向量机算法进行光滑处理，得到

$$\min_{w,b,C} 1/2 * C \|e - D(Aw - eb)\|_2^2 + 1/2 * (w^T w + b^2) \quad (2)$$

在式(2)中，目标函数中存在具有强凸性的非可微函数，但是光滑性不能得到满足，Lee等人对其进行了改进处理，得到

$$p(x, k) = x + 1/k * \lg(1 + e^{-kx}) \quad (3)$$

结合式(2)和式(3)得到光滑支持向量机目标函数，如式(4)。

$$\min_{w,b,C} 1/2 * C \|p(e - D(Aw - eb), k)\|_2^2 + 1/2 * (w^T w + b^2) \quad (4)$$

在光滑支持向量机提出后，很多学者和专家构建了光滑函数来代替非可微函数<sup>[18-20]</sup>，本文在前人的基础上构建了一个新的光滑函数，见式(5)。

$$M(x, k) = \begin{cases} 0, & x < -\frac{1}{7k} \\ 7k^2 x^3 + \frac{7}{4} kx^2 + \frac{1}{2} x + \frac{1}{14k}, & -\frac{1}{7k} \leq x < 0 \\ -7k^2 x^3 + \frac{7}{4} kx^2 + \frac{1}{2} x + \frac{1}{14k}, & 0 \leq x \leq \frac{1}{7k} \\ x, & x > \frac{1}{7k} \end{cases} \quad (5)$$

本文构建的光滑支持向量机算法为

$$\min_{w,b,C} 1/2 * C \|M(e - D(Aw - eb), k)\|_2^2 + 1/2 * (w^T w + b^2) \quad (6)$$

本文构建的光滑函数  $M(x, k)$  定义如式(5)， $x_+ = \max\{0, x\}$  表示非可微函数，那么光滑函数  $M(x, k)$  满足如下：

- (1)  $M(x, k)$  关于  $x$  二阶光滑；
- (2)  $M(x, k) \geq x_+$ ；
- (3) 对任意给定的  $x \in R$ ， $M(x, k)^2 - x_+^2 \leq \frac{1}{821k^2}$ 。

证明：

(1) 当  $x = \pm \frac{1}{7k}$  和  $x = 0$  时，函数  $M(x, k)$  分别有以下条件成立。  
 $M(-\frac{1}{7k}, k) = 0$ ， $\lim_{x \rightarrow 0^-} M(x, k) = \lim_{x \rightarrow 0^+} M(x, k)$ ， $M(\frac{1}{7k}, k) = \frac{1}{7k}$ ，  
 $M'(-\frac{1}{7k}, k) = 0$ ， $\lim_{x \rightarrow 0^-} M'(x, k) = \lim_{x \rightarrow 0^+} M'(x, k)$ ， $M'(\frac{1}{7k}, k) = 1$ ，  
 $M''(-\frac{1}{7k}, k) = 0$ ， $\lim_{x \rightarrow 0^-} M''(x, k) = \lim_{x \rightarrow 0^+} M''(x, k)$ ， $M''(\frac{1}{7k}, k) = 0$ 。  
 因此， $M(x, k)$  是关于  $x$  二阶光滑。

(2) 在  $x \geq \frac{1}{7k}$  和  $x \leq -\frac{1}{7k}$  时， $M(x, k) \geq x_+$  是显然成立的；  
 $x \in (0, \frac{1}{7k})$ ， $M(x, k) - x_+$  的特性是递减的，  
 $M(x, k) - x_+ \geq M(\frac{1}{7k}, k) - x_+ = 0$ ；  
 $x \in (-\frac{1}{7k}, 0)$ ， $M(x, k) - x_+$  的特性是递增的，  
 $M(x, k) - x_+ \geq M(-\frac{1}{7k}, k) = 0$ ；因此， $M(x, k) \geq x_+$ 。

(3) 在  $x \geq \frac{1}{7k}$  和  $x \leq -\frac{1}{7k}$  时， $M(x, k)^2 - x_+^2 \leq \frac{1}{821k^2}$  是显然成立的；  
 $x \in (-\frac{1}{7k}, 0)$ ， $M(x, k)^2 - x_+^2 = M(x, k)^2 \leq M(0, k)^2 < \frac{1}{821k^2}$ ；  
 $x \in (0, \frac{1}{7k})$ ，令  $f(x) = M(x, k)^2 - x_+^2 = M(x, k)^2 - x^2$ ， $t = 7kx \in (0, 1)$ ，

