

DOI: 10.16538/j.cnki.fem.20240409.101

战略管理研究中的机器学习: 研究述评与展望

吴建祖, 郑朝杰

(兰州大学 管理学院, 甘肃 兰州 730030)

摘要: 机器学习为战略管理研究带来了新的方法论和机遇。本文基于1999—2022年间发表在国际顶尖战略管理期刊的27篇文献,系统梳理了机器学习在战略管理研究中的四个应用:变量测量、特征选择、模型估计和因果推断。本文分析了每个维度的代表性文献,详述了其所用机器学习算法的原理和应用,并探讨了未来的研究方向。研究发现,自然语言处理技术和分类算法不仅提高了变量测量精度,也增强了特征选择方面的能力,从而加深了研究者对战略管理研究对象的理解。同时,随机森林、支持向量机等机器学习算法,在提升模型估计的稳健性方面有显著效果。另外,双重套索、双重机器学习和广义随机森林等方法,为因果关系推断提供了技术支持。本文旨在为运用机器学习技术深化战略管理研究提供指导和建议,同时为战略管理领域的未来研究奠定了方法论基础。

关键词: 机器学习;战略管理;方法综述;未来展望

中图分类号: F270 **文献标识码:** A **文章编号:** 1001-4950(2025)03-0119-18

一、引言

机器学习引领战略管理研究进入数据驱动的新纪元,其卓越的数据处理能力不仅拓展了数据来源、创新了建模方法,也显著提升了研究结果的稳健性。这些前沿进展共同推动了研究者对复杂商业与管理现象的深层理解和精准洞察,为战略管理研究注入了新的理论活力和实践指导价值。机器学习(machine learning)是指一系列能从数据中学习并通过经验改进性能的模式(Mitchell, 1997)。本文关注的机器学习涵盖从基础算法到复杂技术的广泛领域,包括监督学习、无监督学习、强化学习以及深度学习等多个分支。这些技术在战略管理研究中的应用广泛,覆盖了变量测量、模式识别、因果推断等多个方面(Choi等, 2021; Choudhury等, 2021a; Choudhury等, 2019; Miric等, 2023)。本文采用广义的机器学习定义,旨在囊括所有相关算法与技术,全面分析机器学习方法在战略管理研究的不同维度。

机器学习在揭示数据模式和优化决策过程方面展现出卓越的能力(Krogh等, 2023),其优

收稿日期: 2024-01-17

基金项目: 国家自然科学基金项目(71972094)

作者简介: 吴建祖(1970—),男,兰州大学管理学院教授,博士生导师(通信作者, jzww@lzu.edu.cn);

郑朝杰(1996—),男,兰州大学管理学院博士研究生。

势体现在处理、分类、预测和检验高维大规模数据方面(Ghoddusi等,2019)。因此,机器学习在财务会计(Ding等,2020)、市场营销(Chen等,2017)、运营管理(Hoehn等,2011)等多个管理学研究领域得到广泛应用。

尽管已有研究探讨了机器学习在管理学中的应用,但本文在战略管理领域的应用研究上更为专注且深入。Ghoddusi等(2019)对机器学习在能源经济研究中的应用进行了批判性回顾,Ngai和Wu(2022)通过文献综述为市场营销研究中机器学习的应用提供了清晰的概念框架。这些研究为机器学习在特定管理学子领域的应用提供了宝贵的见解和分析框架,但在战略管理研究中的应用亟待进一步探讨。

最近,机器学习在战略管理研究中引起了广泛关注。企业如何在其复杂的内外部环境制定、实施战略并提升绩效这一核心议题的研究方式正在经历深刻变革。战略管理研究涵盖竞争策略、创新管理、资源分配、国际化战略及企业社会责任等多个方面,涉及组织决策的各个层面。机器学习技术在处理、分类、预测和检验高维大规模数据方面的优势(Ghoddusi等,2019)为研究者挖掘更为深入的洞见提供了新的技术(Choudhury等,2021a),如变量测量、模式识别和模型估计等方面(Choi等,2021;Choudhury等,2021a;Choudhury等,2019;Miric等,2023)。这些工具和方法不仅展示了机器学习为战略管理研究提供新视角和探索可能性的巨大潜力,也标志着战略管理研究方法论的一次深刻的变革。

本文聚焦于探讨机器学习技术和算法在战略管理研究中的具体应用,而非对战略管理研究主题的全面分析。Choudhury等(2021a)的研究在这方面提供了有价值的参考,特别是在特征选择方面的应用,为使用该工具的研究者提供了实用的指导。由于特征选择是本文的核心内容之一,Choudhury等(2021a)的研究可视为对本综述的重要补充。此外,刘景江等(2023)的研究探讨了机器学习如何赋能管理学研究,并提出了在工商管理 and 会计财务领域使用机器学习的关键策略。与这些研究不同,本文的目标并非全面覆盖管理学的所有领域,而是通过全面审视机器学习在战略管理研究中的应用,深入其在该领域的发展潜力。

机器学习弥补了传统战略管理研究方法的不足,并为解释复杂战略管理研究问题提供了新视角。首先,机器学习在处理海量非结构化数据(洪永森和汪寿阳,2021)以及特征选择方面具有独特优势(Choudhury等,2021a),使战略管理学者能够发现新兴变量并探索更稳健的潜在模式。其次,机器学习在构建和检验假设方面的灵活性,使其能够确定最适合的函数形式(Mullainathan和Spiess,2017),并提供更准确的模型估计(Chou等,2023)。最后,机器学习在处理数据异质性方面的能力,使其在因果推断方面也展现出显著优势(Athey和Imbens,2019)。

尽管机器学习在战略管理研究中的应用具有巨大潜力,关于其应用的系统性回顾仍然不足。本文旨在填补这一空白,专注于机器学习在战略管理研究中的应用,深入理解机器学习在战略管理研究中的应用价值与潜力。我们期望通过这种专注的探讨,促进机器学习在战略管理研究中被更广泛采纳,为该领域提供有力的方法论支持。

本文旨在提供如下贡献:首先,识别并阐述机器学习在战略管理研究中应用的四个核心维度:变量测量、特征选择、模型估计与因果推断,并构建一个综合性技术参考框架。该框架详细阐述了各维度的内容、关联以及如何共同构成战略管理实证研究的基础。其次,分析机器学习方法在战略管理研究中的实际运用,包括其途径、原理以及在特定研究场景中的应用,为战略管理学者提供全面的指导原则,促进对这些方法的深入理解和应用。最后,比较战略管理研究的限制与机器学习的技术优势,讨论机器学习在未来战略管理研究中的关键应用方向,为突破现有研究限制、推进研究深化提供新的视角和路径。

二、文献检索与分析

本章旨在阐述本文采用的文献检索与分析方法,以确保对机器学习在战略管理研究中应用的全面理解。我们采取了多步骤、系统性的方法来选择和评估相关文献,以保证研究的质量和广度。

(一)期刊与时间区间选择

首先,我们集中关注在学术界公认的10本国际权威战略管理期刊上发表的论文。这些期刊由于其在战略管理研究中的重要性和对机器学习应用研究的影响力而被选定。我们参考了陈冬梅等(2020)的研究,以确保选取的期刊具有较高的学术标准和专业性(如表1所示)。

表1 文献评估来源

文献评估来源类型	文献评估来源期刊
战略管理学术期刊	<i>Strategic Management Journal</i>
	<i>Strategy Science</i>
	<i>Organization Science</i>
	<i>Academy of Management Journal</i>
	<i>Academy of Management Review</i>
	<i>Administrative Science Quarterly</i>
	<i>Journal of International Business Studies</i>
	<i>Journal of Management</i>
	<i>Journal of Management Studies</i>
	<i>Management Science</i>

考虑到机器学习方法最早的实质性应用文章发表于1999年(刘景江等,2023),因此,本研究将样本时间范围确定为1999—2022年。这一时间区间的选择旨在捕捉机器学习在战略管理研究中应用的历史发展轨迹和最新趋势。

(二)关键词和搜索策略

接下来,我们设计了关键词搜索策略。关键词选取基于对机器学习定义、实证分析,以及理论阐述的深入理解,确保覆盖机器学习在战略管理研究的多个方面。我们运用了“Machine Learning”“Supervised Learning”“Unsupervised Learning”“Nearest Neighbor”“Support Vector Machine”“Random Forest”等关键词,并根据搜索结果不断调整和扩展这些关键词,以确保文献的全面性(如表2所示)。

