

河北省应用基础研究计划重点基础研究项目 项目编号: 12YJA880116
河北省人才培养工程经费资助
河北省高校学科拔尖人才选拔与培养计划资助
河北省高等学校人文社会科学重点研究基地项目

中国商业银行全面风险 评价研究

基于粗糙集的数据挖掘算法在金融风险防范中的应用



王宪明 胡继成 田媛 等著

ZHONGGUO SHANGYEYINHANG QIANMIAN FENGXIAN
PINGJIA YANJIU

JITU CUCAOJI DE SHUIJU WAJUE SUANFA ZAI
JINRONG FENGXIAN FANGXIAN ZHONGDE YINGYONG

河北省应用基础研究计划重点基础研究项目 项目编号: 12YJA880116

河北省人才培养工程经费资助

河北省高校学科拔尖人才选拔与培养计划资助

河北省高等学校人文社会科学重点研究基地项目

中国商业银行全面风险评价研究

——基于粗糙集的数据挖掘算法在金融风险防范中的应用

王宪明 胡继成 田 媛 等著

经济科学出版社

图书在版编目 (CIP) 数据

中国商业银行全面风险评价研究：基于粗糙集的数据挖掘算法在金融风险防范中的应用 / 王宪明等著. —北京：经济科学出版社，2015. 1

ISBN 978 - 7 - 5141 - 5434 - 4

I. ①中… II. ①王… III. ①商业银行 - 风险评价 - 研究 - 中国 IV. ①F832. 33

中国版本图书馆 CIP 数据核字 (2015) 第 016642 号

责任编辑：高进水 刘 颖

责任校对：曹 力

责任印制：潘泽新

中国商业银行全面风险评价研究

——基于粗糙集的数据挖掘算法在金融风险防范中的应用

王宪明 胡继成 田 媛 等著

经济科学出版社出版、发行 新华书店经销

社址：北京市海淀区阜成路甲 28 号 邮编：100142

总编部电话：010 - 88191217 发行部电话：010 - 88191522

网址：www. esp. com. cn

电子邮件：esp@ esp. com. cn

天猫网店：经济科学出版社旗舰店

网址：http://jjkxchs. tmall. com

北京汉德鼎印刷有限公司印刷

华玉装订厂装订

710 × 1000 16 开 19.5 印张 270000 字

2015 年 1 月第 1 版 2015 年 1 月第 1 次印刷

ISBN 978 - 7 - 5141 - 5434 - 4 定价：38.00 元

(图书出现印装问题，本社负责调换。电话：010 - 88191502)

(版权所有 侵权必究 举报电话：010 - 88191586)

电子邮箱：dbts@ esp. com. cn



序 言

金融是现代经济的核心，金融风险则贯穿金融活动的全过程，对现代金融机构而言，积极地对金融风险进行管理已成为其核心业务，识别、评价、防范并控制金融风险已经成为现代金融机构的主要职能。商业银行作为最重要的金融机构，主动或被动承担了多种类型的风险，因而风险管理对其自身发展也至关重要。商业银行自成立以来，风险一直与之相生相伴，我们进行风险管理的目的不是为了消除风险，而是通过风险管理获取长远的收益。有效的风险管理可以降低商业银行的损失，为银行提供产品定价依据，给商业银行带来更高的回报，并有助于商业银行在行业中获得竞争优势。商业银行能否正确识别、评价、有效管理和控制风险，直接决定了商业银行未来的经营成败。

1952年，美国学者哈里·马科维茨（Harry Markowitz）在《金融杂志》上发表了《资产组合选择——投资的有效分散化》一文，标志着现代金融理论的创立。马科维茨创造性地将数理统计的方法引入到金融投资组合当中，利用方差（或标准差）来对投资风险进行量化，这是金融史上的一次创举。此后，在金融风险管理领域的历次创新基本上都是将一些先进的数学理论和方法引入到该领域，促进了金融风险管理理论和方法的不断进步。因此，提升风险管理水平，对金融风险进行准确的评价和有效的控制，需要将先进的理论和方法引入风险管理领域。

