



Research on Credit Evaluation of Listed Companies in Science and Technology Sector Based on SVM

Su-juan Xu^{1,2}, Mu Zhang^{1,*}

¹ School of Big Data Application and Economics, Guizhou University of Finance and Economics, Guiyang (550025), Guizhou, China

² Guizhou Institution for Technology Innovation & Entrepreneurship Investment, Guizhou University of Finance and Economics, Guiyang (550025), Guizhou, China

*Corresponding author. Email: zhangmu01@163.com

ABSTRACT

In recent years, the state has strongly supported the development of scientific and technological enterprises. Scientific and technological enterprises occupy an increasingly important position in China's economic development. However, scientific and technological enterprises are in the growth period, and there are many risks in the process of development and expansion, so there are some problems in financing. Based on the establishment of the credit evaluation index system of listed companies in the science and technology sector, this paper calculates the IV value of each index to screen the indicators, and uses SVM to classify the selected sample enterprises, and compares the classification accuracy of the samples before and after the index screening. The results show that the classification accuracy of both training samples and test samples is improved after removing the indicators with little information value, which also shows the feasibility and effectiveness of the model.

Key words: Science and Technology Enterprises, Support Vector Machine, Credit Evaluation.

基于 SVM 的科技板块上市公司信用评估研究

许素娟^{1,2}, 张目^{1,*}

¹ 贵州财经大学 大数据应用与经济学院, 中国 贵州 贵阳 550025

² 贵州财经大学 贵州科技创新创业投资研究院, 中国 贵州 贵阳 550025

* 通讯作者. 电子邮箱: zhangmu01@163.com

摘要:

近年来, 国家大力支持科技企业的发展, 科技型企业在我国的经济发展中占据着越来越重要的地位, 但是科技型企业正处于成长期, 在发展壮大的过程中存在不少风险, 所以在融资方面有一定问题。本文在建立了科技型企业信用评价指标体系的基础上, 计算各个指标的 IV 值从而对指标进行筛选, 并运用支持向量机对所选取的样本企业进行分类, 比较样本在指标筛选前后的分类准确率。结果表明, 不管是训练样本还是测试样本, 剔除

信息价值不大的指标后，其分类准确率均有所提升，这也表明了该模型的可行性和有效性。

关键词：科技型企业,支持向量机,信用评估.

1. 引言

科技型企业是我国经济高质量发展阶段培育壮大新动能的重要载体，具有较大的市场活力和发展动力，不少企业已经占据了产业、行业的核心技术制高点。国家大力支持发展科技型中小企业，近年来快速发展，在国家的经济发展中起着非常重要的作用，但是仍处于发展初期，面临着许多挑战和风险。由于信息不对称，商业银行对于科技型企业的营运情况、盈利状况等信息掌握不完全，融资过程存在较大不确定性，容易引发信用风险。

加强信用风险管理对降低银行信用风险、提高资产质量具有重要意义。企业信用评级一方面降低了银行的信用风险，使银行充分发挥资金中介的作用，另一方面缓解了企业融资困难的问题，使企业能取得所需资金从事其各项生产经营活动，为双方的信息缺口开辟通道，增强市场的整体效率，这对交易双方都是有利的。因此，如何有效提高银行的信用风险评估水平，降低贷款风险的发生，是科技型企业持续健康发展的关键。

目前，信用评估最常用的方法是分类方法，即根据借款人的财务、非财务状况将其分为“正常”和“违约”两类，这样信用评估问题就转化为统计学上的某种分类问题。本文引入了统计学基于小样本学习理论的通用学习算法——支持向量机 SVM，选取了科技板块上市公司的样本和相关财务指标，利用 SVM 算法对指标筛选前后的分类正确率进行对比，以此来验证该模型的有效性。

2. 文献综述

支持向量机能较好的解决小样本、非线性、高维数和局部极小点等实际问题，赵春晖和陈万海等（2007）将支持向量机的二分类问题应用到多分类问题上^[1]；龙玉峰与崔和（2008）介绍了支持向量机的标准训练算法及其分解算法、变形算法等，重点描述了最小二乘支持向量机算法^[2]；侯夏晔和陈海燕等（2022）提出了一种适用于半监督环境的基于支持向量机的主动度量学习算法^[3]；在个人信用评价中，也经常用到支持向量机，刘潇雅和王应明（2020）将集

