

自然语言处理与经济学研究范式变革：

一个文本分析、理解与生成的框架

杨国超 院 茜 龚 强

目录

附录 I 文献检索过程及结果	1
附录 II 文本预处理的步骤	3
附录 III 早期文本分类方法介绍：朴素贝叶斯和支持向量机	4
附录 IV 词嵌入方法的实例介绍：Word2Vec 和 GloVe 模型	5
附录 V 大语言模型的评估方法	7
附录 VI 研究中利用 NLP 方法的研究场景	8
参考文献	14

附录 I 文献检索过程及结果

本文以“文本分析”、“自然语言处理”、“词典法”、“词嵌入”、“主题分类模型”、“朴素贝叶斯法”、“支持向量机”、“可读性”、“语调”、“文本情绪”、“文本相似度”、“Computational Linguistics”、“Textual Analysis”、“Natural Language Processing”、“Diction”、“Dictionary”、“Word Embedding”、“Word2Vec”、“Topic Model”、“Latent Dirichlet Allocation”、“Naïve Bayes”、“Support Vector Machine”、“Readability”、“Tone”、“Sentiment”、“Similarity”、“NLP”、“LDA”、“SVM”、“BERT”等为关键词，以经济学五大刊及各子领域顶刊、管理学 FT50 期刊中的金融学、会计学、企业管理、市场营销等领域的期刊为英文期刊检索范围，以经管领域 AMI 顶尖及权威期刊为中文期刊检索范围，截止 2022 年共检索得到相关文献 344 篇。

表 I 1 期刊名称及文献数量

英文期刊	数量	中文期刊	数量
经济学期刊			
<i>American Economic Review</i>	1	经济研究	7
<i>Quarterly Journal of Economics</i>	2	经济学（季刊）	7
<i>Review of Economic Studies</i>	2	世界经济	4
<i>Econometrica</i>	2	中国工业经济	11
<i>Journal of Political Economy</i>	1	数量经济技术经济研究	8
<i>The Economic Journal</i>	2		
<i>Journal of Development Economics</i>	2		
<i>Journal of Econometrics</i>	5		
<i>Journal of International Economics</i>	3		
<i>Journal of Labor Economics</i>	1		
<i>Journal of Monetary Economics</i>	2		
<i>Journal of Public Economics</i>	4		
合计	27	合计	37
管理学期刊			
<i>Management Science</i>	16	管理世界	13
<i>Journal of Marketing</i>	9	南开管理评论	12
<i>Journal of Marketing Research</i>	14	管理科学学报	20
<i>The Academy of Management Journal</i>	2		
<i>Academy of Management Review</i>	0		
<i>Strategic Management Journal</i>	6		
<i>Journal of International Business Studies</i>	0		
合计	47	合计	45
会计学期刊			
<i>Review of Accounting Studies</i>	16	会计研究	25
<i>The Accounting Review</i>	17	审计研究	18
<i>Journal of Accounting Research</i>	14		
<i>Contemporary Accounting Research</i>	9		
<i>Journal of Accounting and Economics</i>	19		

合计	75	合计	43
金融学期刊			
<i>The Review of Financial Studies</i>	15	金融研究	14
<i>Journal of Financial Economics</i>	18		
<i>The Journal of Finance</i>	10		
<i>Journal of Financial and Quantitative Analysis</i>	13		
合计	56	合计	14
总计	205	总计	139

附录 II 文本预处理的步骤

文本预处理是开展自然语言处理的起点，通常包含以下步骤：第一，文本切分，以句号等标点符号为分隔符将长文本切分为短句子；第二，文本清洗，利用正则表达式匹配并删除页码、标点符号等无实际意义的字符；第三，文本分词（Tokenization），利用分词工具如 Jieba、NLTK 等对原始文本进行分词；第四，去停用词（Stop words）。停用词表通常由两部分构成，一部分是通用的无意义词汇，如“的”“了”等，实践中通常使用通用的停用词表，如哈工大停用词表、百度停用词表、四川大学机器智能实验室停用词表等；另一部分则根据研究问题，自行设置部分停用词汇，以去除与研究问题无关的词汇，从而降低噪音；第五，词干化（Stemming）和词形还原（Lemmatization）处理。这一步骤得名于对英文单词的规范化处理，例如，将不同时态、不同词性的单词统一为它们的词根表示。那么，针对中文文本，实践中更常见的是基于研究问题的需要，通过同义归并、术语规范化和词表映射等将部分含义相似的词汇替换为统一的单词表示，以最大程度地减少噪音和冗余信息。

