

**The Research on Financial Performance Prediction
Using The least squares support vector machine**

A Dissertation Submitted for the Degree of Master

Candidate: Li Yue

Supervisor: Prof. Zhao Guanhua

School of Accountancy

Shandong University of Finance and Economics

中图分类号：

密级：公 开

学科分类号：

论文编号：KJ118222009120201027

硕 士 学 位 论 文

最小二乘支持向量机 在财务业绩预测中的应用研究

作者姓名：李玥

申请学位级别：管理学硕士

指导教师姓名：赵冠华

职 称：教 授

学 科 专 业：会计学

研 究 方 向：财务决策支持系统

学 习 时 间：自 2009 年 9 月 1 日 起 至 2012 年 6 月 30 日 止

学位授予单位：山东财经大学

学位授予日期：2012 年 6 月

山东财经大学学位论文独创性声明

本人声明所呈交的学位论文是我个人在导师指导下进行研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得山东财经大学或其它教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

学位论文作者签名：

日期： 年 月 日

山东财经大学学位论文使用授权声明

本人完全同意山东财经大学有权使用本学位论文（包括但不限于其印刷版和电子版），使用方式包括但不限于：保留学位论文，按规定向国家有关部门（机构）送交学位论文，以学术交流为目的赠送和交换学位论文，允许学位论文被查阅、借阅和复印，将学位论文的全部或部分内容编入有关数据库进行检索，采用影印、缩印或其他复制手段保存学位论文。

保密学位论文在解密后的使用授权同上。

学位论文作者签名：

日期： 年 月 日

指导教师签名：

日期： 年 月 日

摘 要

自奥特曼对财务困境展开开创性研究以来，对财务困境预测的研究已经相对成熟，但是财务困境预测以公司是否被 ST 作为预测标准，这难免给人一种为时已晚的感觉。鉴于此，有学者初步提出财务业绩预测并进行了相关研究。财务业绩预测仍然处于一个起步阶段，它通过预测公司未来业绩的变动情况，为利益相关者提供决策所需信息。本文的研究目的就是希望在以往研究的基础上提出一种适用于制造业业绩预测的方法，向利益相关者提供公司发展趋势的信息，从而有利于他们做出科学的决策。

本文共分为六个部分：第一部分，对研究背景、研究意义以及相关参考文献进行了阐述。第二部分，介绍了文章研究的理论基础，主要包括统计学习理论和支持向量机相关理论。第三部分，介绍了预测指标的选取以及预处理，本文主要选取了能够反映公司偿债能力、获利能力、营运能力、成长能力以及现金流量能力的 24 个指标，将这 24 个指标依次进行正态分布检验、显著性差异检验以及因子分析处理之后，共提取出 2 个公因子作为模型的输入变量。第四部分，是文章的实证研究部分，作者尝试性的选择半年度作为预测期间，初始选取了 2010 年制造业 642 家上市公司，经过剔除被 ST 的公司以及业绩增减幅度小于 50% 的公司，最终经过处理后的研究样本数量为 264 家（其中 132 家业绩上升的公司，132 家业绩下降的公司），将这 264 家样本分成训练集和样本集，其中训练样本 88 家（44 家业绩上升公司，44 家业绩下降公司），测试样本 44 家（22 家业绩上升的公司，22 家业绩下降的公司），分别构建 Logit 回归模型以及 LS-SVM 模型，并将两种模型的预测结果进行了比较。第五部分，主要是将 LS-SVM 模型加以应用的案例分析。第六部分是文章的结论以及不足。

本文在财务业绩预测过程中引入最小二乘支持向量机，并采用半年度数据进行预测，最终构建的 LS-SVM 模型的预测准确率高于 Logit 回归模型的预测准确率。本文的主要创新点包括以下两点：

第一，将最小二乘支持向量机应用于财务业绩预测。通过研究发现基于最小二乘支持向量机的财务业绩预测模型判断准确率为 64.8%，模型预测准确率比 Logit 回归模型的预测准确率略高。

第二，文章的研究视角方面，文章主要进行半年度业绩预测。文章在进行研究时主要采用的是半年度数据，这就为财务业绩预测提供了一定的及时性。

关键字：财务业绩 最小二乘支持向量机 Logit 回归模型 上市公司

Abstract

Financial distress prediction and research has entered a relatively mature stage since Altman began pioneering research on it. However the research might give the forecasters a feeling of too late because the standard: whether a company is special treated. Because of this, scholars put forward a research: prediction and research of financial performance. The prediction and research of financial performance is still in an initial stage. It provides information required for stakeholders in decision-making by predicting the future performance of the company. The purpose of this study is to choose a performance prediction method for manufacturing to provide information for the stakeholders.

This article is divided into six parts: the first part describes the researching background, researching meaning and references. The second part of the article studies the theoretical basis, including the statistical learning theory and the related theory of support vector machine. The third part describes the selection and pretreatment of the predictors. There are 24 indicators which reflect the company's solvency, profitability, operational capabilities, the ability to grow and cash flow capacity. After taken normal distribution test, significant difference test and factor analysis, two common factors were extracted. Part IV is the empirical research of the article. The paper chose 642 manufacturing listed companies in 2010 initially, after removed the ST companies as well as those whose performance increasing or decreasing less than 50%, the final number of the treated sample is 264 (of which there are 132 performance increase as well as 132 performance decrease), these 264 samples were divided into training set and a sample set. There are 88 training samples (44 results rise in the company, 44 decline in performance company) and 44 test samples (22 increase in performance companies as well as 22 decrease in the performance). Logit regression model and the LS-SVM model were constructed and the predicting results of them were compared. The fifth part is the case of LS-SVM model using. Part VI is the conclusion of the article as well as the shortcomings.

The least squares support vector machine (LS-SVM) was introduced in forecasting process of financial results and Semi-annual data was used to predict. The result is that: the

prediction accuracy of LS-SVM model is higher than that of the Logit regression model. The main innovation of this paper contains the following two sides.

First, the least squares support vector machine was applied to the financial performance forecast. Through the study we found that the prediction accuracy of the LS-SVM model is 64.8%, which is slightly higher than that of the Logit regression model.

Second, the article using semi-annual indicators for forecast. During the study, semi-annual data was used which might provide financial results forecast in time.

Keywords: Financial Performance Least Squares Support Vector Machine Logit Regression Model Listed Company

目录

第 1 章 绪论	1
1.1 研究背景及意义	1
1.1.1 研究背景.....	1
1.1.2 研究意义.....	1
1.2 文献综述.....	3
1.2.1 统计类预测模型.....	3
1.2.2 人工智能型预警模型.....	8
1.3 研究方法和内容	11
1.3.1 研究方法.....	11
1.3.2 研究内容.....	12
第 2 章 最小二乘支持向量机及其相关理论	14
2.1 经验风险最小化原则	14
2.1.1 机器学习问题的简单描述.....	14
2.1.2 经验风险最小化原则及其不足.....	15
2.2 统计学习理论的主要思想	16
2.2.1 函数集学习能力的描述.....	16
2.2.2 结构风险最小化原理.....	16
2.3 支持向量机.....	17
2.3.1 支持向量机原理.....	17
2.3.2 支持向量机分类.....	18
2.3.3 最小二乘支持向量机.....	27
第 3 章 预测指标的选取及预处理	30
3.1 预测指标的选取	30
3.1.1 预测指标选取的原则.....	30
3.1.2 初始预测指标的选取.....	31
3.2 指标的预处理.....	33
3.2.1 指标数据的正态分布检验.....	33
3.2.2 指标数据的显著性差异检验.....	34
3.2.3 指标数据的多重共线性检验.....	35
第 4 章 基于 LS-SVM 的财务业绩预测实证研究	39
4.1 研究样本的选择	39
4.1.1 因变量的选择.....	39
4.1.2 自变量的选择.....	39
4.1.3 数据来源.....	39
4.2 Logit 模型.....	40
4.2.1 Logit 回归模型的形式.....	40
4.2.2 Logit 回归模型的估计.....	41
4.2.3 Logit 回归模型的检验.....	42
4.2.4 半年度预测的 Logit 模型的构建.....	43
4.3 LS-SVM 模型	45
4.3.1 基于 LS-SVM 的财务业绩预测.....	45

4.3.2 基于 LS-SVM 模型构建及实证检验.....	46
4.4 实证结果的分析与比较.....	47
第 5 章 基于 LS-SVM 的财务业绩预测案例分析.....	48
5.1 上市公司简介.....	48
5.1.1 开开实业公司简介.....	48
5.1.2 上海棱光实业公司简介.....	48
5.2 公司的财务状况.....	49
5.2.1 开开实业的财务状况.....	49
5.2.2 棱光实业的财务状况.....	49
5.3 基于 LS-SVM 的财务业绩预测结果分析.....	50
5.3.1 开开实业的预测结果分析.....	50
5.3.2 棱光实业的预测结果分析.....	50
第 6 章 研究结论及不足.....	51
6.1 研究结论.....	51
6.2 研究的不足之处.....	51
参考文献.....	53
附录.....	55
攻读硕士学位期间取得的学术成果.....	59
致谢.....	60

图表清单

图表清单

序号	图题	页码
图 1-1	论文研究框架	13
图 2-1	结构风险最小化原则	17
图 2-2	线性可分支持向量机示意图	19
图 2-3	线性不可分支持向量机示意图	23

序号	表题	页码
表 3-1	财务指标	31
表 3-2	半年度指标数据正态分布检验结果	33
表 3-3	半年度预测指标数据显著性差异检验结果	34
表 3-4	半年度预测指标数据的 KMO 与 Barlett 球形检验值	36
表 3-5	半年预测度指标变量的共同度	36
表 3-6	半年度预测指标变量系数矩阵的特征值、方差贡献率和累积方差贡献率	36
表 3-7	初始因子载荷矩阵	37
表 3-8	旋转后的因子载荷矩阵	38
表 3-9	各公因子命名	38
表 3-10	因子得分系数	38
表 4-1	样本数据分布	40
表 4-2	Logit 回归模型整体显著性检验结果	43
表 4-3	Logit 回归模型拟合优度评价指标	43
表 4-4	半年度预测 Logit 回归系数估计值及显著性检验结果	43
表 4-5	Logit 回归模型回代和预测结果	44
表 4-6	LS-SVM 模型（高斯核）回代及预测结果	47
表 4-7	两种模型预测准确率比较	47
表 5-1	开开实业各项指标数据	49
表 5-2	棱光实业各项指标数据	49

第 1 章 绪论

企业从财务正常到财务危机并非一蹴而就。因此，在企业正常经营的过程中有必要对其财务状况和经营成果进行及时的评价和监督，建立一套科学的财务业绩预测体系，从而为利益相关者提供决策所需信息。本章是论文的绪论部分，首先介绍了财务业绩预测的研究背景及研究意义；其次，对国内外有关预测方法进行了综述；最后，给出了本文的研究方法以及研究内容。

1.1 研究背景及意义

1.1.1 研究背景

信息革命和知识经济强有力地推动了世界各国通信和国际金融超高速运转的国际化，“地球村”的形成在使得世界经济日益紧密相联、兼容的同时也促进了各国之间的经济竞争并使其趋于“白热化”。在如此激烈的竞争环境下，企业能否可持续发展是众多利益相关者关注的焦点。上市公司业已成为支撑我国国民经济增长的第一生力军，然而随着市场竞争越来越激烈，上市公司的经营管理正面临越来越严峻的挑战，财务危机成为影响中国上市公司生存的重要因素。因此，研究中国上市公司经营中存在的财务风险并建立预警系统，成为中国上市公司管理者的一个难题。基于此，国际上的许多著名学者对企业危机及风险管理进行了研究。随着我国市场经济体制改革的不断深化以及证券市场的不断完善，面对日益激烈的市场竞争，在微观经济实体中建立有效的企业预警系统已不再是一个单纯的理论问题。随着研究的逐步深入，学者们逐渐发现：企业从财务正常到财务出现困境甚至破产并非是一蹴而就的，而是一个不断恶化的过程；而当企业被 ST 后再去进行亡羊补牢式的财务困境预测显然已属于扬汤止沸！鉴于此，有的学者开始提出企业财务业绩的分析与预测。通过研究企业业绩的变动趋势，为利益相关者提供决策所需的信息。对上市公司财务业绩进行分析与预测具有重要的理论意义和现实意义。

1.1.2 研究意义

（1）理论意义

①财务业绩预测有利于丰富企业管理理论。企业管理理论的核心是财务管理，财务管理工作的效率直接影响着企业的管理水平。与日常的财务管理不同的是：企业财

务业绩预测和管理的客体拥有不确定性和前瞻性。时至今日，我国多数学者均注重于财务困境预警模型的研究，而对财务业绩预测和分析的研究关注甚少。因此，有关财务业绩的评价体系不是很完善，财务业绩预测也缺乏理论和实践指导，这可谓是财务管理理论的不足。因此，可以说对企业财务业绩预测和分析的研究在一定程度上对财务管理理论进行了充实，对丰富我国的企业管理理论具有非同小可的重要意义。本文以中国上市公司为研究对象，从财务风险的基本理论入手，研究中国上市公司财务风险的种类、发生的原因以及预警问题。本文将支持向量机（SVM）这个目前比较热门的机器学习算法应用到企业财务危机预警领域，建立了企业 LS-SVM 财务预警模型，为研究企业财务预警问题提出了一种新的建模思想和方法，拓宽了该领域的研究范围。

②企业财务业绩预测有利于丰富危机管理理论。危机管理理论得到长足发展是在 20 世纪 70-80 年代，该理论主要包含对危机的事前、事中和事后所有方面的管理以期实现以最小的费用获得最大的效果。资源稀缺性和市场竞争性促使企业推行危机管理以保证企业财产和经营安全、避免出现财务困境。而对企业财务业绩预测能够从新的角度对企业危机管理提供依据。因此，对企业财务业绩预测和分析的研究在一定程度上丰富了危机管理理论。

③有利于丰富委托代理理论。经济权与所有权相分离引致了委托代理问题，而委托代理问题的存在又是人们进行财务业绩预测的动机。利益相关者总是想通过一定的标准来评估管理人员是否尽职尽责、恪尽职守。他们评估的标准既包括对历史财务业绩的考核，也包括对未来业绩的预测。通过对未来业绩的预测，利益相关者对管理者的行为形成一个预期的标准，当经营管理者的行为发生后，利益相关者则将经营管理者的行为效应与其所预期的标准加以比较，以便及时掌握管理者的行为动态。这就有利于从一个更广泛的层次去理解委托代理理论的应用，有利于进一步丰富委托代理理论伸展和健全。

④有利于丰富经济人理论假设的应用。经济人假设将所有参与经济的微观主体视为利益最大化追求者，他们总是期冀以最小的投入获得最大的产出，即达到一个帕累托最优。经济人假设是财务业绩预测行为产生的推动力。正是因为各利益相关者的行为满足经济人的假设，所以他们更倾向于把握未来，预知未来，以便及时做出应急之策。另外，利益相关者的机会主义行为是财务业绩行为存在的另一个推动力。故而通过对财务业绩预测问题的研究可以在一定程度上解释经济人假设和机会主义行为。

(2) 现实意义

①宏观方面。宏观经济是由各微观经济单元组成的。各微观经济实体的不稳定必然会引起整个宏观经济的不稳定。上市公司蝉联不绝的频繁破产不仅会影响到证券市场的稳定繁荣，而且还会引发大量失业乃至财政、经济以及社会的动荡不安，促使整个国民经济体系受到严重影响和破坏，各种资源得不到有效的利用从而增加了国家的社会成本和经济成本，削弱了国家的经济实力，制约了 GDP 的可持续增长。我们通过对企业财务业绩预测和分析，可以针对企业的财务状况以及是否会发生危机进行事前的准确预测，从而有利于我们国家宏观制定出合理的资源配置政策和规避风险的政策方针。

②微观层面。在微观层面上，财务业绩预测和分析可以匡扶襄助利益相关者做出科学而客观的决策。对于企业自身来说，企业管理者通过财务业绩预测可以评价现有和预期财务状况、分析企业是否会发生财务困境并分析原因，找出症结所在，以便及时制定出与之相对应的经营和财务政策；而对于投资者：则可以通过对企业财务业绩预测来分析企业的经营和财务状况，以便做出正确的投资决策；对于债权人：通过分析、预测企业的财务业绩，评价企业的还本付息能力，以便做出合理的信贷政策和及时的收款决策；对于政府监管部门：通过对企业财务业绩的预测和分析，能够制定出合理的资源配置政策；对于税务部门：有利于分析企业经营能力，制定合理的征税政策，保证财政收入的稳定；对于中介机构：比如审计统计、会计师事务所、注册税务师事务所等相关部门、单位、机构，可以通过对企业的财务业绩评价，制定合理的审计核查计划和方案等。

1.2 文献综述

企业财务业绩预测的研究，在国内外才刚刚兴起。但是，对于企业财务困境的预测，国内外学者早已进行了大量的研究。企业财务业绩预测研究可谓是财务困境预测研究的提前化，故而对于在财务困境所应用的研究方法，在很大程度上，可被企业财务业绩预测研究采用。下面着重将相关的研究方法加以综述。

1.2.1 统计类预测模型

传统统计类预测模型主要包括：一元判别分析模型、多元判别分析模型、线性概率分析模型、累积求和模型以及生存分析法等。

(1) 一元判别分析模型

Fitz Patrick 最早将一元判别分析模型应用于财务困境预测，他于 1932 年在“*A Comparison of Ratios of Successful Industrial Enterprises with Those Failed Firms*”一文中，通过单个财务比率把 19 家研究样本划分为两组：破产组和非破产组。通过研究，Fitz Patrick 得出：净资产利润率以及产权比率的预测效果最好^[1]。

Secrist 选用一、二个财务指标，通过简单图表法比较了不同时期经营失败银行和经营成功银行在财务指标上差异，以判断银行是否存在财务困境。

继 Fitz Patrick 之后，Beaver (1966) 提出的一元判别分析模型最具有代表性，他通过选取 1954-1964 年期间的 79 家失败公司以及同行业 79 家近似资产规模的成功公司为样本，利用二分类检定法、剖面分析对样本破产前五年的 29 个财务指标进行逐年比较。通过研究，Beaver 得出：成功样本与失败样本的财务比率存在显著性差异，随着破产日的临近，模型误判率越低；其中总资产利润率的预测效果最好，模型误判率最低，“净收入/总资产比率”次之^[2]。

在我国，财务困境的研究起步较晚。国内学者陈静首先运用上市公司作为样本进行财务困境研究。陈静（1999）以 1998 年的 27 家特殊处理公司和 27 家正常公司为样本，选取资产负债率、净资产收益率、总资产报酬率、流动比率等 6 个财务指标进行了一元判别分析，在分析中发现：模型在 ST 前三年的预测效果很好，其中误判率最低的财务指标是流动比率和负债比率^[3]。

虽然，一元判别分析模型在财务困境预测中起到了一定的成效，但是仍然存在诸多不足。比如单个财务指标在反映企业整体财务状况时表现乏力；不同财务指标对同一企业预测结果可能出现分歧；促使企业经理人为达到目标而粉饰某个财务指标。

(2) 多元判别分析模型

为了规避一元判别分析模型存在的诸多弊端，国内外学者开始将多元判别分析模型引入财务预测研究领域。多元判别分析模型是指在进行财务预测时，将多个财务比率进行汇总，最终得出一个判别值达到预测的目的。

①Z 值模型。1968 年，美国学者 Edward Altman 在《财务比率、判别分析和公司破产预测》一文中首次将多变量判别模型引入到财务困境预测领域。他选取 1946-1965 年间 33 家申请破产的制造企业和 33 家同行业近似规模运营正常企业的作为研究样本，从 22 个财务指标中筛选了 5 个最具影响力指标并建立了著名的 Z 值模型。分析指出：Z 值越小，企业陷入财务困境的可能性越大。Z 值高于 1.8 则可判定这家公司

