

基于数据挖掘技术的虚假财务报告
识别研究

邓庆山 著

江西高校出版社

图书在版编目(CIP)数据

基于数据挖掘技术的虚假财务报告识别研究/邓庆山
著. —南昌:江西高校出版社, 2011.5

ISBN 978 - 7 - 5493 - 0263 - 5

I. ①基... II. ①邓... III. ①数据采集 - 应用 -
会计报表 - 鉴别 - 研究 IV. ①F231.5

中国版本图书馆 CIP 数据核字(2011) 第 084605 号

出版发行	江西高校出版社
社 址	江西省南昌市洪都北大道 96 号
邮 政 编 码	330046
总编室电话	(0791) 8504319
销 售 电 话	(0791) 8508191
网 址	www.juacp.com
印 刷	南昌市光华印刷有限责任公司
照 排	江西太元科技有限公司照排部
经 销	各地新华书店
开 本	850mm × 1168mm 1/32
印 张	5.125
字 数	120 千字
版 次	2011 年 5 月第 1 版第 1 次印刷
印 数	1 ~ 1000 册
书 号	ISBN 978 - 7 - 5493 - 0263 - 5
定 价	28.00 元

赣版权登字 - 07 - 2011 - 96

版权所有 侵权必究

前 言

公司财务造假是一个存在已久、影响极其恶劣的国际性难题。1720年的英国南海公司事件是世界上第一例上市公司财务造假案。南海公司事件直接导致了两个结果：一是股份公司近百年的关闭；二是以注册会计师为主体的民间审计的出现。尽管如此，公司财务造假现象一直没有消失，远的如1929年美国罗宾斯财务造假导致的美国股市的大崩溃，近的如2001年美国的安然公司事件都是典型的财务造假案例。而在我国，继琼民源、PT郑百文等财务造假丑闻之后，又暴露出了银广夏、ST黎明、蓝田股份等财务造假恶性案件。上市公司财务造假问题引起了社会各方面的关注。针对公司的财务造假问题，专家学者进行了大量的定性或者定量的研究，取得了一定的研究成果。本书从舞弊理论、舞弊特征、舞弊识别模型等三个角度充分分析了国内外研究现状，在此基础上，进行了以下几个方面的研究。

一、分析财务造假的原因和手段。

本书从两个角度分析了财务造假的原因，一是客观原因，二是内在动因。客观原因包括：政府职能的问题、公司治理结构的缺陷、证券市场相关制度的不完善、不健全的监管机制等；内在动因有：公司为了取得股票发行资格、提高股票发行价格、取得配股资格、避免被“戴帽”或者争取“摘帽”、为维持和提高股票价格、为了获取信贷资金和商业信用、为了业绩考核、为了满足发行债券的条件、为了减少纳税而粉饰会计报表等。财务造假的手段主要是利用关联方交易、资产重组、选用不当会计政策、会计政策与会计估计变更、资产评估等手段进行利润调节。

二、分析确定舞弊识别变量和实验样本。

我们将 1999—2002 年持续经营、未被取消上市资格的上市公司作为训练样本的选择基础,将 2003—2006 年上市公司作为选择测试样本的基础。我们将审计意见为“拒绝表示意见”或“否定意见”或“无法表示意见”作为选择虚假财务报告样本的基础;另外再根据中国证券市场的历史特点,按照一定的标准选择控制样本。参考前人的研究成果,从偿债能力及管理效率比率、利润—现金创造能力比率、费用配比合理性比率、公司增长潜力比率、公司发展协调性比率等几个角度分析确定了 47 个备选舞弊识别变量。

三、设计基于分类方法的虚假财务报告识别框架,并使用多种分类方法进行虚假财务报告识别研究,引入期望错误分类成本概念,对分类实验结果进行比较。

首先根据大部分分类方法以及财务报表数据的特点,设计了基于分类方法的虚假财务报告的识别框架。并依据此框架,分别使用 Logistic 回归、神经网络、支持向量机、朴素贝叶斯等分类方法进行了实证分析。在进行实证分析时,使用按一定标准选择的训练样本和经过论证分析后选择确定的识别变量对这四种分类模型分析进行了训练,然后使用测试样本检验不同的分类模型分类效果,并基于期望错误分类成本,对这四种分类模型的识别效果进行了综合比较分析,发现神经网络和朴素贝叶斯两种分类方法在识别准确率和稳定性方面要好于另外两种方法。

