

doi:10.19920/j.cnki.jmsc.2023.05.010

大数据时代的管理研究新范式：以 CEO 解聘问题为例^①

王 聪, 易希薇, 张志学

(北京大学光华管理学院, 北京 100871)

摘要: 面对复杂的管理现象, 主流的实证研究通常从某一角度去揭示特定现象. 学者难以基于众多细微研究构建出中层理论, 甚至也无法评价已有研究所探索的诸多影响因素的相对重要程度. 本文提出, 通过机器学习方法分析海量数据, 并基于学界已经积累的研究结论, 学者能够比较准确地把握与现象最相关的因素, 从而建立更具解释力的理论. 本研究以公司治理中 CEO 解聘这一流行话题为例, 通过梳理现有文献, 基于 CEO 大型文本语料库及 CEO 任职期间其个人与公司的相关数据, 采用机器学习方法提取与 CEO 解聘相关的三类因素, 并进一步计算出三类因素单独和彼此结合的预测力. 研究发现证实了近年来这一领域研究话题向个体异质性转变的合理性, 也为学者建立解释 CEO 解聘的整合理论奠定基础. 本研究展示了机器学习与实证研究相结合的优势和前景.

关键词: 大数据; 管理研究; 中层理论; 机器学习; CEO 解聘

中图分类号: C93 **文献标识码:** A **文章编号:** 1007-9807(2023)05-0200-14

0 引 言

近年来, 以云计算、移动互联网、人工智能技术为代表的新兴信息技术正在深刻改变人们的工作和生活方式, 也成为推动中国经济与实现社会可持续发展的重要手段. 人们的经济活动和社会生活轨迹都被以数字化的形式记录下来, 形成了具有超规模、高维度、跨领域、重时效等特点的大数据, 在推进我国向数字经济时代迈进的同时, 也为各领域学者提供了一个催生重大理论创新成果的“富矿”^[1, 2].

理论创新是衡量科学研究贡献的重要标准. 社会科学理论分为宏大理论、细微理论和中层理论^[3]. 宏大理论是抽象的, 包括一组相互联系、涉及诸多情境现象的命题或假设; 细微理论非常具体, 只包括涉及有限情境下少数现象的概念; 中层理论则介于两者之间, 可解释复杂现象背后的一

般规律^[4]. 管理现象是高度情境化的^[5], 很多因素会影响组织管理实践. 学者通常只能研究有限的变量, 导致探讨同一话题的研究往往具有不同的结论, 且所建立的众多细微理论缺乏对管理实践的解釋力. 为此, 管理学者需要建立中层理论, 以便解释及预测复杂的管理现象.

要实现细微理论向中层理论的迈进, 需要对已有研究中涉及的概念或变量进行分析, 找出最具解释力的概念或变量, 从而为建立中层理论打下基础. 为此, 学者多采用元分析 (meta-analysis) 方法^[6]对现有文献中报告的统计指标进行再分析, 然后提出一系列命题或者有关某个现象的新理论框架. 然而, 这类分析性文章仍然受限于作者的理论视角, 带有较大的主观性. 那么, 是否存在一种类似元分析的方法, 能够对于已有研究进行相对客观地整理和总结, 找出与现象最相关的因素或变量, 从而为建立更具整合性和解释力的中

① 收稿日期: 2021-12-31; 修订日期: 2022-12-26.

基金项目: 国家自然科学基金资助重大研究计划集成项目(92146003); 国家自然科学基金资助项目(71802007; 72101007).

作者简介: 王 聪(1991—), 女, 黑龙江鸡西人, 博士, 助理教授, 博士生导师. Email: wangcong@gsm.pku.edu.cn

层理论提供指导呢?

本文认为基于大数据训练的机器学习方法就是一种重要的解决方案。具体而言,传统的管理研究关注在特定场景下“小数据”样本内的拟合能力,形成适用于特定场景的细微理论。因忽略了对样本外情境的泛化能力,难以形成可解释及预测相对复杂现象的中层理论。若要试图构建中层理论,需对某一管理现象进行深入调研,提炼出多维度、跨领域的“大数据”进行分析,并同时追求样本内的解释力及样本外的泛化能力。机器学习方法的优势正在于此。其一,自然语言处理、图像处理、语音识别等方法可从多个途径提取并分析管理场景中海量结构化及非结构化数据,构建出传统管理研究中无法测量的新变量。其二,基于机器学习方法训练而成的模型可用于探测变量间多种类型的组合关系。其三,机器学习算法可通过训练集及测试集的划分兼顾解释能力与泛化能力。其四,新兴的可解释性机器学习框架可用于分析各因素的预测能力,进而为理论构建提供有效借鉴。因此,利用大数据和机器学习技术开展管理研究,将为打通传统管理学研究中的宏观和微观的壁垒,整合企业的战略、组织和行为并建立新的理论提供有效的手段。

本文将以前战略学者长期研究的CEO解聘这一话题为例,详细阐释机器学习方法在组织与战略管理领域研究中实现理论创新的具体应用。首先,通过机器学习方法处理数据库中的结构化数据和与CEO及公司相关的文本等大量非结构化数据,提炼出众多可能影响CEO解聘的变量。随后,根据问题特点训练预测CEO解聘的机器学习模型,进而在分析不同类型因素对模型预测能力影响的基础上,将影响CEO被解聘的众多因素进行排序,为理论构建提供必要线索。

1 文献回顾

1.1 机器学习方法在战略管理领域的应用

战略管理研究多依赖于对财务报表、市场价格、销量等结构化数据的分析,忽略了社会、经济、

市场、产品、人员相关信息的非结构化数据^[7-9]。自然语言处理、图像处理、语音识别等人工智能技术的发展,使原本无法分析的数据和无法测量的变量变得可测可获,大大拓宽了战略管理研究的范围。

自然语言处理方法的一些典型应用,如情感分析、文本分类等已被应用于战略管理研究中,实现对以往无法测量的变量进行测量,从而可以探讨过去无法研究的新问题。例如Gans等^[10]通过对客户推文的情绪分析预测企业行为;Barlow等^[11]应用自然语言处理方法衡量APP相似程度,以帮助APP开发者进行市场定位以获得竞争优势。此外,传统研究中一些测量起来复杂繁琐的变量,也可以借助机器学习方法进行简化。例如,Harrison等^[12]通过机器学习方法分析上市公司季度电话会议文本材料,提供了测量CEO大五人格的方法。Yi等^[13]使用文本分析方法分析了新任CEO与证券分析师的季度电话会议内容,发现对前任CEO的奉承减少了前任CEO留任董事会对前任CEO早期解聘的消极影响。除了构建新变量及新测量方法外,机器学习方法也可用于构建更为复杂的多层次、非线性的模型以用于分析海量、非结构化的数据,从而为战略管理研究提供新的范式^[14]。

1.2 CEO解聘文献回顾

过往研究认为CEO的被动离职是由董事会推动的公司治理机制的重要环节之一。当CEO难以胜任或对企业绩效、声誉等造成负面影响时,代表股东利益的董事会将解聘CEO。过往研究关于CEO解聘的影响因素分为企业绩效下降、企业违规行为、公司治理结构、CEO背景特征以及经营环境特征五大类。

1.2.1 企业绩效下降

早期研究认为企业绩效下降是导致CEO解聘的主要原因。公司业绩是反映CEO能力和努力程度的有效信号,也是利益相关者评价CEO表现的标准。早期研究表明公司业绩越差,CEO离职的可能性越大。具体来说,公司规模越大、资产负债率越高、现金比率越低、多元化程度越高,CEO解聘率更高^[15]。

1.2.2 企业违规行为

企业的违规行为会使企业失去重要利益相关者的支持.企业会通过解聘 CEO 恢复合法性.现有文献对企业违规行为导致 CEO 解聘的现象进行了深入、系统的探讨,发现财务重述^[16,17]、SEC 调查^[18]等违规事件都会显著提高 CEO 解聘的概率.这一观点弥补了之前文献只从经济学视角讨论 CEO 解聘的不足,强调 CEO 的重要责任之一就是建立和维持企业的合法性,当企业的违规行为破坏了企业的合法性时,通过解聘 CEO 隔离违规行为对企业的影响^[17].