的财务状况良好；若 $Z < 1.8$ 则可判定这家公司的财务状况不良，甚至会发生财务危机^[4]。

Altman 得出的最终判别方程是： $Z = 0.012X_1 + 0.014X_2 + 0.033X_3 + 0.006X_4 + 0.999X_5$ 。其中， X_1 反映企业的流动性及生产规模，表示为营运资金(WC)与资产总额(TA)的比值； X_2 反映企业的获利能力，表示为留存收益与资产总额的比值； X_3 反映企业的增长能力，表示为息税前收益(EBIT)与资产总额(TA)的比值； X_4 反映企业的偿债能力，表示为权益市价(MVE)与债务总额(TL)的比值； X_5 反映企业的营运能力，表示为销售额(S)与总资产(TA)的比值。

②ZETA 模型。随着经济环境的不断变化，原来的 Z 模型已无法准确的预测企业的财务危机。1977 年，Altman, Haldeman 和 Narayanan 在 Z 值模型的基础上，提出了 Zeta 模型。他们以美国 1969-1975 年间 53 家破产企业和配对的 58 家非破产企业为样本，选取 27 个初始财务指标进行判别分析，最终选取 7 个财务指标（总资产利润率、收益稳定性、利息保障倍数、留存收益/总资产、流动比率、普通股权益/总资本、普通股权益/总资产）建立 ZETA 模型。研究表明，在企业财务困境发生前五年和前一年的预测准确度分别达到 70% 和 91%，模型预测效果比 Z 模型更好^[5]。

基于在 Beaver 和 Altman 的研究基础，Deakin 运用 1964-1970 年间 32 家正常公司及其配对样本，重新进行了 Beaver 的两分检验，最终得出了经营失败的判别模型。Deakin 指出，运用现金流/总负债比率得出的结果与 Beaver 的研究结果相近。而 Beaver 则认为，在公司破产前一到三年内，净利润可以得到相同的预测准确性^[6]。

在比较上述学者研究成果的基础上，Scott 认为：多元判别分析模型的预测准确性要高于一元判别分析模型，但是至于哪个模型最优，难以得出准确的结论。总的来说，在多元分析判别模型中，ZETA 模型具有较高的分类能力，其预测准确性最好^[7]。

陈肇荣采用中国台湾地区的企业财务资料作为研究样本，建立了多元判别函数模型： $Z = 0.35X_1 + 0.67X_2 - 0.57X_3 + 0.39X_4 + 0.55X_5$ ，其中， X_1 表示速冻比率； X_2 表示营运资金与资产总额的比率； X_3 表示固定资产与资本净值的比率； X_4 表示应收账款与销售净额的比率； X_5 表示现金流入量与现金流出量的比率。但是，该模型没有给出企业破产与否的临界值和警度区间。

③日本开发银行模型。20 世纪 70 年代，日本开发银行利用与 Z 值模型相关的分析方法，选择东京证券交易所 310 家上市公司作为样本，建立破产模型，进行了财务危机预测。模型为： $Z = 2.1X_1 + 1.6X_2 - 1.7X_3 - X_4 + 2.3X_5 + 2.5X_6$ 其中， X_1 表示销售额增长

率； X_2 表示总资本利润率； X_5 表示股利分配率； X_4 表示资产负债率； X_5 表示流动比率； X_6 表示粗附加值生产率，判别函数的Z值越大，表明企业财务状况越是“优秀”；反之则是不良的表现。

④F分数模型。国内学者周首华等（1996）在Z分数模型的基础上引入现金流量变动情况指标，选用1977-1990年间的62家公司为样本建立F分数模型，经验证明F模型的判别准确率高达近70%。F分数模型表示如下：

$Z=0.1774+1.1091X_1+1.1074X_2+1.9271X_3+0.0302X_4+0.4961X_5+0.4961X_5$ ，其中， X_1 、 X_2 及 X_4 与Z计分模型中的 X_1 ， X_2 及 X_4 表达式相同，这里不再赘述； X_3 表示为(税后纯收益+折旧)与平均总负债的比率（即现金净流量/平均总负债）； X_5 表示为(税后纯收益+利息+折旧)与平均总资产的比率。在F分数模型中， X_3 和 X_5 引入了反映企业现金流量的指标。其中， X_3 反映企业的偿债能力，表示的是企业可以用来偿还债务的全部现金流量； X_5 反映企业的获利能力，表示为企业总资产能够产生现金流量的能力。F分数模型中F的临界值是0.0274。当公司的F值小于0.0274时，意味着企业濒临破产；当企业的F值大于0.0274时，该企业可持续经营^[8]。

⑤与Z模型相比，F模型具有其自身的独特性：首先，将现金流量引入F分数模型，提高了模型预测准确性；其次，更新反映企业财务状况的标准，比如流动比率；再者，扩大了研究样本的范围。在F分数模型中，周首华等选取了4160家公司的数据作为研究样本，远远多于Z计分模型中66家研究样本数。

张玲（2000）在沪深两市选择了14个行业的120家上市公司作为研究样本。他将样本划分成两组：开发样本和检验样本。开发样本包括60家公司，其中30家是ST公司，30家是作为配对的绩优公司。他从偿债能力、盈利能力、资本结构状况和营运状况等四个方面选取了15个相关财务比指标进行研究，经过检验后保留了4个变量用以构建模型。通过研究得出，若开发样本为公司被ST前1-5年，则模型判别准确率分别达到100%，87%，70%，60%和22%；模型在公司被ST前四年可以起到预测公司前景的作用^[9]。

高培业、张道奎（2000）以深圳国有企业作为研究样本，筛选出161家制造企业和140家非制造企业。在制造类企业中有79家失败公司和82家成功公司；在非制造业中有66家失败公司和74家成功公司。利用多元判别分析技术，他们分别设计出了制造业与非制造业的判定模型。研究发现，制造企业的模型误判率较低，为6.25%，非制造业模型在对非失败公司的判定上准确度可达到100%；制造业公司、非制造业

公司分别受负债比率和营运资本比率的影响较大^[10]。

向德伟（2002）选取沪深两市的 80 家企业的 A 股作为研究样本，进行了 Z 值模型分析。通过研究，他得出我国上市公司大体上支持“Z 值”的有效性，但也存在个别特殊情况。作者认为 Z 值分析针对于某一行业具有很强的指导性^[11]。

基于 Z 值模型的基础，杨淑娥（2003）选取了 67 家财务失败公司和 67 家配对公司，通过主成分分析法进行研究并提出了 Y 值模型。他得出当 Y 值为 0.5 时，模型的预测准确率可以达到 86%^[12]。

虽然多元判别分析模型在一定程度上对样本公司的财务状况起到预警的作用，而且有些模型的预测率很高，但是这种分析方法往往局限于样本的影响：根据样本 A 预测出的模型可能对样本 B 的预测效果不好。另外，这些模型有用性不强，在实务中缺乏应用和推广。

（3）线性概率分析模型

线性概率分析模型主要包括 Logit 模型和 Probit 两种。两种模型的因变量都是属于分立性型态，Logit 模型和 Probit 模型的区别在于二者具有不同的概率累计函数。

Ohlson（1980）最先采用 Probit 方法进行财务困境预警研究。他选取 1970-1976 年间 105 家破产公司和 2058 家非破产公司组成配对样本，采用极大似然法构造模型。研究发现，资产总额、资产负债率、流动资产/资产总额、流动比率、现金净流量/期末总资产、营运资本/资产总额等指标判断效果非常好，在临界值取 0.038 时，模型的误判率为 14.9%。与此同时，Ohlson 采用 Logistic 方法对其选择的样本进行研究并建立了财务困境预警模型，通过针对企业资本结构和经营绩效以及其规模与流动性等要素深入的研究发现：对这四个因素进行财务困境预警的准确率可达到 96.12%^[13]。

以中国上市公司为研究样本，陈晓和陈治鸿对公司财务困境预测进行多元逻辑回归，研究结果表明：在上市公司可以获得的财务数据中，负债—权益比率、应收账款周转率、总资产利润率以及预留收益总资产比率对我国上市公司的财务困境预测具有显著性影响，而且模型判断准确率可以达到 78.24%^[14]。

以 LR 线性回归模型为工具，国内学者张后奇等针对不同的样本分别建立了不同的预警模型。通过研究，他们得出：所建立的模型的预测准确率能够接近 70%^[15]；姜秀华通过研究提出 Logit 模型的最佳分割点为概率 0.1；齐治平在 Logit 模型中引入二次项和交叉项进行研究^[16]；宋力、李晶对财务数据进行调整后建立 Logit 模型；吴世农、卢贤义 2001 年以 70 家处于财务危机的公司和 70 家配对公司为样本，通过对

Fisher 线性判定分析、多元线性回归分析和 Logistic 回归分析三种方法检验得出 Logistic 回归模型的精度最高。

(4) 生存分析法

生存分析是研究某一事件（即死亡事件）的发生概率，随着时间推移而不断变化的规律。生存分析法不仅关注企业的生存或死亡还关注企业的生存时间。

Lane (1986) 最早应用生存分析模型进行预测，它选取 1978-1984 年间 130 家倒闭银行，以及成立年限和规模相近的 334 家健全银行。从 21 个财务指标中筛选出重要的变量，以 Cox 比例风险模型和多变量判别分析模型建立危机前一年和前二年的财务预警模型并比较两种模型的预测能力。实证结果发现，两种预测模型在总分类正确性上无显著性差异，但 Cox 比例风险模型比多变量判别分析模型有较小的第一类错误。

Chen 和 Lee (1993) 利用生存分析法 (Survival analysis) 选取 1980-1988 年间的 75 家石油企业进行研究，结果表明，流动性比率、财务杠杆比率、营业现金流、开采成功率、企业历史和规模对企业能否存活影响巨大。

国内学者张忙利 2008 年以上市公司 IPO 作为起点事件，根据生存分析的研究特点，遵循市场原貌，采用非配对抽样原则，选取 1996—2003 年之间上市的所有制造公司，剔除财务数据缺失的公司总计共 401 家非 ST 公司和 104 家 ST 公司为研究样本建立模型，结果显示生存分析的财务预警模型展现出与其他分析方法类似的规律，即前一年的预测效果较好，随着时间的向前推移，预测效果递减。

1.2.2 人工智能型预警模型

随着信息化、网络化的不断发展，为了更准确地进行财务困境预测，必须及时获取、处理各种动态的财务信息，许多学者开始考虑将人工智能的技术应用于财务困境预警的模型。所谓人工智能技术 (Artificially Intelligence, AI) 是这样一种技术：它以专家系统 (Expert System, ES) 为基础工具，通过对现有知识的学习及推导，以期实现或接近于人类智能的水平。人工智能技术能够针对不断输入的动态数据，及时并持续地学习和调整学习算法，从而为最终判别结果的可靠性提供一定的保证。为了较为客观地进行财务业绩预测，需要及时提供各种与企业经营状况相关的最新资料和数据，从而对预测模型形成不断地调整及修正，最终实现预测模型准确性的提高。故而，一些学者开始将 AI 技术应用于财务预测。人工智能型预警模型主要包括：人工神经

网络 (Artificial Neural Network, ANN) 模型、遗传算法 (Genetic Algorithm, GA) 模型、粗集理论 (Rough Set Theory, RST) 模型、递归划分分析 (Recursively Partitioned Analysis, RPA) 模型以及支持向量机 (Support Vector Machine, SVM) 模型。

(1) 神经网络模型 (ANN)

神经网络模型是将神经网络的分类方法应用于财务预警。ANN 作为一种平行分散处理模式, 是对人类大脑神经运作的模拟。ANN 具有较好识别能力、容错能力以及学习能力。

Odom (1990) 最早把人工神经网络技术应用于财务困境预测研究, Odom 和 R.Sharda 在研究中比较了 BP 网络模型和 Altman 判别分析法在企业破产前一年的判别能力, 结果表明 BP 神经网络模型的预警效果较佳^[17]。Coats & Fant (1993) 采用 ANN 模型对 47 家财务危机公司和 47 家正常公司预测, 模型的预测准确率达到 91%^[18]。

此外, Tam and Kiang Altman, Marco & Varetto (1994) 等众多学者比较不同方法对财务危机预测的准确度。结果表明: ANN 比以往的线性分析模型预测准确度更高。

国内学者杨保安 (2001) 利用 BP 网络对企业的状态进行分类; 胡燕京 (2002) 通过改进的 BP 神经网络模型建立了关于我国金融风险的预警模型; 李秉祥 (2004) 用对我国 ST 公司采用模糊神经网络工具进行了财务危机预测^[19]; 吴德胜、梁樑、殷尹 (2004) 按照 ROE 将上市公司的财务状态分为四种 ($ROE > 10\%$, 非常安全; $5\% < ROE < 10\%$, 正常; $0 < ROE < 5\%$, 财务困境), 运用神经网络建立财务预警模型, 结果显示模型的短期预测准确率达 81.25%, 中期预测准确率为 56.25%^[20]; 杨淑娥、黄礼 (2005) 以 120 家上市公司的截面财务指标为建模样本, 以同期 60 家公司为检验样本运用 BP 人工神经网络工具, 建立了财务危机预警模型。研究结果表明建模样本的预测准确率为 90.8%, 检验样本的预测准确率为 90%^[21]。

(2) 遗传算法模型 (GA)

遗传算法是一种基于自然选择和基因遗传学原理的优化搜索方法。Varetto Franco (1998) 用遗传算法分别提取线性函数和判别规则, 结果表明遗传算法可获得不受统计约束的最优线性方程, 与 MDA 相比, 该方程省时且受主观影响小, 但结果逊于 MDA^[22]; Shin & Lee (2002) 以财务比率为基础采用遗传算法建立的预测模型, 模型结构清楚、容易理解且能对定性变量进行规则提取^[23]。

(3) 粗糙集理论模型 (RST)

RST 方法可用一组多价值属性的财务比率有效地描述困境与非困境公司,并揭示财务特征与公司失败风险之间的关系。将 RST 应用到财务困境预测研究的学者主要有: Franeis EH、Dimitras AI、Pawlak、Ziarko、Dimitras AI 等。

国内学者张华伦、孙毅(2006)建立了一种基于粗集-模糊神经网络(Rough-Fuzzy-ANN, RFA)的预测模型并给出了相应的算法。通过实证分析得出 RFA 模型预测精度高、学习和泛化能力强,适应性广,能够为企业财务困境的动态预测提供一条新的途径^[24]。

(4) 递归划分分析模型(RPA)

Frydman(1985)首次采用 RPA 建立财务预测模型,研究发现 MDA 模型对研究样本的期望误判成本明显高于 RPA 模型^[25]。DKEE(2000)将熵值和 RPA 结合,通过建立决策树进行企业破产预测,结果发现其预测精度在 95%以上。

(5) 支持向量机模型(SVM)

Fan A etc.在 2000 年以标准支持向量机为基础构建了财务危机预警模型,在处理变量之间的距离这一问题上,他采用的是欧氏距离方法并实现了筛选预测指标的的目的。虽然该方法对支持向量机模型的预测性能是否达到最优难以提供保证,但是同以其他方法对指标进行选择的支持向量机模型相比,该种方式具有相对较低的计算成本^[26]。2003 年 Van Gestel T etc.以最小二乘支持向量机理论为基础,构造了基于最小二乘支持向量机的财务困境预警模型,研究结果表明该模型的预测准确率可以高达 89.91%^[27]。2005 年 Jae H Min 和 Young-Chan Lee 对 SVM 的核函数进行了优化,优化的方法是借助表格搜索技术,通过优化处理后所构建的 SVM 模型的预测效果显著优于 MDA、LR 和 BP-ANN 模型^[28]。Kyung-Shik Shin etc.于 2005 年构建了基于 SVM 的财务危机预警模型,并利用该模型对韩国的中型制造业企业进行了实证研究,通过研究发现:SVM 模型的预测准确率要显著优于 BP-ANN 模型;2005 年 Wolfgang Hardle etc.以美国 2001-2002 年间 84 家企业为研究样本,构建 SVM 财务困境预警模型,研究结果表明支持向量机模型在分类问题上具有理想的研究效果。

2005 年李贺、冯天谨以我国烟酒行业的 50 家上市公司作为研究样本,分别利用 SVM 模型和 ANN 模型进行财务危机预警并将研究结果加以比较,结果表明:在预测效果及预测准确性方面,SVM 模型要明显优于 ANN 模型^[29]。徐晓燕于 2006 年将 Logit 回归模型与 SVM 模型加以结合,提出了一套新的集成预测方法,即 LR-SVM。研究结果表明:利用该方法进行预测,其效果要显著优于一般的支持向量机^[30]。

2006 年姚宏善对我国的上市企业财务困境预测进行了比较细致的研究, 并通过研究对 SVM 模型和 MDA 模型在以下两方面进行了比较: 预测企业的准确率以及预测企业的鲁棒性。通过研究得出启示: MDA 模型无论是在准确性还是在鲁棒性方面都稍逊于 SVM 模型, 换句话说, SVM 模型在对财务困境的识别问题上表现的更加准确^[31]。

宋姣于 2007 年以中国上市公司为研究样本构建了基于支持向量机财务危机预警模型。通过研究得出: 支持向量机在小样本情况下具有较好的预测效果, 它克服了人工神经网络等方法的不足, 预测准确性远远高于其他预测方法^[32]。

蒋艳霞和徐程兴于 2009 年利用 AdaBoost 算法对支持向量机进行了改进, 并通过实验法来实现对支持向量机核函数的选择, 并对各个核函数及其相关参数的预测效果进行了逐一核算和比较, 研究结果表明支持向量机方法在模型预测准确性方面远远高于其它预测方法^[33]。

赵冠华于 2009 年首次提出了一种基于 Renyi 熵的最小二乘支持向量机的增长记忆算法, 并且在企业财务困境分析与预测过程中引入了基于生物遗传机理的遗传算法参数优化技术。通过研究得出结论, 遗传算法在更大的范围内能够实现自动寻优的功能, 而且在提高模型预测准确性问题上具有较大成效。特别是在基于 Renyi 熵的最小二乘支持向量机增长记忆算法模型中引入遗传算法, 有利于实现在小样本数据条件下, 获得较高的预测正确率^[34]。

通过诸多学者的研究可知, SVM 模型具有如下几点优势: 首先, 适用于小样本研究, 受样本量限制较小; 其次, 在一定程度上可以规避共线性影响。支持向量机对于变量之间是否存在共线性的影响不敏感, 即可在一定程度上规避变量之间共线性对预测结果的影响。再者, SVM 模型规避了维数问题。分类问题多设计维数问题, 而 SVM 模型避开了维数问题, 算法的复杂与否与维数无关。第四, SVM 模型的泛化能力较强。虽然支持向量机模型具有诸多优点, 但是金无足赤, 人无完人, SVM 模型难以避免存在一定的局限性, 比如, SVM 对特征集和核参数提出了较严格的要求, 倘研究样本等难以达到要求势必会影响到模型的预测效果。

1.3 研究方法和内容

1.3.1 研究方法

文章的研究目的是期冀建立一种符合我国上市公司财务业绩预测与分析的模型, 从而向广大利益相关者提供有助于决策的信息, 比如哪些公司财务业绩下滑, 有发生

危机的可能从而引起相关者的警觉；哪些公司财务业绩上升或持平，值得投资。特别是对于那些处于财务业绩下滑状态的企业，可以为公司内部管理人员提供信息，使其做到未雨绸缪、防患于未然。

在文章的研究过程中，综合运用了以下几种研究方法：

(1) 在企业财务业绩预测指标体系的选取和评价方面，文章主要采用决策分析和统计分析的方法。这是由于，财务业绩预测的内涵亦即通过对一系列指标的运用而做出决策之过程。故而，在预测指标体系的选取上应当考虑多目标决策的要求，也就需要运用决策分析的方法。同时，财务业绩预测是基于历史样本数据，故而文章还采取了统计分析的方法。