表2 检索关键词表

机器学习关键词
Machine Learning, Supervised Learning, Unsupervised Learning, Nearest Neighbor, Clustering, Support Vector Machine, Random Forest, Classification Tree, Deep Learning, Neural Networks, Convolutional Neural Network, Wavelet-Based Neural Networks, Artificial Neural Network, Recurrent Neural Network, Long-Short Term Memory, Ensemble Methods, Radial Basis Function Network, Kernel-based Extreme Learning, Feed-Forward Deep Network, Genetic Algorithm, Particle Swarm Optimization, Agent-Based Algorithmic Learning, Ensemble Empirical Mode Decomposition, Data Fluctuation Network, Soft Computing, Simulated-Based Neural Network, AI, artificial intelligence, Natural Language Processing, NLP, Chatbot, Machine Vision, ML, Recommendation Engine

(三)文献筛选过程

基于前文确定的10本国际权威战略管理期刊及机器学习关键词,在Web of science数据库中检索了2023年以前的所有文献,共得到118篇文献。为确保研究的高相关性,两位研究员独立审阅了文献的摘要、关键词和标题。通过对文献的摘要和正文部分的仔细阅读,我们筛选出

68篇使用机器学习方法作为研究方法的文献样本。进一步地,我们筛选出与战略管理话题相关并排除方法讨论的文章。经过比对和汇总,首轮检索得到21篇文献。为减少遗漏文献,本研究还通过手动交叉引用的方式,对样本文献的参考文献进行了逆向追踪。最终确定的样本共包含27篇英文文献,样本筛选的具体流程如图1所示。

经过上述严格的筛选和评估过程,我们汇集了1999—2022年间发表的具有代表性的27篇文献。此过程确保了所选文献的质量和相关性,为深入理解机器学习在战略管理研究中的应用提供了坚实基础。尽管我们的搜索可能会忽略一些发表在非主流战略管理期刊的论文,鉴于我们的目标是对多数论文的方法进行审查,而非提供统计分析,因此这一限制并不会影响我们的主要结果。

(四)文献编码与分析

文献编码与分析旨在确立研究框架并提炼关键维度。在编码的第一步,参考以往理论和实证研究结果(刘景江等,2023;Ngai和Wu,2022),确定了战略管理研究中机器学习方法的应用维度。我们认为,机器学习在战略管理研究中的应用可以被理解为一个由四个相互依赖的维度构成的有机整体:变量测量、特征选择、模型估计以及因果推断(如表3所示)。

表3 机器学习在战略管理研究中应用

维度	描述	途径	算法
变量测量	量化和描述数据特征的过程,确保观测值准确性和一致性,为后续分析奠定基础	文本挖掘 预测建模	Word2Vec, SVM, GloVe, FastText, LDA, ANN, NB, KNN, DT, SVM, RF
特征选择	从数据集中识别并保留对预测目标最相关的属性,以提高模型效率和性能	主题提取 特征权重评估 深度学习	LASSO, GBM, SVM, RF, SVM, PCA, LSTM, LDA, LSA, DTM, CTM
模型估计	基于选定特征,使用算法对数据进行拟合,推断模型参数的过程	非线性关系建模 数据质量优化 双重套索	RF, KNN, K-means
因果推断	分析变量间因果关系的方法,旨在确定一个变量是否直接影响另一个变量	双重机器学习 通用机器学习 广义随机森林	Double ML, Double Lasso, DT, Gradient Boosting, RF, ANN, RF

注:因篇幅所限,表中所列算法的全称及其对应的中文名称作为附录备索,下同。

首先,变量测量为战略管理研究提供了量化基础,确保研究对象的特征和行为能以数值形式表达,为后续分析和检验奠定基础。其次,特征选择通过识别关键因素优化数据结构,提高模型估计的效率和准确性。再次,模型估计基于筛选出的关键特征揭示变量之间的内在联系及其

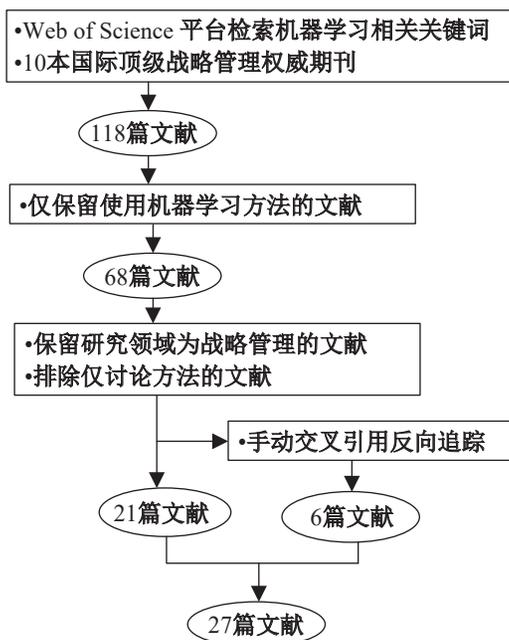


图1 机器学习方法在战略管理研究中应用文献筛选过程

对结果变量的影响,提高因果推断的可靠性。最后,因果推断验证模型估计的准确性,并深化理论构建,明确不同战略选择对组织绩效的效应。综上,这四个维度相互关联,形成紧密的逻辑链条,推动战略管理研究向更深层次的理论构建与实践应用发展。

在编码的第二步,由两位研究者分别阅读文献的摘要和正文,依据第一步确定的四个应用维度,对筛选出的文献进行编码。为确保研究的客观性,编码人员积极讨论,力求达成一致的观点,以最大限度地避免编码者主观选择对结果的偏差产生影响。

在编码的第三步,我们在深入阅读文献的基础上,精心梳理出文献涉及的机器学习技术、算法与结论。基于这些分析,本文从变量测量、特征选择、模型估计和因果推断这四个应用维度出发,系统阐述在战略管理研究中应用机器学习的具体途径,并对未来战略管理研究的机会进行深入探讨。

三、变量测量

运用机器学习处理非结构化数据生成代理变量已成为战略管理研究的主流(Gamache等, 2014; Boudt和Thewissen, 2019; Eklund和Mannor, 2021; Menon等, 2018)。战略管理研究中的历史数据通常样本量较小且为定性数据(Jones和Khanna, 2006),这使得数据应用成为一大挑战。同时,战略管理研究强调对高层管理者的关注,但正如Hambrick(2007)所述,获得高层管理者数据十分困难,因为“它需要与大量高管进行深入接触……而他们通常不愿意接受学术性的审视和探究”。在此背景下,机器学习方法为传统战略管理研究提供了新视角和工具,使研究者能够基于历史数据量化变量,并从多个非结构化数据源中提取有价值的信息(Kaplan和Vakili, 2015)。

根据编码分析结果,文本挖掘和预测建模被识别为战略管理研究中,使用机器学习进行变量测量的两种主要途径(如表4所示)。这些途径能够通过深入分析如电话会议纪要、社交媒体互动和致股东的信等丰富的文本数据和历史数据,揭露关键的管理信息,并展现出远超传统研究方法的效能和精度。接下来,我们将具体阐述这两种途径如何在量化变量的同时推动战略管理研究领域的发展边界。

表 4 用机器学习方法进行变量测量

途径	原理	算法	应用示例
文本挖掘	通过对非结构化的文本数据(如财报、新闻、社交媒体等)进行主题模型、词向量和情感分析等,可以提取出反映企业战略导向、利益相关者等变量量化指标	Word2Vec, GloVe, FastText, LDA	Guzman和Li(2022); Harrison等(2019); Kaplan和Vakili(2015); Bellstam等(2021); Choi等(2021) Choudhury和
预测建模	运用有监督的机器学习算法,基于已标记数据获连续数据构建模型,以识别规则并准确预测未标记数据的类别	ANN, NB, KNN, DT, SVM, RF	Kim(2019); Ng和Sherman(2022); Harrison等(2022)