1.2.3 公司治理结构

公司治理结构主要包括所有权结构、董事会组成、CEO 权力和第三方利益相关者,这些都是影响 CEO 是否会被解聘的重要机制.

第一,所有权结构包括所有权集中度、股东类型以及股权性质等.早期文献强调当公司股权较为集中时,大股东具有监督 CEO 的强烈动机以及决定 CEO 去留的权力,加大了 CEO 因绩效下降被解聘的可能性^[19].近期研究关注股东类型对 CEO 解聘的影响.机构投资者具有雄厚的资本和有效的信息参与公司的治理活动,可以很好地监督 CEO 的行为,提高了 CEO 在企业绩效下降时被解聘的概率^[20].

第二,董事会是公司治理的重要组成部分.早期文献探讨了董事会规模对 CEO 解聘的影响,随着董事会规模的扩大,CEO 因企业绩效下降而被解聘的概率更低^[21,22].此外,围绕董事会的监督作用对 CEO 解聘的影响,多数文献认为董事会独立性越高,CEO 的解聘率越高^[22,23].近期文献开始探讨女性董事对公司治理的影响,相关研究认为董事会女性董事比例增加能够提升企业对女性 CEO 的包容性,降低其离职风险^[24].

第三,CEO 拥有的权力为自身提供了保护伞.首先,CEO 二元性、CEO 任命董事比例以及 CEO 与董事的任期重叠均象征着 CEO 在董事会中的影响力.研究表明,CEO 拥有两职合一的职位头衔、经 CEO 推荐的董事会成员任职比例越高或者 CEO 与董事的任职期间重叠越高,CEO 被解聘概率越低^[25].其次,CEO 薪酬、CEO 股权也代

表着 CEO 权力的大小.CEO 的薪酬越高、持有公司的股权越高,其在公司中的权力相对越大,被解聘的风险越低^[26].

第四,企业的第三方利益相关者也扮演了重要的监督角色.股票分析师通过从公司报告、业绩说明会等收集相关信息来分析公司的业绩和未来前景,其建议对投资者的决策进而对公司股价产生重大影响^[27].现有文献关注了分析师数量、分析师评级对 CEO 解聘的影响,分析师数量越高、分析师评级越高,CEO 解聘的风险越低^[27].

1.2.4 CEO 个人背景特征

CEO 个人背景特征也是影响 CEO 解聘的重要因素.过往研究主要聚焦于 CEO 的性别、职业经历等背景特征对 CEO 解聘的影响.研究表明,女性 CEO 容易受到性别偏见的影响,与男性 CEO 相比具有更高的解聘率^[28].与内部擢升的 CEO 相比,外部遴选的 CEO 的继任过程往往包含更多信息不对称性,因而董事会做出错误选择的可能性更高,相应地,外部 CEO 通常面临着更高的解聘风险^[29].CEO 上任之前作为继承人的身份意味着其能力经受住了董事会的考验,因而具有较低的解聘率^[30].此外,CEO 过往在其他企业担任 CEO 的经历赋予了他们管理企业的必要知识和经验,降低了被解聘的概率^[31].

1.2.5 经营环境特征

企业所在的行业环境特征对 CEO 解聘产生影响.复杂的经营环境对 CEO 的能力提出了挑战,而丰裕的环境为 CEO 的战略决策落实提供了支援^[32].近期研究提出,环境变化是导致 CEO 能力不再胜任原职的重要原因,主要表现为行业环境动态性^[16].

1.3 CEO 解聘文献评述

现有研究为 CEO 解聘前因提供了丰富见解,但存在两点不足:第一,在 CEO 个人背景特征对其解聘风险的影响上,过往研究聚焦于 CEO 职业经历等背景特征的影响^[28-31],忽视了 CEO 人格、关注焦点、领导风格等在其中发挥的作用.第二,过往研究倾向于将 CEO 视为解聘决策中的被动接受者,较少关注 CEO 如何通过语言沟通主动管理与组织重要利益相关者的关系进而降低解聘风

险的行为.这两类变量之所以较少被研究,主要是因为它们需要基于非结构化的文本、音频、视频等数据进行测量,而大样本处理此类非结构化数据十分困难.

随着文本分析和机器学习等方法的进步,CEO的人格特征以及语言行为开始受到研究者的关注.首先,学者们开始探究CEO的人格特征如何影响企业各类战略决策,包括并购、战略变革、创新创业、企业社会责任等.在并购方面,Chatterjee和Hambrick^[32]的研究表明,自恋型CEO偏好通过大胆行动来博取他人的关注,因此会给企业带来巨大的盈利或亏损.Malhotra等^[33]发现,外向型CEO更擅长寻找、发现收购机会,积极看待这类机会,并将其转化为集体行动.因而,他们更有可能参与收购,且收购规模更大、更容易获得成功.Gamache等^[34]发现,CEO的促进焦点(对收益的敏感性以及对进步和成长的渴望)会提升企业并购的数量和价值,而预防焦点(对损失的敏感性以及对稳定和安全的渴望)会降低并购的数量和价值.在战略变革方面,Herrmann和Nadkarni^[35]发现,CEO发起战略变革与提升变革绩效所需的大五人格特质存在差异.在创新创业方面,Simsek等^[36]发现,具有高核心自我评价的CEO对公司创业有着积极的影响.此外,在社会责任方面,Tang等^[37]发现自恋型CEO更关心企业社会责任,而自大型CEO则不太关心企业社会责任.当自恋型CEO观察到同行公司比自己的公司更多地参与企业社会责任时,他们往往会反其道而行之.

此外,CEO的语言行为也开始受到关注.CEO经常通过语言行为表达自己的战略主张,以期与重要的利益相关者建立良好的关系.Ocasio等^[38]等提出CEO与利益相关者之间的交流沟通可以引导利益相关者的注意力,进而影响他们对战略变革的理解.Fanelli等^[39]发现,CEO魅力性愿景表达有利于提升股票分析师对公司的推荐,使得不同分析师的评级趋于一致,但也会提高分析师对公司未来业绩预测错误的风险.Yi等^[13]发现,CEO在季度收益电话会议的公开发言中表现出对前任CEO的迎合,可以降低前任CEO留

任董事会主席对其地位的威胁.Park等^[40]提出,CEO经常对企业良好的业绩表现进行内部归因,宣称良好的业绩归功于他们的战略选择.然而,这种内部归因给财务分析师创造了一种期望,即在该CEO的领导下,良好的绩效结果会持续存在.因而,当企业绩效下降时,财务分析师也会作出内部归因,认为CEO的战略决策是导致绩效下降的原因.

考虑到CEO的人格特征和语言行为在塑造战略决策、维持利益相关者关系中所扮演的重要角色,本文计划通过机器学习的方法,来探究CEO的人格特征以及语言行为是否以及如何影响其解聘风险.本文将探究这些影响的程度相较于传统研究所发现的因素而言,是否值得进一步探索以及有无构建新的中层理论的可能.