粗糙集 (Rough Sets) 理论是 1982 年波兰学者帕拉克 (Z. Pawlak) 提出的, 它是一种刻画不完整和不确定性的数学工具, 能有效地分析不精确、不一致、不完整等各种不完备的信息, 还可以对数据进行分析和推理, 从中发现隐含的知识, 揭示潜在的规律。粗糙集理论最初的原型来源于比较简单的信息模型, 它的基本思想是通过关系数据库分类归纳形成概念和规则, 通过等价关系的分类以及分类对于目标的近似来实现知识发现。由于思想新颖、方法独特, 粗糙集理论已成为一种重要的智能信息处理技术, 该理论已经在机器学习与知识发现、数据挖掘、决策支持与分析等方面得到了广泛应用。针对粗糙集是一种能有效地分析不精确、不一致、不完整等各种不完备的信息的先进数学理论, 我们将其引入到金融风险管理, 可以有效地解决金融机构与监管者之间以及金融机构内部之间信息不对称问题, 能在数据微量的变化间捕捉到潜在的风险因素, 特别是商业银行在构建全面风险管理领域具有重大现实意义。

本书是河北省应用基础研究计划重点基础研究项目, 也是课题组全体成员集体智慧的结晶, 由于目前国内不同的金融主体风险的评价方式不同, 例如国际金融监管体系采用的基本是《巴塞尔协议》有关规定, 我国银监会也有一套自身监管指标, 不同商业银行也有不同的内部风险和外部风险评价体系, 为了尽可能完善我们自己建立的全面风险监管体系, 研究过程中, 我们对国际、国内不同监管主体的评价指标体系进行了系统梳理, 同时也参照了目前国际上比较先进的第三方评价体系 (例如标准普尔公司、穆迪投资者公司等资信评级公司) 和证监会对上市企业的监管要求, 通过多维度的综合比较, 以相关研究成果中各项指标的出现频率、全面性、重要性等为依据, 并充分考虑实证研究的数据可得性问题之后, 提出适合我国商业银行全面风险评价的三级指标体系。

本书写作安排首先是系统回顾了商业银行风险管理理论与实践, 参照不同评价主体的风险评价标准, 并基于 16 家上市银行内部变量指标的数据分析

和宏观经济数据的外部变量指标，建立了我国商业银行全面风险评价的指标体系，在此基础上运用粗糙集理论对我国商业银行风险评价进行了实证研究，同时还运用人工神经网络的方法对实证结果进行了全面验证，最终应用粗糙集理论构建了我国商业银行全面风险评价体系。本书写作具体分工如下：第一章由胡继成负责，第二章由谷晓飞和田媛负责，第三章由高海龙和马一宁负责，第四章由杜光辉和何志强负责，第五章由柳凌燕负责，第六章由王宪明和胡继成负责，第七章由王宪明负责，最终由王宪明教授负责全书总纂定稿。

本书在研究和写作过程中得到了河北金融学院领导汤生玲教授和陈尊厚教授的大力支持，中国人民银行石家庄中心支行、河北省银监局、河北省证监局、河北省金融办以及河北省科技厅、教育厅有关领导通过不同方式对研究进行关心与支持，在此一并表示感谢！同时也感谢经济科学出版社高进水副总编以及刘颖编辑对图书出版的支持；也感谢课题组全体同事以及家属对我们的支持与谅解。由于目前粗糙集理论主要应用于工程领域，将粗糙集应用于金融风险管理，特别是引入全面风险管理尚属首次，书肯定会有很多不足之处，因此恳请广大读者批评和指正。

