DOI:10.19641/j.cnki.42-1290/f.2018.06.020

基于国家治理的上市公司舞弊审计实证检验

张 莉(副教授)

【摘要】为了揭示违规上市公司共同的倾向性因素,基于国家治理视角定义了舞弊内涵,选取A股上市公司1993~2015年度4461例样本数据,选取公司治理、财务指标、持续经营能力三类指标下33个变量对舞弊进行解释,采用四种抽样方法构建决策树模型,并运用Bagging机器学习算法对模型进行优化。结果表明:抽样方法会影响模型预测精确度,采用Bagging自助抽样算法可改善模型预测效果;内控有效性、内控意见、董事会会议次数、董事会人数、监事会人数、审计收费、息税前利润、股利分配率、每股企业自由现金流量、总资产净利润率等指标揭示舞弊效果显著;经样本外数据测试,Bagging算法优化后模型总体准确度比优化前提高了35.87%,召回率比优化前提高了21.1%,表明模型有较好的预测力与稳定性。

【关键词】国家治理;舞弊审计;决策树;Bagging算法;机器学习

【中图分类号】F234.9 【文献标识码】A 【文章编号】1004-0994(2018)06-0138-9

一、引言

上市公司违规的信息披露扰乱了证券市场的良序,加大了市场风险。自1993年以来,共有1532家上市公司存在3702次信息违规披露问题并受到处罚,极大地损害了投资者、债权人的利益以及他们对证券市场的信心。上市公司信息披露违规对注册会计师审计影响巨大,审计可在一定程度上揭露舞弊,但由于审计的固有限制,即使注册会计师按照审计准则的规定恰当地计划和执行了审计工作,也不可避免地存在财务报表中的某些重大错报未被发现的风险。那么审计应如何有效地识别舞弊影响因素,揭示舞弊风险并加强监管呢?

本文基于国家治理视角识别舞弊,利用机器学习建模揭示倾向性、苗头性问题,以更好地识别上市公司舞弊信号,为监管部门制定有针对性的舞弊治理策略提供理论支持,使审计更好地发挥"免疫系统"职能。本研究的独特之处在于:①从国家治理的广义审计视角定义舞弊内涵,除了财务指标,还考虑了公司治理、持续经营类指标;②跳出经典统计的舞弊寻因与抽样方法,基于大样本、多变量的大数据视角,减少了分析模型对样本及前提假设的约束;③研究了抽样方法对模型评价结果的影响;④采用自助汇聚抽样的机器学习算法优化决策树模型,提高了预测准确度。

二、文献综述

即使各国政府都出台了诸多法规来规范上市公司的行为,但舞弊事件仍层出不穷。国内外学者致力于舞弊影响因素、信号识别研究,从分析企业财务报表之间的内在逻辑冲突来发现违规行为,转向利用数据挖掘技术来识别违规行为。刘懿(2007)将上市公司违规与审计意见结合建模,发现违规是影响审计质量的显著因素。陈关亭(2007)建立了Logistic模型,发现舞弊压力主要来源于避免被ST处理、退市;股权集中度较高、独立董事比例较低、董事会会议次数较少、董事会成员持股量较少、董事长兼任总经理、监事会无效、变更主审会计师事务所则为财务报

【基金项目】国家自然科学基金项目"面向国家治理的云计算环境下联网审计流数据处理关键技术研究"(项目编号:61572079)

告舞弊提供了机会，内部控制对舞弊也有一定影响。吴革、叶陈刚（2008）对违规披露分类后，排除了业务舞弊、表外关联交易、延迟披露等违规类别，选择被证监会处罚的公司作为舞弊样本，从财务报表和公司治理方面建立Logistic回归模型，发现股权集中度、每股净资产差异率、非主营业务利润率、存货占流动资产的比重等因素会对舞弊行为产生影响。余玉苗、吕凡（2010）从发生财务违规公司的前一年与违规当年的财务指标的动态增量信息视角入手，建立Logistic识别模型，发现固定资产增长率、每股收益等会对财务违规产生重要影响。洪文洲等（2014）通过建立舞弊Logistic回归模型，发现折旧率变高、股权比较分散的上市公司更有可能发生财务违规行为。顾宁生和冯勤超（2009）、张秋三等（2014）运用神经网络建立了上市公司财务违规识别模型并验证了模型的有效性。高媛媛（2014）应用决策树模型对违规识别指标进行特征选择，并在此基础上构建神经网络模型，发现决策树—神经网络组合模型在识别违规方面的精确性和稳定性更好。

Glen L. Graya、Roger S. Debrecenyb（2014）建立了基于模式识别的分类数据挖掘模型，对文本及邮件数据进行舞弊分析。Amer Alhazaimeh等（2014）使用动态面板系统GMM估计模型，对约旦上市公司的公司治理与董事会结构进行研究，发现董事会活动、外国所有权、非执行董事和股东对自愿性信息披露有显著影响。Norazida Mohamed、Moorison Handley-Schachelor（2014）发现，管理的完整性、完善的内部系统可以降低财务舞弊发生的概率。Ismajli H.等（2017）采用调查问卷方式，得出了内部审计可以作为发现财务报告中舞弊和错误检测的起点的结论。