通过变换可以得到： $f(t) = \left(-\frac{1}{49}t^3 + \frac{1}{28}t^2 + \frac{1}{14}t + \frac{1}{14}\right)^2 - \frac{1}{49}t^2$ ，

在  $t \in (0, 1)$  上，求得  $f(t)$  的最大值点为  $t = 0.1705$ ，代入到  $f(t)$  中，可以得到  $f(0.1705) = 0.0012/k^2 < \frac{1}{821k^2}$ 。因此，对任意

给定的  $x \in R$ ， $M(x, k)^2 - x_+^2 \leq \frac{1}{821k^2}$ ，本文构建的二阶函数是光滑的。

### 2.4 属性约简—光滑支持向量机中小企业信息化水平评价模型

基于属性约简—光滑支持向量机信息化水平评价模型，首先利用属性约简将中小企业信息化体系的指标进行预处理，剔除掉不是核心属性的指标，减少了光滑支持向量机的

运算维度，降低了时间复杂度，然后利用层次分析法和熵值法计算得到各个指标权重，最后利用属性约简—光滑支持向量机模型对中小企业的信息化水平进行组合评价。

### 3 实证分析(The empirical analysis)

#### 3.1 评价指标体系的构建

本文首先对影响中小企业信息化水平的各种关键因素进行搜集，主要数据来源于<sup>[21-23]</sup>，借鉴国家信息化测评中心和众多研究学者的成果，综合分析，得到四个较通用的综合指标，分别是软硬件保障水平、信息化组织水平、信息技术应用与盈利水平、信息化能力水平。下面将根据这四个指标进行中小企业信息化程度进行评价，评价体系如表1所示<sup>[14]</sup>。

表1 中小企业信息化水平评价指标体系

Tab.1 Informationization level evaluation index system of small and medium-sized enterprises

一级指标	二级指标	三级指标
软硬件保障水平	软件水平	固定资产水平
		流动资产水平
		计算机普及水平
		数据库建设水平
	网络水平	网络性能水平
		网络覆盖范围
		网站建设水平
	安全保障水平	安全投入比例水平
		安全制度建设水平
	人力保障水平	员工学历水平
信息化技能掌握水平		
信息化组织水平	信息化地位水平	学习的电子化水平
		高层关注水平
		国家扶持水平
		信息化投入水平
	机构设置水平	信息化规划水平
		信息职能部门设置水平
	市场预测水平	业务流程覆盖水平
		市场调查预测水平
	管理信息化水平	设计制造水平
		ERP水平
供应链管理		
客户关系管理		
信息技术应用与盈利水平	营销与信息化水平	OA水平
		营销信息化水平
	盈利水平	决策信息化水平
		库存占比水平
		资金周转水平
	运营速度水平	
	效益增长水平	

(续表)

一级指标	二级指标	三级指标
信息化能力水平	信息化管控能力水平	信息化计划能力水平
		信息化领导能力水平
		信息化组织能力水平
		信息化控制能力水平
		信息化创新能力水平
	信息化模块管理水平	设计生产能力水平
		销售能力水平
		财务能力水平
		组织文化能力水平
		技术系统能力水平

#### 3.2 属性约简

从表1可知，此指标体系共包含40个具体指标。如信息化管控能力水平，其包含信息化计划能力水平、信息化领导能力水平、信息化组织能力水平、信息化控制能力水平、信息化创新能力水平等五个指标。然后针对表1所涉及的原始指标体系对该市各种类型的90家企业进行调研，取得一手数据。以此指标中的信息化管控能力水平情况为例，五个子指标作为条件属性，将取得的原始数据进行离散化处理，按照属性约简，将它们分为四个不同的等级{4,3,2,1}，等级分别表示的意思为{优,良,可,差}。同时对企业的评价准则也分为四类，作为粗糙集的决策属性D={4,3,2,1}，等级分别表示的意思为{好,较好,一般,差}。其余的11个处理方式与此相同。由于篇幅所限，本文将10个中小企业做为(U)，最终建立信息化管控能力水平离散化决策表如表2所示。

表2 某市10个企业的信息化管控能力水平

Tab.2 Informationization management and control capability of 10 enterprises in a city

U	信息化计划能力水平	信息化领导能力水平	信息化组织能力水平	信息化控制能力水平	信息化创新能力水平	等级(D)
1	3	3	4	3	3	3
2	2	2	3	4	3	3
3	3	3	3	4	4	4
4	2	3	3	4	3	3
5	1	3	2	3	3	3
6	3	4	3	4	4	4
7	2	3	1	3	2	2
8	2	3	3	2	2	2
9	2	2	3	3	2	2
10	1	2	2	2	1	1