(2) 在对企业财务业绩预测方法和研究部分，文章主要采用实证研究方法。特别是对中国制造业上市公司的财务数据的研究。此举之目的在于提出一种符合中国国情、适用于制造业财务业绩预测的研究方法。

另外，在文章的第四部分还穿插应用了比较分析法。主要是将 Logit 回归模型预测准确率与 LS-SVM 模型的预测准确率进行了比较。

1.3.2 研究内容

国内外学者多专注于财务困境的研究，且在一定程度上颇有成效，但是对于财务业绩预测方面的研究，尚处于起步阶段。因为我国股市成立的时间较短，可以获取的样本数据数相对匮乏，而通过一般的方法不易获取较好的财务业绩预测效果。当前我国学者对于财务业绩预测的相关研究尚不深入，预测方法相对单一，模型的泛化能力不强。文章在结合我国实际国情的基础上，选用了最小二乘支持向量机作为预测模型，并主要以中国信息技术业、房地产业以及批发零售业上市公司的财务数据为基础进行了实证研究。

第1章，是整篇文章的绪论，主要介绍文章研究的框架结构。文章首先介绍了研究背景与研究意义；其次，对预警模型相关文献进行综述；最后，对研究内容、研究方法以及文章的研究框架进行了阐述。

第2章，对统计学习理论和支持向量机理论进行了概述。首先，之于经验风险最小化原则进行阐述；其次，针对统计学习理论的主要思想进行了概述；最后，对支持向量机原理以及最小二乘支持向量机用于预警的过程进行了描述。

第3章，主要是构建财务业绩预测的指标体系。首先，文章以其他学者对财务困

境预测及财务业绩预测过程中所选取的指标为基础，给出了本文所采用的预测指标体系，主要以财务指标为主；其次，通过对研究样本的初始财务指标进行正态分布检验、显著性差异检验和多重共线性检验，剔除了无显著性差异的指标、消除了多重共线性的影响，并最终形成了用于业绩预测的财务指标体系。

第 4 章，主要是运用最小二乘支持向量机的方法对财务业绩预测进行实证研究。在对研究样本进行了选择之后；Logit 回归分析以及 LS-SVM 模型，并对实证结果进行了分析与比较。

第 5 章，主要选取制造业所有上市发行 A 股的公司，通过运用最小二乘支持向量机模型进行预测，对模型的可用性及预测准确性做了进一步说明。

第 6 章，主要综述本文的主要成果及不足，并对该课题将来的发展动向进行了简述。

本文的研究框架如图 1-1 所示。

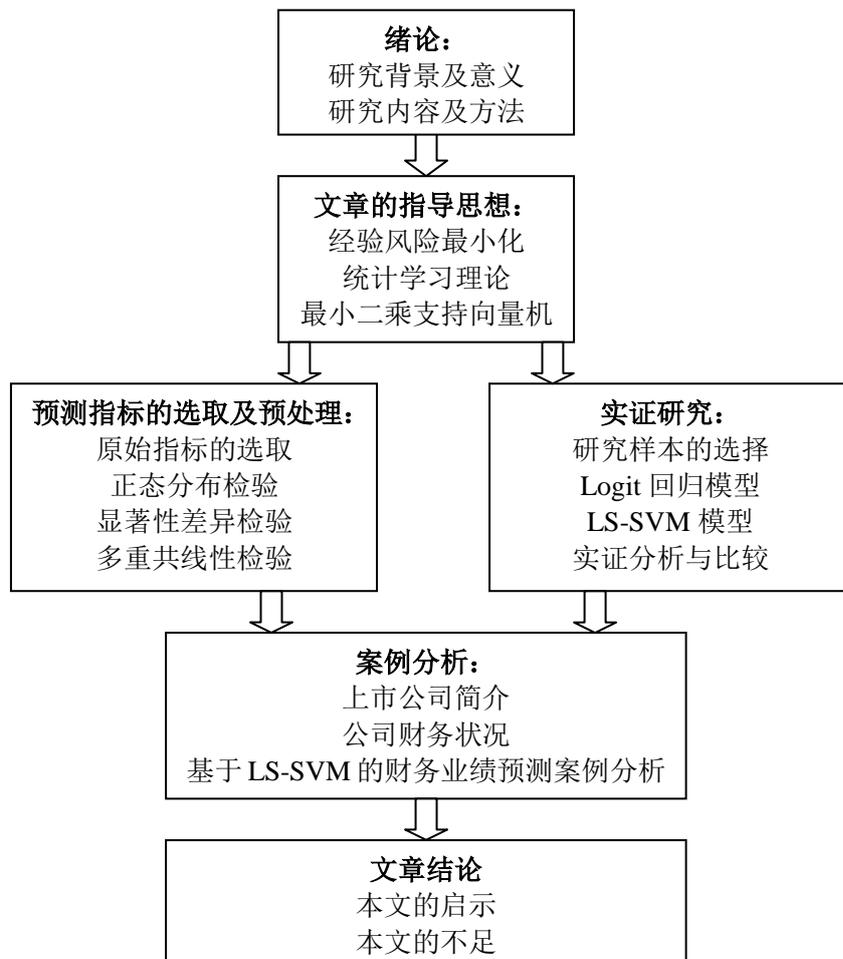


图 1-1 论文研究框架

第 2 章 最小二乘支持向量机及其相关理论

任何实证研究都需要一定的理论作为指导，企业财务业绩预测和研究概莫能外。因为本文主要研究最小二乘支持向量机在财务业绩预测中的应用，所以，在进行财务业绩预测实证研究之前，应当对最小二乘支持向量机及其相关理论加以阐述。首先，对经验风险最小化原则及其存在的不足进行了阐述；其次，对统计学习理论的主要思想进行了介绍；最后，阐述了最小二乘支持向量机的相关理论。

2.1 经验风险最小化原则

2.1.1 机器学习问题的简单描述

学习是一个不断获取知识的过程，是人类认识过程中的必要手段。人类的认知能力在不断的学习过程中得到迅速提升。机器学习，简单地说，就是机器模仿人类进行学习的能力。人工智能学科诞生于 1956 年，随着各方面的不断发展，机器学习理论应运而生。机器学习理论是人工智能学科发展到一定阶段的产物，是人工智能研究内容的重要组成部分。随着研究的深入，机器学习理论越来越受到关注。不少学者也试图给机器学习进行定义，但是至今尚未形成统一的结论。H.A.Simon 对“机器学习”的观点是：认为机器学习是“系统在不断重复的工作中对本身能力的增强或改进，使得系统在下次执行同样的任务或者类似的任务时，会比现在做得更好或效率更高”。从经济学的角度而言，机器学习是系统在重复任务的过程中所带来的边际效益递增。

机器学习的实质是寻找一个函数 $f(x)$ 作为 y 的预测值，并使得这一预测值与 y 的差距最小。机器学习理论的数学定义为：给定一组独立同分布样本， $(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)$ ， $x_i \in R^n, y_i \in R, i = 1, 2, \dots, l$ ，在选定的函数集 S 中寻找一个函数 $f(x)$ ，并使其经验风险最小。经验风险可以表述为：

$$R(f) = \frac{1}{2} \int C(x, y, f(x)) dP(x, y) \quad (2.1)$$

其中 C 为损失函数， $f(x)$ 为决策函数，函数 $P(x, y)$ 的概率分布是未知的。

机器学习主要包含三类：回归估计、概率密度估计和模式识别，而这三类问题又分别对应着不同的损失函数。

其中回归估计的损失函数被表述为：

$$C(x, y, f(x)) = (y - f(x))^2 \quad (2.2)$$

概率密度估计问题的损失函数为：

$$C(x, y, p(x)) = -\log p(x) \quad (2.3)$$

模式识别问题的损失函数可以描述为：

$$C(x, y, f(x)) = \begin{cases} 0, & y = f(x) \\ 1, & y \neq f(x) \end{cases} \text{其中 } f(x) \in S \quad (2.4)$$

本文所进行的财务业绩预测实质上是一个模式识别问题。

2.1.2 经验风险最小化原则及其不足

(1) 经验风险最小化原则

上文已经说到利用机器学习的目的是希望实现期望风险最小化，然而由于事先假定的概率分布是未知的，所以只能用经验风险来代替期望风险。经验风险的定义为：给定训练样本集 $T = \{(x_i, y_i), i = 1, 2, \dots, l\}$ ， $x_i \in R^n, y_i \in \{+1, -1\}, i = 1, 2, \dots, l$ ，及其损失函数 C ，则经验风险可以定义为：

$$R_{emf}(f) = \frac{1}{l} \sum_{i=1}^l C(x_i, y_i, f(x_i)) \quad (2.5)$$

经验风险最小化的基本原则为：对于任意的训练集 T 和假设集 $S = \{f : x \in R^n \rightarrow y \in \{+1, -1\}\}$ ，在 S 中寻找能够使 $R_{emf}(f)$ 最小化的决策函数 f 。

(2) 经验风险最小化的不足

经验风险最小化原则的不足主要体现在以下两点：

①缺乏客观性。采用经验风险替代期望风险来选择决策函数，缺乏客观的论证，仅是想当然的方法；

②易导致“过学习”问题。以经验风险最小化为原则容易产生盲目追求最小化的现象，从而导致模型推广能力受到限制。

③假设集 S 缺乏限制。举例而言：给定样本集 T ，设

$$f(x) = \begin{cases} y_i, & x = x_i \\ 1, & x \neq x_i, i = 1, 2, \dots, l \end{cases} \quad (2.6)$$

很明显，该假设也在 $S = \{f : x \in R^n \rightarrow y \in \{+1, -1\}\}$ 之中，而且 $R_{emf}(f) = 0$ ，但是该假设不能作为决策函数。

如前所述,经验风险最小化与期望风险最小化是不同的,我们需要寻找一种理论,该理论适宜于小样本前提下的有效学习以及广泛应用,即统计学习理论。

2.2 统计学习理论的主要思想

在传统统计学不断发展以及延伸之后,逐渐形成了统计学习理论,该理论专门用于研究小样本情况下的学习和推广,是支持向量机的理论基础。

2.2.1 函数集学习能力的描述

由上文可知我们要寻找决策函数 f ,需要对假设集 S 加以选择,而选择 S 的关键在于 S 的大小,也就是 S 的学习能力(也叫表达能力),而对这种学习能力加以描述的正是 VC 维。

VC 维主要用于描述假设集 S 的表达能力。其定义为:对于假设集 S ,若其中的假设能够将存在的 h 个样本点按所有可能的 2^h 种形式区分开,则称 S 能够将 h 个样本点打散。如果对于任意的样本都存在假设可以将其打散,那么这个假设集 S 的 VC 维将是无穷大。VC 维越大,说明假设集的表达能力越强,学习机器也就越复杂。比如在 n 维实数空间中,函数:

$$f(x,w) = \sin(wx), w \in R \quad (2.7)$$

的 VC 维是无穷大,这意味着在实数范围内,该函数可以将所有的样本点打散。

2.2.2 结构风险最小化原理

在此,令假设集 S 的 VC 维为 h ,训练样本集的个数为 l ,那么,当下列表达式:

$$l > h \quad (2.8)$$

$$h \left(\ln \frac{2l}{h} + 1 \right) + \ln \frac{4}{\delta} \geq \frac{1}{4} \quad (2.9)$$

成立时,对于任意给定的概率分布 $P(x,y)$ 以及任意的 $\delta \in (0,1]$,任意假设 $f(f \in S)$ 皆能够使如下不等式以至少 $1-\delta$ 的概率成立:

$$R(f) \leq R_{emf}(f) + \sqrt{\frac{8}{l} \left(h \left(\ln \frac{2l}{h} + 1 \right) + \ln \frac{4}{\delta} \right)} \quad (2.10)$$

特别情况,当 $R_{emf}(f)=0$ 时,有

$$R(f) \leq \sqrt{\frac{8}{l} \left(h \left(\ln \frac{2l}{h} + 1 \right) + \ln \frac{4}{\delta} \right)} \quad (2.11)$$

其中 $\sqrt{\frac{8}{l} \left(h \left(\ln \frac{2l}{h} + 1 \right) + \ln \frac{4}{\delta} \right)}$ 为置信风险， $R_{emf}(f) + \sqrt{\frac{8}{l} \left(h \left(\ln \frac{2l}{h} + 1 \right) + \ln \frac{4}{\delta} \right)}$ 为结构

风险，结构风险是期望风险的上界。经验风险依赖于决策函数 f 的选择，随着 VC 维 h 的增加而减少，置信风险则随着 VC 维 h 的增加而增加。结构风险最小化原则的示意图如图 1。

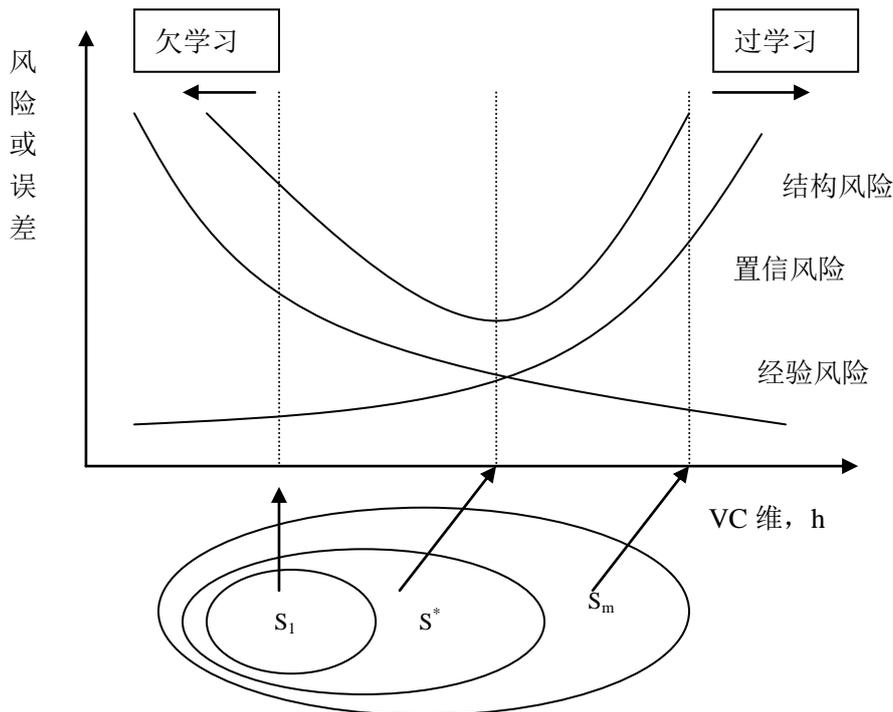


图 2-1 结构风险最小化原则

结构风险最小化的基本原则是寻找假设 $f (f \in S)$ ，使得式(2-10)右端成为最小的风险结构，也就是说风险结构最小化必须控制结构风险最小化的同时，还要密切关注 VC 维大小。

2.3 支持向量机

2.3.1 支持向量机原理

瓦普尼克同其领导的贝尔实验室研究小组最早对支持向量机进行研究和开发，支持向量机是一种新型的机器学习技术，是统计学习理论的发展和丰富。作为一种新型的多项式神经网络或径向神经网络，支持向量机以结构风险最小化原理为基础，将推广误差的上界减到最小。

支持向量机被称为一种通用的学习机器，它所依赖的基本原理是统计学理论中的结构风险最小化原理，它所针对的问题是模式识别问题。即给定训练样本 $T = \{(x_i, y_i), i = 1, 2, \dots, l\}, x_i \in R, y_i \in \{+1, -1\}$ ，寻找决策函数：

$$f(x) = \text{sgn}(g(x)) \quad (2.12)$$

推断对任意给定的样本点的 $y = f(x)$ 的值，其中 $\text{sgn}(\cdot)$ 为符号函数。在机器学习问题中，我们经常把求解分类的问题称为分类学习机。如果 $g(x)$ 是线性函数，那么对应的分类问题是线性分类学习机；倘若 $g(x)$ 是非线性函数，则与之相对应的分类问题是非线性分类学习机。针对不同的分类问题，可以把支持向量机划分为三类：线性可分支持向量机、线性不可分支持向量机以及非线性支持向量机。

2.3.2 支持向量机分类

(1) 线性可分支持向量机

支持向量机可以用于求解有唯一解的二次规划问题，而且所需要的复杂性而不是问题的维数决定了解的复杂性。

对于事先给定的训练样本集： $(x_1, y_1), \dots, (x_i, y_i), \dots, (x_l, y_l), i = 1, 2, \dots, l$ ，其中 $x_i \in R^n$ 被称为 n 维投入，亦即 n 个解释变量，在本文中 $x_i \in R^n$ 指的是 n 个财务指标； $y_i \in \{+1, -1\}$ 是被解释变量。在文章中， $y = 1$ 表示公司财务业绩上升， $y = 0$ 表示公司财务业绩下降。

假如训练样本集可以被一个分类超平面：

$$w^T x + b = 0 \quad (2.13)$$

分开。其中， $w = (w_1, w_2, \dots, w_l)^T$ 垂直于分类超平面，被称为权重向量； $x = (x_1, x_2, \dots, x_l)^T$ 是训练集向量， b 是常数。

倘若该超平面可以把训练集中的所有向量准确无误的分为两类，并且能够使该超平面同距其最近的向量之间的距离达到最大，那么就称该训练集向量被该最优超平面分开。图 2-2 给出了线性可分支持向量机示意图，图中的笑脸和四角星分别表示两类样本（在本文中笑脸表示财务业绩上升的样本；四角星表示财务业绩下降的样本），粗实线表示的是分类超平面，虚线表示穿过两类样本中距离分类面最近的样本的平

面，该平面与分类面平行，这两条虚线之间的间隔被称为分类间隔，大小等于 $2/\|w\|$ ，其中 $\|w\|$ 是向量 w 的模。而位于图 2-2 虚线上的样本点即是支持向量。最优分类超平面是这样一个平面：它不仅可以把两类样本准确的分开，而且还能使分类的间隔 $2/\|w\|$ 最大化。

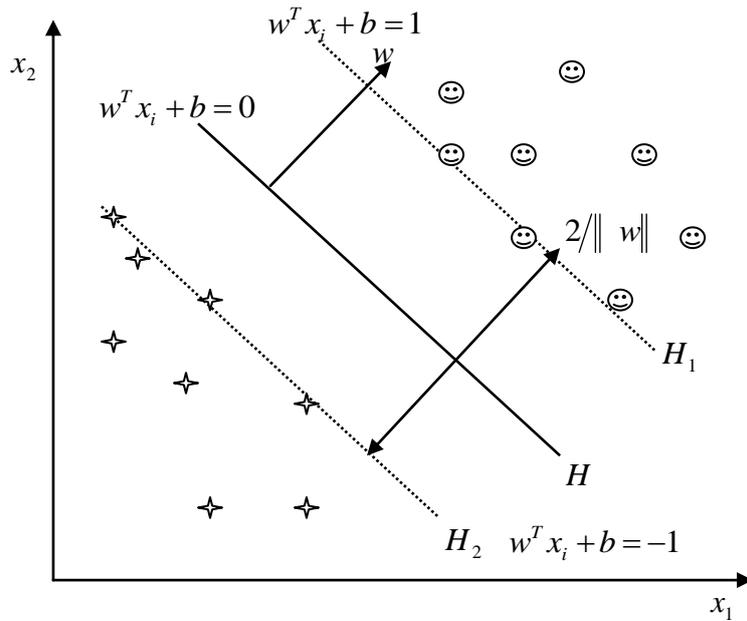


图 2-2 线性可分支持向量机示意图

为了进一步对分类超平面加以描述，在此假定训练样本集数据满足：

$$w^T x_i + b \geq 1, \text{ 若 } y_i = +1 \quad (2.14)$$

$$w^T x_i + b \leq -1, \text{ 若 } y_i = -1 \quad (2.15)$$

那么它们可以用下式加以表达：

$$y_i(w^T x_i + b) - 1 \geq 0 \quad (2.16)$$

对于满足不等式(2.14)的样本点，其中一部分恰好落在超平面 $H_1: w^T x_i + b = 1$ 上，其他则在 H_2 外；同样，对于满足不等式(2-28)的样本点，其中一部分落在超平面 $H_2: w^T x_i + b = -1$ 上，其他则在 H_2 外。超平面 H_1 和 H_2 相互平行，它们之间不存在任何样本点，而且它们之间的距离是 $2/\|w\|$ 。而最大化分类间隔 $2/\|w\|$ 就是最小化的 $\frac{1}{2}\|w\|^2$ 。由此可知，最优分类超平面即是可以将训练样本点准确无误的区分开而且使得 $\frac{1}{2}\|w\|^2$