王宪明

二〇一五年元月

目 录

Contents

1. 绪论 / 1

1.1 研究背景与意义 / 1

1.2 项目研究的主要问题 / 3

1.3 项目研究的主要内容与基本框架 / 5

1.4 项目的研究方法 with 可行性分析 / 7

2. 商业银行风险管理理论与实践 / 9

2.1 风险管理概述 / 9

2.1.1 风险管理目的 / 9

2.1.2 风险管理的发展 / 10

2.1.3 风险的分类 / 12

2.2 银行业风险监管的发展 / 14

2.2.1 国际银行业监管的发展 / 15

2.2.2 我国银行业风险监管的发展 / 18

2.3 商业银行的信用风险 / 21

2.3.1 信用风险的定义 / 21

2.3.2 信用风险的来源 / 21

2.3.3 信用风险的种类 / 22

2.3.4 信用风险的度量 / 22

2.3.5 信用风险测量的几种方法 / 24

2.4 商业银行的市场风险 / 29

2.4.1 市场风险的定义 / 29

2.4.2 市场风险的分类 / 29

2.4.3 市场风险的度量和管理 / 32

2.5 商业银行的操作风险 / 41

2.5.1 操作风险的定义 / 41

2.5.2 操作风险的分类与特点 / 42

2.5.3 操作风险监管的发展 / 43

2.5.4 操作风险的计量 / 44

2.5.5 操作风险研究综述 / 52

3. 我国商业银行全面风险评价指标体系研究 / 55

3.1 商业银行全面风险评价概述 / 55

3.1.1 商业银行全面风险评价的概念 / 55

3.1.2 商业银行全面风险评价的主体 / 56

3.2 基于不同评价主体的评价标准 / 56

3.2.1 国际评价标准 / 56

3.2.2 国内的相关评价标准 / 61

3.2.3 第三方评价标准 / 69

3.3 我国商业银行全面风险评价指标体系的确立 / 72

3.3.1 指标体系确立的原则 / 72

3.3.2 指标体系的主要内容 / 74

3.4 基于 16 家上市银行内部变量指标的数据分析 / 84

3.4.1 财务指标 / 84

3.4.2 监管指标 / 95

3.5 基于宏观经济数据的外部变量指标分析 / 109

4. 基于粗糙集理论的我国商业银行风险评价实证研究 / 120

4.1 利用粗糙集理论评价商业银行金融风险的意义及国内外研究动态 / 120

4.1.1 利用粗糙集理论评价商业银行金融风险的意义 / 120

4.1.2 国内外研究动态 / 121

4.2 粗糙集理论基础 / 122

4.2.1 粗糙集的基本概念 / 122

4.2.2 典型约简算法 / 130

4.2.3 进行属性离散化 / 134

4.3 粗糙集理论在商业银行金融风险评价中的应用研究与实证分析 / 137

4.3.1 基于粗糙集理论的金融风险评价模型的建立 / 137

4.3.2 样本数据的选取与确定 / 138

4.3.3 数据补全及离散化处理 / 139

4.3.4 属性约简 / 142

4.3.5 规则提取 / 143

4.3.6 模型应用测试结果 / 143

4.3.7 模型特点及总结 / 145

5. 基于神经网络的商业银行全面风险评价模型验证 / 147

5.1 神经网络理论及其相关应用介绍 / 147

5.1.1 人工神经网络的起源发展及其分类 / 148

5.1.2 BP神经网络在风险评价方面的适用性分析 / 149

5.1.3 BP神经网络相关理论基础 / 150

5.2 商业银行风险评价模型的设计与建立 / 154

5.2.1 商业银行风险评价模型相关指标的确定 / 154

5.2.2 商业银行风险评价模型的建立 / 155

5.3 基于粗糙集的商业银行风险评价模型的神经网络仿真与验证 / 156

5.3.1 数据的选取和预处理 / 157

5.3.2 MATLAB 仿真软件介绍 / 168

5.3.3 模型的仿真与验证 / 169

5.4 本章小结 / 174

6. 基于粗糙集理论的商业银行风险评价体系构建 / 176

6.1 建立商业银行风险预警机制 / 176

6.1.1 商业银行风险预警机制功能 / 176

6.1.2 商业银行风险预警流程 / 177

6.2 以粗糙集理论为核心的商业银行风险评价体系 / 178

6.2.1 商业银行风险评价体系的设计原则 / 178

6.2.2 基于粗糙集模型的商业银行风险评价体系设计 / 180

6.2.3 基于粗糙集模型的商业银行风险评价体系实施方案 / 186

7. 总结 / 191

附录 本研究中的相关数据 / 195

参考文献 / 273

1.