对决策表进行属性约简。由表2中数据可求出哪些是冗余属性，哪些是核属性。属性约简后的指标体系如表3所示。

此指标体系约简去了流动资产水平、网站建设水平、安全制度建设水平、学习的电子化水平、国家扶持水平、客户关系管理水平、决策信息化水平、运营速度水平、信息化创新能力水平、信息化组织能力水平、信息化控制能力水平、组织文化能力水平、信息化规划水平、设计制造水平、技术系统能力水平等15个指标。这样就剔除了不是核心属性的评价指标，减少了指标权重的计算复杂度。

表3 约简后的制造企业信息化水平评价指标体系

Tab.3 The reduced evaluation index system of informatization level of manufacturing enterprises

一级指标	二级指标	三级指标
软硬件保障水平	软硬件水平	固定资产水平
		计算机普及水平
		数据库建设水平
	网络水平	网络性能水平
		网络覆盖范围
		安全保障水平
人力保障水平	人力保障水平	员工学历水平
		信息化技能掌握水平
	信息化地位水平	高层关注水平
		信息化投入水平
信息化组织水平	机构设置水平	信息职能部门设置水平
		业务流程覆盖水平
	市场预测水平	市场调查预测水平
信息技术应用与盈利水平	管理信息化水平	ERP水平
		供应链管理
		OA水平
	营销与信息化水平	营销信息化水平
		库存占比水平
	盈利水平	资金周转水平
信息化能力水平	信息化管控能力水平	信息化计划能力水平
		信息化领导能力水平
	信息化模块管理	设计生产能力水平
		销售能力水平
		财务能力水平

利用层次分析法和熵值法得到指标体系权重值为{0.068, 0.032, 0.066, 0.092, 0.062, 0.030, 0.011, 0.065, 0.035, 0.044, 0.044, 0.046, 0.072, 0.039, 0.045, 0.019, 0.021, 0.034, 0.014, 0.011, 0.019, 0.022, 0.031, 0.028, 0.020}。用上述指标体系和权重值评

价某省22家中小企业的信息化水平，限于篇幅，这里不给出指标数据。表4是用线性加权法算出的22家企业的信息化综合得分及排序。

表4 企业信息化综合得分及排序

Tab.4 Comprehensive score and ranking of enterprises' informationization

企业	信息化评价得分值	信息化水平排序	企业	信息化评价得分值	信息化水平排序
1	63.48737	5	12	31.49973	21
2	63.48736	6	13	33.26548	18
3	54.01222	9	14	53.12536	10
4	70.01982	4	15	54.68252	8
5	33.00245	19	16	78.10595	1
6	58.55261	7	17	36.04678	17
7	73.27225	2	18	50.15325	13
8	47.84138	15	19	49.00562	14
9	52.61829	11	20	72.23141	3
10	51.24625	12	21	31.88007	20
11	25.07297	22	22	41.89744	16

### 3.3 算法比较分析

本文光滑支持向量机算法的实现采用MATLAB2014a编程，为了验证支持向量机算法在综合评价中的适应性，同时引入了BP神经网络算法<sup>[24]</sup>对最后五家企业的信息化水平进行评价，并与原始数据相比较，最后结果如表5所示。

表5 五家企业的评价结果对比

Tab.5 Comparison of evaluation results of 5 enterprises

	原排序	18	19	20	21	22
线性加权法得分		50.153	49.005	72.231	31.880	41.897
排序		2	3	1	5	4
传统BP神经网络得分		62.118	62.111	51.936	78.168	22.730
排序		2	3	4	1	5
属性约简—光滑支持向量机得分		56.232	50.189	76.853	40.231	46.366
排序		2	3	1	5	4

从表5可以看出，BP神经网络出现了三家企业错判的情况，预测效果不是很理想。用光滑支持向量机评价时，五家企业的信息化水平排序均正确。因此光滑支持向量机算法更好地获取并存储了信息化评价专家的知识和经验，使得评价结果更加客观、真实。