最小的分类面，即：

$$\min \Phi(w) = \frac{1}{2} \|w\|^2 \quad (2.17)$$

由式(2.16)可知，对于每一对 w 和 b 均存在一个与之相应的超平面方程 $w^T x_i + b = 0$ ，我们的目的即是寻找一对最优的超平面 (w, b) ，使得存在某些样本点 x_i 恰好落在图 2-2 中的虚线上，即满足 $y_i(w^T x_i + b) - 1 = 0$ ，而这些 x_i 就是“支持向量”。

在此，欲构造最优的分类超平面，必须用 $\|w\|$ 最小的超平面将两个不同类的样本集 $(x_1, y_1), \dots, (x_l, y_l), \dots, (x_l, y_l)$ ， $i = 1, 2, \dots, l, y_i \in \{-1, +1\}$ 中向量 x_i 分开。

按照统计学习理论，选择一个规范分类超平面构成的预测函数集：

$$f(x) = \text{sgn}(w^T x + b) \quad (2.18)$$

的 VC 维 h 满足：

$$h \leq \min\left\{\left[R^2 A^2\right], n\right\} + 1 \quad (2.19)$$

在这里， n 是向量空间的维数， R 是覆盖样本向量的超球半径， $\|w\| \leq A$ 。因此，我们能够运用增大分类间隔来降低 VC 维，并根据结构风险最小化原则做出函数复杂性的遴选。支持向量机的出发点是经验风险固定前提下，将期望风险最小化问题转化为最小化 $\|w\|$ 。因此，支持向量机即是在满足(2.18)的前提下，将分类间隔最大化。故而，分类间隔最大化问题又可以转化为下式的最优化问题：

$$\min \Phi(w) = \frac{1}{2} \|w\|^2 \quad (2.20)$$

$$s.t. y_i(w^T x_i + b) \geq 1, i = 1, 2, \dots, l \quad (2.21)$$

问题(2-20)和(2-21)的最优解即是下式 (2.22) 拉格朗日函数的鞍点：

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^l \alpha_i [y_i(w^T x_i + b) - 1] \quad (2.22)$$

其中 $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_l)^T \in R^l$ ， α_i 是与约束条件(2.22)相对应的拉格朗日乘子。倘若 α 的分量 $\alpha_i > 0$ ，那么与之相对应的输入样本点 x_i 就是支持向量。

因为式 (2.21) 在鞍点处的 w 和 b 的梯度为零，所以：

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^l \alpha_i y_i x_i = 0 \Rightarrow w = \sum_{i=1}^l \alpha_i y_i x_i \quad (2.23)$$

$$\frac{\partial L}{\partial b} = \sum_{i=1}^l \alpha_i y_i = 0 \Rightarrow \sum_{i=1}^l \alpha_i y_i = 0 \quad (2.24)$$

将式(2.23)和式(2.24)代入到(2.22)中，同时对它关于 α 求最大，则可得到问题(2.20)和(2.21)的对偶问题：

$$\max_{\alpha} \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j x_i^T x_j \quad (2.25)$$

$$s.t. \sum_{i=1}^l \alpha_i y_i = 0 \quad (2.26)$$

$$\alpha_i \geq 0, i, j = 1, 2, \dots, l \quad (2.27)$$

倘若 α 是问题(2.25)~(2.27)的解，将 α 代入(2.23)，那么：

$$w = \sum_{i=1}^l \alpha_i y_i x_i \quad (2.28)$$

按照 KKT 定理，分类超平面满足下式条件：

$$\alpha_i [y_i (w^T x_i + b) - 1] = 0, i = 1, 2, \dots, l \quad (2.29)$$

是最优超平面的充要条件。

大部分样本点对应的 α_i 是等于 0 的，而那些对应的 α_i 值不为 0 得样本点就是支持向量。依照公式(2.29)，唯独支持向量是对 w 有贡献的， α_i 同支持向量共同决定了最优超平面，这便是“支持向量”称谓的由来，而与之对应的学习方法则被称为支持向量机。对于 b 的求解，我们完全能够任意选择一个支持向量，按照(2.21)求出，在这里约束条件的表达式是取等号的。由(2.29)可知，当 $\alpha_i \neq 0$ ，有 $y_i (w^T x_i + b) = 1$ ，即 $w^T x_i + b = 1$ ，从而可知 $b = (1 - y_i w^T x_i) / y_i$ ，将(2.28)代入可得：

$$b = y_i - w^T x_i = y_i - \sum_{i=1}^l \alpha_i y_i x_i^T x_i \quad (2.30)$$

在实际应用过程中，为了可靠计算，通常首先分别计算与所有支持向量相对应的 b 值，然后再对所有的 b 值求平均，即：

$$b = \frac{1}{N_{SV}} \sum_{x_i \in SV} (y_i - \sum_{i=1}^l \alpha_i y_i x_i^T x_i) \quad (2.31)$$

所以，给定未知样本点 x ，我们只需计算： $f(x) = \text{sgn}(w^T x + b)$ 就能够判别 x 所属的类别。当 $f(x) = 1$ 时，将该样本点划归到 $y = 1$ 这一类中，（本文中将其判定为财务业绩上升的公司）；当 $f(x) = 0$ 时，则将该样本点划归到 $y = 0$ 这一类中（本文中将其判定为财务业绩下降的公司）。

(2) 线性不可分支持向量机

在实务操作过程中，多数我们所遇到的问题并不满足线性可分的条件，同时，即便问题是线性可分的，但是由于各方面原因，也会出现“野点子”，比如出现被标错的样本点，这些点最终会对最优超平面产生严重影响。此时，面对线性不可分的问题，我们往往通过增加一个非负的松弛变量以实现约束条件的放宽。也就是说对于每一个训练样本点，引入一个大于零的松弛变量 $\xi_i \geq 0, i = 1, 2, \dots, l$ ，从而约束条件即变为

$y_i(w^T x_i + b) \geq 1 - \xi_i$ ， $\sum_{i=1}^l \xi_i$ 代表了训练集被错分的程度。所以，在求解线性不可分问题时，我们需要达到两个目标：首先，保持分类间隔 $2/\|w\|$ 尽可能大；其次，错分程度 $\sum_{i=1}^l \xi_i$ 尽可能小。为了实现以上两个目的，在此我们引入了一个惩罚因子 $C > 0$ ，这就得到了下面的优化问题：

$$\min_{w, b, \xi} \Phi(w) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i \quad (2.32)$$

$$s.t. y_i(w^T x_i + b) \geq 1 - \xi_i \quad (2.33)$$

$$\xi_i \geq 0, i = 1, 2, \dots, l \quad (2.34)$$

其中， ξ_i 代表样本点离该类的距离（如图 2-3），用来反映样本点 x_i 违反约束 $y_i(w^T x_i + b) \geq 1$ 的程度， C 表示对误差的惩罚因子， C 越大，表示对错误的惩罚越严重。

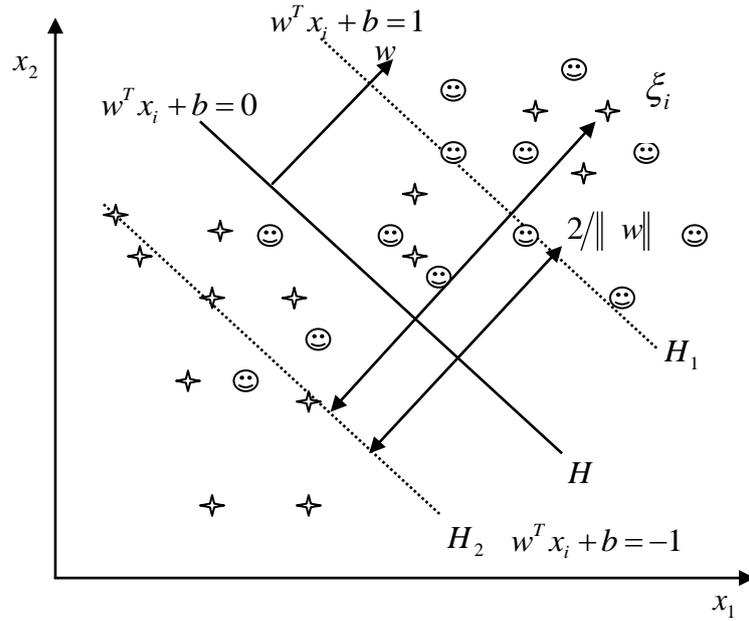


图 2-3 线性不可分支持向量机

与线性可分问题相同，(2.32) ~ (2.34) 亦存在一个二次型规划问题，引入拉格朗日函数，该问题的最优解则成为下面拉格朗日函数的鞍点：

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i - \sum_{i=1}^l \alpha_i [y_i (w^T x_i + b) + \xi_i - 1] - \sum_{i=1}^l \beta_i \xi_i \quad (2.35)$$

上式中， $\alpha_i \geq 0$ ， $\beta_i \geq 0$ 是拉格朗日乘子，因为在鞍点处的 w 、 b 和 ξ_i 的梯度等于零，所以存在：

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^l \alpha_i y_i x_i = 0 \Rightarrow w = \sum_{i=1}^l \alpha_i y_i x_i \quad (2.36)$$

$$\frac{\partial L}{\partial b} = \sum_{i=1}^l \alpha_i y_i = 0 \Rightarrow \sum_{i=1}^l \alpha_i y_i = 0 \quad (2.37)$$

$$\frac{\partial L}{\partial \xi_i} = C - \alpha_i - \beta_i = 0 \quad (2.38)$$

将式(2.36) ~ (2.38)代入式(2.35)，然后对其关于 α 求最大值，得到(2.32) ~ (2.34) 的对偶问题：

$$\max_{\alpha} \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j x_i^T x_j \quad (2.39)$$

$$s.t. \sum_{i=1}^l \alpha_i y_i = 0 \quad (2.40)$$

$$0 \leq \alpha_i \leq C, i = 1, 2, \dots, l \quad (2.41)$$

依照 KKT 定理可知, 在最优点上, 拉格朗日乘子 α_i 和约束条件的乘积是:

$$\alpha_i [y_i (w^T x_i + b) + \xi_i - 1] = 0, i = 1, 2, \dots, l \quad (2.42)$$

$$(C - \alpha_i) \xi_i = 0, i = 1, 2, \dots, l \quad (2.43)$$

求解 (2.39) ~ (2.41) 可以得到 α_i , 其中 α_i 的取值有以下三种情况:

①当 $\alpha_i = 0$ 时, 由 (2.43) 可得 $\xi_i = 0$ 。此时约束条件不发挥作用, 所有的样本点可以被正确地分类;

②当 $0 < \alpha_i < C$ 时, 由 (2.43) 可得 $\xi_i = 0$, 并且由 (2.42) 可知 $y_i (w^T x_i + b) = 1$ 。在该种情况下的 x_i 即为标准支持向量;

③当 $\alpha_i = C$ 时, 由 (2.43) 可知, ξ_i 取值不确定, 可能等于 0, 也可能不等于 0。当 ξ_i 不等于 0 时, 根据式 (2.42), $y_i (w^T x_i + b) < 1$, 此时与之相应的样本点被称为误判点。当 ξ_i 等于 0 时, 与之对应的样本点被称为边界支持向量。

根据 (2.49), 可以求出 $w = \sum_{i=1}^l \alpha_i y_i x_i$, 再根据式 (2.41) 可得, 对于任一支持向量

满足 $y_i (w^T x_i + b) = 1$, 故而, $b = y_i - w^T x_i = y_i - \sum_{i=1}^l \alpha_i y_i x_i^T x_i$ 。

在实际应用过程中, 为了计算可靠, 通常采用的是 b 的平均值, 即

$$b = \frac{1}{N_{SV}} \sum_{x_i \in SV} (y_i - \sum_{i=1}^l \alpha_i y_i x_i^T x_i)。$$

(3) 非线性支持向量机

前面的讨论都是假定训练样本集是线性的, 线性问题的求解是通过内积 $x_i^T x_j$ 来表示的。倘若训练集样本是非线性的, 那么我们则需要寻求一个非线性映射, 以将输入空间映射到某一特征空间, 记: $\varphi: R^n \rightarrow H$ 。这样便可以实现在特征层用线性可分支持向量机求解非线性问题了。

倘若训练集样本是非线性的, 那么不再存在能够完全区分两类样本点的超平面,

此时我们需要寻求一个最优的超曲面。为了解决这一问题，我们引入了非线性映射，通过非线性映射把输入空间映射到某一个具有更高维数的特征空间。将原来输入空间的非线性问题进行映射到特征空间之后，便可以在特征空间中利用线性可分支持向量机对问题进行求解了。

对于非线性分类问题，在此引入一个非线性映射函数 以将输入空间映射到一个更高维的特征空间：

$$\varphi: x \in R^n \rightarrow \varphi(x) \in H \quad (2.44)$$

在特征空间中对输入变量进行线性分类，这时候分类超平面为：

$$w^T \varphi(x) + b = 0 \quad (2.45)$$

$b \in R$ ， $w \in H$ ，求解分类超平面即是对下面的最优化问题的求解：

$$\min_{w, b, \xi} \Phi(w) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i \quad (2.46)$$

$$s.t. y_i (w^T x_i + b) \geq 1 - \xi_i \quad (2.47)$$

$$\xi_i \geq 0, i = 1, 2, \dots, l \quad (2.48)$$

式 (2.46) ~ (2.48) 是一个二次规划问题，为了便于求解，在此引入拉格朗日函数，式 (2.46) ~ (2.48) 的最优解便是下面拉格朗日函数的鞍点：

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i - \sum_{i=1}^l \alpha_i [y_i (w^T x_i + b) + \xi_i - 1] - \sum_{i=1}^l \beta_i \xi_i \quad (2.49)$$

其中， $\alpha_i \geq 0$ ， $\beta_i \geq 0$ 是拉格朗日乘子，因为在鞍点处的 w 、 b 和 ξ_i 的梯度等于零，所以可得：

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^l \alpha_i y_i x_i = 0 \Rightarrow w = \sum_{i=1}^l \alpha_i y_i x_i \quad (2.50)$$

$$\frac{\partial L}{\partial b} = \sum_{i=1}^l \alpha_i y_i = 0 \Rightarrow \sum_{i=1}^l \alpha_i y_i = 0 \quad (2.51)$$

$$\frac{\partial L}{\partial \xi_i} = C - \alpha_i - \beta_i = 0 \quad (2.52)$$

将 (2.50) ~ (2.52) 代入到 (2.49)，对它求关于 α 的极值，便可得到 (2.46) ~ (2.48) 的对偶问题为：

$$\max_{\alpha} \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j x_i^T x_j \quad (2.53)$$

$$s.t. \sum_{i=1}^l \alpha_i y_i = 0 \quad (2.54)$$

$$0 \leq \alpha_i \leq C, i=1,2,\dots,l \quad (2.55)$$

其中，称 $K(x_i, x_j) = \varphi^T(x_i)\varphi(x_j)$ 为核函数。根据 KKT 定理可知，在最优点上，拉格朗日乘子和约束条件的乘积为：

$$\alpha_i [y_i (w^T x_i + b) + \xi_i - 1] = 0, i=1,2,\dots,l \quad (2.56)$$

$$(C - \alpha_i) \xi_i = 0, i=1,2,\dots,l \quad (2.57)$$

求解问题 (2.50) ~ (2.52) 可以得到 α_i ， α_i 的取值包括三种情况： $\alpha_i = 0$ 、 $0 < \alpha_i < C$ 、 $\alpha_i = C$ ，正如在线性不可分支持向量机问题中的讨论，只有满足 $0 < \alpha_i < C$ 的 x_i 才是支持变量。

根据 (2.50)，可以求出 $w = \sum_{i=1}^l \alpha_i y_i \varphi(x_i)$ ，再由 (2.69) 可以得知，对于任一个支持

向量，均满足： $y_i (w^T \varphi(x_i) + b) = 1$ ，故而可知 b 为：

$$b = y_i - w^T \varphi(x_i) = y_i - \sum_{i=1}^l \alpha_i y_i \varphi^T(x_i) \varphi(x_i)。$$

为了计算可靠，在实际应用过程中，我们通常使用 b 的平均值，即

$$b = \frac{1}{N_{SV}} \sum_{x_i \in SV} (y_i - \sum_{i=1}^l \alpha_i y_i \varphi^T(x_i) \varphi(x_i))，其中，K(x_i, x_j) = \varphi^T(x_i)\varphi(x_j) 为核函数。所以，$$

对于给定的未知样本点 x ，我们只需要计算：

$$f(x) = \text{sgn}(w^T \varphi(x) + b) = \text{sgn}\left(\sum_{i=1}^l \alpha_i y_i K(x_i, x) + b\right)$$

便可以判别 x 所属的类别。

我们可以求证，函数 $K(x_i, x)$ 如果可以满足 Mercer 的要求，即能够将它视作核函数，设若可以寻找到比较适宜的核函数 $K(x_i, x)$ 那么则可以将输入空间中的非线性分类问题加以转化，形成在特征空间中求解线性可分的问题。

基于如上分析我们总结出如下结论: 通过非线性函数的应用固然可以把样本数据映射到高维空间甚至是无穷维的特征空间, 也可以于特征空间里面构建超平面最优分类, 可是对于最优化问题的求解以及计算决策函数方面, 就已经没有必要再针对非线性函数开展显示计算, 只需研究核函数即可, 从而规避了特征空间维数灾难的发生。

由此可知, 我们在对待非线性支持向量机的分类问题时, 仅需对特征空间核函数 $K(x_i, x_j)$ 的具体形式加以考虑, 而不需要对非线性转换函数 $\varphi(x)$ 的显式进行设定, 即没有必要再对 $\varphi^T(x_i)\varphi(x_j)$ 对应内积一个个的加以核算, 这样便在很大程度上减少了支持向量机的核算难度。因此, 我们得出: 支持向量机中一个非常重要的组成部分便是核函数。

当然, 在支持向量机最优化问题的求解之前, 我们首先必须对核函数进行选择。而核函数选择的标准即是是否满足 Mercer 定理, 只有那些满足 Mercer 定理要求的函数才可以成为核函数。由于 Mercer 定理仅给我们提供了一个函数可否成为支持向量机的内积核函数的要求, 故而, 研究者在基本要求已经满足的前提下, 还应当对核函数的具体形式加以选择。研究过程中常用的核函数形式包括如下三类:

$$\textcircled{1} \text{多项式核函数: } K(x_i, x) = (1 + x_i^T x)^d \quad (2.58)$$

$$\textcircled{2} \text{Guass 径向基核函数: } K(x_i, x) = \exp(-|x_i - x|^2 / 2\sigma^2) \quad (2.59)$$

$$\textcircled{3} \text{Sigmoid 核函数: } K(x_i, x) = \tanh(kx_i^T x - \delta) \quad (2.60)$$