绪 论

1.1 研究背景与意义

2008 年以来的经济危机影响在逐步消退,但国际经济形势依然错综复杂,世界经济整体低迷态势仍在延续。一方面,美、日等主要发达经济体的持续宽松货币政策对全球经济影响存在不确定性,财政政策的不确定性也将继续影响全球经济。2012 年下半年以来,主要发达经济体持续量化宽松货币政策对全球经济影响的溢出效应已经开始显现,全球跨境资本流动大幅波动,国际大宗商品价格剧烈起伏,全球性资产泡沫加剧。与此同时,发达经济体的高额财政赤字,已经影响到其内部的财政平衡,对福利和税收的影响显现,并直接影响到发达经济体的经济复苏,并埋下了全球金融风险的种子。

另一方面,全球经济复苏艰难曲折,贸易保护主义倾向上升。部分发达经济体采取贸易保护措施来转嫁内部经济与政治压力,发达经济体与新兴发展中国家的贸易摩擦增加,并有可能进一步危及经济复苏进程。此外,部分地区的政治形势趋于恶化,地区局势紧张,地缘政治风险所引发的军事冲突,对正常的双边经贸关系和区域经济合作造成严重干扰。

虽然面临外部环境的高度不确定性,但总体而言我国宏观经济运行呈现缓中趋稳的态势。我国以促进经济结构调整和发展方式转变为目标,以积极的财政政策和稳健的货币政策为指引,取得了较为良好的成效。近年来,我国财政收入和货币信贷持续增长,物价水平和就业情况总体较为稳定,内需

对我国经济增长的拉动作用正在逐步提升。

2013 年是我国全面深入贯彻落实党的十八大精神的开局之年，是为全面建成小康社会奠定坚实基础的重要一年，我国处于重要战略机遇期的基本格局没有改变，经济社会平稳较快发展仍具备有利条件，但面临的国内外环境十分复杂，平稳运行与隐忧风险并存。^①我国正在逐步推进利率市场化和人民币汇率形成机制改革，优化信贷结构，并全面提高金融服务实体经济的质量。现今我国银行业整体运行稳健，在保持规模、盈利持续增长的同时，组织体系更加完善，资本充足水平稳步提升，拨备较为充足，但不可忽视的是各项风险也在增强。某些行业、领域和地区的信用风险已显现出来，具有融资功能的非金融机构和民间借贷风险也在凸显，银行流动性风险管理的重要性在增强，其理财产品、表外业务同样存在着较高的风险，需要监管机构和银行机构的共同关注。

随着中国加入 WTO 后，金融领域逐步开放融合，加之信息技术的广泛运用，使得近十多年来的金融全球化迅速发展。作为金融全球化进程中重要的微观主体——我国商业银行快速发展，银行业的跨区域运营，并与各种类型的非银行金融机构开展业务单元合作，风险的传导机制日趋复杂多元，高度集成的系统性风险日益成为银行业关注的焦点，对银行业综合性、系统性风险的监控与测度已成为当前经济金融领域的一项重要前沿课题；从 2008 年以来美国“次贷危机”所衍生的系统性经济危机和金融风险对我国商业银行的影响就能看出端倪。