四、提出一种聚类方法进行虚假财务报告的聚类分析。

在目前的审计环境下,单纯以公司财务报告的注册会计师审计结果来判断该财务报告的真假是有风险的(尤其是判断为真的时候)。所以有监督分类方法的学习结果可能使得分类存在一定的偏差。由于聚类是一种无监督学习的数据挖掘技术,无监督的学习与有监督的学习不同,它不要求学习样本带有类别的信息,或者说不需要有先验知识,所以聚类方法对纷繁复杂的、真假难辨的

会计信息进行真假识别或许更有帮助。基于此,本书设计了一种基于聚类有效性测度的自组织映射和 K 均值聚类相结合的聚类模型 V - KSOM。此模型继承了自组织映射自主学习的优点;将 K 均值聚类运用于训练完成后的自组织映射平面网格结点,可以有效地划分聚类边界,克服了自组织映射聚类边界不清晰的缺点;由于自组织映射受结点初始值(随机产生)及样本学习顺序(随机的)的影响,聚类结果不稳定,为了选择最佳的聚类结果,本模型引入一种有效性测度来评判每次聚类结果。使用该聚类模型对虚假财务报告进行聚类分析,实验结果表明该模型是行之有效的。

本书的主要内容来自本人的博士毕业论文。在此我要特别感谢我的博士生导师梅国平教授,在梅老师的精心指导下,我才能在学术研究的道路上取得一点成绩,本书也才能得以定稿。此外,对本书提供过帮助的人还有关爱浩博士、毛小兵博士、陈辉博士等,在此一并致谢。

邓庆山

2011 年 2 月

目 录

第一章 绪论.....	1
第一节 本书研究背景.....	1
第二节 虚假财务报告识别研究的意义.....	2
第三节 相关概念界定.....	4
一、什么是会计信息.....	4
二、什么是会计信息失真.....	4
第四节 技术路线与内容安排.....	5
一、技术路线.....	5
二、论文内容安排.....	6
第二章 文献综述.....	8
第一节 国外财务舞弊研究综述.....	8
一、舞弊理论研究.....	8
二、舞弊特征研究.....	10
三、舞弊识别模型研究.....	14
第二节 国内财务舞弊研究综述.....	24
一、舞弊理论研究.....	24
二、舞弊特征研究.....	25
三、舞弊识别模型研究.....	29
第三节 数据挖掘技术及其应用现状.....	32
一、数据挖掘的含义与功能.....	32
二、数据挖掘在经济管理领域的典型应用.....	33
第四节 评述.....	34
第三章 财务造假分析.....	36

第一节 造假原因分析	36
一、客观环境分析	36
二、内在动因分析	39
第二节 造假手段分析	43
一、关联方交易	43
二、资产重组	45
三、选用不当会计政策	46
四、会计政策、会计估计变更	46
五、资产评估	47
六、其他手段	47
第四章 研究样本与变量选择	49
第一节 样本选择	49
一、虚假财务报告样本	49
二、控制样本	50
三、测试样本	54
第二节 变量选择	56
一、选择标准	56
二、变量分析	56
第三节 指标计算及其标准化	67
一、中心化	67
二、标准化	68
第五章 基于分类方法的虚假财务报告识别研究	69
第一节 基于分类方法的虚假财务报告识别框架	69
第二节 Logistic 回归分析及其应用	71
一、Logistic 回归简介	71
二、实验过程与分析	73
第三节 神经网络及其应用	78
一、人工神经网络的特点	78

二、神经元及其特性	79
三、BP 神经网络模型	80
四、基于神经网络的虚假财务报表的识别	83
第四节 支持向量机及其应用	90
一、支持向量机理论	90
二、基于支持向量机的会计信息失真的识别	95
第五节 贝叶斯分类及其应用	100
一、朴素贝叶斯分类简介	100
二、朴素贝叶斯分类的应用	102
第六节 几种分类方法实验结果的比较	107
第六章 虚假财务报告的聚类研究	111
第一节 自组织映射	112
第二节 K 均值聚类方法	114
第三节 聚类有效性	114
第四节 V - KSOM 模型	117
第五节 V - KSOM 模型的应用	119
一、实验样本选择	119
二、变量选择	119
三、实验过程与结果分析	120
第七章 总结与展望	123
第一节 研究总结	123
第二节 研究的局限和今后研究方向	126
参考文献	127
附录 1: 分类训练样本中 88 家公司列表	137
附录 2: 分类测试样本中 172 家公司列表	142
附录 3: 聚类实验样本中 100 家公司列表	150