本文的创新性在于:第一,过往文献通常采用大样本回归方法检验研究假设,受制于样本规模及测量方法,研究结论的准确性难以保证.本文采用机器学习的方法对海量数据进行分析,能够更加准确地把握与现象最有关联的影响因素,进而建立更具解释力的理论.第二,受数据来源所限,过往研究对于CEO解聘原因的探讨多基于结构化数据.本文提出文本数据作为重要的非结构化数据来源,与结构化数据相比涵盖数量更多、范围更广的信息,进而能够为该研究问题提供更为丰富的见解.第三,过往研究仅关注了CEO的个人背景特征对其解聘的影响,本文通过对文本数据的挖掘,探究了CEO的人格和领导风格等特征对其解聘风险的影响.与个人职业背景特征相比,人格特征和领导风格等特点更为直接地影响CEO的战略决策偏好及其与利益相关者关系的处理模式,从而能够实现对CEO解聘风险更为准确的预测.第四,过往研究将CEO作为解聘的被动接受者,忽视了其主观能动性.本文探究了CEO特定的语言特征如何影响其自身地位和解聘风险,极大地弥补了过往研究的不足.

2 数据来源及变量提取

根据文献回顾,本文将影响CEO解聘的因素

划分为结构变量、CEO 人格变量、CEO 语言变量三方面. 通过汇总多数据源, 建立了一个包含 5 484 位 CEO 的多维度数据集, 其中有 868 位 CEO 曾被解聘, 占比 15.8%. 数据集描述性统计信息如表 1 所示.

表 1 数据集描述性统计

Table 1 Descriptive statistics of the dataset

年份	企业数量	CEO 数量/ (人次)	解聘 CEO 数量/ (人次)
2002	1 849	1 849	62
2003	1 906	1 906	43
2004	1 859	1 859	55
2005	1 781	1 781	56
2006	1 937	1 937	62
2007	2 253	2 253	75
2008	2 176	2 176	73
2009	2 147	2 147	49
2010	2 113	2 113	50
2011	2 072	2 072	56
2012	2 034	2 034	55
2013	2 002	2 002	43
2014	1 970	1 970	70
2015	1 894	1 894	54
2016	1 789	1 789	67

以下是对数据来源及三个维度变量提取方式的详细介绍.

2.1 结构变量

企业绩效、企业违规构成了 CEO 解聘的重要原因. 为此, 本文引入了与企业绩效状况相关的变量. 企业规模通过公司总资产的自然对数衡量. 资产负债率通过公司总负债与总资产的比值计算得到. 现金比率通过公司现金及现金等价物与流动负债的比值计算得到. 多元化程度通过 1 减去赫芬达尔指数计算得到, 该指数的计算方法为业务部门销售额的平方和除以公司销售额的平方. 相关数据均来源于 Compustat 数据库. 此外, 本文考虑了企业违规的两种情形: 企业违反会计准则所

造成的财务重述以及企业受到 SEC 的调查. 具体来说, 企业违规 - 财务重述为虚拟变量, 若公司存在违反会计规则 (GAAP/FASB) 的重述行为则取值为 1, 反之则为 0. 企业违规 - SEC 调查为虚拟变量, 若公司存在被 SEC 调查则取值为 1, 反之则为 0. 为了综合考虑以上两种违规行为的影响, 本文进一步引入了企业违规 - 总和这一变量, 若公司存在上述两种违规行为中至少一种则该变量取值为 1, 反之则取值为 0. 相关数据来源于 Audit Analytics 数据库.

公司治理结构是影响 CEO 解聘的另一重要原因. 在所有权结构层面, 本文考虑了机构投资者的影响. 股权集中度通过赫芬达尔 - 赫希曼指数 (Herfindahl-Hirschman Index, HHI) 计算. 机构投资者股权占比通过机构投资者持有本公司的股份与公司已发行股份的比值计算得到. 上述数据来源于 ISS 数据库. 在董事会方面, 引入了董事会规模、独立董事以及女性董事占比情况. 董事会规模通过董事会中董事的数量衡量. 独立董事占比通过董事会中独立董事的数量与所有董事数量的比值计算得到. 女性董事占比通过董事会中女性董事的数量与所有董事数量的比值计算得到. 相关数据来源于 Boardex 数据库. 在 CEO 权力层面, 本文引入了 CEO 二元性、CEO 薪酬、CEO 股权、CEO 任命董事占比以及 CEO - 董事会任期重叠. CEO 二元性为虚拟变量, 若 CEO 同时担任董事会主席则值为 1, 反之则值为 0. 数据来源于公司公开信息、新闻报道等. CEO 薪酬由 CEO 年度工资、奖金、被授予的限制性股票总价值及股票期权总价值、长期激励等相加得到. CEO 股权通过 CEO 持有本公司的股份与公司已发行股份的比值计算得到. 以上数据来源于 Execucomp 数据库. CEO 任命董事占比即 CEO 所任命的董事数量与董事会总人数的比值, 而 CEO - 董事任期重叠通过 CEO 与董事会成员任期重叠的平均年数衡量. 数据来源于 Boardex 数据库. 此外, 本文还考虑了股票分析师的影响. 股票分析师作为企业重要的外部监督力量, 为其他外部利益相关者了解企业经营状况进而监督企业提供了重要的借鉴^[27]. 为此, 本文引入了分析师数量和评级两个变量. 分析师数

量通过本年度关注本公司的分析师平均数量衡量.分析师评级通过本年度分析师对本公司的平均评级衡量.原始评级采用五点量表,1代表强烈建议买入(strong buy),5代表强烈建议卖出(strong sell),分数越高代表推荐程度越低.为便于解释,本文通过从6中减去原始评级进行反向编码,以使得分数越高代表推荐程度越高.相关数据来源于I/B/E/S数据库.

本文还控制了CEO的背景特征.其中,CEO性别为虚拟变量,若CEO为男性则取值为1,为女性则取值为0.CEO年龄为连续变量.外部CEO为虚拟变量,若CEO从公司外部继任则取值为1,从公司内部继任则取值为0.过往CEO经历为虚拟变量,若CEO在继任前曾有在其他公司担任CEO的经历则取值为1,反之则为0.内部制定继任人经历为虚拟变量,若CEO在继任前曾作为本公司指定的继任人则取值为1,反之则为0.上述背景信息收集于Execucomp数据库、公司公开信息以及新闻报道等.

此外,本文控制了不同的行业类型,并引入了行业复杂度、行业丰裕度和行业动态性三个行业特征变量.行业虚拟变量通过公司所在行业的四位行业标注代码(Standard Industry Classification Code, SIC)衡量.行业复杂度通过行业销售集中度指数(McNa-mara)衡量.本文将公司所在行业过去5年内($t-4$, t)的销售收入对年份进行回归,回归系数用以衡量行业丰裕度.而行业动态性则通过公司所在行业过去5年内($t-4$, t)行业销售增长的标准差衡量.相关数据来源于Compustat数据库.

2.2 CEO人格变量

CEO大五人格数据通过机器学习算法对美国上市公司季度收益电话会议文本进行分析得到.本文沿用Harrison等^[12]采用的对CEO大五人格的测量方法,基于以下步骤使用机器学习方法对CEO大五人格进行测量:a)文本向量化;b)模型训练及选择;c)人格得分预测.

CEO内部焦点和CEO外部焦点.CEO内部焦点表示CEO对公司内部事务的关注,CEO外部焦点表示CEO对公司外部环境的关注.本文选择

了企业季度收益电话会议文本作为测量数据源.参照先前研究,本文挑选了CEO公开发言环节的文本进行分析^[13].在操作中,本文使用DICTION 7.0软件进行文本分析.DICTION是一个带有内置字典的单词计数软件,亦可通过用户自定义词典实施计数.该软件通过搜索文本中与预设词典一致的单词,计算匹配数量并转化为标准化分数.DICTION已在先前的战略和组织研究中得到广泛应用,被证明内部可靠且外部有效^[41,42].通过使用DICTION 7.0软件,参照Yadav等^[43]开发的词典对企业季度收益电话会议中CEO公开发言进行文本分析,本文得到了CEO空间焦点的得分.