### 4 结论 (Conclusion)

本文依据中小企业的信息化指标体系，设计了一种基于属性约简—光滑支持向量机的评价模型，通过实例分析，本文算法对中小企业的信息化评价是客观有效地，与比BP神经

网的评价效果相比,本文算法更加准确和简便。因此在信息化飞速发展的背景下,本文的研究方法在理论和实际应用中都具有重要的意义。

## 参考文献(References)

- [1] 马立川,裴庆祺,冷昊,等.大数据安全研究概述[J].无线电通信技术,2015,41(1):01-07.
- [2] 数博会组委会.2016中国大数据产业峰会暨中国电子商务创新发展峰会[J].中国信息安全,2016(5):4.
- [3] 李国杰,程学旗.大数据研究:未来科技及经济社会发展的重大战略领域—大数据的研究现状与科学思考[J].中国科学院院刊,2012,27(6):647-657.
- [4] Saunders C, Stitson M O, Weston J, et al. Support vector machine[J]. 2002, 1(4): 1-28.
- [5] Lin K C, Chien H Y. CSO-based feature selection and parameter optimization for support vector machine[C]. Pervasive Computing. DOI:10.1109/JCPC.2009.5420080.
- [6] Shalev Shwartz S, Singer Y, Srebro N, et al. Pegasos: primal estimated sub-gradient solver for SVM[J]. Mathematical Programming, 2011, 127(1): 3-30.
- [7] Shom Prasad Das, Sudarsan Pa dhy. A novel hybrid model using teaching learning based optimization and a support vector machine for commodity futures index forecasting[J]. International Journal of Machine Learning & Cybernetics, 2018, 9(1): 97-111.
- [8] Yang Z, Lin H Y. A protein-protein interaction extractor for biomedical literature using SVM and rich feature sets [J]. Journal of Biomedical Informatics, 2010, 43(1): 88-96.
- [9] Biswajeet Pradhan. A comparative study on the predictive ability of the decision tree, support vector machine and neuro fuzzy models in landslide susceptibility mapping using GIS [J]. Computers & Geosciences, 2012, 51(2): 350-365.
- [10] Qian H, Mao Y, Xiang W, et al. Recognition of human activities using SVM multiclass classifier[J]. Pattern Recognition Letters, 2010, 31(2): 100-111.
- [11] 谢宏,程浩忠,牛东晓,等.基于信息熵的粗糙集连续属性离散化算法[J].计算机学报,2005,28(9):1570-1574.
- [12] Dun Liu, Tianrui Li, Junbo Zhang. A rough set based incremental approach for learning knowledge in dynamic incomplete information systems[J]. International Journal of Approximate Reasoning, 2014, 55(9): 1764-1786.
- [13] Siyuan Jing. A hybrid genetic algorithm for feature subset selection in rough set theory[J]. Soft Computing, 2013, 18(7): 1373-1382.
- [14] Richard Jensen, Andrew Tuson. Finding rough and fuzzy rough set reducts with SAT[J]. Information Sciences, 2014, 255: 100-120.
- [15] Phophalia A, Rajwade A, Mitra S K. Rough set based image denoising for brain MR images[J]. signal processing, 2014, 103(10): 24-35.
- [16] Burges C. A tutorial on support vector machines for pattern recognition[J]. Data Mining and Knowledge Discovery, 1998, 2(2): 121-167.
- [17] Lee Y J, Hsieh W F, Huang C M. A smooth support vector machine for epsilon insensitive regression [J]. IEEE Transactions on Knowledge & Data Engineering, 2005, 17(5): 678-685.
- [18] 袁玉波,严杰,徐成贤,等.多项式光滑的支撑向量机[J].计算机学报,2005,28(1):9-17.
- [19] Wu Q. A new class of piecewise smooth support vector regressions[J]. Journal of Information Science Engineering, 2015, 31(5): 1813-1828.
- [20] Qing Wu, Bo Li, Wa Wang, et al. Face recognition method based on multi-class classification of smooth support vector machine[J]. Journal of Computer Applications, 2015, 19(5): 688-696.
- [21] 肖素梅,殷国富,汪永超,等.企业信息化水平评价指标与评价方法研究[J].计算机集成制造系统,2005(08):106-114.
- [22] 何永琴,王燕嘉,毛文静,等.中小企业信息化水平调查研究[J].内蒙古财经大学学报,2015(2):60-67.
- [23] 李鹏飞.制造企业信息化水平的评价研究[D].青岛:中国海洋大学,2013.
- [24] 李钢,田丽娜,齐二石,等.基于BP神经网络和遗传算法的企业信息化评价研究[J].科技进步与对策,2018,25(5):129-132.

## 作者简介:

戴强(1972-),男,本科,工程师.研究领域:人工智能,信息化评估.