综合来看，学者们对舞弊特征进行研究的样本大多低于200个，视角多集中于财务指标、股权集中度、股东规模、审计意见类型等，最常用的方法为Logistic回归方法。

三、理论基础与研究设计

（一）基于国家治理的舞弊内涵界定

我国2010年修订的《中国注册会计师审计准则第1141号——财务报表审计中与舞弊相关的责任》第四条认定：舞弊是一个宽泛的法律概念，但注册会计师关注的是导致财务报表发生重大错报的舞弊，与财务报表审计相关的故意错报包括编制虚假财务报告导致的错报和侵占资产导致的错报。

上市公司信息违规披露形式包括虚构利润、虚列资产、虚假记载（误导性陈述）、推迟披露、重大遗漏、披露不实（其他）、欺诈上市、出资违规、擅自改变资金用途、占用公司资产、内幕交易、违规买卖股票、操纵股价、违规担保、一般会计处理不当及其他等16类。从注册会计师审计视角来看，推迟披露、内幕交易、违规买卖股票、操纵股价等违规行为并不一定与财务报表审计相关；但基于国家治理，从审计作为防范整个经济运行安全的“免疫系统”职能来看，该类违规是主观故意的行为，其危害远高于会计报表舞弊，政府审计更有必要、有责任、有义务通过历史数据研究发现违规的苗头性问题，以揭示内幕交易、违规买卖股票、操纵股价等舞弊行为。因此，对于上市公司的违规行为，本研究均界定为舞弊范畴。

（二）理论分析与研究假设

目前被广泛认可的经典舞弊三角理论认为，压力、机会、借口是舞弊行为发生的三大因素。舞弊者基于盈利、债务、股利分配、现金流等压力的异常需要是舞弊行为产生的根源。组织缺乏完善的内部控制制度、信息不对称、违规成本低、制度不健全、管理者无知无能等为舞弊行为的实施提供了途径。

1. *完善内部控制评价报告体系可抑制舞弊。*完善内部控制评价与披露制度是防范舞弊的一项重要措施。2008～2010年，财政部、证监会、审计署、银监会、保监会五部委先后联合颁布了《企业内部控制基本规范》及18项配套指引，在境内外同时上市的公司需于2011年起执行，这标志着我国企业内部控制规范体系基本建成。2014年证监会、财政部颁布了《公开发行证券的公司信息披露编报规则第21号——年度内部控制评价报告的一般规定》，进一步规范了上市公司内部控制评价报告编制规则。上市公司内控制度评价报告和内控审计报告披露机制要求上市公司披露内控是否存在缺陷、内控是否有效、内控审计意见等关键因素，这些因素也逐步显示出了其不可替代的作用。

上市公司内部控制通过两种作用机制实现其对舞弊的抑制：第一，内部控制实现了组织内部的权力制衡，减少了舞弊机会、错报和欺诈；第二，有效的内部控制可以抑制企业会计政策和会计估计的滥用，提升了员工的道德认知水平，削弱了舞弊动机。但是，仍有相当数量的违规上市公司的内部控制存在问题。经统计分析，1993～2015年A股上市公司年报数据中，共有1485家上市公司累计4502个年度存在

13165类违规行为。其中：有532家公司内部控制存在缺陷，28家公司内部控制无效；106家公司被出具了非标准内部控制审计意见。因此，提出假设：

H1：完善的内部控制可抑制舞弊。

2. *财务报告审计意见及审计收费对舞弊有揭示作用*。财政部2010年修订的《中国注册会计师审计准则第1141号——财务报表审计中与舞弊相关的责任》第六条明确规定：注册会计师在按照审计准则的规定执行审计工作时，有责任对财务报表整体是否不存在由舞弊或错误导致的重大错报获取合理保证。经统计分析，在过去23年间，注册会计师勤勉地履行了审计监督职能，约有85%的公司被出具了标准审计意见。但这并不意味着这些公司不存在舞弊，由于审计的固有限制，即使注册会计师按照审计准则的规定恰当地计划和执行了审计工作，也不可避免地存在财务报表中的某些重大错报未被发现的风险。共有562家被出具无保留审计意见的公司存在1826类违规行为且遭到处罚，其中不乏重大遗漏、虚假记载（误导性陈述）、披露不实（其他）、占用公司资产、虚构利润、违规担保、欺诈上市、虚列资产等审计应该揭示的舞弊行为。由于舞弊可能涉及行为者的精心策划、虚假隐瞒、串通合谋，在舞弊导致错报的情况下，固有限制的潜在影响尤其重大。

证监会[2016]126号文件《上市公司股权激励管理办法》要求最近一年财务报表被审计师出具无法表示意见和否定意见的公司不得实行股权激励。因此，财务报表审计意见类型一方面成为公司舞弊的压力指标，另一方面是审计揭示上市公司舞弊的有力手段。审计人员应提高对舞弊的识别能力，降低固有风险对舞弊识别的影响。