CEO促进焦点和CEO预防焦点.CEO促进焦点表示CEO对收益的敏感性以及对进步和成长的渴望,CEO预防焦点表示CEO对损失的敏感性以及对稳定和安全的渴望.本文通过使用DICTION 7.0软件,参照Gamache等^[34]开发的词典,对企业季度收益电话会议中CEO公开发言部分进行文本分析,得到了CEO调节焦点的得分.

类似地,CEO魅力型领导、CEO自信以及创业导向均通过使用DICTION 7.0软件分析企业季度收益电话会议中CEO公开发言文本测量.衡量CEO魅力型领导的词典由Fanelli等^[39]开发,包括三个维度:否定过去、目标明确以及实现途径,三个维度得分相加得到最终得分.衡量CEO自信的词典由Loughran和McDonald^[44]开发.衡量创业导向的词典由Short等^[45]开发,包括五个维度:进攻性、自主性、创新性、主动性、冒险性,五个维度得分相加得到最终得分.此外,CEO认知复杂性通过使用DICTION 7.0软件分析企业收益电话会议中间答环节测量^[46].衡量CEO认知复杂性的词典由Graf-Vlachy等^[46]开发,包含思维差异性、细微性、比较性三个维度,三个维度得分标准化后取平均值得到最终得分.

2.3 CEO语言变量

借助DICTION 7.0软件对CEO在季度收益电话会议公开发言中的语言表达特征进行分析,本文得到了下述变量,如表2所示.

表 2 CEO 语言变量及其测量

Table 2 CEO language variables and measurement

CEO 语言变量	测量
总分析词数	CEO 发言文本的总词数
CEO 语义独特性	使用独特性术语的频率
CEO 语义数量化	使用数字、数字运算和其他定量术语的频率
CEO 语义犹豫性	使用表示犹豫或不确定的术语的频率
CEO 语义自我性	使用包含整体特征的术语的频率
CEO 语义整体性	使用包含整体特征的术语的频率
CEO 语义集体性	使用忽略个体差异、建立完整感和保证感的术语的频率
CEO 语义绝对性	使用与类别有关的术语的频率
CEO 语义肯定性	使用表示对某个人、团体或抽象实体肯定的术语的频率
CEO 语义积极性	使用与积极的情感状态、娱乐、胜利以及关怀相关的术语的频率
CEO 语义道德性	使用与可取的道德品质、有吸引力的个人品质以及社会和政治理想相关的术语的频率
CEO 语义邪恶性	使用表示社会性不恰当和邪恶的名词,以及表示不幸和意外变迁的形容词的频率
CEO 语义困难性	使用表示自然灾害、敌对行为、可谴责的人类行为以及令人讨厌的政治结果的术语的频率
CEO 语义攻击性	使用描述人类竞争和暴力行动的术语的频率
CEO 语义完成性	使用表达任务完成和有组织的人类行为的术语的频率
CEO 语义沟通性	使用表示沟通交流的术语的频率
CEO 语义认知性	使用表示功能性和想象性的认知过程的术语的频率
CEO 语义消极性	使用表示妥协、顺从、停止、惰性、不感兴趣的术语的频率
CEO 语义普遍性	使用常用介词、代词、助词和连词的频率
CEO 语义空间性	使用表示地理实体、物理距离和测量模式的术语。CEO 语义时间性统计了与时间相关的术语的频率
CEO 语义时间性	使用与时间相关的术语频率
CEO - 现时导向	使用推断现在时态的一般性及抽象性术语的频率
CEO - 过去导向	使用推断过去时态的一般性及抽象性术语的频率
CEO 语义人性化	使用与人及其活动相关的术语的频率
CEO 语义具体性	使用表示有形性和物质性的术语的频率
CEO 语义制度性	使用表示制度规律或关于核心价值的实质性协议的术语的频率
CEO 语义亲和性	使用描述人群之间态度相似性的术语的频率
CEO 语义互动性	使用表示行为互动的术语的频率
CEO 语义多元性	使用描述与规范不同的个人及群体的术语的频率
CEO 语义排他性	使用描述社会孤立的来源和影响的术语的频率
CEO 语义自由性	使用描述个人选择最大化和拒绝社会习俗的术语的频率
CEO 语义负面性	使用标准的否定词汇、否定功能词和指定空集的术语的频率
CEO 语义动态性	使用表示人类运动、物理过程、旅程、速度以及交通方式的术语的频率
CEO 语义重复性	统计了所有出现 3 次及以上的名词或名词衍生形容词,并按照公式[重复出现的词数 × 出现次数]/10 计算得分
CEO 语义修饰性	通过计算形容词与动词的比值得到
CEO 语义词汇多样性	计算段落中不同单词的数量与段落总单词数的比值得到
CEO 语义复杂性	通过计算每个单词的平均字符数得到

3 模型构建及预测效果

3.1 深度学习模型构建

CEO 解聘问题可视为一个基于 CEO 在职近 T 年的相关特征对其面临解聘的可能性进行预测的问题. 模型输入为 CEO 在职近 T 年的结构、人格及语言特征, 而模型输出为 CEO 在 T 年是否被解聘(0/1). 由于 CEO 在职近 T 年的特征可以被视为一种时序型的输入, 本文采用机器学习中常用于处理时序数据的长短时记忆神经网络 (Long Short-Term Memory, LSTM) 对相关特征进行嵌入表示处理. 由于 CEO 被解聘常常与其当年的表现高度相关, 本文考虑在模型中单独引入最近一年的相关特征, 经全连接模型处理后与经 LSTM 处理形成的嵌入表示连接, 再经由全连接模型处理, 进而经 Softmax 层得到模型输出. 具体模型架构如图 1 所示.

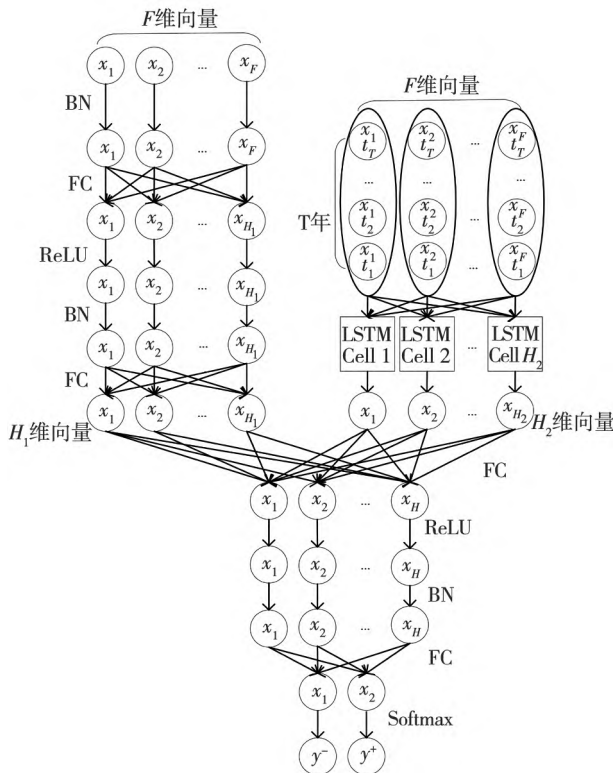


图 1 深度学习模型架构

Fig. 1 Deep learning model structure

其中 LSTM Cell 表示一个典型的 LSTM 单元, 包含输入门 i_t 、遗忘门 f_t 、输出门 o_t 和记忆单元 C_t . 通过拟合式(1)所示的遗忘函数得到 f_t , 其值越

接近于 0 越表示要遗忘, 反之则越是要保留.

$$f_t = \text{sigmoid}(V_f h_{t-1} + W_f a_t + b_f) \quad (1)$$

其中 V_f 和 W_f 为权重矩阵, b_f 为偏置参数, h_{t-1} 表示上一时刻的状态变量, a_t 表示输入数据. f_t 将被作为遗忘系数和上一时刻留下的信息 C_{t-1} 经过 Hadamard 乘积后作为上一时刻遗留表示信息, 与当前时刻的信息 i_t . C'_t 相加后得到该层的信息 C_t , 用来输出到下一时刻, 如式(2) ~ 式(4)所示.