金融体系的安全、高效、稳健运行对经济全局的稳定和发展至关重要。随着金融领域市场化程度和市场开放度空前提高，我国的金融机构面对复杂的市场环境和激烈的内外部竞争，如何实现自身高效、安全运行，成为商业银行的基本要求。此外，国内金融市场环境的变化和金融需求导向的细化，商业银行业在陆续研发多种金融服务产品，这些产品的运行状况、安全性对

^① 中国人民银行金融稳定分析小组：《中国金融稳定报告（2013）》，中国金融出版社 2013 年版。

于整个金融体系的安全运行也会带来不同程度的影响。由于商业银行和其他金融机构、非金融机构间建立了广泛的联系,金融产品数量多、信息庞杂,数据庞大,如何基于创新性的理论应用,在海量数据中抽丝剥茧,建立一个有效的全面风险分析评价模型,通过对商业银行相关业务数据监控和模型运用中识别并及时给出潜在的风险水平,是目前金融风险防范领域关注的热点问题,也是本项目研究的关键所在。

1.2 项目研究的主要问题

西方发达国家的金融业伴随其工业化进程,经历了多次衰败、复苏、繁荣的周期性波动,例如20世纪80年代拉美债务危机,2007年以来的次贷危机和主权债务危机都对全球金融业造成了巨大冲击。因而金融专业学者和研究人员普遍将风险管理及风险预警作为金融企业的核心问题之一。在风险识别方面,美国的罗杰·博斯玛(A. Roger Bosma)、乔什·梅(Josh W. May)以及爱德华·里德(Edward W. Reed)等人做了大量的研究工作,得出了许多有价值的结论,其中,美国第一富达银行(First Fidelity Bank)的高级管理人员罗杰·博斯玛(A. Rogor Bosma)通过对资产负债表、收益表以及现金流量表等进行考察,提出根据借款人的经营状况预测其违约的方法。^①莱蒂宁(Laitinen)和庄(Chong)研究了中小企业预警系统,他们对芬兰的公司和美国银行的决策者就中小企业失败进行了因素调查,分析了导致这些企业失败的最关键因素。^②

历史地看,金融企业尤其是商业银行的风险评估方法的发展可分为三个阶段:第一个阶段是在1970年以前,以专家系统法为主要评价信用风险,如5C分析法;第二个阶段的方法主要是建立以财务指标为自变量的模型,以计算出可以代表被评价人信用情况的结果,如Logit、Probit回归模型、美国学

① 季海:《基于人工智能的商业银行信贷风险预警研究》,1999年。

② 余丛国、席西民:《我国企业预警研究理论综述》,载于《预测》2003年第2期,第23~29页。

者阿尔特曼 (Altman) 提出的 Z 评分模型和 ZETA 评分模型、神经网络模型等;第三个阶段是在 20 世纪 90 年代以后,现代金融理论和数学、计算机工具被引入信用风险评价领域,并由此建立了以风险价值 (VAR) 为基础,以违约概率 (PD)、预期损失 (EL) 为核心指标的度量模型,如 JP 摩根开发的信用风险度量模型 (CreditMetrics)、麦肯锡公司提出的信用组合观点模型 (Credit Portfolio View) 等。

目前,随着金融企业信息化程度的进一步加深,信息技术在金融企业风险防范方面的作用日益显现,金融企业对此也进行了大量投入,利用现代最新的科学技术如数据库技术、人工智能技术构建风险评估体系。近几年来,随着我国金融体制改革的不断深化,我国的金融企业在借鉴国外先进经验的基础上不断摸索、更新风险防范体系,或者引进采用国外较为先进的评估算法,但总体来看我国国内金融企业目前的风险评估体系总体水平仍较为落后,相比于国外的先进金融企业,数理统计模型、金融工程和人工智能方面均有较为明显的差异。