另外，异常的审计收费可能暗示着舞弊发生的机会或对舞弊揭示的能力不足。其中：过低的审计收费可能暗示着会计师事务所规模不足、低价竞争导致注册会计师不够勤勉；过高的审计收费可能暗示着注册会计师或会计师事务所合谋、对舞弊风险的估计不足、舞弊识别能力欠缺等。因此，提出假设：

H2：财务报告审计意见及审计收费对舞弊有揭示作用。

3. *提高治理层监管能力是预防舞弊的有效手段*。董事会人数、监事会人数反映了公司的决策效率和对管理层的监管作用。过少的人数可能会导致高管权力凌驾于内部控制之上；过多的人数可能会降低监管效率。董事会及监事会的履职频率从一定程度上反映了治理层的执行力。恰当的股权集中度有利于公司股东利益产生趋同效应；过高的集中度可能使中小股东的利益被侵占，在一定程度上为大股东实施舞弊行为提供了途径。因此，提出假设：

H3：恰当的治理层规模、履职频率及股权集中度可预防舞弊。

4. *盈利及持续发展压力催生舞弊动机*。我国对上市公司业绩的评价高度依赖财务数据，证监会2015[119]、2015[122]、2016[127]号文件等对首次执行上市、重大资产并购重组及配股、增发等规定了相应的业绩要求，若公司在持续经营中达不到盈利或业绩要求，则会面临ST处理或退市等处罚。因此，当上市公司可能面临盈利及可持续经营压力时，可能发生舞弊行为。

H4：盈利能力、可持续发展能力类指标能反映舞弊压力。

（三）变量选择

本文在参考国内研究上市公司违规征兆相关文献的基础上，将上市公司舞弊设为因变量。自变量基于舞弊三角理论，考虑影响舞弊的压力、机会、借口等因素指标，从公司治理、财务指标、持续经营三个维度十四个方面选择了33个特征指标，如表1所示。

1. *预测变量*。上市公司舞弊为预测变量，包括非舞弊（0）、舞弊（1）两种状态。

2. *公司治理类变量*。公司治理类变量包括环境特征、上一年度的内部控制及审计意见、治理层规模、履职频率、股权集中度等10个变量。

3. *财务指标类变量*。选择盈利能力、偿债能力、经营能力、发展能力等四类13个常用财务指标，并将其纳入模型进行分析。发展能力和盈利能力会受到国家政策、经济环境、行业发展趋势及公司经营状况的影响，既是评价管理者业绩的重要指标，又是反映公司内部压力的代表性指标。

4. *持续经营能力类变量*。持续经营能力是影响上市公司发展前景的关键因素，选择四类10个指标纳入模型进行分析。其中：风险水平和股利分配是来自债权人和股东的外部压力指标；发展水平和现金流是影响公司持续经营能力、评价管理者业绩的重要因素。

（四）数据预处理

本研究选取GSMAR财经数据库中1993～2015年A股上市公司年报相关的多库数据，初步选择A

表 1　　特征指标及含义

指标类别		指标名称	代码	公式/备注
预测变量		舞弊	Fraud	分类变量(0,1),0为非舞弊,1为舞弊
公司治理	环境特征	行业	Hy	分类变量(多分类)
	上一年度的内部控制	内控是否存在缺陷	DIC	分类变量(0,1),0为无缺陷,1为存在缺陷
		内控是否有效	EIC	分类变量(0,1),0为有效,1为无效
		内控意见	ICOP	分类变量(0,1),0为标准,1为非标准
	上一年度的审计意见	审计收费	ARO	分类变量(0,1),0为标准,1为非标准
		审计费用合计	FEE	连续变量
	治理层规模	董事会人数	BDN	上市时董事会人数
		监事会人数	BSN	上市时监事会人数
	履职频率	董事会会议次数	BOME	每年召开董事会会议的次数
	股权集中度	十大持股Cr指数	Cr	前十大股东持股比例之和
财务指标	盈利能力	净资产收益率	ROE	净利润/股东权益平均余额
		息税前利润	EBIT	净利润+所得税费用+财务费用
		总资产净利润率	ROA	净利润/总资产平均余额
	偿债能力	权益乘数	EM	资产总计/所有者权益合计
		资产负债率	DAR	负债合计/资产总计
		流动比率	CR	流动资产/流动负债
		利息保障倍数	TIER	(净利润+所得税费用+财务费用)/财务费用
		经营活动产生的现金流量净额/带息债务	CFBDR	经营活动产生的现金流量净额/(非流动负债合计平均余额+短期借款平均余额+一年内到期的非流动负债平均余额)
	经营能力	应收账款与收入比	ARIR	应收账款/营业收入
		总资产周转率	TAT	营业收入/资产总额期末余额
	发展能力	基本每股收益增长率	BESGR	(基本每股收益本年本期单季度金额-基本每股收益上一个单季度金额)/基本每股收益上一个单季度金额
		净利润增长率	NPGR	(净利润本年本期单季度金额-净利润上一个单季度金额)/净利润上一个单季度金额
		可持续增长率	SGR	股东权益收益率×(1-股利支付率)/[1-股东权益收益率×(1-股利支付率)]
持续经营能力	发展水平	每股收益	EPS	净利润本期值/实收资本本期期末值
		息税前每股收益	EPSB	(净利润本期值+所得税费用本期值+财务费用本期值)/实收资本本期期末值
		每股企业自由现金流量	FCFS	(现金及现金等价物净增加额本期值-筹资活动产生的现金流量净额本期值)/实收资本本期期末值
	风险水平	财务杠杆	FLI	(净利润+所得税费用+财务费用)/(净利润+所得税费用)
		综合杠杆	DTL	(净利润+所得税费用+财务费用+固定资产折旧、油气资产折耗、生产性生物资产折旧+无形资产摊销+长期待摊费用摊销)/(净利润+所得税费用)
	股利分配	股利分配率	DDR	每股派息税前/(净利润本期值/实收资本本期期末值)
		收益留存率	RR	1-每股派息税前/(净利润本期值/实收资本本期期末值)
		每股股利变动值	DPS	每股派息税前本期-每股派息税前上年同期
	现金流	净利润现金净含量	NCP	经营活动产生的现金流量净额/净利润
		全部现金回收率	TCR	经营活动产生的现金流量净额/资产总计期末余额