$$i_t = \text{sigmoid}(V_i h_{t-1} + W_i a_t + b_i) \quad (2)$$

$$C'_t = \tanh(V_c h_{t-1} + W_c a_t + b_c) \quad (3)$$

$$C_t = f_t \circ C_{t-1} + i_t \circ C'_t \quad (4)$$

最后计算表示向量 h_t 作为输出, 其表达式如式(5) ~ 式(6)所示.

$$o_t = \text{sigmoid}(V_o h_{t-1} + W_o a_t + b_o) \quad (5)$$

$$h_t = o_t \circ \tanh(C_t) \quad (6)$$

本研究中所采用的全连接模型是深度学习模型中常用的一类网络结构, 其中 BatchNorm 层主要用于对数据进行归一化处理, 加速神经网络收敛速度及稳定性. 具体数学表示如式(7) ~ 式(10)所示.

$$\mu_\beta \leftarrow \frac{1}{m} \sum_{i=1}^m x_i \quad (7)$$

$$\sigma_\beta^2 \leftarrow \sum_{i=1}^m (x_i - \mu_\beta)^2 \quad (8)$$

$$\hat{x}_i \leftarrow \frac{x_i - \mu_\beta}{\sqrt{\sigma_\beta^2 + \epsilon}} \quad (9)$$

$$y_i \leftarrow \gamma \hat{x}_i + \beta \quad (10)$$

其中 μ_β 和 σ_β^2 分别表示每批数据的均值和方差, 而 \hat{x}_i 则表示经规范化处理后的输入, 进而经过缩放及平移得到本层输出 y_i .

FC 层则是一个全连接层, 对前一层的输入进行加权求和, 如式(11)所示.

$$f_{fc} = W_{FH_1} X + b_{H_1} \quad (11)$$

ReLU 层采用如式(12)所示的激活函数, 减少参数间的依存关系, 从而防止模型过拟合.

$$f_{\text{ReLU}} = \max(0, X) \quad (12)$$

3.2 模型训练及预测

研究中将数据划分为 80% 和 20% 两部分, 其中 80% 数据用于对模型进行训练, 而余下 20% 则用于对模型预测效果进行测试. 模型训练过程中采用五折交叉验证的方式在训练集上对模型参数进行充分调优, 参数最优取值如表 3 所示.

表 3 最优参数取值

Table 3 Optimal parameter value

参数	最优取值
T	10
F	88
H_1	32
H_2	32
Learning rate	0.001
Batchsize	64

为衡量模型预测准确度,研究中选取受试者工作特征曲线下面积(AUC)及预测准确度(Accuracy)两种常用指标.其中 Accuracy 衡量的是模型对样本进行正确分类的比例,AUC 则可理解为,对随机选出的两个标签分别为正、负的样本,分类器对前者给出的分数高于后者的概率.由两者的描述可见,若分类问题存在样本不均衡问题,Accuracy 指标可能无法衡量模型分类效果,而 AUC 指标则不受影响.CEO 解聘预测是个二分类问题,解聘样本比例约为 15.8%,即存在类别不平衡问题,因而 AUC 指标可用于更好地衡量模型预测效果.本文同时采用了常用于进行分类预测的三种机器学习模型作为基准模型,分别是逻辑回归模型(LR)及两种集成学习模型,随机森林模型(RF)及梯度提升树(GBDT).四个模型的预测效果如表 4 所示.可见本研究中提出的基于 LSTM 的模型在预测准确性上相较于基准模型具有更优的预测效果.

表 4 模型预测效果

Table 4 Model prediction performance

模型	AUC	Accuracy
LSTM	0.818	0.882
LR	0.655	0.845
RF	0.684	0.856
GBDT	0.712	0.854

为评估对序列数据和最后一年数据进行融合建模的合理性,本文进一步尝试对两类数据进行分别建模.即对图 1 展示的双塔结构的两侧进行拆分,分别为输入仅包含时间序列数据的 LSTM 模型和仅包含最后一年数据的模型,两个模型预测效果及其与全模型的对比如表 5 所示.可见,对两类数据进行分别建模的效果均弱于对两者进行

综合建模的全模型,说明综合考虑时间序列数据及最近年度数据进行建模具有优越性.

表 5 不同年度数据消融实验结果

Table 5 Data ablation results of different years

数据	AUC	Accuracy
仅时间序列数据	0.796	0.859
仅最后一年数据	0.799	0.848
全模型	0.818	0.882

3.3 不同类特征预测能力比较

为进一步评估上文总结的三类主要变量的预测能力,本研究还通过消融实验分别探究了不同类特征及其组合对预测准确率的影响(见表 6).在三类变量中,结构变量对于 CEO 解聘预测作用较大,而 CEO 性格及语言特征在单独使用时预测效果较差.将 CEO 性格或语言特征与结构特征进行融合均取得更优的预测效果,其中语言特征的影响又比性格特征影响更大.将三类特征全部融合取得了最佳预测效果,说明在构建 CEO 解聘的理论时,不应仅聚焦于传统的结构变量,而应充分考虑 CEO 语言及性格特征,及其与传统结构特征间的关系.

表 6 各类型变量预测准确性

Table 6 Prediction performance of various types of variables

变量类型	AUC	Accuracy
结构特征	0.782	0.867
性格特征	0.670	0.721
语言特征	0.683	0.758
结构特征 + 性格特征	0.785	0.872
结构特征 + 语言特征	0.805	0.873
性格特征 + 语言特征	0.699	0.804
全模型	0.818	0.882

4 模型可解释性分析

机器学习算法因为“黑箱”问题而备受诟病,其可解释性近来受到了广泛关注,相关可解释性框架为深入理解输入变量对预测结果的解释力度提供了可能性.本研究希望从上文构建出的预测模型入手,分析各输入特征对 CEO 解聘的影响,从而为形成相关理论创新奠定基础.为全面考虑各特征对于最终预测结果的贡献率,本研究采用

Lundberg 和 Lee^[47] 提出的 SHAP(Shapley Additive exPlanations) 框架计算每个特征对预测的贡献. 其主要原理为基于博弈论而提出的 Shapley 值可用于衡量各特征分别对于模型预测效果的作用^[48], 具有较大的绝对 Shapley 值的特征相对更为重要. 本文对数据中每个特征的绝对 Shapley 值进行平均, 并按照重要性由高到底对特征进行排序, 以展示不同特征对 CEO 解聘预测的重要性, 如图 2 所示.