成方法应用到信用评估领域，提出改进 DS 证据理论的支持向量机集成个人信用评估模型^[4]。

有相关文献对科技型企业信用风险进行研究，构建了信用评估模型。王少英和兰晓然等（2019）以科技型中小企业为研究对象，运用带有非凸惩罚的 SVM 模型对影响中小企业的信用风险因素进行研究^[5]；毛茜和赵喜仓构建了 Logistic-BP 神经网络组合模型，对科技型小微企业进行分阶段信用评估^[6]；曾梓铭（2020）选取了 58 家上市农业中小企业的财务数据构建信用风险评估体系，在因子分析基础上，对比 Logistic 模型和支持向量机的信用风险评估效果^[7]；甄彤和范艳峰（2006）将支持向量机非线性分类器应用于信用风险的评估中，分析对比了选取不同核函数和参数的实验结果，认为 SVM 在有限样本情况下具有良好的泛化能力^[8]；陈战勇（2020）对特征变量进行筛选，并建立融入证据权重的 Logistic 信用评分卡模型对客户进行信用评分，进而预测客户借款到期时的违约情况^[9]。

综上所述，众多学者运用支持向量机对企业进行分类预测，并与不同参数的模型或其他分类方法进行正确率比较。本文对评价指标体系进行筛选，运用支持向量机对公司进行分类预测，通过对比指标筛选前后的分类正确率来判断该模型是否有效。

3. 指标筛选理论

3.1 分箱

分箱的本质其实就是连续变量离散化，将不同属性的变量为不同的类别打上不同的分数。卡方分箱具体步骤：首先，分箱数选取的最大区间数为 n 。其次，对于无序变量先转化，对于连续数据进行顺序转化。再次，基于空值占比观察是否需要单独分箱。最后，计算每一对相邻区间的卡方值，将卡方值最小的一对区间合并。重复以上步骤直到满足条件。

$$x^2 = \sum_{i=1}^2 \sum_{j=1}^2 \frac{(A_{ij} - E_{ij})^2}{E_{ij}} \quad (1)$$

在公式（1）中， A_{ij} 表示第 i 区间的实例数量； E_{ij} 表示 $E_{ij} = \frac{N_i}{N} \times C_j$ ， N 是合并区间样本数， N_i 是第

i 组样本数, C_j 是第 j 类样本数在合并区间的样本数。

3.2 WOE 编码

分箱之后需要对变量进行编码, WOE 称为证据权重 (weight of evidence), 其公式为:

$$WOE_i = \log \frac{P_{i1}}{P_{i0}} = \log \frac{*B_i/*B_T}{*G_i/*G_T} \quad (2)$$

在公式 (2) 中, P_{i1} 是第 i 箱中坏样本占所有坏样本比例, P_{i0} 是第 i 箱中好样本占所有好样本比例; $*B_i$ 是第 i 箱中坏样本人数; $*B_T$ 是第 i 箱中好样本人数; $*G_i$ 是所有坏样本人数; $*G_T$ 是所有好样本人数。

3.3 IV 值变量筛选

离散化连续变量必然伴随着信息的损失, 并且箱子越少, 信息损失越少。为了衡量指标变量上的信息量以及特征对预测函数的贡献度, 利用信息增益 IV (information value) 表示, IV 值越大, 模型的预测效率越高, IV 信息增益对照表见表 1。

表 1 信息增益对照表

IV	预测能力
<0.02	无效
0.02~0.1	弱预测
0.1~0.2	中预测
>0.2	强预测

4. 应用实例

4.1 指标体系与样本数据

本文参照国家财政部统计评价司的企业绩效评价指标体系和中国工商银行企业资信评估指标体系, 并根据科技型企业的特点在指标体系中加入了创新能力来评判科技型企业的创新发展能力, 在遵循指标选取的系统性、科学性、客观性、可比性及可操作性等的原则上, 从盈利能力、偿债能力、营运能力、成长能力和创新能力 5 个方面构建科技型企业信用评估指标体系。变量的描述性统计见表 2。