股年报数据30000余条记录，采用SQL Server 2012数据库进行数据预处理：①浏览数据。对特征指标变量进行描述性统计分析，观察分布情况。②缺失值处理。对来自15个不同数据库的几十万条数据进行清理、转换，导入SQL Server 2012数据库进行处理，删除缺失值样本。③重新标记，将分类变量用数字0、1

进行重新编码，将审计更为关注的类别定义为1。如正常公司定义为0，违规公司定义为1；标准审计意见定义为0，非标准审计意见定义为1。④分离属性。将原数据库中某一属性的多种水平分离处理为多个样本。⑤多表关联，构造舞弊特征指标数据集。将舞弊数据表与公司治理、财务指标、持续经营能力三大类指标一一匹配后形成样本数据。

（五）样本选择

舞弊数据样本来源于1993年1月1日～2015年12月31日因违规而受到上交所、深交所、证监会、财政部等处罚的3702条上市公司数据。非舞弊样本选自数据库中正常公司A股年报数据。经数据预处理后共形成4461个正常与舞弊的样本数据，其中舞弊样本714个。

信赖过度风险经常会导致严重的审计后果，因而审计人员更为关注对舞弊公司的正确分类。研究表明，对于一些基分类器而言，均衡的数据集可以更有效地提高全局的分类性能。由于舞弊数据（714个）仅占有效样本总数（4461个）的16%，相对于非舞弊数据（3747个）过于稀少，因此本研究分别采取1∶1配对样本及1∶2、1∶3、1∶4三种过度抽样比例进行研究，选择70%的数据作为训练样本，另外30%为测试样本，样本规模如表2所示。

表2　抽样比例与样本量

抽样比例	原始样本	抽样规模	训练样本	测试样本
1∶1	4461	1428	1000	428
1∶2	4461	1071	750	321
1∶3	4461	952	666	286
1∶4	4461	893	625	268

四、实证分析

（一）描述性统计分析

1. 舞弊公司分析。在选定的原始样本中，对舞弊公司的数据进行分析，其中2012年违规交数最多（见表3）。计算机、通信和其他电子设备制造业，化学原料及化学制品制造业，电气机械及器材制造业这三个行业的违规次数最多（见表4）。由于存在同公司、同年度多类违规的情况，在714家上市公司违规样本中共有1862类违规行为，违规次数最多的前三类为其他、重大遗漏和推迟披露。

表3　舞弊公司分年度统计数据

年度	2008	2009	2010	2011	2012	2013	2014	合计
违规次数	4	17	34	99	199	197	164	714

表4　舞弊行业及类别统计

行业	违规次数	违规类别	违规次数
计算机、通信和其他电子设备制造业	66	其他	509
化学原料及化学制品制造业	54	重大遗漏	284
电气机械及器材制造业	49	推迟披露	279
医药制造业	43	虚假记载（误导性陈述）	206
专用设备制造业	43	违规买卖股票	176
软件和信息技术服务业	36	一般会计处理不当	156
汽车制造业	28	披露不实（其他）	105
房地产业	23	占用公司资产	34
土木工程建筑业	23	虚构利润	29
橡胶和塑料制品业	23	擅自改变资金用途	28
有色金属冶炼及压延加工业	22	违规担保	25
批发业	22	内幕交易	24
通用设备制造业	19	欺诈上市	4
电力、热力生产和供应业	17	虚列资产	2
其他	246	操纵股价	1
合计	714	合计	1862