数据挖掘 (Data Mining) 是近几年兴起的新型应用技术,是数据库知识发现 (Knowledge Discovery in Database, KDD) 的核心,是指从数据库中提取潜在的、有用的、最终可理解的知识的过 程。关联规则算法则是数据挖掘的一个重要研究方向,其侧重于确定数据库中不同领域间的联系,找出满足给定支持度和可信度的多个域之间的相互关系。由于数学基础成熟、无须先验知识,且应用较为简便,目前作为关联规则算法之一的粗糙集理论与方法对于处理不精确 (imprecise)、不一致 (inconsistent)、不完整 (incomplete) 等各种不完备的信息较为有效。粗糙集理论的主要思想是利用已知的知识库,将不精确或不确定的知识用已知的知识库中的知识来 (近似) 刻画。^① 它与概率方法、模糊集方法和证据理论方法等其他处理不确定性问题理论的最显著区别是它无须提供问题所需处理的数据集合之外的任何先验信息,而且与

^① 张文修、吴伟志:《粗糙集理论介绍和研究综述》,《模糊系统与数学》,2000 年。

其他处理不确定性问题的理论有很强的互补性。^① 所以粗糙集理论是十分适用于金融行业的数据挖掘算法。

国外迪米特里斯 (Dimitras) 等 (1999) 首次利用粗糙集理论对企业破产进行了预测, 他们以 40 家破产公司和 19 家健康公司作为训练样本, 利用粗糙集数据约简的功能对财务指标进行了约简, 并对约简后的财务数据决策表进行规则提取, 最终得到了发生破产前五年的预测规则, 实证检验证明其预测效果非常好。^② 此外, 粗糙集在信息安全、医学诊疗、股票分析、决策分析等多个领域得到了应用, 并取得了很好的效果。从 20 世纪 90 年代末期开始, 粗糙集理论被我国学者逐渐接触并认识到了其自身的巨大应用价值, 很多学者开展粗糙集领域的研究, 并取得了显著成果, 如采用粗糙集神经网络实现财务预警, 基于 BP 网络建立的公司财务预警模型等。

本书研究项目将以模糊处理的经典理论——粗糙集理论为基础, 以我国已上市商业银行为研究样本, 详细分析 16 家已上市商业银行的多维度业务运营数据, 针对商业银行所反映出的典型风险特征进行量化分析和研究, 力图构建完善的综合风险评价指标体系, 进一步利用现代编程工具和科学的模糊算法设计风险评价模型和预警系统, 通过已用的数据对模型进行测试检验, 最终得出具有较强实用性和可操作性的风险评价系统, 进而提升商业银行的内部风险评价水平和外部监管效果。

1.3 项目研究的主要内容与基本框架

项目研究的主要内容为: 分析目前商业银行面临的潜在风险, 并对风险进行概念化、抽象化及评价指标的量化; 选择合理的风险评价指标, 将收集

^① 张文修、吴伟志、梁吉业、李德玉:《粗糙集理论与方法》, 科学出版社 2001 年版。

^② Dimitras A, Slowinski R, Susmaga R, Zopounidis C. Business failure prediction using rough sets. *European Journal of Operational Research*, 1999 (114): 263-280.

到的数据按风险指标进行预处理；利用粗糙集的预处理算法对处理数据进行离散化，并对离散化的数据进行属性约简，导出问题的决策或分类规则；利用分类可信度和覆盖度定义生成精确度较高的决策规则，并使用测试数据对决策规则进行检验。

具体研究框架如下：

1. 明确金融企业面临的风险，并对风险进行分类和抽象，设计不同类型风险的定量分析方法和评价指标。

2. 选取风险评价指标。所选择的风险评价指标是否得当会严重影响到评估模型的工作效能，因此，对评价指标的选取应在风险管理理论知识的指导下进行，指标要能够充分反映银行的运行状况。完成指标选择后，就可以将收集到的数据按照指标进行预处理，保证数据的完整性。

3. 对预处理好的样本数据进行连续属性离散化。连续属性离散化的根本出发点，是在尽量减少决策表信息损失的前提下，得到简化和浓缩的决策表，以使用粗糙集理论分析，获得决策所需要的知识。^① 在离散化时可利用粗糙集理论计算连续数据的断点，通过计算的断点将各样本的连续数据序数化，得到简化和浓缩的决策表。