其中, $d > 0$, $\sigma > 0$, $\delta > 0$, $k > 0$ 视为核参数。

2.3.3 最小二乘支持向量机

为了进一步提高支持向量机的预测精确度, 许多学者均关注着支持向量机的科研工作。瓦普尼克等人通过研究指出: 因为广义研究样本和支持向量机预测准确度密切相关, 那么由于研究样本范畴的逐步扩展加大, 二次规划的求解问题则变得愈来愈复杂, 计算速度变慢, 而且不利于进行动态模型的构建, 而解决该问题的有效方法即是小二乘支持向量机。小二乘支持向量机是对标准支持向量机的一个改进模式。苏伊肯斯对小二乘支持向量机进行了最早的解释。他指出小二乘支持向量机主要是针对不同优化问题的目标函数, 推出一系列相异的等式约束。下文给出了小二乘支持向量机的一般性描述。

对于给定的 l 个样本 $\{(x_i, y_i), i=1, 2, \dots, l\}$ 的训练集合, 第 i 个输入数据为 $x_i \in R^n$, 第 i 个输出变量为 $y_i \in \{-1, +1\}$, 是一个二分变量。支持向量机模型的目标即是构建一个如下类别的分类器:

$$f(x) = \text{sgn}(w^T \varphi(x) + b) \quad (2.61)$$

使得样本可以被 $f(x)$ 准确的分类。最小二乘支持向量机分类即是要求解如下类别的最优化问题。

$$\min \frac{1}{2} \|w\|^2 + \frac{1}{2} C \sum_{i=1}^l \xi_i^2 \quad (2.62)$$

$$\text{s.t. } y_i(w^T x_i + b) = 1 - \xi_i, \quad i=1, 2, \dots, l \quad (2.63)$$

其对偶问题的拉格朗日多项式为:

$$L(w, b, \xi, \alpha) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i^2 - \sum_{i=1}^l \alpha_i [y_i(w^T \varphi(x_i) + b) + \xi_i - 1] \quad (2.64)$$

其中 α 是拉格朗日乘子, 根据该拉格朗日多项式为等式可知, α 的取值可正亦可负。最优化条件是:

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^l \alpha_i y_i x_i = 0 \quad (2.65)$$

$$\frac{\partial L}{\partial b} = \sum_{i=1}^l \alpha_i y_i = 0 \Rightarrow \sum_{i=1}^l \alpha_i y_i = 0 \quad (2.66)$$

$$\frac{\partial L}{\partial \xi_i} = C \sum_{i=1}^l \xi_i - \sum_{i=1}^l \alpha_i = 0 \quad (2.67)$$

$$\frac{\partial L}{\partial \alpha} = \sum_{i=1}^l [y_i(w^T \varphi(x_i) + b) + \xi_i - 1] = 0 \quad (2.68)$$

该组式子等价于:

$$w = \sum_{i=1}^l \alpha_i y_i x_i \quad (2.69)$$

$$\sum_{i=1}^l \alpha_i y_i = 0 \quad (2.70)$$

$$C \sum_{i=1}^l \xi_i = \sum_{i=1}^l \alpha_i \quad (2.71)$$

$$\sum_{i=1}^l [y_i (w^T \varphi(x_i) + b) + \xi - 1] = 0 \quad (2.72)$$

上述数组可以写成如下形式的方程组：

$$\begin{bmatrix} I & 0 & 0 & -Z^T \\ 0 & 0 & 0 & -Y^T \\ 0 & 0 & CI & -I \\ Z & Y & I & 0 \end{bmatrix} \begin{bmatrix} w \\ b \\ \xi \\ a \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \bar{1} \end{bmatrix} \quad (2.73)$$

$\xi = (\xi_1, \xi_2, \dots, \xi_l)^T$, $a = (a_1, a_2, \dots, a_l)^T$, $\bar{1} = (1, 1, \dots, 1)^T$, $Z = (\varphi(x_1) y_1, \varphi(x_2) y_2, \dots, \varphi(x_l) y_l)^T$, $Y = (y_1, y_2, \dots, y_l)^T$, 其中 I 为单位矩阵。消去式 (2.73) 中的 ξ 和 w , 根据 Mercer 条件:

$$\Omega_{ij} = y_i y_j K(x_i, x_j), i, j = 1, 2, \dots, l \quad (2.74)$$

从而, 式 (2.73) 可以转换成如下表达式:

$$\begin{bmatrix} 0 & -Y^T \\ Y & \Omega + C^{-1}I \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} 0 \\ \bar{1} \end{bmatrix} \quad (2.75)$$

该方程组仅同 b 和 a 有关, 假设 $A = \Omega + C^{-1}I$, 因为 A 是一个对称的半正定矩阵, 故而 A^{-1} 存在, 求解线性方程组 (2.73), 可以得出解:

$$b = \frac{Y^T A^{-1} \bar{1}}{Y^T A^{-1} Y} \quad (2.76)$$

$$a = A^{-1}(\bar{1} - Yb) \quad (2.77)$$

用式 (2.69) 替代式 (2.61) 中的 w , 根据 Mercer 条件式 (2.73), 可以得出:

$$f(x) = \text{sgn}\left(\sum_{i=1}^l a_i y_i K(x_i, x) + b\right) \quad (2.78)$$

此处, a_i, b 可以分别根据式 (2.77) 和 (2.76) 得出; $K(x_i, x)$ 是核函数, 将其代入式 (2.78) 即可求出最优分类结果。

第3章 预测指标的选取及预处理

本章首先根据预测指标和财务业绩两者存在的内在关系,在广泛参考诸多学者的研究成果的前提下,建立了本文的指标体系。然后分别对所选指标进行了显著性差异检验、正态分布检验以及多重共线性检验,最终形成了文章的预测指标体系,即财务业绩预测指标体系。

3.1 预测指标的选取

3.1.1 预测指标选取的原则

公司财务业绩的变动及变动幅度大小,同公司的各个部门和生产的各个环节相关联,同时又通过公司的一系列财务指标体现出来。所以文章在选择预测指标时,在参考前人研究的基础上,主要采用的是财务指标。文章构建财务指标体系时主要遵循了以下原则:

首先,相关性原则。能够选入预测指标体系的财务指标必须具有相关性。一方面所选取的指标必须与财务业绩相关,这些指标能够单独或联系的反映企业的财务状况、经营成果和现金流量。另一方面,所选取的财务指标应当与利益相关者的需求相关,能够充分的满足利益相关者对信息的需求。比如,在我们所选择的财务指标中至少应当包括:能够反映企业偿债能力的指标,以反映企业的财务状况,并满足债权人的信息需求;能够反映企业盈利能力的指标,以反映企业的经营成果,并满足投资人的信息需求;能够反映现金能力和营运能力的指标,以反映企业的财务状况和现金流量,从而满足企业管理者进行经营决策的需求。

其次,重要性原则。重要性原则是指所选择的预测指标应当对财务业绩预测具有重要性影响。公司财务业绩的升降通过一系列财务指标反映出来。可以说这一系列的财务指标均与公司的财务业绩呈现相关性,只不过因为指标所反映的角度不同、侧重点不同,所以导致了指标对财务业绩变动的贡献不同。这就要求我们在选择指标时不能眉毛胡子一把抓,而应当有所侧重。此处重要性的标准主要体现在以下两个方面。一是性质重大。比如反映公司偿债能力的资产负债率。因为公司偿债能力的大小关乎公司的生存与发展。无论是基于收回本息为目的的债权人,还是基于分散风险为目的获取收益的投资人,以及为保住自己职位及“经理人”这项无形资产的管理人,都会对该偿债能力指标加以重视,这就是说这项指标在性质上具有重要性。二是贡献重大。

此处贡献重大主要体现在所选择的财务指标能够引起财务业绩较大幅度的变动。举一特例而言：假设某公司净资产收益率上升 0.21 个百分点，其中销售利润率上升 2 个百分点，总资产周转率上升 1 个百分点，权益乘数没有发生变动，那么此处应当将销售净利润和总资产周转率考虑在内，因为权益乘数此处没有发生变动，则可以不予考虑。

再者，可操作性原则。文章构建指标体系在指标的选取上应当具有可操作性。再完美的指标体系，倘若无法得到执行，相关的数据收集不到，那么也只能是一个没用的花架子。所以在构建预测指标体系时，必须考虑预测指标的可操作性，以及预测数据的可获取性。

3.1.2 初始预测指标的选取

参考国内外学者在财务预警指标上的选取，我们在选择自变量时，主要考虑了能够反映公司财务业绩的财务指标。经过筛选，文章最终选取了 24 个财务指标，如表 3-1 所示。

表 3-1 财务指标

指标性质	财务指标	计算公式
短期偿债能力	流动比率 (X_1)	流动资产/流动负债
	速动比率 (X_2)	(流动资产-存货)/流动负债
	营运资金对资产总额比率 (X_3)	营运资金/资产总额
长期偿债能力	资产负债率 (X_4)	负债总额/资产总额
	长期负债率 (X_5)	长期负债总额/负债总额
	负债对权益比率 (X_6)	负债总额/有形净资产总额
	长期负债对权益比率 (X_7)	长期负债总额/所有者权益总额
营运能力	应收账款周转率 (X_8)	营业收入/应收账款平均占用额
	存货周转率 (X_9)	营业成本/存货平均占用额
	固定资产周转率 (X_{10})	营业收入/固定资产平均净额
	总资产周转率 (X_{11})	营业收入/平均资产总额
盈利能力	营业利润率 (X_{12})	营业利润/营业收入
	资产报酬率 (X_{13})	(利润总额+财务费用)/平均资产总额
	总资产净利润率 (X_{14})	净利润/平均资产总额
	固定资产净利润率 (X_{15})	净利润/固定资产平均净额
	营业收入增长率 (X_{16})	(本年营业收入-上年营业收入)/上年营业收入
成长能力	固定资产增长率 (X_{17})	(本年期末固定资产-上年期末固定资产)/上年期末固定资产
	总资产增长率 (X_{18})	(本年期末总资产-上年期末总资产)/上年期末总资产
	净利润增长率 (X_{19})	(本年净利润-上年净利润)/上年净利润

现金流量	营业收入现金比率 (X_{20})	经营活动现金净流量/营业收入
	每股经营活动现金净流量 (X_{21})	经营活动现金净流量/总股数
	每股现金净流量 (X_{22})	现金及现金等价物净增加额/总股数
	销售收到现金比率 (X_{23})	销售商品提供劳务收到的现金/营业收入
	净利润现金比率 (X_{24})	净利润/现金及现金等价物

(1) 短期偿债能力指标

短期偿债能力反映的是以企业流动资产偿还流动负债的能力，流动资产是偿还流动负债的保障，如果流动资产大于流动负债，则在公司内部形成一个短期债务的缓冲垫，延长短期债务穿透短期资产的时间。债权人关注公司的短期偿债能力主要考虑其可以到期获取应得的本息；投资者通过短期偿债能力决策是否有效运用财务杠杆；管理者通过短期偿债能力权衡投资者和债权人的利益。文章选取的短期偿债能力指标主要有：流动比率 (X_1)、速动比率 (X_2) 以及营运资金对资产总额比率 (X_3)。

(2) 长期偿债能力指标

长期偿债能力揭示了公司对长期债务的偿还能力。公司的利益相关者关注长期偿债能力，其中长期债权人对该指标的关注度尤甚。这是因为长期债券投资人的本金需要经过相对较长的时间方可收回，这就加大了该部分资金的机会成本，而债权人则需担负更大的风险。虽然投资者和管理人有联合起来欺诈债权人的动机，但是他们也会考虑将公司的长期偿债能力控制在一定的范围，否则他们将一分钱也借不到。文章采取的长期偿债能力指标主要有：资产负债率 (X_4)、长期负债率 (X_5) 和有形净值债务率 (X_6)、长期负债对权益比率 (X_7)。

(3) 营运能力指标

营运能力反映的是公司对资产的利用能力，它揭示了公司的生产经营效率。如果公司的营运能力的较强，则意味着公司的生产效率高，同时也对应着相对较高的偿债能力。文章选取的营运能力指标主要包括：应收账款周转率 (X_8)、存货周转率 (X_9)、固定资产周转率 (X_{10}) 以及总资产周转率 (X_{11})。

(4) 盈利能力指标

企业的盈利能力揭示了公司的经营成果以及发展前景。公司盈利，债权人的本息方有保障，投资者的资金也可保值增值。文章主要选取了五个盈利能力指标：营业利润率 (X_{12})、资产报酬率 (X_{13})、总资产净利润率 (X_{14})、固定资产净利润率 (X_{15}) 以及营业收入增长率 (X_{16})。

(5) 成长能力指标

成长能力揭示了公司的长远发展的能力和发展趋势。主要包括公司生产规模的扩大、公司盈利能力的增强等。文章选取的反映成长能力的指标有：固定资产增长率(X_{17})、总资产增长率(X_{18})和净利润增长率(X_{19})。

(6) 现金流量指标

现金流量指标反映了公司产生现金流的能力。如果公司在经营期间能够产生越多的现金流入，那么公司的债务偿还将越有保障，公司股东的收益将会得到提高。文章主要选用营业收入现金比率(X_{20})、每股经营活动现金净流量(X_{21})、每股现金净流量(X_{22})、销售收到现金比率(X_{23})以及净利润现金比率(X_{24})五个财务指标揭示公司产生现金流入的能力。

3.2 指标的预处理

3.2.1 指标数据的正态分布检验

文章采用单样本 K-S 检验，以明确样本资料是否服从正态分布。单样本 K-S 检验时将样本数据的累计分布函数与某个确定的理论分布函数做比较，以检验某个样本是否服从于指定的分布函数。单样本 K-S 检验的原假设 H_0 : 样本的总体分布与某指定的分布函数无明显差异，即样本总体分布服从于该指定的理论分布。因为文章所选择的样本数量较多，属于大样本，故而采用 Z 统计量。

通过 SPSS16.0 可以方便的获得 Z 值以及与之相对应的伴随概率 P，对于给定的显著性水平 0.05，当伴随概率 $P < 0.5$ 时，说明原假设属于小概率事件，不可能发生，否定原假设，接受备选假设，即样本在总体上不服从正态分布；当伴随概率 $P > 0.5$ 时，接受原假设，否定备选假设，即样本在总体上服从正态分布。表 3-2 给出了半年度指标数据正态分布的检验结果。

表 3-2 半年度预测指标数据正态分布检验结果

财务指标	Z 统计量	伴随概率 P	财务指标	Z 统计量	伴随概率 P
X_1	3.871	0.000	X_{13}	2.083	0.000
X_2	4.102	0.000	X_{14}	2.132	0.000
X_3	1.022	0.248*	X_{15}	3.600	0.000
X_4	0.788	0.564*	X_{16}	3.147	0.000
X_5	2.732	0.000	X_{17}	4.259	0.000
X_6	2.505	0.000	X_{18}	3.611	0.000
X_7	2.526	0.000	X_{19}	5.713	0.000
X_8	6.990	0.000	X_{20}	6.709	0.000
X_9	2.896	0.000	X_{21}	2.576	0.000

X ₁₀	3.610	0.000	X ₂₂	2.613	0.000
X ₁₁	2.163	0.000	X ₂₃	3.783	0.000
X ₁₂	6.799	0.000	X ₂₄	2.784	0.000

注：*明该变量在5%显著性水平上通过了K-S正态分布检验，即该变量服从正态分布

通过表3-2可知，对于半年度指标数据而言，所选指标中只有X₃（营运资金对资产总额的比率）和X₄（资产负债率）服从正态分布，其余指标都不服从正态分布。因此，在进行显著性差异检验时，只能选择非参数检验法。

3.2.2 指标数据的显著性差异检验

截至目前，对两组变量间是否存在显著性差异的检验方法主要有：Wilcoxon符号秩检验、Marginal Homogeneity检验、DNemar检验以及符号检验。其中，DNemar检验仅可用于二分变量，Marginal Homogeneity检验仅可用于分类变量，文章选择Wilcoxon符号秩检验进行研究。

Wilcoxon符号秩检验的原假设H₀为：两组变量总体分布无显著性差异；备选假设H₁为：两组变量总体分布存在显著性差异。给定显著性水平 $\alpha = 0.05$ ，当伴随概率 $P > 0.05$ 时，接受原假设，说明两组变量之间不存在显著性差异；当伴随概率 $P < 0.05$ 时，否定原假设，接受备选假设，说明两组变量之间存在显著性差异。

利用SPSS16.0可以很方便的得出两组变量间是否存在显著性差异的检验结果。因为文章所选择的样本数据属于大样本，故而在此我们仍然选择Z统计量。表3-3给出了半年度指标数据显著性差异的检验结果。

表3-3 半年度预测指标数据显著性差异检验结果

财务指标	Z统计量	伴随概率P	财务指标	Z统计量	伴随概率P
X ₁	-0.743	0.457	X ₁₃	-4.282	0.000*
X ₂	-0.363	0.717	X ₁₄	-3.729	0.000*
X ₃	-0.796	0.426	X ₁₅	-3.171	0.002*
X ₄	-0.904	0.366	X ₁₆	-0.226	0.821
X ₅	-0.564	0.573	X ₁₇	-0.832	0.405
X ₆	-1.087	0.277	X ₁₈	-0.891	0.373
X ₇	-1.102	0.281	X ₁₉	-3.771	0.000*
X ₈	-0.355	0.723	X ₂₀	-2.025	0.043*
X ₉	-1.409	0.159	X ₂₁	-2.537	0.011*
X ₁₀	-1.515	0.130	X ₂₂	-0.037	0.970
X ₁₁	-2.478	0.013*	X ₂₃	-1.530	0.126
X ₁₂	-2.201	0.028*	X ₂₄	-0.041	0.982

注：*表示该变量在5%的水平上显著，即两组变量分布存在显著性差异

通过表 3-3 可以得出, 在 5% 的显著性水平上存在显著性差异的财务指标只有 8 个, 它们是: X_{11} 、 X_{12} 、 X_{13} 、 X_{14} 、 X_{15} 、 X_{19} 、 X_{20} 和 X_{21} , 这 8 个指标应当保留, 其余指标应当剔除。

3.2.3 指标数据的多重共线性检验

对于样本数据进行显著性差异检验之后, 剔除了在财务业绩上升和财务业绩下降之间不存在显著性差异的指标。保留了具有显著性差异的指标, 保留下来的这些指标可能还存在多重共线性的影响, 故而, 应当消除这种影响。文章将采用因子分析法消除变量间多重共线性问题。

因子分析主要的目的在于降维、简化数据。它是通过寻找公共因子, 并确定因子得分系数来实现这一目的的。进行因子分析的主要步骤是: 首先, 计算原始变量之间的相关系数, 并对相关系数的显著性差异进行检验。当相关系数显著性差异的伴随概率 $P > 0.05$ 时, 表明变量间相关性较弱, 不宜采用因子分析; 当伴随概率 $P < 0.05$ 时, 说明变量间存在较强的相关性, 应当进行因子分析。其次, 为了检验数据是否适合因子分析, 应当计算 KMO 和 Bartlett 球形检验值。 $0 < KMO < 1$, KMO 越趋近于 1, 意味着变量间相关性越强; KMO 越趋近于 0, 说明变量间相关性越弱。通常情况下, 认为 $KMO < 0.5$ 时不宜进行因子分析。Bartlett 球形检验值较大, 应当进行因子分析; 否则, 不宜进行因子分析。第三, 计算变量的共同度, 并检验每一个变量对公共因子的依赖度。第四, 确定公共因子的数量。第五, 明确各变量与公共因子的关系, 并对公共因子命名。第六, 计算因子得分矩阵并写出表达式。