2. 配对样本t检验和Wilcoxon秩和检验。为了检验舞弊与非舞弊公司在不同变量间有无显著差异，对分类变量采用卡方检验，对连续变量采用配对样本t检验与Wilcoxon秩和检验，检验结果见表5。由表5可知：20个指标在舞弊公司与非舞弊公司间有显著差异。

（二）模型设计与评价

根据历史数据对舞弊进行建模分类，基于四种样本规模，采用R语言建立决策树模型，采用C5.0算法，用信息增益确定分枝规则。

1. 决策树C5.0算法。决策树方法是以一组特征变量为基础来预测二分类因变量的机器学习方法，以树形结构建模，将某一属性作为决策结点并进行分权，从根节点开始至叶节点终止。C5.0算法由计算机科学家J. Ross Quinlan开发，是最知名的决策树算法之一，目前已成为生成决策树的行业标准。具体算法如下：

（1）选择最佳分割点。决策树算法依据一系列特征变量，寻找用来划分二分类因变量的关键特征，即确定哪一个最佳分割变量可使分类最纯。最佳分割点能够实现样本的最佳分组，以使每个组仅由一个类别支配。C5.0算法使用熵值来计算由每一个可能

表 5　　配对样本检验结果

代码	因变量类型	均值	配对样本t检验		Wilcoxon秩和检验	
			t/x^2	Sig.	w	Sig.
Hy		多分类	1295.4	2.2e−16***	分类变量；卡方检验	
DIC	0/1	0表示无缺陷，1表示存在缺陷	1.4617	0.2267		
EIC	0/1	0表示有效，1表示无效	17.283	3.221e−05***		
ICOP	0/1	0表示标准，1表示非标准	25.529	4.357e−07***		
ARO	0/1	0表示标准，1表示非标准	24.366	7.967e−07***		
FEE	0	1931716	−4.5744	5.631e−06***	75512	0.4377
	1	1054801				
BDN	0	9.39	−0.7728	0.4399	105310	0.7586
	1	9.31				
BSN	0	4.16	−4.0267	6.278e−05***	54002	4.897e−05***
	1	3.84				
BOME	0	10.02	−0.74736	0.4551	158810	4.675e−10***
	1	9.84				
Cr	0	58.77	0.7524	0.4521	122990	0.7817
	1	59.42				
ROE	0	0.06	−0.93821	0.3485	149740	3.019e−05***
	1	0.04				
EBIT	0	1892000000	−4.8549	1.481e−06***	176160	2.2e−16***
	1	360700000				
ROA	0	0.04	−1.3631	0.1733	138990	0.01966**
	1	0.01				
EM	0	2.92	−1.5724	0.1163*	143550	0.00194***
	1	2.28				
DAR	0	0.48	0.57353	0.5665	144260	0.00128***
	1	0.53				
CR	0	2.50	1.3437	0.1795	110020	0.9993
	1	2.82				
TIER	0	1.36	−1.7492	0.08069*	134000	0.1238
	1	−46.78				

代码	因变量类型	均值	配对样本t检验		Wilcoxon秩和检验	
			t/x^2	Sig.	w	Sig.
CFBDR	0	9.16	0.88213	0.378	130880	0.2774
	1	36.06				
ARIR	0	0.22	1.5091	0.1317	112910	0.9962
	1	0.24				
TAT	0	0.61	1.3929	0.1641	126140	0.6067
	1	0.65				
BESGR	0	−1.91	0.82507	0.4096	132170	0.1868
	1	0.3				
NPGR	0	2.2	−0.23319	0.8157	135810	0.06884*
	1	1.64				
SGR	0	0.04	−0.92327	0.3562	152060	4.666e−06***
	1	0.02				
EPS	0	0.39	−2.5567	0.01077**	150370	1.848e−05***
	1	0.3				
EPSB	0	0.61	−2.8481	4.525e−03***	153520	1.318e−06***
	1	0.48				
FCFS	0	−0.3087	−2.9414	3.373e−03***	149620	3.313e−05***
	1	−0.49722				
FLI	0	1.2558	−0.00236	0.9981	132220	0.2023
	1	1.2548				
DTL	0	1.886	−0.18759	0.8513	132350	0.1959
	1	1.669				
DDR	0	0.2265	0.64067	0.5219	107550	0.8634
	1	0.2573				
RR	0	0.7735	−0.64067	0.5219	118580	0.1366
	1	0.7427				
DPS	0	−0.01216	−1.4756	0.1405	110520	0.0598*
	1	−0.02409				
NCP	0	2.7388	−0.89757	0.3697	112910	0.9962
	1	1.1357				
TCR	0	0.035237	−2.3623	0.01843**	136100	0.06225*
	1	0.02538				