(一)文本挖掘

基于文本挖掘的变量测量依托自然语言处理技术(NLP),能够将大量非结构化的文本数据转化为用于战略分析和决策支持的结构化信息,从而为战略管理研究提供了一种创新的量化方法和视角(Haans, 2019; Kaplan和Vakili, 2015)。NLP通过自动化和深入分析公司报告、新闻报道、社交媒体及其他公开资料,为处理和分析大规模文本数据集提供了高效的方法(Choudhury

等,2019;Choi等,2021)。具体而言,NLP能够通过主题建模和词向量等算法和技术,精准提取文本的语义和结构信息,将非结构化文本转化为可量化的变量。这一能力极大地增强了战略管理研究的分析能力,为揭示复杂的管理和决策问题提供了新的分析工具。

战略管理研究中,基于词向量和主题模型的测量策略在提取文本的语义内容和主题结构方面发挥着至关重要的作用。这两种策略在实际应用中通常互为补充,共同推动了对文本数据的语义挖掘和主题识别流程的改进和优化。因此,本文将详细探讨这两种策略在变量测量中的作用,凸显其在推进战略管理研究方法论发展中的关键贡献。

1. 基于词向量的测量策略

词向量技术由Mikolov等(2013a)首次提出,已成为探索大规模文本数据中词语共现关系的重要工具。该技术通过分析文本数据,学习词汇间的深层语义联系,将词语表示为高维空间中的向量。这些向量的距离和方向揭示了词语的语义相似度,为研究者提供了量化工具,探究变量词与其他词之间的语义关系。在战略管理研究中,这种工具意味着能够精确地捕捉并量化关键概念,如“可持续性”“社会责任”等语义

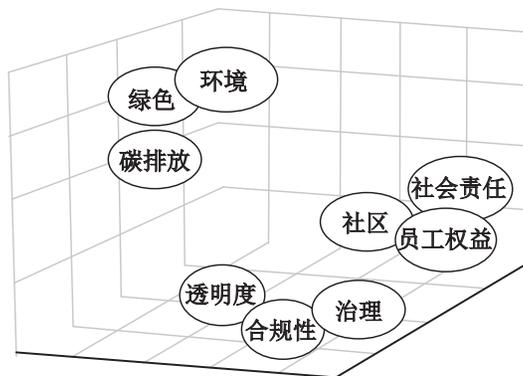


图2 词向量空间

关联,衡量变量词与其他词之间的语义联系。例如,如图2所示,词向量空间中,语义相近的词汇如“绿色”“环境”和“碳排放”聚集在一起,而“社会责任”“社区”和“员工权益”形成另一聚类,显示词向量技术在捕捉语义关联上的精确性。

词向量技术已成为战略管理研究中量化变量的重要工具(Aceves和Evans,2023),如领导力感知和CEO特征等(Bhatia等,2022;Harrison等,2019)。词向量技术在提升战略管理实证研究的测量准确性方面起着关键作用,例如,Guzman和Li(2022)运用词向量技术评估创业公司的战略差异化,探讨其对绩效的预测作用,而Choudhury等(2021b)结合Word2Vec等机器学习模型,利用词向量技术预测员工情绪。Harrison等(2019)则开发了基于词向量的CEO大五人格特质测量方法,研究CEO人格特质与公司绩效及战略变革的关联。

2. 基于主题模型的测量策略

战略管理研究中,主题模型同样是量化变量的重要工具。正如Kaplan和Vakili(2015)所指出,主题模型能够系统地揭示文本集中的主题分布,能够深刻反映出文档的核心内容。主题建模尤其适用于分析复杂的组织环境和战略配置。例如,通过分析企业年报和市场报告中的主题,研究人员能够识别出企业的战略议题、市场趋势以及潜在的竞争优势等关键因素。

Latent Dirichlet Allocation(LDA)是一种由Blei等(2003)提出的主题模型算法,已广泛应用于战略管理研究中。传统方法多聚焦于定量数据分析,如财务报表和市场数据。LDA则允许研究者探索超越传统定量数据的领域,特别是在评估公司创新方面。例如,Kaplan和Vakili(2015)利用LDA确定专利文本中的主导技术分组,通过主题分析,捕捉技术演变趋势,为技术创新研究提供了新的视角。同样,Bellstam等(2021)对703家公司的665 714份分析师报告进行LDA文本分析,提取关键主题并重新定义了企业创新水平。此外,Choi等(2021)运用LDA测量企业多元化水平,将传统人工判断与机器学习模型相结合,全面评估产品、地理位置等多维度的多样性。

(二)预测建模

有监督的机器学习算法是量化变量和预测建模的核心,在处理标记过的数据集时尤其重要(Gaber等,2007)。例如通过分类算法,模型被训练以识别和预测文本中的特定特征或类别,为战略管理研究提供了精确的文本分类和标注手段,这对量化变量至关重要。例如,准确地识别情绪倾向、主题分类或是对特定事件的关注,并将这些内容转化为可量化的数值变量。

战略管理研究广泛采用了多种分类算法以测量变量,包括人工神经网络(ANN)、朴素贝叶斯(NB)、K近邻(KNN)、决策树(DT)以及支持向量机(SVM)等。(1)ANN,特别是反向传播(BP)神经网络,因其卓越的非线性拟合能力和分类性能,在变量测量(如文本分类)中获得了广泛应用。然而,ANN面临着对大量数据和计算资源的依赖,以及模型可解释性方面的局限性。(2)NB算法源于Maron和Kuhns(1960)的概率论原理,尽管在小规模数据集和文本分类上表现优异(Domingos和Pazzani,1997;Ng和Jordan,2001),但受到条件独立假设的限制,难以处理复杂的变量测量。(3)KNN算法(Cover和Hart,1967),通过考虑最近邻居进行类别预测,在多类别变量测量中表现良好,但可能在处理大型数据集时面临计算复杂度的挑战。(4)DT算法(Breiman等,1984),因其强大的解释性和在高维特征及不平衡类别分布中的适用性而受到认可(Kotsiantis,2013)。虽然存在过拟合风险,但通过适当的剪枝策略可以控制。(5)SVM算法(Cortes和Vapnik,1995),在处理高维度和非线性问题方面表现出色,特别适用于文本分类(Joachims,1998),但其训练过程可能较为耗时。(6)集成分类算法,如随机森林和AdaBoost,通过整合多个分类器来优化性能,在变量测量中考虑关键变量的贡献。

现有研究揭示了分类算法在战略管理研究中的潜力,特别是在量化文本数据方面。例如,Nauhaus等(2021)基于分类算法和NLP,对文本中的专家情绪进行分析,并选择了精度最高、召回率最佳的分类器(NB),发现机器学习的准确率比传统词典方法高达30%。Harrison等(2022)则采用分类算法(RF)扩展之前基于人工编码数据的方法,强调分类算法在生成关于CEO重视度和董事会主席倾向的大规模数据集的应用价值。Miric等(2021)基于文本数据,运用分类算法(NB)成功区分了平台型公司与非平台型公司,并对其并购策略进行了对比分析。Choudhury和Kim(2019)通过NLP分析发明家姓名数据,构建了与种族相关的分类变量。Ng和Sherman(2022)将分类算法应用于用户的职位描述,创建可靠且可管理的职位角色分类法。

四、特征选择

在战略管理研究中,特征选择是构建模型的关键步骤。特征选择涉及从复杂数据集中筛选出对模型最重要的特征,从而揭示驱动组织战略和绩效的关键因素。此外,在处理多维信息数据集时,特征选择能够剔除冗余和不相关的数据,并提高模型的预测准确性。这一过程为构建高效的预测模型奠定了基础,促进了对影响组织战略决策的核心因素的深入理解。

特征选择包括三种主要途径:主题提取、特征权重评估以及深度学习(参见表5)。首先,在处理文本数据时,主题提取为揭示文本中的主题和相互关联提供了有力支持。其次,特征权重评估在消除数据冗余方面发挥关键作用,确保模型的简洁性和有效性。最后,深度学习专注于挖掘数据中隐藏的规律和结构,为理解复杂的战略管理问题提供新的视角。下面进一步探讨这三种途径,并将这些方式确定为特征选择在战略管理研究中的关键组成。