表 2 变量的描述性统计

	样本数	均值	标准差	最大值	最小值
年化总资产报酬率(%)	200	5.56	4.28	22.87	-10.55
销售净利率(%)	200	7.87	10.83	37.14	-69.04
营业利润/营业总收入(%)	200	8.77	10.62	40.07	-58.56
管理费用率(%)	200	7.64	5.03	35.19	0.50
净利润(万元)	200	28235.94	57744.75	402961.76	-262497.16
流动比率(%)	200	2.34	1.54	10.54	0.30
速动比率(%)	200	1.95	1.49	10.44	0.17
产权比率(%)	200	81.06	66.09	671.69	7.57
有形资产/负债合计(%)	200	176.81	166.61	1263.69	6.13
净资产负债率(%)	200	84.09	66.96	618.72	7.57
净负债率(%)	200	-10.88	52.24	146.89	-529.15
现金流动负债比(%)	200	0.14	0.22	0.87	-0.62
货币资金/短期债务(%)	200	1080.13	2557.61	17123.40	16.46
营业利润率(%)	200	8.78	10.62	40.07	-58.56
每股收益增长率(%)	200	32.80	75.49	308.01	-197.31
营业收入增长率(%)	200	20.69	23.15	115.70	-52.75
营业利润增长率(%)	200	36.60	81.49	411.03	-191.55
利润总额增长率(%)	200	34.76	76.55	285.19	-203.64
净利润增长率(%)	200	37.00	78.07	285.37	-204.24
净资产增长率(%)	200	17.86	30.32	223.67	-38.58
存货周转率(%)	200	6.50	10.24	123.79	0.41

	样本数	均值	标准差	最大值	最小值
应付账款周转率(%)	200	6.59	8.77	63.38	1.06
营运资金周转天数(天/次)	200	132.52	119.91	689.59	-132.04
流动资产周转率(%)	200	1.07	0.76	8.47	0.15
研发投入占营业收入比例(%)	200	9.01	6.84	40.31	0.11
申请专利数量(个)	200	291.80	429.58	3655.00	1.00
营业总成本同比增长(%)	200	21.28	22.43	100.55	-48.89
存货周转天数(天/次)	200	106.85	87.52	887.50	2.91

选取沪、深股市中科技型的上市公司作为实验样本, 样本的区间选定为 2021 年 1 月 1 日至 2021 年 12 月 31 日, 数据来源于 RESSET 数据库。剔除异常数据样本后, 最终选取了 200 家样本企业, 其中有 100 家为“非 ST”企业, 其财务状况良好, 违约的可能性较小, 信用风险较低; 其余 100 家为 ST 企业或 *ST 企业, 其财务状况较差, 违约的可能性较大, 信用风险较高。所选取公司的行业分类如表 3 所示。将实验样本集划分为训练样本集和测试样本集。在保持两类样本数目相当的前提下, 随机抽取 70% (140 家) 作为训练样本, 用于构造 SVM 模型, 剩余的 30% (60 家) 作为测试样本, 用于检验模型的泛化能力。数据分组情况如表 4 所示。

表 3 样本行业分布统计

行业分类	数量 (个)
新型材料	92
互联网	21
公共安全	9
BAT	59
智能产业	6
新科技	6
制造 2025	5
新模式	2

表 4 数据分组情况

	正常企业	违约企业	合计
训练集	70	70	140
测试集	30	30	60
合计	100	100	200

4.2 指标筛选结果

表 5 筛选后的指标及其 IV 值

	指标	IV 值
--	----	------

通过上述指标筛选的方法, 对所选取的 28 个指标运用 python 进行卡方分箱, 最后计算各个指标的 IV 值, 可以知道各个指标对模型预测效率的贡献度。当指标的 IV 值小于 0.02 时, 指标的预测能力是无效的, 这时可以剔除 IV 值小于 0.02 的所有指标, 剩下的指标中选择 IV 值较高的, 模型预测效果较好, 最终筛选出来的一共有 12 个指标, 如表 5 所示。

4.3 SVM 模型构造及结果

本文一共选取了 200 家企业样本, 表示为 (x, y) , 其实 x 的维数为指标的个数, y 是样本的类别属性, 维数为 1, 对于正常企业 $y=1$, 对于违约企业 $y=-1$, 对于企业信用风险评估这样的非线性分类问题, 需选择用内积核函数代替原空间中的内积。本文运用根据实际分类效果, 选择径向基函数作为 SVM 模型的内积核函数。