注：*、**、***分别表示在10%、5%和1%的水平上显著。

特征的分割所引起的同质性变化（信息增益）度量的分类纯度，如式（1）所示。

$$\text{Entropy}(S)=\sum_{i=1}^{c}-p_i\log(p_i) \qquad (1)$$

熵表示分类值是如何混杂在一起的。在熵的公式中：S代表给定的分割；常数C代表分类水平（本预测变量的水平为2）；p_i代表落入某一分类中的特征值的比例。对于特征F，信息增益是分割前的数据分区（S_1）的熵值与由分割产生的数据分区（S_2）的熵值差，如式（2）所示。

$$\text{InfoGain}(F)=\text{Entropy}(S_1)-\text{Entropy}(S_2) \qquad (2)$$

决策树经历一次分割后，数据被分到多个分区中，计算由分割产生的熵值时需要考虑所有分区熵值的总和，如式（3）所示。

$$\text{Entropy}(S)=\sum_{i=1}^{n}w_i\text{Entropy}(p_i) \qquad (3)$$

因此，从一个分割得到的总熵是根据落入每一分类的比例 w_i 加权的 n 个分类的熵值的总和。信息增益越高，根据某一特征分类后创建的分组越均衡。除了信息增益分割标准，其他常用的评估决策树的最佳分割点的标准还包括“基尼系数（Gini index）”、卡方统计量（Chi-Squared statistic）和增益比（Gain ratio）等。

（2）修剪决策树。决策树在无限分割中易使决策过于具体，产生过度拟合问题，修剪决策树可以更好地预测未知数据，这是有效提高决策树预测能力的环节。通常使用预剪枝法或后剪枝法抑制树的过度生长。C5.0 算法采用自动修剪技术，先生成一个过度拟合训练数据的大树，通过事后修剪法，修剪掉对分类误差影响不大的节点和分枝，利用子树提升和子树替换的方法完成修剪。决策树 C5.0 算法易于调整训练方案，适用于大多数问题，学习过程高度自动化，更易于理解和部署，因而具有更强的适用性。

2. 模型评价标准。模型采用精确率、召回率、F1 分数、KAPPA 值及 ROC 曲线（Receiver Operating Characteristic Curve，受试者工作特征曲线，简称“ROC 曲线”）进行评价。评价标准的计算依赖于表达预测值与真实值间关系的混淆矩阵（见表 6），其中：T 表示舞弊；F 表示非舞弊；TN 表示对非舞弊公司的正确预测（真阴性）；FP 表示实际为非舞弊公司，但被预测为舞弊公司（假阳性）；FN 表示实际为舞弊公司，但被预测为非舞弊公司（假阴性）；TP 表示对舞弊公司的正确预测（真阳性）。

表 6　　评价标准混淆矩阵

	预测	
实际	F	T
F	TN	FP
T	FN	TP

精确率是指所有预测为舞弊的公司中真正舞弊公司的比例，用来评价分类模型的质量，计算方法为真阳性的数目除以真阳性和假阳性的总数，见式（4）。召回率是实际舞弊公司中预测准确的比例，见式（5），用来评价分类模型的完整性，是对舞弊公司正确分类的比例，是审计所关注的犯第二类错误即误受风险被正确揭示的概率。F1 分数是用来衡量模型精确度的一种指标，是模型精确率和召回率的一种加权平均，计算方法见式（6），介于 0～1 之间。

$$精确率=\frac{TP}{FP+TP} \quad (4)$$

$$召回率=\frac{TP}{FN+TP} \quad (5)$$

$$F1分数=\frac{2TP}{2TP+FP+FN} \quad (6)$$

KAPPA 统计量代表了预测值和真实值之间的一致性，一般 0.2～0.4 代表一致性尚可，0.4～0.6 表示中等的一致性，1 表示完全一致。ROC 曲线又被称为感受性曲线（Sensitivity Curve），是以召回率为纵坐标，（1−特异度）为横坐标，在单位面积为 1 的正方形内绘制的曲线。ROC 曲线下的面积（AUC）代表预测效果，曲线越凸向左上角的顶点，AUC 面积越大，则模型预测效果越好。

3. 实证结果分析。抽样模式对模型预测精度有较大影响。与 1∶1 配对抽样相比，过度抽样模式有效提高了模型预测的召回率，降低了舞弊识别的误受风险。在 1∶3 过度抽样模式下，决策树 C5.0 算法对舞弊预测的召回率最高，比 1∶1 配对抽样提高了 32.82%（见表 7）。但 KAPPA 值与 1∶1 抽样模式相比有显著降低，说明过度抽样模式在提高舞弊公司识别的召回率的同时会导致对非舞弊公司的误判风险增大。

表 7　　不同抽样模式下的模型比较

模型	训练样本					预测样本			
	评估指标	1∶1	1∶2	1∶3	1∶4	1∶1	1∶2	1∶3	1∶4
决策树 C5.0	精确度（%）	97.50	95.77	92.19	97.47	56.94	78.35	80.33	87.26
	召回率（%）	93.41	97.54	98.80	99.40	57.75	79.74	90.57	87.68
	F1 分数	0.954	0.966	0.954	0.984	0.573	0.790	0.851	0.875
	KAPPA 值	0.910	0.903	0.727	0.897	0.145	0.269	0.081	0.292