(一)主题提取

战略管理研究中,有效提取和理解大量文本数据中的信息对于揭示企业环境、竞争态势和战略选择至关重要。在LDA中,文档代表输入数据的样本,主题表示不同的类别集合,而词汇则用于表示样本的特征,这些元素之间存在深刻的关系。LDA通过分析文档中词汇的出现模

式,学习得到主题—词汇分布和文档—主题分布的信息,为战略管理研究中的主题提取任务提供了强大支持。

表 5 用机器学习方法进行特征选择

途径	原理	算法	应用示例
主题提取	通过分析文本数据的潜在结构和模式,以自动识别和提取文本中的关键主题和特征	LDA,LSA,DTM,CTM	Joseph等(2020);Bao和Datta(2014) Giardili等(2022); Sen和Puranam(2022);
特征权重评估	利用机器学习模型本身的特性来评估特征的重要性。例如,Lasso回归通过L1正则化来实现稀疏化,自动将不重要的特征系数压缩为0	LASSO,GBM,SVM,RF,PCA,LSTM	He等(2020);He等(2020);Tidhar和Eisenhardt(2020); Sen和Puranam(2022)
深度学习	运用机器学习算法归纳方式自动检测和学习隐藏在庞大数据集中的复杂模式和规律	SVM,PCA,LSTM	He等(2020),Tidhar和Eisenhardt(2020); Sen和Puranam(2022)

主题模型在战略管理研究领域的广泛运用凸显了其在主题提取方面的核心价值。学者利用主题模型探索文本数据中的主题结构,从而构建更为精确的预测模型。例如,Joseph等(2020)运用了包括LDA在内的多种机器学习算法来提取主题,并进一步运用所得数据集探究董事会领导层、公司绩效以及首席执行官解雇之间的相互关系。Bao和Datta(2014)则基于LDA技术,研究了风险披露对投资者风险感知的影响,并开发了一种改进的LDA算法变体,能够同时识别和量化文本中的风险类型。此外,主题模型为战略管理研究提供了深入理解特定议题的新途径。它们不仅促进了学术研究的深度和广度,还为学术综述构建了一个有力的分析框架,进一步丰富和完善了战略管理学的知识体系。以邹波等(2024)的研究为例,他们通过基于LDA主题模型的方法提炼出20个主题,并结合人工编码,围绕“生成性”这一概念确定了其内涵、成因、作用机制及产出结果等四个核心研究内容,进行了深入的概念性评述。

(二)特征权重评估

战略管理研究中,提高模型的预测准确性至关重要,是增强模型泛化能力的核心环节。基于特征权重评估的特征选择旨在识别对模型预测贡献最大的变量,同时排除那些冗余或不相关的因素。这种方法在处理大规模数据集时,不仅能够提高模型的效率和精度,而且是确保预测准确性和防止过拟合的关键策略。

战略管理研究中,各种机器学习算法展现出其在模型特征选择上的独特优势:(1)套索回归在战略管理研究中的应用得到了广泛认可(Belloni等,2014;Guzman和Stern 2020;Miric等,2019)。该算法通过引入L1正则化实现稀疏化的模型参数,能够有效剔除冗余变量,进而避免过拟合。(2)随机森林(RF)算法通过特征重要性分数,基于节点分裂频率或对模型准确性的贡献识别关键变量。(3)梯度提升机(GBM)则侧重通过多个树模型中变量的频繁使用来衡量其重要性。(4)支持向量机(SVM)在变量选择方面的特点是分析支持向量的权重,较大的权重通常意味着该变量对分类边界至关重要。(5)深度学习不直接进行变量选择,而是通过层级学习和特征融合的方式,自动学习并提取特征,为不同的输入特征赋予不同的重要性。

战略管理研究中,机器学习通过特征权重评估为研究者进行特征选择提供了新的工具,丰富了研究者在探索和解释驱动企业战略因素的实证手段。例如,Giardili等(2022)通过套索回归选择最相关的变量来预测生产变化,通过正则化过程,使模型中一些系数趋向于零,有效排除了不相关或次要的变量。这种方法不仅简化了模型结构,还增强了模型的解释性,从而更准确地理解管理变更对生产的影响。Sen和Puranam(2022)基于套索回归,检查所选择的预测因

素与附加交易可能性之间的正向或负向关联。而He等(2020)将套索回归与神经网络结合,使用神经网络作为套索回归的预处理,以检测和解释社区治理纠纷数据中的模式问题。

(三)深度学习

基于深度学习的模式识别为战略管理研究者提供了一种从大量复杂数据中提取有价值信息的手段。基于深度学习的特征选择核心在于其归纳性质,这在某种程度上类似于传统定性研究中的扎根理论方法。该技术的基本原理主要体现在其能够通过数据驱动的方法,自动检测和学习隐藏在大规模数据集中的复杂模式和规律。

战略管理研究中,机器学习能够帮助研究者识别和预测潜在的企业行为模式:(1)DT算法通过递归地分割数据集,揭示了特定战略情境下对企业行为起决定性作用的关键特征。(2)RF算法通过集成多个决策树,增强了模型的预测能力和对异常值的鲁棒性,这对于在不确定性高的战略环境中做出决策尤为重要。(3)SVM算法通过寻找最优分类面来处理战略管理中的分类问题,尤其是在处理高维数据和非线性问题时表现出色,有助于揭示复杂的战略决策模式。(4)ANN,尤其是深度学习模型,能够自动从数据中学习特征表示,对于处理非结构化数据如文本和图像,在战略分析中识别潜在的模式和关联具有独特的优势。(5)主成分分析(PCA)等降维技术通过减少数据的维度来揭示数据中的主要趋势和模式,使得战略管理研究者能够更容易地理解复杂的数据结构。而长短期记忆网络(LSTM)等序列模型在处理时间序列数据方面表现出色,能够识别和预测战略管理中的动态变化模式。

机器学习算法在揭示复杂战略管理问题方面具有优势,明显区别于传统计量经济学演绎式检验。例如,Chou等(2023)提供了机器学习在模式识别方面比传统方法更有效的证据。机器学习进行模式识别应被视为传统计量经济学模型估计的补充(Mullainathan和Spiess, 2017; Choudhury等, 2021a)。He等(2020)探究了社区如何解决治理纠纷问题,通过收集有关软件许可决策的讨论数据,并结合机器学习和定性方法来解释数据中的模式。这一研究为理论构建提供了新视角,强调了机器学习在战略管理研究中的价值。Tidhar和Eisenhardt(2020)采用App Store上的数据,结合多案例方法和大规模机器学习分析,验证和扩展了这些案例研究的发现。Sen和Puranam(2022)使用机器学习技术记录稳健的经验模式,溯因性地建立理论。他们使用了29个可能解释附加交易采用的潜在因素来构建理论,为学者进行测试理论提供了示范。

五、模型估计

战略管理实证研究中,机器学习作为一种新途径,是传统统计方法的补充。Hofman等(2017)解释了基于监督机器学习的预测建模如何补充回归建模,以进行理论解释和模型估计。机器学习技术与传统计量经济学方法的深度融合将是未来管理研究的新趋势(陈冬梅等, 2020; Athey和Imbens, 2019)。相较于传统的统计经济学方法,机器学习在处理大规模数据方面展现出更高的灵活性和效率。一方面,非线性模型以其灵活性和对复杂数据结构的适应能力,为战略管理研究提供了深入分析和预测变量的可能性。另一方面,高质量和完整的数据确保模型估计的准确性,使得分析结果更加稳健。因此,基于这两个维度的讨论,不仅有助于提升模型的估计效果,也为战略管理的研究和实践提供了重要的视角和工具(参见表6)。

(一)非线性关系建模

非线性模型是输出与输入之间的关系不能通过线性方程简单描述的模型。与线性模型相比,非线性模型能够更好地捕捉数据中的复杂模式和动态变化,是理解和预测战略管理研究中非线性关系的有力工具。机器学习方法特别适用于探索变量间的非线性关系,尤其是在传统统计技术难以处理的高维数据集中。

表6 用机器学习方法进行模型估计

途径	原理	算法	应用示例
非线性关系建模	通过构建一个非线性函数关系估计非线性关系,通过优化算法估计模型参数,从而使得模型预测值与实际观测值之间的差异最小化	XGBoost	Hasan等(2022)
数据质量优化	通过异常值识别和缺失值插补等技术,以确保模型训练和预测的准确性	RF, KNN, K-means	Hasan等(2022); Jones(2017); Chou等(2023)