将数据转化为 SVM 模型所需的数据格式后, 在 Matlab 中选择 70% 的样本构造训练模型, 再将剩下的 30% 样本作为测试集来检测模型的预测效果, 在指标筛选前和指标筛选后做了对比, 两次结果中, 训练样本和测试样本的准确率均有所提高, 说明了数据筛选及该模型的有效性。

指标筛选前, 一共有 28 个指标, 将数据和指标带入 SVM 模型中, 其结果以及第一类错误率和第二类错误率如表 6 所示。计算出了各个指标的 IV 值后, 提出了 IV 值小于 0.02 的指标以及多余重复的指标, 最后用剩下的 12 个指标构建 SVM 模型, 其结果如表 7 所示。

盈利能力	营业利润/营业总收入	0.0503
	营业利润率	0.0503
偿债能力	流动比率	0.0671
	速动比率	0.0864
	净资产负债率	0.0864
营运能力	流动资产周转率	0.1383
	存货周转率	0.0326
成长能力	每股收益增长率	0.1534
	营业利润增长率	0.1031
	净利润增长率	0.1031
创新能力	研发投入占营业收入比例	0.4436
	申请专利数量	0.1330

表 6 指标筛选前分类结果

	正确率	第一类错误率	第二类错误率
训练样本	87.8571%	11.43%	4.29%
测试样本	83.3333%	20%	13.33%

表 7 指标筛选后分类结果

	正确率	第一类错误率	第二类错误率
训练样本	89.2857%	17.1428%	4.2857%
测试样本	85%	20%	10%

从表 6 和表 7 两个表格可以知道, 指标筛选前训练样本的正确率为 87.8571%, 测试样本的正确率为 83.3333%, 而指标筛选后训练样本的正确率为 89.2857%, 测试样本的正确率为 85%, 对比指标筛选前后的分类结果可以看出, 指标筛选后 SVM 模型的预测准确率有所提高, 说明对指标进行卡方分箱计算指标的 IV 值可以对科技型企业的信用评估评价指标进行筛选, 避免异常值和极端值的影响, 从而选出信息价值较高的指标, 这样可以提高分类正确率, 有利于银行对企业进行信用评估, 有效识别风险。

第一类错误率是指原假设是正确的但是却拒绝了原假设, 第二类错误率是指原假设是错误的但是却接受了原假设。在本文科技板块上市公司信用评估的模型中, 第一类错误率就是将正常企业判断为违约企业的错误率, 这样就会导致银行不会贷款给这一类企业, 一方面银行自身失去了盈利机会, 另一方面这一类企业也会面临融资困难的问题, 这都是第一类错误率所带来的机会成本; 第二类错误率是将违约企业判断成正常企业的错误率, 如果银行对这一类企业的信用评估失误, 将会面临较大的信用风险, 因为这一类企业有较大的可能性违约, 不能按时还款, 银行就会遭受损失。

5. 结语

支持向量机是一种有坚实理论基础的新颖的小样本学习方法, 其最终决策函数只有少数的支持向量所确定。该方法具有较好的“鲁棒”性, 例如增加或者删除非支持向量样本时对模型没有影响; 在有些成功的应用中, SVM 对核的选取不敏感。将 SVM 应用到企业信用风险评估中, 比传统方法更简单、有效, 具有更好的泛化能力。SVM 不仅可以用于两分类问题, 而且可以推广到更为复杂的多分类的信用等级评估问题上, 以更好地反映企业的信用情况。

本文在对信用评估方法、SVM 算法、企业信用评估相关文献进行总结的基础上, 运用卡方分箱、WOE 编码以及计算指标的 IV 值来计算每个指标的信息价值, 从科技型中小企业的 28 个指标中筛选出来 12 个指标, 剔除信息价值低的指标, 从而在一定程度上提高了 SVM 的预测精度, 证明了本文的信用评估模型的可行性和有效性。银行在对企业进行信用评估时, 要考虑各方面的因素, 从不同的角度出发, 建立全面规范的评价指标体系, 但是有的指标对于评价模型的信息价值并不大, 这些指标可以舍去, 使模型的预测更加准确, 这样有利于对企业的贷款及融资