与因变量舞弊关联性由强至弱的前十个变量依次为内控是否有效、息税前利润、董事会会议次数、内控意见、董事会人数、监事会人数、股利分配率、每股企业自由现金流量、总资产净利润率、审计收费，如表 8 所示。

表 8　　舞弊的重要影响因素综合排序（前十）

变量	EIC	EBIT	BOME	ICOP	BDN	BSN	DDR	FCFS	ROA	FEE
重要程度（%）	100	98.90	82.30	78.10	71.70	65.00	40.70	38.90	36.80	35.40

采用信息增益剪枝的C5.0决策树模型建立了58条决策规则，可由模型或决策树图得出。比如：约有11%的内控无效的企业被揭示出舞弊；约有35%的内控无效、息税前利润小于682715900元且董事会会议次数小于9.5次的企业被揭示出舞弊；约有26%的内控无效、息税前利润小于682715900元、董事会会议次数高于9.5次、董事会人数小于7.5人、应收账款收入比高于76%且利息保障倍数低于-17的企业被揭示出舞弊；约有9%的内控有效、息税前利润高于682715900元、可持续增长率不高于1.12%且董事会人数小于15人的企业被揭示出舞弊；约有1%的内控有效、息税前利润小于682715900元且可持续增长率低于1.12%的企业被揭示出舞弊等。

（三）模型校正与优化

过度抽样预测虽然有效提高了对舞弊分类预测的召回率，但审计需要获取更多证据来排除对非舞弊公司的舞弊误报，极大地影响了审计效率。配对抽样模式通过随机抽取训练样本与测试样本的方法，不能全面地代表总体分布情况。而利用机器学习算法，可有效优化抽样及预测效果。

1. Bagging优化算法。为了提高分类模型的预测准确率，通常将多个分类方法聚集在一起进行集成学习，首先由训练数据构建一组基分类器，然后通过对每个基分类器的预测进行投票来分类。自助汇聚法（简称“Bagging算法”）是得到广泛认可的最好的集成学习方法之一。Bagging算法对原始训练数据使用自助抽样的方法，根据均匀概率分布从数据中重复抽样，产生多个训练数据集，使每个训练数据集训练一个基分类器，多个数据集各自使用单一的机器学习算法产生多个模型，然后采用投票的方式来组合预测值。

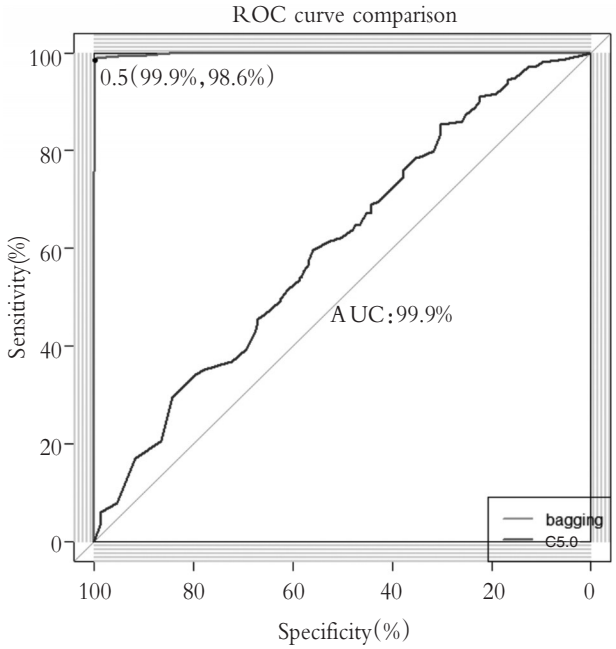
Bagging算法是一种相对简单的集成学习器，与相对不稳定的学习器结合使用可得到很好的效果。决策树算法是一种不稳定的分类器，会随着数据抽样的不同而产生较大的差别。因此决策树算法与Bagging算法组合使用，可确保即使自助抽样数据集之间的差异很微小，集成学习器的投票结果也具有多样性。

Bagging算法在抽样中由于采用重复抽样，每个样本被选中的概率相同，因此Bagging并不侧重于训练数据集中的任何特定实例，其性能依赖于基分类器的稳定性，如果基分类器不稳定，Bagging算法会通过降低基分类器的方差，缩小泛化误差，有效提高基分类器的准确率。

2. Bagging算法优化结果。对决策树C5.0模型使用Bagging集成学习算法进行优化。采用R语言中ipred包，选取25个决策树进行投票，训练结果表明对训练数据的拟合效果非常好，预测准确度为99.23%，召回率为99.86%，KAPPA值为0.985，如表9所示。与C5.0算法比较的ROC曲线如下图所示，Bagging算法优化后曲线下面积为0.999。

表9　Bagging集成学习算法优化后模型验证与评估

优化与验证方法	总体准确率	精确度	召回率	F1分数	KAPPA值
优化前的决策树C5.0模型	57.24%	56.94%	57.75%	0.573	0.145
Bagging优化后	99.23%	78.71%	99.86%	0.72	0.985
十折交叉验证（均值）	66.5%	–	–	–	0.33
样本外测试（344）	93.11%	77.36%	78.85%	0.78	0.740