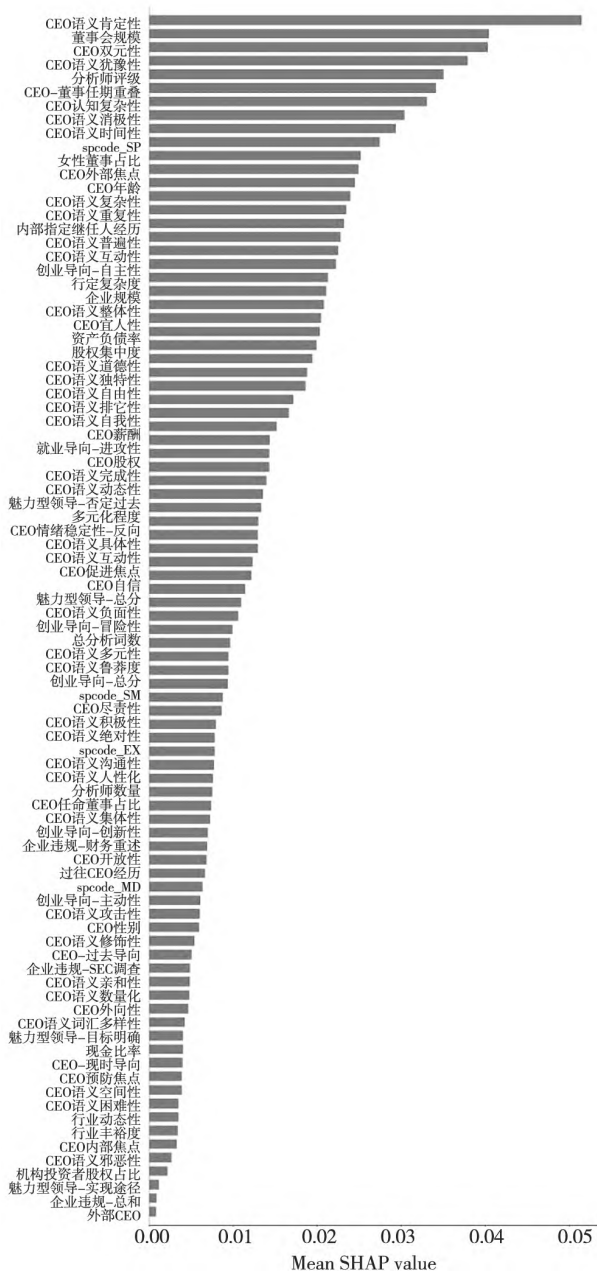


图2 变量预测能力

Fig. 2 Predictive power of different variables

可见, 结构变量中的 CEO 年龄、总收入、CEO

与董事任期重叠时间为影响 CEO 解聘的重要因素, 这与以往研究相一致, 但难以提供新理论构建的必要依据. 而 CEO 的语言使用情况, 如 CEO 使用包含自信和整体性词汇、CEO 使用不恰当及邪恶词汇等方面都会对 CEO 解聘产生重要影响, 超过了许多传统研究中常用的结构化变量. 此外, CEO 性格变量, 如 CEO 开放性、尽责性、宜人性等对预测效果的影响高于部分语言变量. 可见基于机器学习方法, 从多领域、跨模态的大数据中提炼的新变量的可为分析管理问题提供新的视角, 为拓宽原有理论的边界提供了可能性.

为进一步分析上述重要语言变量对 CEO 解聘的影响, 本文进一步分析了历史语言行为变量及当前语言行为变量对 CEO 解聘的影响, 具体如图 3 所示. 历史贡献度及当前贡献度的比例对照表明, 相较于历史语言特征而言, CEO 在当期使用更多犹豫词汇、肯定词汇及消极性词汇会对其当期被解聘产生较大的预测效果. 相比之下, 对于 CEO 语义邪恶性而言, 历史的表现贡献了 CEO 被解聘的绝大部分. 这一发现表明从印象管理视角分析 CEO 语言特征对其解聘的影响颇有潜力^[49]. 机器学习模型做出的细颗粒的发现, 有助于学者思考造成过去和当前言语行为解释力出现显著差别的机理.

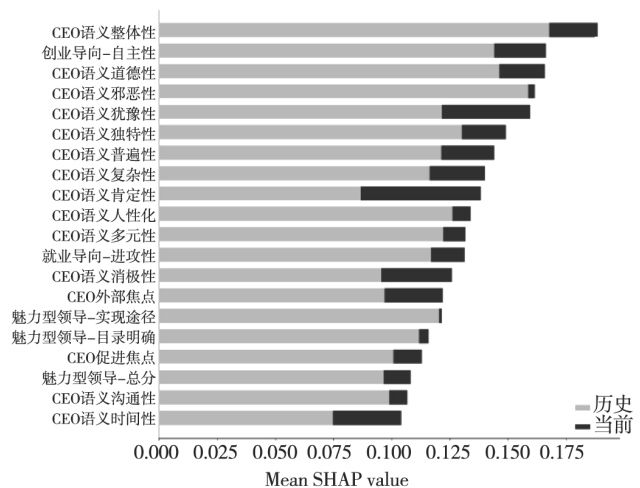


图3 CEO 主要语言特征变量历史贡献及当前贡献对比

Fig. 3 Contribution comparison of CEO's historical and current language features

为进一步探究语言变量对 CEO 解聘影响的方向性, 本文采用 SHAP 框架下的散点图将各样本表示在二维平面上, 其中纵轴表示各变量, 取值

大小由颜色深浅表示,横轴表示 Shapley 值,即变量对模型输出的影响,由此可表示变量取值与其对模型影响的大小,进而可揭示其对模型预测的方向性影响.大部分 CEO 语言特征变量对结果并没有明显的方向性影响,表明不能仅关注此类语言类变量与 CEO 解聘之间的线性关系.这也与现实情况相符,即 CEO 的语言特征与其被解聘之间的关系,受到其所处行业、公司营收状况、公司治理情况等因素的综合影响.因而,在试图基于此类变量构建理论时,应着重思考语言变量与其他变量之间的交互作用对 CEO 解聘的影响.

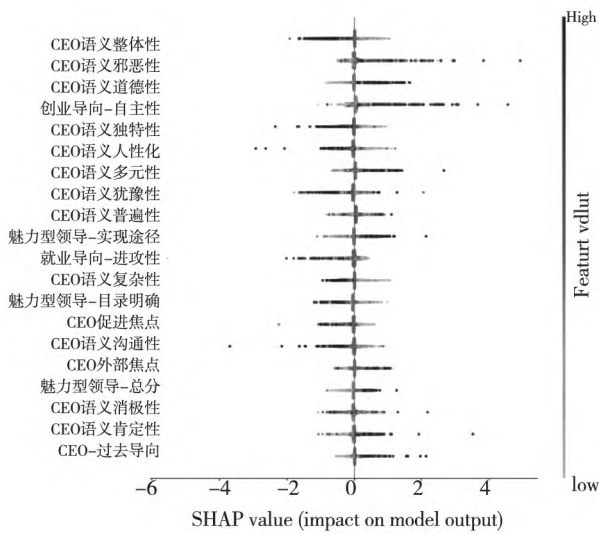


图4 CEO 语言特征对其解聘的影响

Fig. 4 Influence of language features on CEO dismissal

5 结束语

本研究通过机器学习方法对原本难以直接测量的 CEO 行为变量进行了分析,在此基础上构建 CEO 解聘的预测模型,通过可解释性机器学习框架发现预测 CEO 解聘的重要变量.然而,仅通过机器学习方法本身结果的输出无法确切知道这些变量是通过什么途径影响到 CEO 被解聘的概率.有了机器学习的发现,学者可以通过回溯推理的方法搜集定性和定量的数据,揭示某些重要变量与 CEO 被解聘之间的关系.例如,为什么 CEO 的某种语言特征与被解聘之间的关系方向并不明朗?可能的原因在于,数据库并没有记录 CEO 被解聘的原因,可能是工作不称职导致公司业绩持续下滑,也有可能是由于 CEO 背后还有一位强有

力的前任存在.通过回溯推理的方法从事研究或许可以发现,在前一种情况下 CEO 的某类语言特征与被解聘是正向关系,而后一种情况则是负向关系.由此可见,预测型研究可对传统的基于心理测量和计量分析的组织战略领域研究提供必要的补充,并提供了新的视角^[50].学者可以在机器学习发现的基础上,更为精准地明确下一步探索的方向,从而建立有解释力的中层理论.因此,机器学习方法可以为研究者提供新颖而可靠的观察,有助于学者基于数据建立更好的假设,并通过归纳或者溯因的方式建立新的理论^[51, 52].