作出准确可靠的评估,对于企业的融资及风险管理有着重要意义。

信用评估准确率的优劣直接影响着互联网金融机构的利益和投资者的资金安全,影响着整个行业的健康发展。银行对科技型企业作出合理的信用评估是非常必要的,不仅会影响着自身的相关利益,银行在承担最小风险的情况下给科技型企业融资,降低了贷款给信用较差的客户的违约风险,同时减小了因拒绝给正常企业贷款融资的机会成本;另一方面科技型企业正处于成长期,在发展的过程中需要大量金融支持,信用风险评估区分了信用好的企业与信用状况较差的企业,可以让这些企业得到相应的资金支持,进一步推动了全民经济的发展,加强了技型企业作为国民经济支撑的重要性。

在此基础上,银行应该加强信用评估体系的建设。合理的信用评估有助于银行建立完善的内部控制制度和内在激励机制。在外部约束机制作用下促使商业银行加强内部管理,加快规范化经营步伐,获取客观公正的信用等级,已经成为银行树立稳健形象、拓展业务、降低交易成本、提高市场竞争力的一种重要策略与手段。通过信用评级,银行可以从中找到差距,改进管理,对同业交易而言,交易对手的信用高低,直接影响到交易的授信额度及其资金安全性。因此,银行对企业进行信用评估是非常必要的。

致谢

本研究得到了国家自然科学基金地区项目“基于文本信息的科技型中小企业信用风险识别机理研究”的资助,项目编号为 71861003。

参考文献

- [1] 赵春晖,陈万海,郭春燕.多类支持向量机方法的研究现状与分析[J].智能系统学报,2007(02):11-17.
- [2] 崔和,龙玉峰.支持向量机学习算法的研究现状与展望[J].信息与电子工程,2008(05):328-332.
- [3] 侯夏晔,陈海燕,张兵,袁立罡,贾亦真.一种基于支持向量机的主动度量学习算法[J].计算机科学,2022,49(S1):113-118.
- [4] 刘潇雅,王应明.基于支持向量机集成的个人信用评估研究[J].计算机工程与应用,2020,56(14):272-278.

- [5] 王少英,兰晓然,刘丽英.基于非凸惩罚的SVM模型对科技型中小企业信用风险评估[J].数学的实践与认识,2019,49(03):307-312.
- [6] 毛茜,赵喜仓.创新价值链视角下信用评估研究——基于科技型小微企业的实证分析[J].华东经济管理,2018,32(05):134-139. DOI:10.19629/j.cnki.341014/f.180316014.
- [7] 曾梓铭.基于支持向量机的农业供应链金融信用风险评价[J].吉林金融研究,2020(12):46-49.
- [8] 甄彤,范艳峰.基于支持向量机的企业信用风险评估研究[J].微电子学与计算机,2006(05):136-139. DOI:10.19304/j.cnki.issn1000-7180.2006.05.040.
- [9] 陈战勇.珠联璧合:基于机器学习的网络借贷信用评分卡模型研究[J].武汉金融,2020(03):42-50.
- [10] 张雷.基于混合遗传算法-支持向量机的中小型企业信用评估模型[J].河南师范大学学报(自然科学版),2022,50(02):79-85. DOI:10.16366/j.cnki.1000-2367.2022.02.010.
- [11] 孙治河,张雷.基于机器学习算法的中小企业信用评估研究[J].无线互联科技,2021,18(12):94-95.
- [12] 胡海青,张琅,张道宏.供应链金融视角下的中小企业信用风险评估研究——基于SVM与BP神经网络的比较研究[J].管理评论,2012,24(11):70-80. DOI:10.14120/j.cnki.cn11-5057/f.2012.11.015.
- [13] 丁欣.国外信用风险评估方法的发展现状[J].湖南大学学报(社会科学版),2002(S1):140-142.
- [14] 张目,周宗放.企业信用评估中的一种样本预处理方法[J].商场现代化,2010(23):86-87.
- [15] 孟杰.企业信用风险评估方法综述[J].中国国际财经(中英文),2018(02):290. DOI:10.19516/j.cnki.10-1438/f.2018.02.232.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