通过显著性差异检验之后, 剔除在财务业绩上升样本组和财务业绩下降样本组之间不存在显著性差异的变量, 最后保留下来 8 个指标可能存在多重共线性问题。将这 8 个指标数据代入到 SPSS16.0 中, 即可得出这 8 个变量的相关系数矩阵, 详见附录 1。附录 1 的上半部分显示的初始变量的相关系数矩阵, 下半部分显示的是相关系数显著性检验的 p 值。通过附录 1 可以得出: 矩阵中有大部分指标之间具有较强的相关性, 说明指标相关性较强。另外, KMO 统计量为 0.55, 大于 0.5; 球形检验值为 2354.90, 数值较大。这些迹象表明, 用于半年度预测的指标数据比较适合做因子分析处理。表 3-4 给出了指标数据的 KMO 检验统计量与 Bartlett 球形检验值。

表 3-4 半年度指标数据的 KOM 与 Bartlett 球形检验值

KMO 检验值		0.55
Bartlett 球形检验值	渐进的 χ^2 值	2354.9
	显著性概率 P	0.00

表 3-5 给出了用于财务业绩预测的半年度指标的共同度。该表揭示了每个指标对公共因子的依赖程度。

表 3-5 半年度预测指标变量的共同度

变量	因子方差初始值	变量的共同度
X ₁₁	1.000	0.211
X ₁₂	1.000	0.975
X ₁₃	1.000	0.907
X ₁₄	1.000	0.921
X ₁₅	1.000	0.514
X ₁₉	1.000	0.201
X ₂₀	1.000	0.977
X ₂₁	1.000	0.207

从表 3-5 可以看出，变量共同度在 0.9 以上的有 4 个，分别是 X₁₂、X₁₃、X₁₄ 和 X₂₀；变量共同度在 0.5 和 0.9 之间的有 1 个，是 X₁₅。另外 X₁₁、X₁₉、X₂₁ 这三个变量的共同度分别为 0.211、0.201 和 0.207，变量共同度较低，但是总体而言通过因子分析所提取的公因子能够解释变量特征的大部分。

表 3-6 揭示了半年度指标变量的相关系数矩阵的特征值、方差贡献率、累积方差贡献率以及因子旋转前后的特征值、方差贡献率和累积方差贡献率的信息。

表 3-6 半年度预测指标变量系数矩阵的特征值、方差贡献率和累积方差贡献率

因子	初始解			提取因子的总体描述			旋转后的因子总体描述		
	特征值	方差贡献率 (%)	累计方差贡献率 (%)	特征值	方差贡献率 (%)	累计方差贡献率 (%)	特征值	方差贡献率 (%)	累计方差贡献率 (%)
1	2.871	35.887	35.887	2.871	35.887	35.887	2.865	35.814	35.814
2	2.042	25.522	61.409	2.042	25.522	61.409	2.048	25.596	61.409
3	0.923	11.538	72.947						
4	0.895	11.194	84.141						
5	0.844	10.554	94.695						
6	0.405	5.059	99.754						
7	0.015	0.189	99.943						
8	0.005	0.057	100.000						

分析表 3-6 可知，表的前三列显示的是原始变量的特征值、方差贡献率以及累积方差贡献率；中间三列是提取因子后的特征值、方差贡献率和累积方差贡献率；最后三列揭示的是旋转后因子的特征值、方差贡献率和累积方差贡献率。在所提取的 2

个公共因子中，累积方差贡献贡献率达到 61.409%，这意味着公共因子可以解释原始变量 61.409%的方差。表 3-7 给出了初始因子的载荷矩阵。

表 3-7 初始因子载荷矩阵

原始变量	公共因子	
	F ₁	F ₂
X ₁₁	0.394	0.236
X ₁₂	0.138	-0.978
X ₁₃	0.950	0.069
X ₁₄	0.958	0.058
X ₁₅	0.697	-0.166
X ₁₉	0.448	0.008
X ₂₀	-0.025	0.988
X ₂₁	0.435	0.134

通过表 3-7 可以写出各个初始变量的因子表达式：

$$X_{11}=0.394F_1+0.236F_2$$

$$X_{12}=-0.138F_1-0.978F_2$$

...

$$X_{21}=0.435F_1+0.134F_2$$

从表 3-7 可以得出：各因子对应原始变量上的载荷并无明显不同，不利于进行因子的命名。因此，我们需要将因子载荷矩阵加以旋转，以对各公因子命名。表 3-8 给出了初始载荷矩阵经过方差最大化正交旋转后的因子载荷矩阵。

表 3-8 旋转后的因子载荷矩阵

初始指标	公共因子	
	F ₁	F ₂
X ₁₁	0.4124	0.2023
X ₁₂	0.0554	-0.9858
X ₁₃	0.9524	-0.0117
X ₁₄	0.9593	-0.0232
X ₁₅	0.6806	-0.2244
X ₁₉	0.4475	-0.0300
X ₂₀	0.0590	0.9867
X ₂₁	0.4447	0.0966

通过表 3-8 可知，经过旋转后的公因子载荷矩阵系数已经具有显著差异。第一个指标 在指标 X₁₁、X₁₃、X₁₄、X₁₅、X₁₉ 和 X₂₁ 上具有较大载荷，意味着这六个指标之间具有较强的相关性，可以划归为一类。其中 X₁₁ 代表总资产周转率，X₁₃ 代表的是资产报酬率，X₁₄ 代表的是总资产净利润率，X₁₅ 固定资产净利润率，X₁₉ 代表的是净利润增长率，X₂₁ 代表的是每股经营活动现金净流量，故而，可以把第一个公共因子

F₁命名为“营运-盈利-现金因子”；第二个公因子 F₂在指标 X₁₂ 和 X₂₀ 上具有较大载荷，应当划分为一类。其中 X₁₂ 代表营业利润率，X₂₀ 代表营业收入现金比率，所以，把第二个公因子 F₂ 定义为“营业因子”。表 3-9 给出了各公因子的命名。

表 3-9 各公因子命名

公因子	高载荷指标	计算值	因子命名
F ₁	X ₁₁ : 总资产周转率	0.4124	营运-盈利-现金因子
	X ₁₃ : 资产报酬率	0.9524	
	X ₁₄ : 总资产净利润率	0.9593	
	X ₁₅ : 固定资产净利润率	0.6806	
	X ₁₉ : 净利润增长率	0.4475	
	X ₂₁ : 每股经营活动现金净流量	0.4447	
F ₂	X ₁₂ : 营业利润率	-0.9858	营业因子
	X ₂₀ : 营业收入现金比率	0.9867	

表 3-10 给出了半年度预测因子的得分系数矩阵 W。

表 3-10 因子得分系数矩阵

初始指标	公共因子	
	F ₁	F ₂
X ₁₁	0.1465	0.1038
X ₁₂	0.0076	-0.4812
X ₁₃	0.3325	0.0056
X ₁₄	0.3348	0.0001
X ₁₅	0.2351	-0.1016
X ₁₉	0.1560	-0.0094
X ₂₀	0.0323	0.4830
X ₂₁	0.1565	0.0525

根据表 3-10 写出经旋转后的因子得分表达式：

$$F_1 = 0.1465X_{11} - 0.0076X_{12} + 0.3325X_{13} + 0.3348X_{14} + 0.2351X_{15} + 0.1560X_{19} + 0.0323X_{20} + 0.1565X_{21}$$

$$F_2 = 0.1038X_{11} - 0.4812X_{12} + 0.0056X_{13} + 0.0001X_{14} - 0.1016X_{15} - 0.0094X_{19} + 0.4830X_{20} + 0.0525X_{21}$$

第 4 章 基于 LS-SVM 的财务业绩预测实证研究

本章将 LS-SVM 引入财务业绩预测，并构建 LS-SVM 模型。首先，文章阐述了研究样本的选择；其次简要阐述了 Logit 回归模型的建模过程；最后，阐述了 LS-SVM 模型的构建过程，并对 Logit 回归模型和 LS-SVM 模型进行了比较。

4.1 研究样本的选择

4.1.1 因变量的选择

考虑到投资者的投资目的，文章在选择因变量时，选用净资产收益率(ROE)作为基本标准。净资产收益率 ROE 是指当期净利润与投资者享有的剩余权益之间的比率。它意味着投资者每 1 单位投入所能获取的净收益，能够更加可靠的衡量投资者投资回报。故而，文章财务业绩预测模型在选择因变量时以净资产收益率的变化为准。倘若下一年的净资产收益率上升，则赋值 1；倘若下一年的净资产收益率下降，则赋值 0。考虑趋势因素的影响，我们所选择的净资产收益率升降幅度是经过处理的。主要的处理方法是：对于第 i 个样本，其第 t 期的因变量为 $ROE_{it}-ROE_{it-1}-ave_{it}$ 。其中 ave_{it} 是指第 $t+1$ 期、第 t 期、第 $t-1$ 期的净资产收益率相对于其前一期净资产收益率变化的平均值。

4.1.2 自变量的选择

关于自变量的选择，文章选取的主要是对公司财务业绩产生影响的财务指标，共 24 个。首先，用 Excel 对预测数据进行预处理；然后利用 SPSS16.0 软件对经过 Excel 处理后的数据进行正态分布检验、显著性差异检验以及多重共线性检验；最终得到了可以代入模型进行财务业绩预测的输入变量。详细过程见第 3 章。

4.1.3 数据来源

我们所选取的财务数据主要来自于 CSMAR(China Stock Market& Accounting Research)数据库，样本为沪深两市所有上市的制造业公司，数据区间为 2010 年半年度。有两点需要解释：第一，之所以选择制造业，一方面是因为该行业具有大样本的性质，且经营风险相对较小，财务数据一般不会发生大幅波动；另一方面是因为本文主要想通过研究，构建一种适用于制造业财务业绩预测的模型。第二，选择半年度为预测区间，主要是考虑到短期间的预测可能更有及时性和可比性，同时半年度的财务

数据是可以获得的。

考虑到在计算因变量时需要跨几个中期会计期间，比如要预测 2010 年的净资产收益率的变化，需要采用 2011 年半年度、2010 年下半年度以及 2009 年下半年的净资产收益率数据，而通过 CSMAR 数据库所导出的数据不能直接利用，需要重新计算。剔除被 ST 的公司以及业绩增减幅度小于 50% 的公司（之所以剔除变动幅度小于 50% 的公司，笔者认为这部分公司的指标数据对公司业绩的影响不显著），最终经过处理后的研究样本数量为 264 家（其中 132 家业绩上升的公司，132 家业绩下降的公司），在此，我们将这 264 家样本分成训练集和样本集，其中训练样本 176 家（88 家业绩上升公司，88 家业绩下降公司），测试样本 88 家（44 家业绩上升的公司，44 家业绩下降的公司），具体情况如表 4.1 所示：

表 4-1 样本数据分布

样本属性	业绩上升	业绩下降
训练样本	88	88
测试样本	44	44

4.2 Logit 模型

4.2.1 Logit 回归模型的形式

通常情况下，可将线性回归模型的形式表述如下：

$$y_i = \beta_0 + \beta_j x_{ij} + \xi_i \quad (4.1)$$

其中， $y_i \in \{0,1\}$ 为离散型随机变量； (x_{ij}) ， $i=1, 2, \dots, n$ ， $j=1, 2, \dots, k$ ，为 n 组研究数据， k 是每组数据中变量的个数； (β_0, β_j) ， $i=1, 2, \dots, n$ 是回归系数。

$$\text{令 } P = \begin{cases} p_i & , y_i = 1 \\ 1 - p_i & , y_i = 0 \end{cases} \quad (4.2)$$

则有 $E(y_i) = 1 \times p_i + 0 \times (1 - p_i) = p_i$ ，因为 $E(y_i) = E(x_i \beta + u_i) = x_i \beta$ 可得：

$$p_i = x_i \beta \quad (4.3)$$

由式 (4.3) 可知，可以将线性概率模型看成是被解释变量 y_i 取值为 1 时的回归模型。并且，线性概率模型要求被解释变量取值为 0 和 1。为了提高模型的可用范围，对该模型进行了如下变化：

假定存在一个与 x_i 存在线性关系的潜在变量 y_i^* ， $y_i^* \in (-\infty, +\infty)$ 。那么，

$y_i^* = x_{ij}\beta + u_i^*$, 其中 $\beta = (\beta_0, \beta_1, \dots, \beta_k)$, $i = 1, 2, \dots, n$, $j = 1, 2, \dots, k$, 并且:

$$y_i = \begin{cases} 1, & y_i^* > 0 \\ 0, & y_i^* \leq 0 \end{cases} \quad (4.4)$$

y_i 的取值取决于 y_i^* 是否超过某个临界值, 所以 y_i 的分布为:

$$P(y_i = 1 | x_i, \beta) = P(y_i^* > 0) = P(u_i^* > -x_i\beta) = 1 - F(-x_i\beta) \quad (4.5)$$

$$P(y_i = 0 | x_i, \beta) = P(y_i^* \leq 0) = P(u_i^* \leq -x_i\beta) = F(-x_i\beta) \quad (4.6)$$

这样, 原来的线性概率模型可以写成下面的形式。

$$p_i = 1 - F(-x_i\beta) \quad (4.7)$$

当 $F(x) = \frac{e^x}{1 + e^x}$ 时, 上述模型可以表述为:

$$p_i = 1 - F(-x_i\beta) = 1 - \frac{e^{-x_i\beta}}{1 + e^{-x_i\beta}} = \frac{e^{x_i\beta}}{1 + e^{x_i\beta}} = \frac{e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik}}}{1 + e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik}}} \quad (4.8)$$

式 (4.8) 即为 Logit 回归模型的形式。

4.2.2 Logit 回归模型的估计

Logit 回归模型的参数估计方法采用的是最大似然估计法。因为 y_i 服从 (0,1)

分布, 且 y_i 的概率是 $p_i = \frac{e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik}}}{1 + e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik}}}$, 所以, y_i 的分布函数是:

$$f(y_i) = p_i^{y_i} (1 - p_i)^{1 - y_i} = \left(\frac{e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik}}}{1 + e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik}}} \right)^{y_i} \left(1 - \frac{e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik}}}{1 + e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik}}} \right)^{1 - y_i} \quad (4.9)$$

模型的似然函数是:

$$L = \prod_{i=1}^N f(y_i) = \prod_{i=1}^N \left[\left(\frac{e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik}}}{1 + e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik}}} \right)^{y_i} \left(1 - \frac{e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik}}}{1 + e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik}}} \right)^{1 - y_i} \right] \quad (4.10)$$

模型的对数似然函数是:

$$\ln L = \sum_{i=1}^N \left[y_i \ln \left(\frac{e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik}}}{1 + e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik}}} \right) + (1 - y_i) \ln \left(1 - \frac{e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik}}}{1 + e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik}}} \right) \right] \quad (4.11)$$

最大似然函数法就是通过对 β_i 值的估计, 选择 $\hat{\beta}_i$ 的值, 使得式 (4.11) 的取值最大化。

4.2.3 Logit 回归模型的检验

(1) Logit 回归模型整体显著性检验

Logit 回归模型的显著性检验就是检验 $\ln(p_i/(1-p_i))$ 是否与被解释变量之间存在显著性。原假设： H_0 : 任意 $\beta_i=0$ ；备选假设： H_1 : 存在 $\beta_i \neq 0$ 。常用的检验方法是对数似然比法和 Hosmer-Lemeshow 检验。

①对数似然比检验。如果方程中不存在解释变量，那么，方程的似然函数值等于 L_0 ，如果在方程中引入变量，则方程的似然函数值是 L_1 ，那么似然比为 L_0/L_1 ($0 \leq L_0/L_1 \leq 1$)。似然比越趋近于 0，说明解释变量对模型总体贡献比较显著；似然比越趋近于 1，说明变量对模型缺乏显著贡献。通常，为了实现显著性检验，构造似然比卡方统计量：

$$-\ln(L_0/L_1)^2 = 2\ln(L_0/L_1) \quad (4.12)$$

如果原假设满足，该统计量服从自由度为 K 的卡方分布。通过 SPSS 统计软件可以自动获取对数似然比卡方统计量的伴随概率。给定显著性水平 α ，如果 $P < \alpha$ ，则模型整体显著；否则，模型不显著。

②Hosmer-Lemeshow 检验。Hosmer-Lemeshow 检验的基本思想是：若模型整体显著，则实际值等于 1 的预测概率相对较高，而实际值等于 0 的预测概率较低。如果原假设能够满足，Hosmer-Lemeshow 服从自由度为 $(m-2)$ 的卡方分布。利用 SPSS 软件，可以方便的获取该统计量及与之相对应的 P 值。给定显著性水平 α ，如果 $P < \alpha$ ，则模型整体显著；否则，模型不显著。

(2) Logit 模型回归系数显著性检验

对于回归系数的显著性检验，Logit 回归模型通常通过构造 Wald 统计量来判别。Wald 统计量原假设： $\beta_j=0$ ，Wald 统计量具体形式为：

$$Wald = \left(\frac{\hat{\beta}_j}{S_{\hat{\beta}_j}} \right)^2 \quad (4.13)$$

其中， $\hat{\beta}_j$ 为第 j 个解释变量的回归系数的估计值， $S_{\hat{\beta}_j}$ 为 $\hat{\beta}_j$ 的标准差。利用 SPSS 软件，可以得出 Wald 统计量以及与之相对应的伴随概率 P 。给定显著性水平 α ，如果 $P < \alpha$ ，第 j 个回归系数显著不为 0；反之，第 j 个回归系数，显著为 0。

(3) Logit 回归模型拟合优度检验

一般用 Cox-Snell R^2 统计量与 Nagelker R^2 统计量来对 Logit 回归模型进行拟合优度检验。

①Cox-Snell R^2 统计量。Cox-Snell R^2 统计量的值越大说明模型的拟合优度越高；反之，则说明模型的拟合优度越小。

②Nagelker R^2 统计量。Nagelker R^2 统计量是对 Cox-Snell R^2 的修正。 $0 < \text{Nagelker } R^2 \text{ 值} < 1$ ，Nagelker R^2 越接近 1，说明模型具有较高的拟合优度；Nagelker R^2 越接近 0，说明模型拟合优度越低。

4.2.4 半年度预测的 Logit 模型的构建

表 4-2 显示的是利用半年度财务指标进行财务业绩预测 Logit 回归模型显著性检验的结果。通过该表可知，对数似然比卡方统计量是 3.702，与其相对应的伴随概率 p 值为 0.883，大于显著性水平 0.05，因此，回归模型在整体上不显著。

表 4-2 Logit 回归模型整体显著性检验结果

项目	对数似然比检验法
卡方统计量	3.702
自由度	8
相应 P 值	0.883

表 4-3 给出的是财务业绩预测 Logit 回归模型拟合优度评价的指标。-2Log likelihood 统计量与对数自然函数值成反比，是对数自然函数值的-2 倍，该值越小意味着模型的拟合优度较高，由表知，-2Log likelihood 统计量为 215.64；而 Nagelkerke R Square 为 0.198，这说明模型能够解释被解释变量 19.8%的变动。

表 4-3 Logit 回归模型拟合优度评价指标

-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
215.640	0.149	0.198

表 4-4 半年度预测 Logit 回归系数估计值及其显著性检验结果

公共因子	回归系数估计值	估计值的标准差	Wald 统计量	自由度	相应 p 值	Exp(B)
F1	-.915	.391	5.492	1.000	.019	.400
F2	.196	1.333	.022	1.000	.883	1.216
Constant	.211	.178	1.419	1.000	.234	1.235

表 4-4 依次给出了模型输入变量、回归系数估计值、估计值标准差、Wald 统计量、自由度、伴随概率以及第 i 个变量的边际影响。由表可知，公因子 F_1 与上市公司业绩上升成负相关， F_2 与上市公司业绩上升成负相关。从 Wald 统计量和对应的伴随