通过本研究的过程及发现可以预见,人力资源会是大数据分析应用最具前景的领域之一.传统人力资源的选、育、用、留的四大职能在企业组织中还会存在.然而,由于传统企业是由管理者或者相关人员履行这几大职能,而人受到动机和认知的原因很容易出现判断偏差,导致选育留用的精准度不高.尽管领域内学者已经从理论视角上分析了解决该方法,但由于无法全面地获取客观的数据,人力资源的工作没有实质的进展.但是,随着企业数字化系统的完善及大数据采集、存储技术的进步,传统企业人力资源领域的多种数据日益可测可获,可为大数据分析提供丰富的原材料,进而借助机器学习、人工智能等方法解决选育用留中的绝大部分问题.例如,通过分析面试官与申请人的面试记录,结合申请人进入企业后的业绩表现,可以反推面试官评估和决策是否合适.又如,对于人员流转率高的行业,结合以往有关人员离职前因分析的文献及员工在工作中的多维数据,可构建机器学习模型对员工离职可能性及主要影响因素进行预测,进而提前对相关因素进行干预,从而降低人员的离职率.

但大数据分析在人力资源领域发挥价值也存在几个层级的挑战,一是企业能否实现人力资源过程的数字化.只有实现过程的数字化,才能够获得有关人力资源工作的大数据,而目前真正实现过程数字化的企业凤毛麟角;二是已有理论与大数据分析技术的有机融合.即便有了人力资源过程的大数据,仅通过暴力计算找到合理的选育用留模式也是不可取的,是对资源的巨大浪费.借鉴本研究的思路,可以基于已有的文献,更精准地开展大数据分析,从而提升工作效率和精准性.所

以,公司内部从事大数据的人员与人力资源专家之间能否做到相互欣赏和合作,是个重要的组织问题。三是数据价值与隐私保护之间的权衡取舍。随着《数据安全法》、《个人信息保护法》等法律法规的出台,企业应该在多大程度上采集与员工行为有关的数据,并且在多大程度上可以去分析这些数据是相关政策制定者及企业需要回答的重要问题,需要了解大数据在人力资源当中应用,且对于科学和工作伦理有深入思考的专家的持续探索。

本研究回应了早期学者对于人工智能方法的判断。Simon^[53]曾指出,“没有什么能够有效地替代问题解决者,后者在任何特定问题情境下知道实证场景中有效的假定是什么。”从这个意义上说,人工智能只是模仿了人类的认知过程,却不具备人类的灵活性、适应性和生成性。随着人工智能

技术的应用的推进,人与机器之间是替代还是补充的关系成为近来讨论的热点,争论的核心是人工智能以更高效的自动化(automation)方式取代人,还是人与机器互动后获得增强的判断(augmentation)。本文认为,这二者虽然存在冲突,但在组织和管理场景下,自动和增强是相互依赖和加强的。如果说,在工作场景下将很多智慧外包给机器既达不到预定的效果、在道德上也不正确^[54],那么,学者则可以借助机器学习更好地提升自己的专长。计算机科学家和管理学者可以整合各自不同的观点、理论和方法,共同研究人工智能在管理中的应用^[55]。本研究在知识创造或者理论建构的领域,回应了这个重大的争论,也希望能够启迪同行如何利用机器学习的发现来实现对于复杂管理现象的认识。

参 考 文 献:

- [1]陈国青,张 瑾,王 聪,等。“大数据-小数据”问题:以小见大的洞察[J]. 管理世界,2021,37(2):203-213.
Chen Guoqing, Zhang Jin, Wang Cong, et al. The “big data: Small data” problem: Insights for the big through the small [J]. Journal of Management World, 2021, 37(2): 203-213. (in Chinese)
- [2]陈国青,吴 刚,顾远东,等.管理决策情境下大数据驱动的研究和应用挑战——范式转变与研究方向[J]. 管理科学学报,2018,21(7):1-10.
Chen Guoqing, Wu Gang, Gu Yuandong, et al. The challenges for big data driven research and applications in the context of managerial decision-making: Paradigm shift and research directions[J]. Journal of Management Sciences in China, 2018, 21(7): 1-10. (in Chinese)
- [3]Merton R K, Merton R C. Social Theory and Social Structure[M]. New York: Simon and Schuster, 1968.
- [4]陈昭全,张志学,沈 伟.管理研究中的理论建构[M].北京:北京大学出版社,2018.
Chen Zhaoquan, Zhang Zhixue, Shen Wei. Theory Construction in Management Research[M]. Beijing: Peking University Press, 2018. (in Chinese)
- [5]徐淑英,张志学.管理问题与理论建立:开展中国本土管理研究的策略[J].南大商学评论,2005,(4):1-18.
Xu Shuying, Zhang Zhixue. Problems in management and theory construction: Strategies of indigenous management research in China[J]. Nanjing Business Review, 2005, (4): 1-18. (in Chinese)
- [6]Glass G V. Primary, secondary, and meta-analysis of research[J]. Educational Researcher, 1976, 5(10):3-8.
- [7]George G, Haas M R, Pentland A. Big data and management[J]. Academy of Management Journal, 2014, 57(2):321-326.
- [8]George G, Osinga E C, Lavie D, et al. Big data and data science methods for management research[J]. Academy of Management Journal, 2016, 59(5):1493-1507.
- [9]陆 瑶,张叶青,黎 波,等.高管个人特征与公司业绩——基于机器学习的经验证据[J].管理科学学报,2020,23(2):120-140.
Lu Yao, Zhang Yeqing, Li Bo, et al. Managerial individual characteristics and corporate performance: Evidence from a machine learning approach[J]. Journal of Management Sciences in China, 2020, 23(2): 120-140. (in Chinese)
- [10]Gans J S, Goldfarb A, Lederman M. Exit, tweets, and loyalty[J]. American Economic Journal: Microeconomics, 2021, 13(2):68-112.
- [11]Barlow M A, Verhaal J C, Angus R W. Optimal distinctiveness, strategic categorization, and product market entry on the Google Play app platform[J]. Strategic Management Journal, 2019, 40(8):1219-1242.
- [12]Harrison J S, Thurgood G R, Boivie S, et al. Measuring CEO personality: Developing, validating, and testing a linguistic tool[J]. Strategic Management Journal, 2019, 40(8):1316-1330.
- [13]Yi X, Zhang Y A, Windsor D. You are great and I am great (too): Examining new CEOs' social influence behaviors during leadership transition[J]. Academy of Management Journal, 2020, 63(5):1508-1534.