机器学习提供了多种用于建立和估计非线性模型的算法,如DT、SVM、ANN等。这些算法能够通过从数据中自动学习到的复杂模式来预测或分类数据,从而揭示战略变量之间的深层次关系。DT通过递归地分割数据集,构建树状模型拟合非线性关系。SVM通过寻找不同类别之间的最优边界,处理高维数据,并且捕捉复杂的非线性关系。ANN通过模拟人脑神经网络的结构和功能,能够学习到数据之间复杂的非线性关系和模式。ANN由多个层次组成,每一层包含多个神经元,通过激活函数实现非线性映射。

战略管理研究中,战略变量之间存在复杂的非线性关系,上述算法可以灵活拟合特征与目标之间的非线性函数。例如,DT可用于刻画企业战略在不同市场环境下的差异化绩效表现,SVM可用于预测企业信用评级,ANN可用于拟合CEO特质、战略决策与企业绩效之间的复杂映射。Hasan等(2022)的研究通过应用XGBoost算法,展现了在处理高维数据结构和识别变量间非线性关系上的优越性能,尤其是在预测品牌资本与未来股价崩盘风险之间关联性方面取得了创新性进展。

(二)数据质量优化

模型估计过程中,机器学习凭借其在处理复杂数据方面的优势,在异常值检测及缺失数据插补这两个关键领域为战略管理研究提供了有效策略。这些技术增强了模型估计的准确性和研究结果的完整性。异常值识别通过精确识别并处理数据中的异常值,确保了模型估计的稳定性和研究结果的可信度;缺失值插补则通过对缺失数据的有效估计,保障了分析结果的完整性。两种技术的综合应用,体现了机器学习在解决战略管理研究中复杂问题上的关键作用。

1. 异常值识别

战略管理研究中,机器学习为识别历史数据中的异常值提供了新的途径。历史数据可能包含非典型的观测值,这些异常值对于建立准确的模型至关重要,可能揭示关键现象数据收集中的错误。

机器学习算法结合Z分数等手段与传统的模型估计,使得数据分布的符合性检验更为全面和灵活。通过度量数据点相对于其均值的偏离程度能够评估数据是否符合特定的统计分布假设,并为异常模式的识别提供可靠的统计依据,因此可以为战略管理研究中的数据分析和统计推断增添新的价值。

2. 缺失值插补

战略管理研究中,处理数据中的缺失值是确保分析结果有效性的关键步骤。不当的缺失值处理可能会造成偏差,影响模型的准确性和研究结论的可靠性。机器学习在这一领域展现出显著优势,通过捕捉数据中的复杂模式和变量间的微妙关系,提供精细的缺失值插补策略。与传统的统计插补技术(如均值或中位数插补)相比,机器学习算法能够利用数据的整体结构和变量间的相互作用,为缺失值提供更为合理的估计。

尽管机器学习技术在模型估计应用上仍处于早期阶段,但在缺失值处理和异常值检测方

面已经展现出显著潜力。例如,Chou等(2023)的研究证明了RF算法在分析复杂数据集和挖掘潜在模式方面的强大功能,为战略管理理论的构建和实证验证开辟了新的途径。而Jones(2017)的研究表明,Boosting模型在提高预测准确率方面具有显著优势,不仅有效地弥补了传统OLS及Logit回归在处理复杂战略决策变量时的局限性,还为理解复杂战略决策的影响因素提供了更深层次的洞察。Boosting模型通过逐步优化的方式,能够有效地识别和处理数据中的异常值,减少它们对模型预测性能的负面影响(Jones, 2017)。这些研究成果体现了机器学习技术在提升战略管理研究方法精度和理论深度方面的巨大潜力,为未来的研究提供了结合高级统计方法与机器学习算法的分析范式,推动战略管理研究的理论创新与实证研究深度。

六、因果推断

因果推断在战略管理研究中处于核心地位,尽管传统统计方法奠定了方法论基础,但在处理高维数据、复杂交互作用以及内生性问题时有其局限性。随着机器学习技术的兴起,因果推断的方法论得到显著提升。洪永森和汪寿阳(2021)指出,机器学习通过精细化评估反事实和处理效应,开辟了因果估计的新视角。研究者开始运用机器学习方法,如DT、RF、GBM和ANN,更精确地估计因果效应,构建优化的匹配样本(Rathje和Katila, 2021),探索实验中的异质性处理效应(Miric和Jeppesen, 2020),从而深化对因果关系的理解。

现有研究为优化因果估计提供了一系列前沿解决方案,包括双重套索、双重机器学习、通用机器学习以及广义随机森林等(Athey等, 2019)。本部分将专注于探讨这些技术在战略管理研究中的应用,并揭示它们在实现因果推断中的作用(参见表7)。

表7 用机器学习方法进行因果推断

途径	原理	算法	应用示例
双重套索	通过两阶段Lasso回归分别筛选处理效应相关的控制变量和结果变量密切相关的控制变量,以降低模型选择偏差,提高因果估计的精确性和可靠性	Double Lasso	Bach等(2021); Feng等(2020)
双重机器学习	结合机器学习建立预测模型和去偏误的统计估计,克服大规模数据集下传统统计方法的模型选择偏差和过度拟合问题	Double Lasso, GBoost, DML	Bach等(2021); Feng等(2020); Yang等(2020)
通用机器学习	整合多种机器学习方法,识别非线性关系和交互效应,提升因果推断的准确性和可靠性	RF, DT, ANN, GML	Chernozhukov等(2018)
广义随机森林	通过递归分裂寻找处理效应在不同子群体间差异最大的分裂点,每个叶节点上的平均处理效应就反映了该子群体的异质性效应	RF, GF	Athey等(2016); Cabras和Tena(2023); Lee等(2022)

(一)双重套索

双重套索(double lasso)是处理高维数据环境下因果推断问题的有效工具(Belloni等, 2014),为战略管理研究提供了因果估计的手段。基于双重套索选择因子并解决遗漏变量偏差问题以改进因果推断的做法已经成为学者们的共识(刘景江等, 2023)。它通过两阶段正则化线性回归过程,筛选出关键变量并估计其对策略决策的因果效应,同时控制遗漏变量偏差。双重套索的应用为战略管理领域的因果推断提供了一种新的分析框架,有助于研究者在复杂的数据环境中做出更加科学的决策和推断。

战略管理研究中,双重套索为处理复杂问题提供了强大的工具。例如, Bach等(2021)使用双重套索方法度量了性别工资差距的异质性,估计数据集中每个女性的性别工资差距,即每个

女性与具有相同社会经济特征的男性之间的相对薪资损失。这一方法的优势在于能够考虑调查参与者之间的社会经济特征差异,包括可用变量之间的双向交互作用。双重套索方法有效缓解了大量回归系数带来的问题,特别是由社会经济特征的差异引起的问题,进而提高因果估计的精确性和可靠性。Feng等(2020)也采用了双重套索方法,系统评估了新因素对资产定价的贡献,有效地缓解了由于遗漏变量而引起的模型选择错误,提高了定价模型的准确性和可靠性。

(二)双重机器学习

战略管理研究中,因果推断的精确性对于揭示企业战略决策与绩效之间的真实关系至关重要。Chernozhukov等(2016)提出双重机器学习(double machine learning, DML)方法,结合了机器学习的预测能力和传统统计学的因果推断理论,通过两个阶段的学习过程来提高估计因果效应的准确性。

双重机器学习已广泛应用于解决运营和财务领域中的因果估计。以Yang等(2020)的研究为例,他们验证了Chernozhukov等(2016)提出的双重机器学习方法的稳健性,并应用梯度增强的双机器学习方法检验了会计文献中的审计质量效应。随后,Alley等(2022)提出了一种针对分类问题的双重机器学习方法,用于混杂情况下估计具有异质处理效果的需求。他们以二级票务市场门票的交易与定价为背景,研究通过消除门票和市场特征的处理效应推断价格对结果的因果影响。在战略管理研究中,双重机器学习在提升因果推断估计的精准性方面同样具有价值。鉴于战略管理研究常涉及多重影响因素及其相互作用,双重机器学习有效地克服了传统统计方法的局限。例如,Bach等(2021)运用双重机器学习方法量化了多种因素对性别工资差异的贡献,为政策制定和战略管理提供了数据支持。该研究凸显了双重机器学习在揭示复杂经济和社会现象中因果关系的能力,为战略管理研究带来了新的洞察。