4. 属性的简约。由于选取的指标中，并非所有的指标都至关重要，还存在冗余属性，简约就是从中删除不必要或不相关的属性，进而简化操作。

5. 生成决策系统。在删除决策表中不相关或不重要的条件属性，即进行属性约简后，会获得一个简化的决策表，该决策表中每一个对象就代表一条分类规则，大量的决策表就构成了决策规则集合。对生成的规则集合进行约简后，利用分类可信度和覆盖度定义生成最终的决策规则。之后利用测试数据对规则进行测试并对规则进行改进。

项目研究的基本框架如图 1-1 所示。

^① 黄丽萍：《基于粗糙集的属性约简与规则提取》，厦门大学，2007 年。

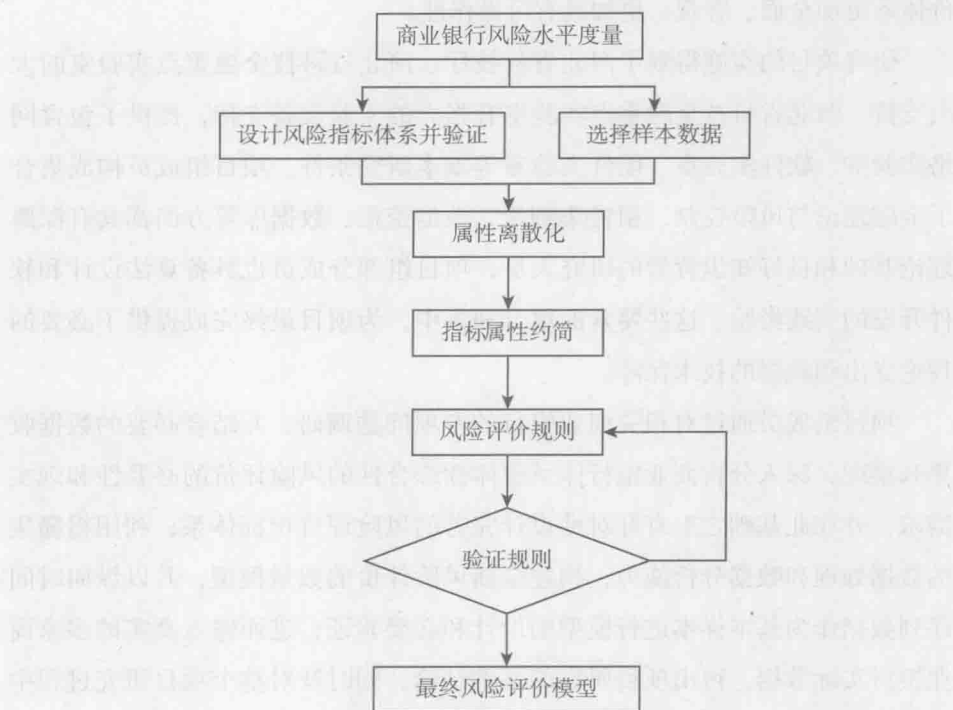


图 1-1 项目研究的主要技术路线

1.4 项目的研究方法 with 可行性分析

项目将基于粗糙集理论的数据挖掘算法应用于金融风险建模和金融风险防范，基于粗糙集理论建立金融业务风险模型，并使用金融企业的业务数据对模型进行验证，通过计算指标和决策的关联度来确定风险评估模型。

在研究方法的选择和应用上，将综合数学、经济学、管理学、统计学、信息技术等学科的相关理论，通过规范分析与实证分析相结合的方法，研究样本将以模拟数据和实际数据相结合，理论推导和实际应用相结合的方法开展研究工作。通过对金融业务进行深入分析，在此基础之上设计完善的风险指标体系；利用粗糙集的数据处理能力，构建金融业务风险模型，使风险评