- [14] Guo G. Demystifying variance in performance: A longitudinal multilevel perspective[J]. *Strategic Management Journal*, 2017, 38(6): 1327–1342.
- [15] Wovak A J, Hambrick D C, Henderson A D. Do CEOs encounter within-tenure settling up? A multiperiod perspective on executive pay and dismissal[J]. *Academy of Management Journal*, 2011, 54(4): 719–739.
- [16] Peterson K. Accounting complexity, misreporting, and the consequences of misreporting[J]. *Review of Accounting Studies*, 2012, 17(1): 72–95.
- [17] Arthaud-Day M L, Certo S T, Dalton C M, et al. A changing of the guard: Executive and director turnover following corporate financial restatements[J]. *Academy of Management Journal*, 2006, 49(6): 1119–1136.
- [18] Wiersema M F, Zhang Y. Executive turnover in the stock option backdating wave: The impact of social context[J]. *Strategic Management Journal*, 2013, 34(5): 590–609.
- [19] Renneboog L. Ownership, managerial control and the governance of companies listed on the Brussels stock exchange[J]. *Journal of Banking & Finance*, 2000, 24(12): 1959–1995.
- [20] Chhaochharia V, Kumar A, Niessen-Ruenzi A. Local investors and corporate governance[J]. *Journal of Accounting and Economics*, 2012, 54(1): 42–67.
- [21] Lau J, Sinnadurai P, Wright S. Corporate governance and chief executive officer dismissal following poor performance: Australian evidence[J]. *Accounting & Finance*, 2009, 49(1): 161–182.
- [22] Deutsch Y. The impact of board composition on firms' critical decisions: A meta-analytic review[J]. *Journal of Management*, 2005, 31(3): 424–444.
- [23] Knyazeva A, Knyazeva D, Masulis R W. The supply of corporate directors and board independence[J]. *The Review of Financial Studies*, 2013, 26(6): 1561–1605.
- [24] Zhang Y, Qu H. The impact of CEO succession with gender change on firm performance and successor early departure: Evidence from China's publicly listed companies in 1997–2010[J]. *Academy of Management Journal*, 2016, 59(5): 1845–1868.
- [25] Khanna V, Kim E H, Lu Y. CEO connectedness and corporate fraud[J]. *The Journal of Finance*, 2015, 70(3): 1203–1252.
- [26] Fredrickson J W, Hambrick D C, Baumrin S. A model of CEO dismissal[J]. *Academy of Management Review*, 1988, 13(2): 255–270.
- [27] Wiersema M F, Zhang Y. CEO dismissal: The role of investment analysts[J]. *Strategic Management Journal*, 2011, 32(11): 1161–1182.
- [28] Gupta V K, Mortal S C, Silveri S, et al. You're fired! Gender disparities in CEO dismissal[J]. *Journal of Management*, 2020, 46(4): 560–582.
- [29] Zhang Y. Information asymmetry and the dismissal of newly appointed CEOs: An empirical investigation[J]. *Strategic Management Journal*, 2008, 29(8): 859–872.
- [30] Zhang Y, Rajagopalan N. When the known devil is better than an unknown god: An empirical study of the antecedents and consequences of relay CEO successions[J]. *Academy of Management Journal*, 2004, 47(4): 483–500.
- [31] Graffin S D, Boivie S, Carpenter M A. Examining CEO succession and the role of heuristics in early-stage CEO evaluation[J]. *Strategic Management Journal*, 2013, 34(4): 383–403.
- [32] Chatterjee A, Hambrick D C. It's all about me: Narcissistic chief executive officers and their effects on company strategy and performance[J]. *Administrative Science Quarterly*, 2007, 52(3): 351–386.
- [33] Malhotra S, Reus T H, Zhu P, et al. The acquisitive nature of extraverted CEOs[J]. *Administrative Science Quarterly*, 2018, 63(2): 370–408.
- [34] Gamache D L, McNamara G, Mannor M J, et al. Motivated to acquire? The impact of CEO regulatory focus on firm acquisitions[J]. *Academy of Management Journal*, 2015, 58(4): 1261–1282.
- [35] Herrmann P, Nadkarni S. Managing strategic change: The duality of CEO personality[J]. *Strategic Management Journal*, 2014, 35(9): 1318–1342.
- [36] Simsek Z, Heavey C, Veiga J J. The impact of CEO core self-evaluation on the firm's entrepreneurial orientation[J]. *Strategic Management Journal*, 2010, 31(1): 110–119.
- [37] Tang Y, Mack D Z, Chen G. The differential effects of CEO narcissism and hubris on corporate social responsibility[J]. *Strategic Management Journal*, 2018, 39(5): 1370–1387.
- [38] Ocasio W, Laamanen T, Vaara E. Communication and attention dynamics: An attention-based view of strategic change[J]. *Strategic Management Journal*, 2018, 39(1): 155–167.
- [39] Fanelli A, Misangyi V F, Tosi H L. In charisma we trust: The effects of CEO charismatic visions on securities analysts[J]. *Organization Science*, 2009, 20(6): 1011–1033.
- [40] Park S H, Chung S H, Rajagopalan N. Be careful what you wish for: CEO and analyst firm performance attributions and CEO dismissal[J]. *Strategic Management Journal*, 2021, 42(10): 1880–1908.
- [41] Craig R, Mortensen T, Iyer S. Exploring top management language for signals of possible deception: The words of Satyam'

- s chair Ramalinga Raju[J]. *Journal of Business Ethics*, 2013, 113(2): 333–347.
- [42] Short J C, Palmer T B. The application of DICTION to content analysis research in strategic management[J]. *Organizational Research Methods*, 2008, 11(4): 727–752.
- [43] Yadav M S, Prabhu J C, Chandy R K. Managing the future: CEO attention and innovation outcomes[J]. *Journal of Marketing*, 2007, 71(4): 84–101.
- [44] Loughran T, McDonald B. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks[J]. *The Journal of Finance*, 2011, 66(1): 35–65.
- [45] Short J C, Ketchen Jr D J, Combs J G, et al. Research methods in entrepreneurship: Opportunities and challenges[J]. *Organizational Research Methods*, 2010, 13(1): 6–15.
- [46] Graf-Vlachy L, Bundy J, Hambrick D C. Effects of an advancing tenure on CEO cognitive complexity[J]. *Organization Science*, 2020, 31(4): 936–959.
- [47] Lundberg S M, Lee S I. A unified approach to interpreting model predictions[C]. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017: 4768–4777.
- [48] Shapley L S. *A Value for N-person Games*[M]. Princeton: Princeton University Press, 2016.
- [49] 于李胜, 蓝一阳, 王艳艳. 盛名难副: 明星 CEO 与负面信息隐藏[J]. *管理科学学报*, 2021, 24(5): 70–86.
Yu Lisheng, Lan Yiyang, Wang Yanyan. The curse of reputation: Superstar CEOs and bad news hoarding[J]. *Journal of Management Sciences in China*, 2021, 24(5): 70–86. (in Chinese)
- [50] Yarkoni T, Westfall J. Choosing prediction over explanation in psychology: Lessons from machine learning[J]. *Perspectives on Psychological Science*, 2017, 12(6): 1100–1122.
- [51] Choudhury P, Allen R T, Endres M G. Machine learning for pattern discovery in management research[J]. *Strategic Management Journal*, 2021, 42(1): 30–57.
- [52] Shrestha Y R, He V F, Puranam P, et al. Algorithm supported induction for building theory: How can we use prediction models to theorize? [J]. *Organization Science*, 2021, 32(3): 856–880.
- [53] Simon H A. The structure of ill-structured problems[J]. *Artificial Intelligence*, 1973, 4(3–4): 181–201.
- [54] Von Krogh G. Artificial intelligence in organizations: New opportunities for phenomenon-based theorizing[J]. *Academy of Management Discoveries*, 2018, 4(4): 404–409.
- [55] Raisch S, Krakowski S. Artificial intelligence and management: The automation-augmentation paradox[J]. *Academy of Management Review*, 2021, 46(1): 192–210.

A new paradigm of management research in the big data era: Taking CEO dismissal as an example

WANG Cong, YI Xi-wei, ZHANG Zhi-xue

Guanghua School of Management, Peking University, Beijing 100871, China

Abstract: Given the complexity of managerial phenomena, current research tends to explain specific phenomena from a certain theoretical perspective. It is challenging for researchers to build a middle-range theory with adequate explanatory power based on potentially contradicting results to assess the relative significance of multiple influencing elements, because they might lack a holistic perspective. This paper proposes that analyzing big data through machine learning could enable scholars to identify factors most related to the specific phenomenon, therefore establishing a holistic view of the various theoretical perspectives. This study takes an important topic in corporate governance, CEO dismissal, as an example. Based on a thorough review of existing literature, structural variables identified by prior research as well as a large-scale CEO text corpus during their tenure are collected. Three different sorts of factors impacting CEO dismissal are identified, by applying machine learning method, and their individual and collective predictive power are analyzed. Our research findings not only support the current trend that emphasizes the individual heterogeneity of CEOs but also lay a solid foundation for scholars to establish an integrated theory on CEO dismissal. This study sheds light on the advantages and prospects of utilizing machine learning as a new empirical method.

Key words: big data; management research; middle-range theory; machine learning; CEO dismissal