概率来看,公因子 F_1 的系数的伴随概率为 0.019, 小于 0.05, 通过了显著性检验, 说明公因子 F_1 对模型具有显著性影响。而公因子 F_2 的系数则没有通过显著性检验, 对模型的影响不显著。

将表 4-4 计算出的回归系数估计值代入到 Logit 回归模型之中, 可以得出样本的概率值 p_i , 如果 p_i 大于 0.5, 公司被判定为业绩上升, 如果 p_i 小于 0.5, 公司被判别为业绩下降。表 4-5 给出了上述回归模型回代及预测的结果。

表 4-5 Logit 回归模型回代和预测结果

实际观测值		回代结果			预测结果		
		业绩上升或下降		正确率	业绩上升或下降		正确率
		+1	-1		+1	-1	
业绩上升或	+1	62	26	70.5%	33	11	75.0%
下降	-1	39	49	55.7%	24	20	45.5%
总体预测准确率		63.1%			60.2%		

通过表 4-4 可知, 回代结果中犯第一类错误的概率为 29.5%, 犯第二类错误的概率为 44.3%, 犯第二类错误的概率远高于犯第一类错误的概率, 总体预测准确率为 63.1%; 预测结果中犯第一类错误的概率为 25.0%, 犯第二类错误的概率为 54.5%, 远高于犯第一类错误的概率, 模型整体预测准确率为 60.2%。

为了获得较准确的预测结果, 文章先后对初始数据进行了不同的处理和 Logit 回归模型的构建, 在进行比较的基础上, 文章选择了上述的样本及预测模型。现将研究过程大致叙述如下:

首先, 以不加剔除的初始指标进行研究。在文章首次进行实证研究时, 所选用的初始指标中, 没有剔除 ST 公司以及财务业绩变动幅度小于 50% 的公司。通过进行正态分布检验可知, 所选初始指标不符合联合正态分布, 只有 X_3 和 X_4 通过了显著性检验, 而其余指标皆不服从正态分布。通过进行显著性检验可知, 文中 X_1 、 X_2 、 X_3 、 X_4 、 X_5 、 X_6 、 X_7 、 X_8 、 X_{10} 、 X_{16} 、 X_{17} 、 X_{18} 、 X_{22} 以及 X_{24} 均未通过显著性检验, 应当剔除; 将保留的指标进行因子分析可知 KMO 检测值为 0.596, 大于 0.5, Bartlett 球形值为 5769, 数值较大, 适宜做因子分析处理。通过因子分析, 保留指标中共提取出三个公共因子, 公因子的累计贡献率为 69.11%, 说明公因子能够解释初始指标 69.11% 的特征。在根据训练样本所构建的 Logit 回归模型中, 模型犯第一类错误的概率为 29.9%, 犯第二类错误的概率为 52.5%, 模型总体预测准确率为 58.8%; 在预测结果中, 模型犯第一类错误的概率为 30%, 犯第二类错误的概率为 53.8%, 模型总体

预测准确率为 56.6%，小于文章所采用模型的预测准确率。

其次，以未经过趋势因素处理的净资产收益率作为被解释变量进行研究。为了构建比较合理的预测模型，文章继而以未经过剔除趋势因素的净资产收益率作为被解释变量，依次进行了正态分布检验、显著性检验以及因子分析处理。通过正态分布检验可知，只有 X_3 和 X_4 通过了正态分布检验，其余指标皆不服从正态分布，样本总体不服从联合正态分布。通过显著性检验可知，只有 X_2 、 X_5 、 X_9 、 X_{11} 、 X_{12} 、 X_{13} 、 X_{14} 、 X_{15} 、 X_{19} 、 X_{20} 以及 X_{23} 共 11 个指标在 10% 的显著性水平上通过了检验，应当保留，其余指标应当剔除。将这 11 个指标进行因子分析处理：其中 KMO 检验值为 0.547，大于 0.5，Barlett 球形检验值为 19950，数值很大，适宜做因子分析处理；通过研究，这 11 个指标中共提取出 5 个公共因子，累计贡献率为 73.23%，说明这五个公因子能够解释 73.23% 的初始指标变动度。通过训练样本得出的 Logit 回归模型整体不显著，模型犯第一类错误的概率为 32.7%，犯第二类错误的概率为 53.4%，模型整体预测准确率为 57%；预测模型犯第一类错误的概率为 34.6%，犯第二类错误的概率为 47.1%，模型整体预测准确率为 59.1%。

通过分析比较可知，在对研究样本进行剔除 ST 公司、剔除变动幅度小于 50% 的公司等预处理之后，构建的 Logit 回归模型具有较高的预测准确性。所以，文章在后续构建最小二乘支持向量机模型时，采用的是经过上述处理之后的样本。

4.3 LS-SVM 模型

4.3.1 基于 LS-SVM 的财务业绩预测

前已述及，最小二乘支持向量机是对标准支持向量机一种改进模式，它用仅包括等式约束的最优化问题替代含有不等式约束的二次规划问题，求解速度较快。利用 LS-SVM 模型进行财务业绩预测的过程如下所示：

第一，给定一个训练集样本 $T = \{(x_i, y_i), i = 1, 2, \dots, l\}$, $x_i \in R^n$, $y_i \in \{1, 0\}$;

第二，选择适当的核函数 $K(x_i, x)$ 以及惩罚因子 C ，构建如下最优化的问题，并对其求解：

$$\min_{\alpha} \frac{1}{2} \|w\|^2 - \frac{1}{2} C \sum_{i=1}^l \xi_i^2$$

$$s.t. y_i (w^T \varphi(x) + b) = 1 - \xi, i = 1, 2, \dots, l$$

第三, 对式 $L(w, b, \xi, a) = \frac{1}{2} \|w\|^2 + \frac{1}{2} C \sum_{i=1}^l \xi_i^2 - \sum_{i=1}^l a_i [y_i w^T \varphi(x_i) + b + \xi - 1]$ 关于

a, b, w, ξ 求偏导, 并令其为 0, 可以得出:

$$\begin{bmatrix} 0 & -Y^T \\ Y & \Omega + C^{-1}I \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} 0 \\ \bar{1} \end{bmatrix}$$

其中

$\Omega_i = y_i y_j \varphi^T(x_i) \varphi^T(x_j) = y_i y_j K(x_i, x_j), a = (a_1, a_2, \dots, a_l)^T, \bar{1} = (1, 1, \dots, 1)^T, i = 1, 2, \dots, l, I$ 是单位矩阵, $Y = (y_1, y_2, \dots, y_l)^T$;

第四, 求解上一步的现行方程组, 得出 b 值和 a 值;

第五, 选定测试样本集的数据 $Q = \{(x_i, y_i), i = 1, 2, \dots, p\}, x_i \in R^n, y_i \in \{1, 0\}, p < l$;

第六, 把测试样本集数据代入到决策函数 $f(x) = \text{sgn}\left(\sum_{i=1}^l \alpha_i y_i K(x_i, x) + b\right)$, 求出相

应的 $f(x)$ 值。倘若 $f(x) = 1$, 便把这个点划归为 $y=1$ 这一类中 (与之相对应的是财务业绩上升的公司); 倘若 $f(x) = 0$, 那么便把这个点归类于 $y=0$ 这一类中 (与之相对应的是财务业绩下降的公司)

4.3.2 基于LS-SVM模型构建及实证检验

赵冠华 (2010) 通过研究得出: 与传统统计类 Logit 模型和标准支持向量机模型相比, 最小二乘支持向量机模型的预测效果要优于前两者; 与多项式核函数相比, 高斯核更适于做支持向量机的核函数。因此, 文章在此构建 LS-SVM 模型 (高斯核)。模型数据检验包括预测结果和回代结果两种情形。预测结果是将预测样本数据代入模型得到的结果; 回代结果是将训练样本集数据代入模型得出的结果。

回代结果, 当高斯核函数 $\delta^2=1.1$, 惩罚因子 $C=13$ 时, 176 个样本中有 74 个误判, 模型误判率为 42%, 预测准确率为 58%。其中犯第一类错误的概率是 35.2%, 犯第二类错误的概率是 49%, 犯第二类错误的概率要高于犯第一类错误的概率。

预测结果: 当高斯核函数 $\delta^2=1.1$, 惩罚因子 $C=13$ 时, 88 个预测样本中有 31 个误判, 误判率 35.2%, 预测准确率 64.8%。其中, 犯第一类错误的概率是 30%, 犯第二类错误的概率是 41%。具体结果如图 4-6。

表 4-6 LS-SVM 模型（高斯核）回代及预测结果

实际观测值		回代结果			预测结果		
		业绩上升或下降		正确率	业绩上升或下降		正确率
		+1	-1		+1	-1	
业绩上升或	+1	57	31	64.8%	31	13	70%
下降	-1	43	45	51.1%	18	26	59%
总体预测准确率		58%			64.8%		

4.4 实证结果的分析与比较

在上文对指标进行预处理及构建 Logit 回归模型和 LS-SVM 模型时，我们已经阐述过，本文选取 24 个初始财务指标，在进行比较和处理后，提取出两个公共因子。文章最终选取了 264 家制造业上市公司作为研究样本，其中 176 家为训练样本（88 家业绩上升，88 家业绩下降），88 家预测样本（其中 44 家业绩上升，44 家业绩下降），以对不同模型之间进行比较。两个模型的试验结果如表 4-7 所示。

表 4-7 两种模型预测准确率比较

预测模型	回代准确率	预测准确率	预测犯 I 类错误概率	预测犯 II 类错误概率	训练样本数	预测样本数
Logit 模型	63.1%	60.2%	25.0%	54.5%	176	88
LS-SVM	58%	64.8%	30%	41%	176	88

通过表 4-7 可以看出：第一，Logit 回归模型与 LS-SVM 模型存在一个共同现象，即犯第 II 类错误的概率要高于犯第 I 类错误的概率，两模型均存在较高的存伪现象，也就是说两模型在判断财务业绩上升公司的效果要好于判别财务业绩下降公司的效果。第二，LS-SVM 模型预测准确率略高于 Logit 回归模型的准确率。其中 LS-SVM 模型的预测准确率为 64.8%，Logit 回归模型的预测准确率为 60.2%。这说明 LS-SVM 模型相对于 Logit 回归模型而言，已有一定程度的改进。

第 5 章 基于 LS-SVM 的财务业绩预测案例分析

文章研究的目的是希望提出一种适用于制造业上市公司财务业绩预测的方法，通过判别公司未来财务业绩变动趋势，为利益相关者提供信息。本章在前文实证研究的基础上，将实证结果应用于具体公司财务业绩预测。

5.1 上市公司简介

5.1.1 开开实业公司简介

上海开开实业股份有限公司，股票代码 600272，前身是开开百货商店，创建于 1936 年。经过数十年的发展，由一家销售自产衬衫和羊毛衫的小型企业，发展为全国比较知名的服装企业之一。公司于 1992 年 12 月设立为股份有限公司，总股数 3070.33 万股，其中国家股 1120.33 万股，法人股 1450 万股，内部职工股 500 万股。1996 年 12 月 19 日，公司在上海证券交易所向社会公众发行 8000 万股外资股 (B 股)，并于 1997 年 1 月 8 日在上交所上市流通。

公司的主要经营范围包括：生产和销售衬衫、羊毛衫、针纺织品、服装、鞋帽，纺织面料等，并提供相关的售后服务；对内销售日用百货，五金交电、一般工艺品、皮革制品、玻璃制品、雨具；出租自有房屋；对外投资。

5.1.2 上海棱光实业公司简介

上海棱光实业股份有限公司，股票代码 600629，是由原全民所有制企业上海石英玻璃厂改制而成，1992 年 5 月 19 日经批准，采用募集方式设立的股份有限公司。公司股票于 1993 年 2 月 9 日在上海证券交易所正式上市交易。2007 年 9 月 12 日经过证监会批准，按照股东会决议，向上海建筑材料(集团)总公司(以下简称“建材集团”)发行 117622929 股普通股，购买建材集团相关资产，变更之后的注册资本达到 269000527 元。

公司主营业务范围：制造销售石英玻璃，电子仪表，半导体材料，工业气体并提供气体充装、气瓶检验服务；制造和销售化工产品，机电产品；日用百货，针纺织品的销售；进出口业务，房屋、汽车租赁业务；提供旅客运输及货物运输服务等。

5.2 公司的财务状况

5.2.1 开开实业的财务状况

开开实业股份有限公司于 2010 年中期产收益率为 10.55%，实现主营业务收入 43315.87 元，净利润 2714.47 元，净利润率为 6.27%，每股现金流量为 0.16。至 2010 年 12 月，开开实业净资产收益率下降为 8.55%，主营业务收入为 86689.78 元，净利润为 2670.38 元，净利润率为 3.08%，每股现金流量为 0.19。除企业每股现金流量有小幅增加外，企业经营业绩呈现下滑状态。表 5-1 给出了开开实业各项指标数据。

表 5-1 开开实业各项指标数据

指标名称	指标数据	指标名称	指标数据
流动比率	0.729	资产报酬率	0.035
速动比率	0.515	总资产净利润率	0.028
营运资金对资产总额比率	-0.129	固定资产净利润率	0.116
资产负债率	0.708	营业收入增长率	-0.062
长期负债比率	0.328	固定资产增长率	-0.040
有形净值债务率	2.474	总资产增长率	-0.095
长期负债与权益比率	7.537	净利润增长率	-0.305
应收账款周转率	3.647	营业收入现金比率	0.090
存货周转率	3.433	每股经营活动现金净流量	0.160
固定资产周转率	1.771	每股现金净流量	-0.019
总资产周转率	0.432	销售收到现金比率	1.146
营业利润率	0.066	利润现金比率	2.132

5.2.2 棱光实业的财务状况

棱光实业股份有限公司于 2010 年上半年净资产收益率为 6.27%，实现营业利润为 29771249.12 元，比去年同期增长 1007.87%，利润总额为 29746249.86 元。2010 年 12 月 31 日公司净资产收益率为 7.00%，公司总体业绩呈现上升趋势。表 5-2 给出了棱光实业各项财务指标数据。

表 5-2 棱光实业各项指标数据

指标名称	指标数据	指标名称	指标数据
流动比率	3.139	资产报酬率	0.046
速动比率	3.088	总资产净利润率	0.034
营运资金对资产总额比率	0.487	固定资产净利润率	0.311
资产负债率	0.258	营业收入增长率	0.896
长期负债比率	0.116	固定资产增长率	-0.204
有形净值债务率	0.390	总资产增长率	0.030
长期负债与权益比率	3.366	净利润增长率	5.630
应收账款周转率	3.779	营业收入现金比率	-0.447
存货周转率	12.342	每股经营活动现金净流量	-0.318

固定资产周转率	2.702	每股现金净流量	0.012
总资产周转率	0.299	销售收到现金比率	0.205
营业利润率	0.156	利润现金比率	0.315

5.3 基于LS-SVM的财务业绩预测结果分析

5.3.1 开开实业的预测结果分析

经过处理之后，开开实业所提取的两个公因子分别为： $F_1=0.09$ ， $F_2=0.06$ 。将公因子代入到LS-SVM模型中，最终得出当高斯核 $\delta^2=5.5$ ，惩罚因子 $C=15$ 时，得出预测值为-1，这说明LS-SVM模型将开开实业判定为财务业绩下降的公司。这与公司的实际情况相符，说明模型判断正确。

5.3.2 棱光实业的预测结果分析

将棱光实业的初始指标经过处理之后，得出两个公因子，它们分别是： $F_1=0.96$ ， $F_2=-0.36$ 。将公因子代入到LS-SVM模型中，最终得出当高斯核 $\delta^2=5.5$ ，惩罚因子 $C=15$ 时，得出预测值为+1，这说明LS-SVM模型将棱光实业判定为财务业绩上升的公司。这与公司的实际情况相符，说明模型判断正确。

第 6 章 研究结论及不足

6.1 研究结论

财务业绩预测主要是通过利用企业公开发布的财务报表中的一系列财务指标进行分析,采用科学的预测和决策方法,对企业总体上的财务业绩水平进行判断,以期预测企业在未来一段时间内的财务业绩的变动幅度。

自从奥特曼对财务困境发韧开拓性研究以来,对财务困境预测及分析的课题得以空前发展。该发展不仅为投资者、债权人、银行等进行投资决策提供了信息,而且为企业经营者及时调整企业经营战略提供了帮助。

但是,随着对该课题研究的日益成熟,不少研究者发现:财务困境预测固然具有一定的可预见性,但是以公司是否被 ST 作为预测标准的话,难免会为时已晚。所以,有学者开始对财务业绩预测进行研究。希望能够通过公司所提供的一系列财务指标等信息判断公司财务业绩的发展趋势。

现代统计学理论的丰富和发展为统计学习研究提供了扎实的理论基础。建立于统计学理论基础之上的支持向量机理论适用于小样本研究,能够在模型的复杂性及泛化性之间寻求一个均衡,克服了神经网络易陷入极端性的缺点。

本文在前人研究的基础上,在财务业绩预测过程中引入最小二乘支持向量机,并采用半年度数据进行预测,最终得出 LS-SVM 模型的预测准确率比 Logit 回归模型的预测准确率越高。具体而言,文章的主要工作和创新点可综述如下:

第一,将最小二乘支持向量机引入财务业绩预测。通过研究发现基于最小二乘支持向量机的财务业绩预测模型判断准确率为 64.8%,模型预测准确率比 Logit 回归模型的预测准确率略高。这可能与数据处理过程中,提取公因子时丢失了大量信息有关。

第二,研究视角的不同,论文首次研究了半年度财务业绩预测。论文在进行研究时主要采用的是半年度数据,这就为财务业绩预测提供了一定的及时性。利用半年度财务指标进行业绩预测能够在较短的时间内及时发现问题,从而为利益相关者提供决策所需信息。

6.2 研究的不足之处

在本文的研究和写作过程之中,笔者阅读了国内外大量的参考文献,并取得了一

定的研究成果。但是，限于时间以及个人能力的有限，文章尚存在如下几点不足需要后期改进和研究。

首先，研究样本的不足。本文在进行实证研究时，主要选择的是中国上市制造业的公司，没有考虑未上市的中小企业、民营企业以及其他小样本行业的公司。事实上，其他小样本行业及中小民营企业在国民经济中做出了较大的贡献，比如信息技术业的发展在一定程度上代表了一个国家的生产力水平。但是，文章在选择样本时，由于各方面原因，剔除了该部分样本，从而得出的模型仅适合于制造业。

其次，本文研究的目的在于构建一个合适的模型去预测企业的业绩是上升还是下降，其根本在于将企业的财务状况分为两类：财务业绩上升(+1)和财务业绩下降(-1)两大类。事实上，根据公司财务业绩变动幅度的大小可以将公司分为多种种类。

再者，在模型输入变量选择上的不足。文章在选择输入变量时，仅仅考虑了对企业财务业绩产生影响的财务指标。这些指标均可在公司公开发行的财务报表中获得。但是，能够影响和反映公司财务业绩的要素有很多，不仅仅局限于财务指标。比如宏观方面的自然现象、社会现象及国家经济政策的变更；中观行业因素的影响；微观企业内部股东大会成员的构成比例、经理人的职业道德和能力等等。这些资料往往在公司公开发行的财务报告中可能无法全面获知，但是这些因素的确对企业的财务业绩存在影响。如何对这些因素进行定性的分析，如何将这些因素同企业模型的定量预测相结合，限于时间和精力，文章尚未对此进行研究，而这也将是以后的一个研究课题。

参考文献

- [1]Frizpatrick. A Comparison of ratios of Successful Industrial Enterprises with those of Failed Firms[M]. Iview York: Certified Public Accountant, 1932
- [2]Beaver WH. Financial ratios as predictors of failure[J]. Journal of Accounting research, 1966, (Supplement)
- [3]陈静. 上市公司财务恶化预测的实证分析[J]. 会计研究, 1999,6:34-37
- [4]Altman EdwardI.Financial ratios,discriminant analysis and the prediction ofbankruptcy[J].Journal of Finance,1968, 23:589-609
- [5] Altman,Haldeman,Narayanan.a new model to identify bankruptcy risk of corporations[J].Journal of Banking and Finance,1977(9)
- [6] Deakin E.B.,A discriminant analysis of prediction of business failure[J],supplyment to journal of accounting research,1972
- [7] Scott A. Wolpert, Asymptotics of the spectrum and the Selberg zeta function on the space of Riemann surfaces, Communications in Mathematical Physics,1987(2): 283~315
- [8]周首华, 杨济华, 王平. 论财务危机的预警分析—F 分数模型[J].会计研究,1996,8:8-11
- [9]张玲. 财务危机预警分析判别模型[J].数量经济技术经济研究.2000,3:49-51
- [10]高培业, 张道奎. 企业失败判别模型实证研究[J], 会计研究.2000 (10)
- [11] 向德伟. 运用“Z 记分法”评价上市公司经营风险的实证研究[J],会计研究.2002(11)
- [12]杨淑娥,徐伟刚.上市公司财务预警模型—Y 分数模型的实证研究[J],中国软科学,2003,1:56-60
- [13]Ohlson J.Financial Ratio and the Probabilistic Prediction of Bankruptcy[J].Journalof Accounting Research,1980,18:109-131
- [14]陈晓,陈治鸿.中国上市公司的财务困境预测[J].中国会计与财务研究.2000.9
- [15]张后奇,刘月平,江洪波.上市公司财务危机预警系统:理论与实证分析.长江证券课题组,2002
- [16]齐治平,余妙志.Logit 模型在上市公司财务状况评价中的运用,东北财经大学学报,2002,(1)
- [17]Odom M,Sharda R.A. Neural Networks Model for Bankruptcy Prediction[J]. Proceedings of the IEEE International Conference on Neural Network,1990,2:163-168
- [18]Coats Pamela K,Franklin FL. Recognizing Financial Distress Patterns Using aNeural Network tool[J].Financial Management,1993,3:142-166
- [19]李秉成.企业财务困境概念内涵的探讨.山西财经大学学报,2003,12:109—112.