第一章 绪 论

第一节 本书研究背景

近十年来,我国频繁出现上市公司会计造假、财务舞弊等案件,比如,“银广夏”虚构主营业务收入、虚构巨额利润“琼民源”制造虚假会计信息以抬高股价,致使广大股东承受巨大的经济损失;红光实业虚构产品销售、虚增产品库存、实际亏损报表却盈利,以此获取股票发行资格等等。近年来,国家加大了对财务舞弊等违法行为的处罚力度,但是仍然有很多企业为了满足局部群体的利益,不惜铤而走险,以身试法,采用各种手段出具与企业真实经营情况不符的财务报表。上海国家会计学院财务舞弊研究中心对2005年的财务舞弊情况做了调查,调查报告显示:有84家公司在该年度自行披露会计重大差错更正;38家公司因涉及财务有关问题而被证监会限令整改;有50家公司被证监会立案调查;有58家公司被证监会公开处罚或者被上海证券交易所和深圳证券交易所公开谴责。在国外,这种财务造假事件不断出现。据《商业周刊》报道,自上世纪末期至今,美国上市公司财务丑闻被曝光的范围和规模超过了大萧条以来的任何时期。2001年12月Enron(安然公司)因会计丑闻曝光刷新了公司破产规模的纪录;6个月之后WorldCom(世界通信)再度爆发丑闻,再创公司破产规模的纪录;07年日本三洋电机财务造假得以曝光;曾经显赫的朗讯、泰科、施乐、默克涉嫌虚假财务报告而成为会计丑闻家族的一员。虚假财务报告的泛滥不仅给政府机构监管公司运营带来困难,更给整个

国家经济、投资者等利益相关方带来实质性的损害。据某些研究人员估计,虚假财务报告每年给美国经济带来的损失高达 4000 亿美元。

目前我国资本市场动荡不定,上证指数从 2005 年的不到 1 千点到 2007 年 6 千点之上,到 2008 年最低的 1600 点,再到现如今的 3 千点左右。股市的起伏一方面是受国际资本市场大环境的影响,一方面反映了广大股民的不理性投资行为,财富效应总能吸引着越来越多的投资者参与到证券市场中。上市公司为了追求相关群体的利益最大化,依然有着强烈的财务造假的动机。上市公司财务造假必然会损害广大投资者的利益。财务报表舞弊问题是我国证券市场当前和未来一段时期内面临的最大的挑战之一。

第二节 虚假财务报告识别研究的意义

公司财务舞弊的影响具体包括下面几个方面:第一,弱化市场的资源配置功能。市场资源配置功能要得以充分发挥,最重要的前提就是要求信息是真实有效并且公开的,然而公司财务舞弊,出具虚假会计信息必然会误导市场,从而导致资源配置的不合理。第二,误导投资者。投资者不会盲目投资,在进行投资之前,必然会要求了解上市公司的经营情况,他们通常是通过分析公司的财务报表来达到这个目的,但是虚假会计信息使得投资者无法获知企业的真实经营情况,这必然影响投资者的决策和判断,从而使投资者蒙受经济损失。第三,损害相关机构利益。公司为了获取充足的发展资金,会向银行或者其他金融机构贷款,公司也可能会以赊账的方式向其供应商购买相关产品或者服务。公司财务造假势必会影响到上述不同机构的切身利益。

强化会计监督,加大企业会计信息真实性的判断与识别是非常有必要、非常有意义的。

对投资者来说,会计信息是投资者进行决策的重要依据,通过会计信息而反映出的企业经营状况与业绩直接影响投资者的最终决策。虚假会计信息将误导投资者,使得投资者蒙受损失。学会识别上市公司的虚假财务报告,将减少投资者的损失,做出正确的决策。如果能提供一种相对简单的虚假财务报告的识别方法,将大大简化投资者的工作量,从海量的财务数据信息中解脱出来,做出正确的判断和决策。

对监管者来说,掌握一套能有效识别上市公司财务欺诈的方法或者建立一个识别财务欺诈的模型,将极大地有利于他们工作的开展。首先,对于监管者来说,他们的一个重要职责就是监管上市公司,及时发现和查处出具虚假财务报告的上市公司,保护公平有序的市场竞争环境,切实维护广大投资者的切身利益。其次,相关监管人员还要对公司的上市或者再融资进行审核,使得提供真实有效的财务报告、并且达到融资要求的公司能够获批,反之,拒绝那些为达到融资要求而采用各种手段对财务报告进行粉饰的公司。此外,监管者如果能够充分认识和了解公司的各种财务造假的手段和特征,将有利于监管者制定出切实有效的舞弊防范措施和相应的规章制度。

对于注册会计师来说,他们的最主要职责是进行上市公司财务报表的审计。本研究对于辅助识别虚假会计信息,改善注册会计师工作效率、提高判断的准确度、减少其误判的可能性等提供了一定的工具与手段。