(三)通用机器学习

通用机器学习(general machine learning, GML)是一种先进的分析工具,通过融合多种机器学习算法,为战略管理研究中的复杂数据分析提供了新的解决方案。Chernozhukov等(2018)指出,GML在处理高维数据集和揭示处理效果的异质性方面表现出显著优势,这对于战略管理研究尤为重要。这是由于企业战略的实施效果往往因多种内外部因素而呈现异质性,GML能够有效地识别和学习这些复杂数据中的模式和规律。

GML的核心优势在于整合多种算法,提高对战略决策影响因素的预测准确性和可靠性。与依赖特定模型设定和前提条件的传统统计经济学方法相比,GML的灵活性和自适应能力使其能够揭示数据中的复杂结构,提供更深入的见解。在评估诸如国际化、产品创新或企业并购等特定战略决策的效果时,GML可以处理大量多渠道数据,识别影响战略决策的关键因素,为战略管理实证研究提供了新的方法论工具。

(四)广义随机森林

广义随机森林(generalized random forests, GRF),作为一种灵活的非参数方法,在战略管理研究中的应用日益受到重视。GRF适用于多种预测问题,包括回归、分类,以及因果推断(Wager和Athey, 2018; Athey等, 2016),并在适应复杂数据结构和精确估计变量间交互效应方面的核心优势。它在处理高维数据和估计存在潜在混杂因素时的处理效应方面表现出强大的预测能力。在战略管理研究中,GRF能够精确估计个体或企业对于特定战略举措的反应,这对于理解战略决策的效果至关重要。

GRF已被广泛应用于解决营销和运营领域中的因果估计。例如,Tian等(2023)采用从TikTok收集的独特数据集,基于GRF的因果推断方法,估计网红受欢迎程度与视频观看次数的因果关系。Zhang和Luo(2023)也证明了因果森林模型可以作为传统倾向评分方法的合适替代

方案。他们使用GRF促进对最具信息量的预测因子的参数解释,建立了消费者发布的照片与餐厅生存的因果关系。

GRF在因果推断中的应用显著提升了分析的准确性和可靠性。例如, Lee等(2022)讨论了如何使用GRF为领导力效应的预测和因果模型提供信息,并建议将GRF和实验设计相结合,通过引入新近开发的技术来分离“异质处理效应”来得出因果推断。Cabras和Tena(2023)基于足球数据,提出了一种基于对机构对运营决策反应的直接观察来识别隐性组织激励的方法。他们使用GRF估计不同裁判行为的影响,以进一步提升采用深度学习进行因果推断的稳健性。

尽管机器学习方法在因果推断中展现了巨大的潜力,但它们并不能完全替代传统计量经济学的深厚基础和精准的因果推断能力。一个高水平的研究策略是将这两种方法相结合,借助计量经济学中的因果推断理论与机器学习算法在处理数据异质性方面的高效能力,实现相互补充(刘景江等,2023)。例如, Bargagli-Stoffi和Gnecco(2019)和Wang等(2022)在其研究中成功运用了工具变量因果树、工具变量诚实因果树、工具变量树和工具变量森林等模型,展示了机器学习与计量经济学结合在战略管理研究中进行因果推断的巨大潜力。此种跨学科的方法论协作,不仅为理解复杂的经济现象提供了新的视角,也为制定基于数据驱动的、更为精确的战略决策提供了坚实的支撑。

七、未来研究展望

机器学习技术使社会科学研究正处于一个拐点(Athey, 2019)。与此同时,机器学习技术也正在重塑战略管理的研究范式。接下来,本文借助前文形成的技术分析框架,深入讨论机器学习技术在促进战略管理研究中的潜力,特别是如何借此机会深化战略管理的理论基础和实证研究方法。

(一)基于多模态数据的变量测量

机器学习技术为战略管理学者分析多模态数据提供了新机遇,使学者能够综合利用文本、图像及视频等数据,以获取组织行为和战略决策的深入见解。通过对文本的逻辑和语义分析,以及对图像数据中的场景和视觉元素进行精确解析,研究者可以更全面地理解企业战略、品牌形象和市场趋势(Choudhury等,2019)。这种基于机器学习的多模态数据的分析方法,对于深化对组织战略配置的理解至关重要。

在中国市场环境下,机器学习技术为战略管理研究带来了前所未有的机遇。鉴于中国市场的高度复杂性和不断变化的动态特性,研究者需要采纳更加精细和多维的分析手段以适应其需求。机器学习能够处理和分析庞大且多样化数据集,为揭示中国企业如何调整战略以应对全球经济波动和市场多元化挑战提供了深刻见解。机器学习的应用不仅拓展了战略管理理论的深度与广度,而且为中国市场乃至全球市场的战略管理研究与实践提供了新的工具。

(二)特征选择中的模拟预测

战略管理研究中,机器学习推动的模拟预测方法正在成为不可或缺的技术,极大地增强了对组织未来走向的预测能力。这些方法已经在多个领域证明了其有效性,包括提高生产效率、加强网络安全、优化金融投资策略、精确期权定价,以及预测制造业中的潜在故障(Giardili等2022; Ban等2018; Kumar等2022; Dutta 2022; Ciocan和Mistic 2022),展现出其广泛的应用潜力。在战略管理的研究中,这一技术不仅使得以更高精度预测组织发展趋势成为可能,而且丰富了对组织动态的理解。

特别在中国这一特殊市场背景下,基于机器学习的模拟预测技术对战略管理研究的贡献

尤为显著。该技术能够准确地识别对战略决策影响最大的关键因素,并考虑复杂的相互作用关系,为战略管理的决策模拟提供了一种更接近现实的方法。这种方法为分析不同战略选择及其可能带来的后果提供了新的视角。例如,通过模拟不同市场条件下的企业行为,研究者可以更有效地预见竞争动态和市场变化如何影响企业的战略布局,从而为企业提供基于数据的决策支持,加强其市场竞争力。模拟预测技术在中国市场环境下的应用,为理解和预测企业战略提供了强大工具,推进了战略管理理论和实践的发展。

(三)模型估计与因果推断

战略管理研究中,模型估计与因果推断的准确性是确保决策科学和有效的关键。随着数据驱动决策模式的兴起,机器学习技术在特征选择和模型估计方面的应用,为提升模型准确性和因果推断的严谨性提供了新解决方案。传统统计方法与机器学习的结合,尤其是将计量经济学的因果推断理论与机器学习在处理数据异质性方面的优势相融合,已成为推动该领域发展的新趋势(刘景江等2023)。这种跨学科的融合不仅丰富了研究的深度和广度,也极大提高了研究结果的可靠性和实际应用价值。

在中国这一充满活力且日益复杂的市场环境中,机器学习技术的引入,特别是在模型估计和因果推断方面的应用,为研究者提供了新途径。例如,在评估政策变动对企业战略的影响时,机器学习通过分析大量的企业行为数据,能够揭示政策变化对企业战略调整的直接和间接效应。此外,结合自然实验或工具变量方法,机器学习的预测模型能够更准确地估计政策变动的局部和全局效应,为政策制定者和企业管理者提供更为科学的决策支持。这些应用不仅增强了模型的解释力,也显著提升了因果推断的准确性,展现了在中国情境下战略管理理论与实践研究的未来机会。

(四)战略管理理论的构建、泛化与检验

机器学习技术在战略管理理论的构建、泛化与检验中扮演着日益重要的角色(Sen和Puranam, 2022; Tidhar和Eisenhardt, 2020)。通过其强大的数据处理和模式识别能力,机器学习不仅揭示了已知特征间的关系,还能够挖掘出尚未被传统研究所发现的新特征关系。这种能力对于提升理论的综合性和适用性具有重大意义。机器学习的应用使得研究者能够从大量数据中提炼出新的理论概念,从而推动理论的发展和创新。此外,机器学习技术在量化验证方面的应用,为战略管理理论提供了更为严格的检验,使得理论的普适性和局限性能在更广泛的数据集上得到测试(Choudhury等, 2021a)。