参考文献

- [20]吴德胜,梁樛,殷尹.不同模型在财务预警实证分析中的比较研究[J].管理工程学报,2004,2:105-108
- [21]杨淑娥,黄礼.基于 BP 神经网络的上市公司预警模型[J].系统工程理论与实践,2005,1:12-18
- [22]Varetto F. Genetical gorithms applications in the analysis of insolveney risk[J]. Journal of Banking and Finanee,1998,22:1421-1439
- [23]Kyung-Shik Shin,Taik Soo Lee,Hyun-jung Kim. An application of supportvector machines in bankruptcy prediction model.Expert Systems with Applications[J],2005,28:127-135
- [24]张华伦,孙毅.企业财务危机预警 Rough-Fuzzy-ANN 模型的建立及应用[J].运筹与管理,2006,2:23-26
- [25]Frydman H,ALTMAN E I, KAO D L.Introducing recursive partitioning forfinancial classification: the case of financial distress[J].Journal of Finance,1985,40:269-292
- [26]Fan A,Palaniswami M.Selecting bankruptcy predictors using a support vector machine approach[J]. Proceeding of the international joint conference on neural network,2000,6:354-359
- [27]Van Gestel T,Baesens B,Suykens J etc. Bankruptcy prediction with least squares support vector machine classifiers.computational intelligence for financial engineering,In IEEEEE international conference proceeding,2003,1-8
- [28]Jae H Min,Young-Chan Lee. Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters[J].Expert Systems withApplications,2005,28:603-614
- [29]李贺,冯天谨.一种基于 SVM 的多变量企业财务预警模型[J].通讯和计算机,2005,2:31-36
- [30]徐晓燕.企业财务困境预测方法研究[D].合肥:中国科学技术大学,2006
- [31]姚宏善.基于支持向量机的财务困境预测研究[D].武汉:华中科技大学,2006
- [32]宋姣.基于支持向量机的上市公司财务危机预警研究[D].哈尔滨:哈尔滨工业大学,2007
- [33]蒋艳霞,徐程兴. 基于集成支持向量机的企业财务业绩分类模型研究[J].中国管理科学,2009,2:42-51
- [34]赵冠华.企业财务困境分析与预测方法研究[D].天津:天津大学,2009

附录

附录 A 半年度预测指标数据相关系数矩阵

	X ₁₁	X ₁₂	X ₁₃	X ₁₄	X ₁₅	X ₁₉	X ₂₀	X ₂₁
X ₁₁	1.0	-0.1	0.3	0.3	0.1	0.1	0.1	0.1
X ₁₂	-0.1	1.0	0.1	0.1	0.2	0.1	-1.0	0.0
X ₁₃	0.3	0.1	1.0	1.0	0.6	0.3	0.1	0.4
X ₁₄	0.3	0.1	1.0	1.0	0.6	0.4	0.0	0.3
X ₁₅	0.1	0.2	0.6	0.6	1.0	0.1	-0.1	0.1
X ₁₉	0.1	0.1	0.3	0.4	0.1	1.0	0.0	0.1
X ₂₀	0.1	-1.0	0.1	0.0	-0.1	0.0	1.0	0.1
X ₂₁	0.1	0.0	0.4	0.3	0.1	0.1	0.1	1.0
X ₁₁		0.0	0.0	0.0	0.0	0.1	0.0	0.0
X ₁₂	0.0		0.1	0.1	0.0	0.2	0.0	0.5
X ₁₃	0.0	0.1		0.0	0.0	0.0	0.2	0.0
X ₁₄	0.0	0.1	0.0		0.0	0.0	0.2	0.0
X ₁₅	0.0	0.0	0.0	0.0		0.0	0.0	0.1
X ₁₉	0.1	0.2	0.0	0.0	0.0		0.5	0.0
X ₂₀	0.0	0.0	0.2	0.2	0.0	0.5		0.1
X ₂₁	0.0	0.5	0.0	0.0	0.1	0.0	0.1	

附录 B 用于半年度预测的 176 个训练样本

财务业绩下降公司				财务业绩上升公司			
序号	股票代码	公司名称	上市时间	序号	股票代码	公司名称	上市时间
1	000020	深华发 A	1992-04-28	1	000016	深康佳 A	1992-03-27
2	000045	深纺织 A	1994-08-15	2	000032	深桑达 A	1993-10-28
3	000055	方大集团	1996-04-15	3	000158	常山股份	2000-7-24
4	000059	辽通化工	1997-1-30	4	000159	国际实业	2000-9-26
5	000078	海王生物	1998-12-18	5	000400	许继电气	1997-04-18
6	000155	川化股份	2000-9-26	6	000407	胜利股份	1996-7-3
7	000422	湖北宜化	1996-8-15	7	000418	小天鹅 A	1997-03-28
8	000522	白云山 A	1993-11-08	8	000519	江南红箭	1993-10-08
9	000525	红太阳	1993-10-28	9	000541	佛山照明	1993-11-23
10	000528	柳工	1993-11-18	10	000551	创元科技	1994-01-06
11	000530	大冷股份	1993-12-08	11	000565	渝三峡 A	1994-4-8
12	000533	万家乐	1994-01-03	12	000581	威孚高科	1998-09-24
13	000550	江铃汽车	1993-12-01	13	000589	黔轮胎 A	1996-3-8
14	000553	沙隆达 A	1993-12-3	14	000599	青岛双星	1996-4-30
15	000566	海南海药	1994-05-25	15	000615	湖北金环	1996-10-16
16	000597	东北制药	1996-05-23	16	000617	石油济柴	1996-10-22
17	000612	焦作万方	1996-09-26	17	000635	英力特	1996-11-20
18	000625	长安汽车	1997-06-10	18	000637	茂化实华	1996-11-14
19	000627	天茂集团	1996-11-12	19	000678	襄阳轴承	1997-01-06
20	000661	长春高新	1996-12-18	20	000687	保定天鹅	1997-2-21
21	000662	索芙特	1996-12-16	21	000702	正虹科技	1997-3-18
22	000665	武汉塑料	1996-12-10	22	000727	华东科技	1997-05-20
23	000710	天兴仪表	1997-04-22	23	000738	中航动控	1997-06-26
24	000731	燕京啤酒	1997-7-16	24	000739	普洛股份	1997-05-09

附录

25	000760	博盈投资	1997-06-27	25	000786	北新建材	1997-06-06
26	000766	通化金马	1997-04-30	26	000788	西南合成	1997-06-16
27	000792	盐湖股份	1997-9-4	27	000789	江西水泥	1997-09-23
28	000800	一汽轿车	1997-06-18	28	000790	华神集团	1998-03-27
29	000803	金宇车城	1998-3-3	29	000815	美利纸业	1998-6-9
30	000806	银河科技	1998-04-16	30	000816	江淮动力	1997-08-18
31	000813	天山纺织	1998-5-19	31	000823	超声电子	1997-10-08
32	000848	承德露露	1997-11-13	32	000830	鲁西化工	1998-8-7
33	000862	银星能源	1998-09-15	33	000850	华茂股份	1998-10-7
34	000901	航天科技	1999-04-01	34	000877	天山股份	1999-01-07
35	000912	泸天化	1999-6-3	35	000880	山东巨力	1998-04-02
36	000913	钱江摩托	1999-05-14	36	000885	同力水泥	1999-03-19
37	000927	一汽夏利	1999-07-27	37	000902	中国服装	1999-4-8
38	000929	兰州黄河	1999-6-23	38	000903	云内动力	1999-04-15
39	000951	中国重汽	1999-11-25	39	000910	大亚科技	1999-6-30
40	000952	广济药业	1999-11-12	40	000915	山大华特	1999-06-09
41	000980	金马股份	2000-06-16	41	000919	金陵药业	1999-11-18
42	000985	大庆华科	2000-07-26	42	000923	河北宣工	1999-07-14
43	000988	华工科技	2000-06-08	43	000928	中钢吉炭	1999-03-12
44	002017	东信和平	2004-7-13	44	000936	华西村	1999-8-10
45	002018	华星化工	2004-7-13	45	000959	首钢股份	1999-12-16
46	002019	鑫富药业	2004-7-13	46	002004	华邦制药	2004-06-25
47	002020	京新药业	2004-07-15	47	002008	大族激光	2004-06-25
48	002047	成霖股份	2005-05-31	48	002028	思源电气	2004-08-05
49	002057	中钢天源	2006-08-02	49	002037	久联发展	2004-9-8
50	002074	东源电器	2006-10-18	50	002044	江苏三友	2005-5-18
51	002076	雪莱特	2006-10-25	51	002053	云南盐化	2006-06-27
52	002097	山河智能	2006-12-22	52	002054	德美化工	2006-7-25
53	002102	冠福家用	2006-12-29	53	002055	得润电子	2006-07-25
54	002107	沃华医药	2007-01-24	54	002056	横店东磁	2006-08-02
55	002117	东港股份	2007-3-2	55	002058	威尔泰	2006-08-02
56	002137	实益达	2007-06-13	56	002094	青岛金王	2006-12-16
57	002160	常铝股份	2007-08-21	57	002096	南岭民爆	2006-12-22
58	002165	红宝丽	2007-9-13	58	002101	广东鸿图	2006-12-29
59	002173	山下湖	2007-9-25	59	002103	广博股份	2007-1-10
60	002175	广陆数测	2007-10-12	60	002114	恒宝股份	2007-01-10
61	002177	御银股份	2007-11-01	61	002118	紫鑫药业	2007-03-02
62	002182	云海金属	2007-11-13	62	002119	康强电子	2007-03-02
63	002215	诺普信	2008-2-18	63	002127	新民科技	2007-4-18
64	002239	金飞达	2008-5-22	64	002132	恒星科技	2007-04-27
65	200706	瓦轴B	1997-03-25	65	002143	高金食品	2007-7-20
66	600031	三一重工	2003-07-03	66	002149	西部材料	2007-08-10
67	600063	皖维高新	1997-5-28	67	002154	报喜鸟	2007-8-16
68	600073	上海梅林	1997-7-4	68	002162	斯米克	2007-08-23
69	600085	同仁堂	1997-06-25	69	002168	深圳惠程	2007-09-19
70	600129	太极集团	1997-11-18	70	002169	智光电气	2007-09-19
71	600151	航天机电	1998-06-05	71	002180	万力达	2007-11-13
72	600160	巨化股份	1998-6-26	72	002197	证通电子	2007-12-18
73	600166	福田汽车	1998-06-02	73	002198	嘉应制药	2007-12-18

74	600176	中国玻纤	1999-4-22	74	002203	海亮股份	2008-01-16
75	600186	莲花味精	1998-8-25	75	002206	海利得	2008-1-23
76	600191	华资实业	1998-12-10	76	002214	大立科技	2008-02-18
77	600195	中牧股份	1999-1-7	77	002217	联合化工	2008-2-20
78	600197	伊力特	1999-9-16	78	002218	拓日新能	2008-02-28
79	600199	金种子酒	1998-8-12	79	002226	江南化工	2008-5-6
80	600202	哈空调	1999-06-03	80	002227	奥特迅	2008-05-06
81	600212	江泉实业	1999-08-17	81	002235	安妮股份	2008-5-16
82	600218	全柴动力	1998-12-03	82	002246	北化股份	2008-6-5
83	600220	江苏阳光	1999-9-27	83	002265	西仪股份	2008-08-06
84	600230	沧州大化	2000-4-6	84	200771	杭汽轮 B	1998-04-28
85	600232	金鹰股份	2000-6-2	85	600055	万东医疗	1997-05-19
86	600243	青海华鼎	2000-11-20	86	600060	海信电器	1997-04-22
87	600253	天方药业	2000-12-27	87	600069	银鸽投资	1997-4-30
88	600260	凯乐科技	2000-7-6	88	600070	浙江富润	1997-6-4

附录 C 用于半年度预测的 88 个测试样本

财务业绩下降公司				财务业绩上升公司			
序号	股票代码	公司名称	上市时间	序号	股票代码	公司名称	上市时间
1	600272	开开实业	1997-1-8	1	600072	中船股份	1997-06-03
2	600302	标准股份	2000-12-13	2	600090	啤酒花	1997-6-16
3	600315	上海家化	2001-3-15	3	600110	林海股份	1997-07-04
4	600320	振华重工	2000-12-21	4	600112	中科英华	1997-10-7
5	600329	中新药业	2001-06-06	5	600117	长征电气	1997-11-27
6	600332	广州药业	2001-02-06	6	600132	西宁特钢	1997-10-15
7	600363	联创光电	2001-03-29	7	600162	重庆啤酒	1997-10-30
8	600371	万向德农	2002-9-16	8	600177	雅戈尔	1998-11-19
9	600375	星马汽车	2003-4-1	9	600184	光电股份	2003-11-06
10	600391	成发科技	2001-12-12	10	600211	西藏药业	1999-07-21
11	600397	安源股份	2002-7-2	11	600233	大杨创世	2000-6-8
12	600432	吉恩镍业	2003-09-05	12	600249	两面针	2004-1-30
13	600456	宝钛股份	2002-04-12	13	600268	国电南自	1999-11-18
14	600470	六国化工	2004-3-5	14	600282	南钢股份	2000-09-19
15	600478	科力远	2003-09-18	15	600285	羚锐制药	2000-10-18
16	600486	扬农化工	2002-4-25	16	600297	美罗药业	2000-11-16
17	600519	贵州茅台	2001-8-27	17	600305	恒顺醋业	2001-2-6
18	600521	华海药业	2003-03-04	18	600308	华泰股份	2000-9-28
19	600532	华阳科技	2002-10-31	19	600312	平高电气	2001-02-21
20	600549	厦门钨业	2002-11-07	20	600316	洪都航空	2000-12-15
21	600550	天威保变	2001-2-28	21	600319	亚星化学	2001-3-26
22	600555	九龙山	2001-3-28	22	600331	宏达股份	2001-12-20
23	600572	康恩贝	2004-04-12	23	600353	旭光股份	2002-11-20
24	600595	中孚实业	2002-06-26	24	600378	天科股份	2001-1-11
25	600599	熊猫烟花	2001-8-28	25	600379	宝光股份	2002-1-16
26	600605	汇通能源	1992-3-27	26	600382	广东明珠	2001-1-18
27	600618	氯碱化工	1992-11-13	27	600399	抚顺特钢	2000-12-29
28	600623	双钱股份	1992-12-4	28	600419	ST 天宏	2001-6-28
29	600636	三爱富	1993-3-16	29	600429	三元股份	2003-9-15
30	600666	西南药业	1993-07-12	30	600449	赛马实业	2003-08-29
31	600710	常林股份	1996-7-1	31	600466	迪康药业	2001-02-12

附录

32	600713	南京医药	1996-07-01	32	600481	双良节能	2003-4-22
33	600737	中粮屯河	1996-7-31	33	600483	福建南纺	2004-5-31
34	600782	新钢股份	1996-12-25	34	600488	天药股份	2001-06-18
35	600789	鲁抗医药	1997-02-26	35	600493	凤竹纺织	2004-4-21
36	600796	钱江生化	1997-04-08	36	600501	航天晨光	2001-6-15
37	600812	华北制药	1994-01-14	37	600531	豫光金铅	2002-07-30
38	600815	厦工股份	1994-01-28	38	600539	天士力	2002-08-23
39	600818	中路股份	1994-01-28	39	600565	迪马股份	2002-7-23
40	600877	中国嘉陵	1995-10-13	40	600582	天地科技	2002-5-15
41	600963	岳阳林纸	2004-5-25	41	600585	海螺水泥	2002-02-07
42	600978	宜华木业	2004-8-24	42	600586	金晶科技	2002-08-15
43	600991	广汽长丰	2004-6-14	43	600597	光明乳业	2002-8-28
44	600993	马应龙	2004-05-17	44	600629	棱光实业	1993-02-09

攻读硕士学位期间取得的学术成果

- [1]赵冠华,颜雯雯,李玥. LS-SVM financial achievement prediction based on targets optimization of neighborhood rough sets [J]. Communications in Computer and Information Science,2011,152:171-178
- [2]赵冠华,李玥,赵娟. 基于遗传算法参数优化的最小二乘向量机财务困境预测. 科学与管理,2011,5:56-63
- [3]李玥. 会计准则供求关系分析. 科技致富向导杂志,2012.3:80-81

致谢

致谢

岁月如白驹过隙，伴随着毕业论文的完成，转眼间三年的研究生生活即将结束。在这三年的学习中，母校给予了我很多人生的财富。基于此，我要真诚的感谢所有帮助和关心我的老师、同学以及朋友们。

首先，衷心的感谢我的导师赵冠华教授。自我的研究生生涯发轫，导师在学习和生活上就一直给予我帮助。他为人正直、品德高尚、治学严谨、学识渊博，而这亦将使我终生受益。同时，在我求学期间会计学院的诸位老师悉心指导，感谢各位老师在我研究生学习期间的帮助，授予我一生受用的知识，为该文的写作和完成奠定了基础。

其次，感谢山东财经大学的同学、好友，特别是我的三位舍友，正是她们在我最困难的时候伸出援助之手，帮我度过难关。她们思维活跃，理论基础扎实，每次同她们讨论，我都能获得新的收获。与她们在一起的日是我这三年最快乐也最珍惜的日子，我将永远铭记，感谢你们。

再者，真诚地感谢我的家人。自余踏上求学之征程，他们始终给予我坚定的支持及帮助，感谢她们对我的辛苦养育及栽培。

最后，衷心感谢所有参与论文审核及答辩的各位的老师。鉴于笔者学识有限，深感文章存在诸多不足有待改进，敬请各位师长给予批评和指正。