从方法论的角度来看,发展虚假财务报告的模式识别模型更具有理论方法创新的重要意义。企业会计人员使用传统的会计方法所编造的财务数据,其所遵循的原则不外乎会计惯例和会计准则,即使是注册会计师也概莫能外。与自然形成的财务数据相比,他们所编造的虚假数据很可能在数据结构上存在内在的缺陷。也就是说,两类财务报告可能存在模式上的内在差异,而这种差异在

通常的会计人员眼中可能不显著,但通过统计或者数据挖掘等技术可以将这种差异放大。

近年来,数据挖掘技术日益成熟,其在经济管理等领域的应用变得越来越广泛,数据挖掘技术正逐渐成为上市公司财务报表分析、判断财务报表真实性程度的另一有效工具,它能利用海量财务报表数据,挖掘其中潜在的有用信息,为财务报表的分析提供有力的帮助。

第三节 相关概念界定

一、什么是会计信息

关于会计信息的概念有很多种。

葛家澍在其《会计学导论》中将会计信息定义为“会计系统所形成并传递的信息,主要应该是经济信息,是关于微观经济主体的经济信息。在市场经济条件下,经济活动离不开市场,离不开商品交换和货币,我们把货币所体现的关于微观经济主体的经济活动过程中价值运动和增值过程的信息称为财务信息。会计所提供的主要是关于微观经济主体经济活动过程中的财务信息”。

王开田在其《会计行为论》中认为“会计信息是信息的一个种类,它包括财务信息和成本信息,它是对企业经济活动引发的企业资金运动的状态、方式的客观描述,是资金运动在会计行为主体头脑中的映象,或是对企业资金运动的表征。它的作用关键在于使其使用者减少对企业经济活动的不确定性、怀疑性和模糊性,增加对企业经济活动的了解、观察、判断和决策的清晰度、确信度”。

二、什么是会计信息失真

一般意义上的会计信息失真,是指信息提供者通过非法手段或利用会计方法本身的缺陷而有意识地提供误导性的财务报告,使投资者和债权人等信息利用者受到损害的行为。信息提供者利

用其信息优势,采取欺诈手段不合法地调整企业的盈利,其有些手段貌似合乎会计准则和会计制度的要求,但是突破了一定的限度,已成为不合法的操纵行为。

会计信息失真可以分为三类:合法性失真(或者说政策性失真)、技术性差错和故意舞弊。

1. 合法性失真,是由于会计基本假设、会计政策的合法选择等造成的,是我国法律所允许的,它是客观存在的会计信息失真。它是由政策制度方法本身的局限性所造成的。这一类会计信息失真不是我们研究的重点。

2. 技术性差错,是指会计人员的非主观故意的过失或者由于素质差造成的。这类会计信息失真非常明显,容易发现和纠正。另外,对于出差错的当事人来说,非没有任何利益,相反可能还要受到单位的惩罚。这类会计信息失真的危害性较小,只要加强会计内部复核、检查即可。

3. 故意舞弊。这类会计信息失真是在管理当局指使、操纵下,采用合规或不合规的会计程序和方法,在会计信息的确认、计量、记录和报告的过程中,故意歪曲篡改、捏造会计事项,使会计信息背离客观事实,以欺诈企业外部会计信息使用者,实现管理当局的利益。

本书所研究的会计信息失真主要是指第三类会计信息失真。文中提及的“财务欺诈”、“虚假财务报告”、“财务(会计)舞弊”、“财务(会计)造假”、“财务报告粉饰”等概念与“会计信息失真”基本同义。

第四节 技术路线与内容安排

一、技术路线

本书主要基于数据挖掘技术,开展虚假财务报告识别研究。

数据挖掘在识别虚假财务报表方面具有很独特的优点。相对于会计相关专业人士的经验判断,使用数据挖掘技术进行虚假财务报告识别,能减少判断的主观性、提高判断准确率、改善工作效率、缩短反应时间。本书遵循实证研究为主、规范研究为辅的原则,在财务造假原因、造假手段、造假防范等方面主要使用规范研究的方法,在虚假财务报告识别方面主要采取实证研究的手段。图 1-1 是本书研究的技术路线图。