中国快速变化的市场结构和政府政策为企业战略提供了丰富的研究素材。机器学习可以帮助研究者深入分析这些因素如何影响企业战略的制定与推进,从而在战略管理理论中引入新的视角。此外,机器学习在识别文化距离对战略实施效果影响方面的应用,有助于跨文化的战略管理理论中填补关键变量,丰富理论的内涵。这些基于机器学习的理论和模型创新,不仅为战略管理的理论体系注入了新的活力,也为中国企业的战略管理实践探索提供了新的指导和洞见。

八、结 论

随着机器学习技术的快速发展,战略管理研究领域正在经历一场深刻的方法论变革。本文深入探讨了机器学习技术在战略管理研究中的广泛应用,揭示了其深远影响并开辟了全新的研究机遇。我们发现,机器学习技术在变量测量方面已成为一种关键工具。通过运用自然语言处理和分类算法,研究者能够更准确地量化关键变量,从而提升研究的精度和可信度;在特征

选择方面,机器学习为理论创新提供了新的机会。通过主题模型、有监督的机器学习以及深度学习,深化了对战略管理研究中复杂现象的理解和分析能力,揭示了未被充分认识的模式和关系;在模型估计方面,基于数据质量与完整性的模型估计为战略管理学者提供了更为稳健的检验工具,增强了研究结论的稳健性;在因果推断方面,机器学习方法,如双重套索、双重机器学习和广义随机森林,为因果关系的精确推断提供了有力支持。这些方法不仅拓展了战略管理研究的深度和广度,也为未来研究提供了新的思路和方法论。

主要参考文献

- [1]陈冬梅,王俐珍,陈安霓.数字化与战略管理理论——回顾、挑战与展望[J].*管理世界*,2020,36(5):220-236.
- [2]洪永森,汪寿阳.大数据、机器学习与统计学:挑战与机遇[J].*计量经济学报*,2021,1(1):17-35.
- [3]刘景江,郑畅然,洪永森.机器学习如何赋能管理学研究?——国内外前沿综述和未来展望[J].*管理世界*,2023,39(9):191-216.
- [4]Aceves, P, Evans, J A. Mobilizing Conceptual Spaces: How Word Embedding Models Can Inform Measurement and Theory Within Organization Science[J]. *Organization Science*, 2023, <https://doi.org/10.1287/orsc.2023.1686>.
- [5]Alley, M, Biggs, M, Hariss, R, et al. Pricing for Heterogeneous Products: Analytics for Ticket Reselling[J]. *Manufacturing & Service Operations Management*, 2022, 25(2): 409-426.
- [6]Athey, S. The Impact of Machine Learning on Economics[M]. *The Economics of Artificial Intelligence*. University of Chicago Press, 2019.
- [7]Athey, S, Imbens, G W. Machine Learning Methods That Economists Should Know About[J]. *Annual Review of Economics*, 2019, 11(1): 685-725.
- [8]Athey, S, Tibshirani, J, Wager, S. Generalized random forests[J]. *The Annals of Statistics*, 2019, 47(2): 1148-1178.
- [9]Belloni, A, Chernozhukov, V, Hansen, C. Inference on Treatment Effects after Selection among High-Dimensional Controls[J]. *The Review of Economic Studies*, 2014, 81(2): 608-650.
- [10]Bellstam, G, Bhagat, S, Cookson, J A. A Text-Based Analysis of Corporate Innovation[J]. *Management Science*, 2021, 67(7): 4004-4031.
- [11]Bhatia, S, Olivola, C Y, Bhatia, N, et al. Predicting leadership perception with large-scale natural language data[J]. *The Leadership Quarterly*, 2022, 33(5): 101535.
- [12]Blei, D, NG, A, Jordan, M. Latent Dirichlet Allocation[J]. *Journal of Machine Learning Research*, 2003, 14: 601-608.
- [13]Boudt, K, Thewissen, J. Jockeying for Position in CEO Letters: Impression Management and Sentiment Analytics[J]. *Financial Management*, 2019, 48(1): 77-115.
- [14]Breiman, L, Friedman, J H, Olshen, R A. et al. Classification and Regression Trees[J]. *Biometrics*, 1984, <https://doi.org/10.1201/9781315139470>.
- [15]Chen, Y, Iyengar, R, Iyengar, G. Modeling Multimodal Continuous Heterogeneity in Conjoint Analysis - A Sparse Learning Approach[J]. *Marketing Science*, 2017, 36(1): 140-156.
- [16]Chernozhukov, V, Chetverikov, D, Demirer, M, et al. Double machine learning for treatment and causal parameters, 2016, <https://doi.org/10.48550/arXiv.1608.00060>.
- [17]Chernozhukov, V, Demirer, M, Duflo, E, et al. Generic Machine Learning Inference on Heterogeneous Treatment Effects in Randomized Experiments, with an Application to Immunization in India[R]. Working Paper, 2018.
- [18]Choi, J, Menon, A, Tabakovic, H. Using machine learning to revisit the diversification-performance relationship[J]. *Strategic Management Journal*, 2021, 42(9): 1632-1661.
- [19]Chou, Y-C, Chuang, H H C, Chou, P-F, et al. Supervised machine learning for theory building and testing: Opportunities in operations management[J]. *Journal of Operations Management*, 2023, 69(4): 643-675.
- [20]Choudhury, P, Allen, R, Endres, M. Machine learning for pattern discovery in management research[J]. *Strategic Management*

Journal, 2021, a,42(1): 30-57.

- [21]Choudhury, P, Foroughi, C, Larson, B. Work-from-anywhere: The productivity effects of geographic flexibility[J]. *Strategic Management Journal*, 2021, b,42(4): 655-683.
- [22]Choudhury, P, Kim, D Y. The ethnic migrant inventor effect: Codification and recombination of knowledge across borders[J]. *Strategic Management Journal*, 2019, 40(2): 203-229.
- [23]Choudhury, P, Wang, D, Carlson, N, et al. Machine Learning Approaches to Facial and Text Analysis: Discovering CEO Oral Communication Styles[J]. *Strategic Management Journal*, 2019, 40(11): 1705-1732.
- [24]Ciocan, D F & Mistic, V V. Interpretable Optimal Stopping[J]. *Management Science*, 2022, 68(3): 1616-1638.
- [25]Cortes, C, Vapnik, V. Support-vector networks[J]. *Machine Learning*, 1995, 20: 273-297.
- [26]Cover, T, Hart, P. Nearest neighbor pattern classification[J]. *IEEE Transactions on Information Theory*, 1967, 13(1): 21-27.
- [27]Ding, K, Lev, B, Peng, X, et al. Machine learning improves accounting estimates: evidence from insurance payments[J]. *Review of Accounting Studies*, 2020, 25(3): 1098-1134.
- [28]Domingos, P, Pazzani, M. On the Optimality of the Simple Bayesian Classifier under Zero-One Loss[J]. *Machine Learning*, 1997, 29(2): 103-130.
- [29]Dutta, H A Consensus Algorithm for Linear Support Vector Machines[J]. *Management Science*, 2022, 68 (5): 3703-3725.
- [30]Eklund, J C, Mannor, M J. Keep Your Eye on the Ball or on the Field? Exploring the Performance Implications of Executive Strategic Attention[J]. *Academy of Management Journal*, 2021, 64(6): 1685-1713.
- [31]Feng, G, Giglio, S, Xiu, D. Taming the Factor Zoo: A Test of New Factors[J]. *The Journal of Finance*, 2020, 75(3): 1327-1370.
- [32]Gaber, M M, Zaslavsky, A, Krishnaswamy, S. A Survey of Classification Methods in Data Streams[M]. *Data Streams. Advances in Database Systems*, Springer, Boston, MA, 2007.
- [33]Gamache, D, Mcnamara, G, Mannor, M J, et al. Motivated to acquire? The impact of ceo regulatory focus on firm acquisitions[J]. *Academy of Management Journal*, 2014, 58(4): 1261-1282.
- [34]Ghoddusi, H, Creamer, G G, Rafizadeh, N. Machine learning in energy economics and finance: A review[J]. *Energy Economics*, 2019, 81: 709-727.
- [35]Giardili, S, Ramdas, K, Williams, J W Leadership and Productivity: A Study of US Automobile Assembly Plants[J]. *Management Science*, 2022, 69 (3): 1500-1517.
- [36]Guzman, J, Li, A. Measuring Founding Strategy[J]. *Management Science*, 2022, 69(1): 101-118.
- [37]Guzman, J, Stern, S. The State of American Entrepreneurship: New Estimates of the Quality and Quantity of Entrepreneurship for 32 US States, 1988-2014. Working Paper, 2020.
- [38]Haans, R. What's the value of being different when everyone is? The effects of distinctiveness on performance in homogeneous versus heterogeneous categories[J]. *Strategic Management Journal*, 2019, 40(1): 3-27.
- [39]Hambrick, D C Upper echelons theory: An update[J]. *Academy of Management Review*, 2007, 32 (2): 334-343.
- [40]Harrison, J, Josefy, M, Kalm, M, et al. Using supervised machine learning to scale human-coded data: A method and dataset in the board leadership context[J]. *Strategic Management Journal*, 2022, 44(7): 1780-1802.
- [41]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.
- [42]Hasan, M M, Taylor, G, Richardson, G. Brand Capital and Stock Price Crash Risk[J]. *Management Science*, 2022, 68(10): 7221-7247.
- [43]He, V, Puranam, P, Shrestha, Y, et al. Resolving governance disputes in communities: A study of software license decisions[J]. *Strategic Management Journal*, 2020, 41(10): 1837-1868.
- [44]Hoehn, W, Koenig, F G, Moehring, R H, et al. Integrated Sequencing and Scheduling in Coil Coating[J]. *Management Science*, 2011, 57(4): 647-666.
- [45]Hofman, J M, Sharma, A, Watts, D J. Prediction and explanation in social systems[J]. *Science*, 2017, 355(6324): 486-488.