C5.0算法与Bagging优化后的ROC曲线比较图

（四）模型稳定性验证

为了验证模型未来性能方面的表现，采用十折交叉验证法建立Bagging树，并进行了样本外数据测试。十折交叉验证的平均准确度与KAPPA值均高于优化前模型的预测效果。样本外数据测试选择同期样本外数据334条，其中舞弊样本53条，正常公司样本281条。选用Bagging算法优化后的决策树模型预测正确的样本有311个，其中：对舞弊公司预测正确

的有41个，对非舞弊公司预测正确的有270个；预测总体准确率为93.11%，比优化前提高了35.87%；召回率为78.85%，比优化前提高了21.1%；KAPPA值为0.74，比优化前提高了0.595（见表9）。这说明优化后的算法对样本外数据依然有较好的预测效果，该模型对未来数据预测有较好的稳健性。

五、结论与展望

（一）主要结论

为了揭示上市公司舞弊情况，选取了我国上市公司1993～2015年度舞弊与正常的4461例数据为样本，选取公司治理、财务指标、持续经营等三维十四类33个变量对舞弊情况进行解释，采用四种抽样方法构建了决策树C5.0预警模型，并结合Bagging机器学习算法对决策树模型进行了优化。为验证模型的稳定性，采用十折交叉验证并对334例样本进行了样本外测试，结论如下：

1. 内部控制评价报告体系对舞弊揭示效果显著。H1得到验证。影响舞弊最重要的十个因素依次为：内控是否有效、息税前利润、董事会会议次数、内控意见、董事会人数、监事会人数、股利分配率、每股企业自由现金流量、总资产净利润率、审计收费。其中内控是否有效、内控意见类型分别位列第一与第四，说明内控评价报告与审计披露制度对揭示舞弊有显著效果。完善的内部控制有助于抑制舞弊，内控薄弱的上市公司更易产生舞弊机会。

2. 财务报告审计意见及审计收费对舞弊揭示作用显著。H2部分得到验证。在影响舞弊最重要的十个因素中，审计收费被证明是影响舞弊的十大关键因素之一，异常的审计收费可能暗示着存在审计师合谋或外部监督不力的情况。上一年度财务报告审计意见对当年舞弊揭示作用不显著，可能是因为若公司上一年度被出具非标准审计意见，则会对下一年度财务报告披露产生负面的影响，从而企业会进行整改，使得其与舞弊无显著关系。上一年度若被出具标准审计意见，则无显著证据表明与本年度是否舞弊有直接的因果关系。

3. 恰当的治理层规模、履职频率及股权集中度可预防舞弊。H3部分得到验证。公司治理类变量对舞弊预测的影响比财务类、可持续发展类指标更为重要。在影响舞弊的前十大因素中，除了内部控制和审计收费，还有董事会会议次数、董事会人数、监事会人数为公司治理类指标，说明董事会的无效监管为舞弊提供了机会和借口。

4. 盈利能力、可持续发展能力类指标能反映舞弊压力。H4部分得到验证。代表盈利能力的息税前利润、总资产净利润率指标对舞弊有显著影响。代表可持续经营能力的股利分配率、每股企业自由现金流量指标催生了财务灵活性压力，对舞弊有显著影响。

5. 抽样模式会影响模型预测的准确率和召回率。基于机器学习的Bagging算法优化决策树模型，对舞弊识别的召回率提升了42.11%。样本外数据测试显示出模型预测力与稳定性良好。

（二）研究展望

舞弊风险预测可有效提高审计疑点发现能力，未来应基于历史数据，考虑不同舞弊类型下的审计策略，进行多模型组合建模，以探索出更有效的审计策略响应机制。

主要参考文献：

刘懿. 上市公司违规与审计意见特征的实证研究[D]. 成都：西南财经大学，2007.

陈关亭. 我国上市公司财务报告舞弊因素的实证分析[J]. 审计研究，2007(5).

吴革，叶陈刚. 财务报告舞弊的特征指标研究：来自A股上市公司的经验数据[J]. 审计研究，2008(6).

余玉苗，吕凡. 财务违规风险的识别——基于财务指标增量信息的研究视角[J]. 经济评论，2010(4).

洪文洲，王旭霞，冯海旗. 基于Logistic回归模型的上市公司财务报告违规识别研究[J]. 中国管理科学，2014(S1).

顾宁生，冯勤超. 基于LVQ神经网络的财务违规识别模型实证研究[J]. 价值工程，2009(10).

张秋三，张磊，张宁，蔡玖琳. 基于数据挖掘的上市公司财务违规识别研究[J]. 科技和产业，2014(11).

Brett Lantz著. 李洪成，许金炜，李舰译.机器学习与R语言[M]. 北京：机械工业出版社，2015.

高媛媛. 基于数据挖掘的财务违规识别研究——决策树—神经网络组合模型的构建[J]. 科技经济市场，2014(11).

作者单位：北京信息科技大学信息管理学院，北京100192