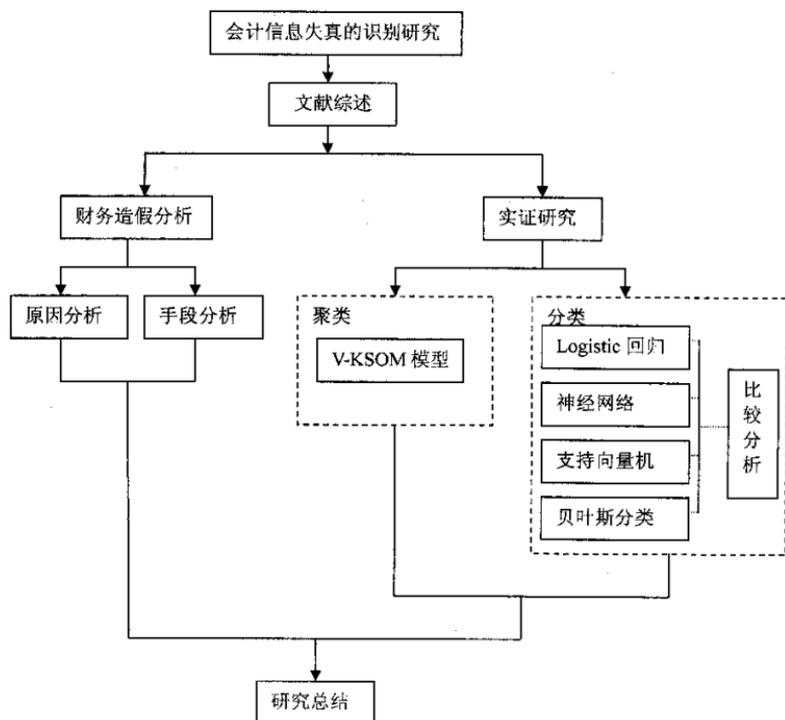


图 1-1 研究的技术路线图

二、论文内容安排

本书共分七章,主要研究内容包括:

第一章 绪论。主要内容包括论文的研究背景、研究意义、相关的概念界定以及研究的技术路线与内容安排。

第二章 文献综述。主要内容是从舞弊理论、舞弊特征研究、舞弊识别模型与方法等角度,综述了国外、国内的研究现状,并对国内外研究现状进行了评述。

第三章 财务造假分析。主要分析了财务造假的客观因素和内在动因,并分析了财务造假的各种手段。了解了财务造假的动因和具体手段,才有利于我们找到财务造假的特征,进而进行识别。

第四章 研究样本与变量选择。由于本书主要是以实证研究为主,实证研究就要有实验样本和变量,所以本章主要介绍了用于本书实证中的虚假财务报告样本和真实财务报告样本的选择标准,识别变量的选择和分析过程。

第五章 基于分类方法的虚假财务报告识别研究。本章将首先给出基于分类方法的虚假财务报告识别框架,然后简要介绍 Logistic 回归、神经网络、支持向量机、朴素贝叶斯等分类方法,建立相应的识别模型,使用第四章所介绍的样本和识别变量进行实证分析,并对实验结果进行分析和比较。

第六章 虚假财务报告的聚类研究。聚类分析是一种无监督的数据挖掘技术。它不需要先验知识,能进一步降低主观判断对识别过程的影响。本章提出了一种基于有效性测度的,将自组织映射与 K 均值聚类方法相结合的一种聚类方法,并将其运用于虚假财务报告识别,取得了较好的实验结果。

第七章 总结与展望。对本书的研究进行了总结,给出本书的研究结论,指出本书的创新点和有待进一步改进的工作。

第二章 文献综述

在财务舞弊识别方面,国内外都开展了一定的研究。本章从舞弊理论、舞弊特征、舞弊识别模型等三个角度,对国外、国内研究的研究现状进行综述,并对数据挖掘技术的当前的应用现状做了一个介绍。

第一节 国外财务舞弊研究综述

一、舞弊理论研究

国外在舞弊理论研究方面代表性的理论主要有二因素论、三因素论、四因素论、企业舞弊风险因子理论等。^[3-4]

1. 二因素论

二因素论,也称为冰山理论。二因素论把导致舞弊行为的各种因素划分为两类,一是舞弊结构部分、二是舞弊的行为部分。二因素论将舞弊比喻为海面上的一座冰山,大家都能看到的外在的部分只是露出海平面的一角,属于舞弊的结构部分,其中包含的内容主要是组织内部管理方面出现的问题。二因素论认为更危险的因素是潜藏在海面下的部分,这些部分是更主观、更个性化的内容,更容易被掩饰起来。这部分内容主要有行为人的感情、态度、价值观念、满意度等等,属于舞弊的行为部分。我们不仅要重视舞弊的结构部分内容,更应注重个体行为方面的考虑。二因素论强调,在舞弊风险因素中,个性化的行为方面更为危险,相比于组织的内部管理方面需要花更多的注意力。^[5]

2. 三因素论