- [46]Joachims, T. Text categorization with Support Vector Machines: Learning with many relevant features. In Proceedings of the ECML. Chemnitz, Germany, 1998, 137-142.
- [47]Jones, G, Khanna, T. Bringing history (back) into international business[J]. *Journal of International Business Studies*, 2006, 37(4): 453-468.
- [48]Jones, S. Corporate bankruptcy prediction: a high dimensional analysis[J]. *Review of Accounting Studies*, 2017, 22(3): 1366-1422.
- [49]Joseph, J, Borland, H, Orlitzky, M, et al. Seeing Versus Doing: How Businesses Manage Tensions in Pursuit of Sustainability[J]. *Journal of Business Ethics*, 2020, 164(2): 349-370.
- [50]Kaplan, S, Vakili, K. The double-edged sword of recombination in breakthrough innovation[J]. *Strategic Management Journal*, 2015, 36(10): 1435-1457.
- [51]Kotsiantis, S B. Decision trees: a recent overview[J]. *Artificial Intelligence Review*, 2013, 39(4): 261-283.
- [52]Krogh, G, Roberson, Q, Gruber, M. Recognizing and Utilizing Novel Research Opportunities with Artificial Intelligence[J]. *Academy of Management Journal*, 2023, 66(2): 367-373.
- [53]Lee, A L, Inceoglu, I, Hauser, O, et al. Determining causal relationships in leadership research using Machine Learning: The powerful synergy of experiments and data science[J]. *Leadership Quarterly*, 2022, 33(5): 101426.
- [54]Maron, M E, Kuhns, J L. On Relevance, Probabilistic Indexing and Information Retrieval[J]. *Journal of the Acm*, 1960, 7(3): 216-244.
- [55]Menon, A, Nave, G, bhatia, S. Emotional Expressions Predict Risky Decisions by S&P 500 Executives. SSRN Electronic Journal, 2018, 10.2139/ssrn. 3303000.
- [56]Miric, M, Boudreau, K J, Jeppesen, L, B. Protecting their digital assets: The use of formal & informal appropriability strategies by App developers[J]. *Research Policy*, 2019, 48(8): 103738.
- [57]Miric, M, Jeppesen, L B. Does piracy lead to product abandonment or stimulate new product development?: Evidence from mobileplatform-baseddeveloper firms[J]. *Strategic Management Journal*, 2020, 41(12): 2155-2184.
- [58]Miric, M, Jia, N, Huang, K. Using Supervised Machine Learning for Large-scale Classification in Management Research: The Case of Identifying Artificial Intelligence Patents[J]. *Strategic Management Journal*, 2023, 44(2): 491-519.
- [59]Miric, M, Pagani, M, Sawy, O. When and Who Do Platform Companies Acquire? Understanding the Role of Acquisitions in the Growth of Platform Companies[J]. *MIS Quarterly*, 2021, 45(4): 2159-2174.
- [60]Mitchell, T M. Machine Learning, McGraw-Hill, Inc, 1997.
- [61]Mullainathan, S, Spiess, J. Machine Learning: An Applied Econometric Approach[J]. *Journal of Economic Perspectives*, 2017, 31(2): 87-106.
- [62]Nauhaus, S, Luger, J, Raisch, S. Strategic Decision Making in the Digital Age: Expert Sentiment and Corporate Capital Allocation[J]. *Journal of Management Studies*, 2021, 58(7): 1933-1961.
- [63]Ng, A, Jordan, M I. On Discriminative vs. Generative Classifiers: A comparison of logistic regression and naive Bayes[J]. *Neural Information Processing Systems*, 2001, 2: 841-848.
- [64]Ng, W, Sherman, E L. In Search of Inspiration: External Mobility and the Emergence of Technology Intrapreneurs[J]. *Organization Science*, 2022, 33(6): 2300-2321.
- [65]Ngai, E W T, Wu, Y Y. Machine learning in marketing: A literature review, conceptual framework, and research agenda[J]. *Journal of Business Research*, 2022, 145(2): 35-48.
- [66]Rathje, J M, Katila, R. Enabling Technologies and the Role of Private Firms: A Machine Learning Matching Analysis[J]. *Strategy Science*, 2021, 6(1): 5-21.
- [67]Sen, P, Puranam, P. Do Alliance Portfolios Encourage or Impede New Business Practice Adoption? Theory and Evidence from the Private Equity Industry[J]. *Strategic Management Journal*, 2022, 43(11): 2279-2312.
- [68]Tidhar, R, Eisenhardt, K. Get Rich or Die Trying... Finding Revenue Model fit Using Machine Learning and Multiple Cases[J]. *Strategic Management Journal*, 2020, 41(7): 1245-1273.

- [69]Wager, S, Athey, S. Estimation and Inference of Heterogeneous Treatment Effects using Random Forests[J]. *Journal of the American Statistical Association*, 2018, 113(523): 1228-1242.
- [70]Yang, J-C, Chuang, H-C, Kuan, C-M. Double machine learning with gradient boosting and its application to the Big N audit quality effect[J]. *Journal of Econometrics*, 2020, 216(1): 268-283.

Machine Learning in Strategic Management Research: A Review and Prospects

Wu Jianzu, Zheng Chaojie

(School of Management, Lanzhou University, Lanzhou 730030, China)

Summary: Machine learning brings new methodologies and opportunities to strategic management research. Based on 27 papers published in top international strategic management journals between 1999 and 2022, this paper systematically summarizes four dimensions of machine learning applications in strategic management research: variable measurement, feature selection, model estimation, and causal inference. This paper analyzes representative literature on each dimension, describes the machine learning algorithms, principles, and steps they employ, and explores future directions for strategic management scholars to utilize machine learning for empirical research. It is found that natural language processing techniques and classification algorithm not only improve the accuracy of quantifying key variables, but also enhance the capabilities of variable selection, topic extraction, and pattern recognition, thus deepening the understanding of the object of strategic management research. In addition, machine learning algorithms, such as random forests and support vector machine, have significant application value in improving the validity and robustness of model estimation. Methods such as double lasso, double machine learning, and generalized random forest provide strong technical support for causal inferences. The purpose of this paper is to provide guidance and suggestions for using machine learning techniques to deepen strategic management research, as well as laying a theoretical and methodological foundation for future research and applications in the field of strategic management.

Key words: machine learning; strategic management; methodological review; future prospects

(责任编辑:宋澄宇)