附录III 早期文本分类方法介绍：朴素贝叶斯和支持向量机

文本分类是一种从原始文本数据中提取特征，并基于这些特征预测文本所属类别的方法。早期研究中常使用朴素贝叶斯 (Naïve Bayes) 和支持向量机 (Support Vector Machine, SVM) 对文本进行二分类。具体地，朴素贝叶斯法沿袭了贝叶斯定理的思想，贝叶斯定理考察条件 X 下 Y 事件出现的概率，朴素贝叶斯法则通过计算某特征条件下文本数据属于某类的概率来进行文本分类，其中特征条件即为样本的标签，类即为样本的类别，实践中通常对已标注数据划分类别进行训练，从而预测未标注数据的分类。贝叶斯定理和朴素贝叶斯法的原理如式 (III1) 和 (III2) 所示。例如，Li (2010) 随机从企业年报 (10-K) 的管理层讨论与分析 (MD&A) 部分选择了 3000 个句子作为训练集，将其标注为积极或消极语调，然后采用朴素贝叶斯法根据训练样本的文本语调情况对测试样本进行分类，从而实现文本语调的自动判别。

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)} \quad (\text{III1})$$

$$P(\text{"属于某类"}|\text{"具有某特征"}) = \frac{P(\text{"具有某特征"}|\text{"属于某类"})P(\text{"属于某类"})}{P(\text{"具有某特征"})} \quad (\text{III2})$$

支持向量机则通过寻找数据间的边界来划分文本类别，图 III1 绘制了采用支持向量机进行文本分类的过程示意图。支持向量机首先将输入数据映射到高维空间，图中的点即为文本数据在向量空间中的向量表示；然后在高维空间中寻找一个能够最大化两类数据点间隔的超平面 (Hyperplane) 来进行文本分类。随着标注特征的增加，向量空间的维度也会相应增加，那么用于分隔数据的超平面会由二维转变为多维平面。在实践中，朴素贝叶斯和支持向量机也可以用于对文本进行多分类。例如，Antweiler and Frank (2004) 基于雅虎网站上 150 万条投资者发帖数据，采用朴素贝叶斯法将投资者情绪分为买入、卖出和持有，考察投资者情绪分歧对股票交易量的影响。马黎琚等 (2019) 采用支持向量机来识别分析师报告文本中的前瞻性语句。具体地，他们首先随机抽取部分分析师报告进行人工标注，若句子中出现了表示预测未来时点或暗示企业即将开展行动的单词^①，则标注为前瞻性语句，然后采用支持向量机对训练数据进行学习并构建文本分类模型，最后利用该模型按照是否前瞻性对大样本分析师报告语句进行自动分类。杨国超等 (2021) 利用支持向量机将媒体报道分为正面、负面和中性三类，考察哪一类媒体报道更能发挥治理作用从而抑制高新技术企业的研发操纵行为。

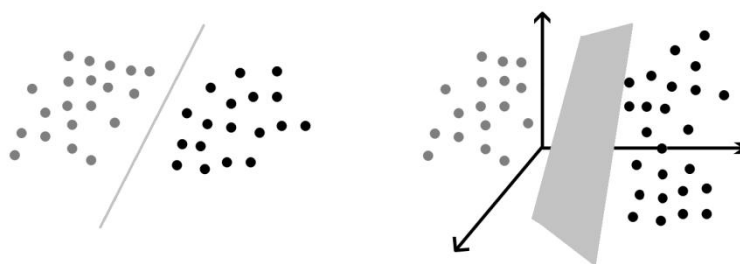
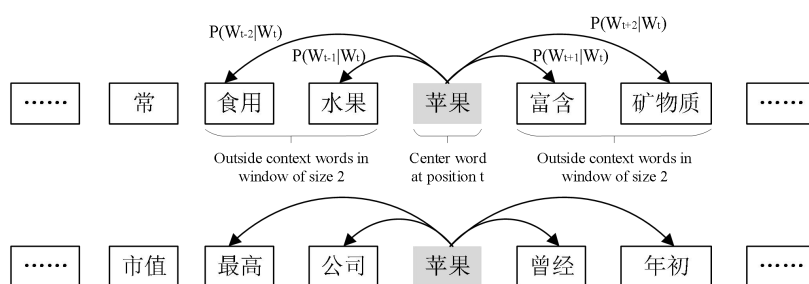


图 III1 支持向量机二维和三维平面模型示意图

① 其中，表示预测过程的表述包括“我们预计/预期/预测/估计/希望/看好”等；表示未来时点的单词包括“届时、未来、将来、今后、明年、两年内、2年内、3年后、2020年”等；表示企业即将发生动作的表述包括“企业将/拟/有望/计划/承诺/打算”等。

附录IV 词嵌入方法的实例介绍：Word2Vec 和 GloVe 模型

图IV1展示了利用 Word2Vec 根据上下文语境推断目标单词的含义的过程。如图所示，“苹果”一词在两个不同的文本中分别表述为“人们最常食用的水果之一，苹果是一种富含矿物质和维生素的水果”“全球市值最高的公司之一，苹果曾经在 2022 年初市值突破 3 万亿美元”，利用 Word2Vec 可以分别计算“苹果”这一单词在不同文本中的词向量表示，在第一个句子中“苹果”是一种水果，而第二个句子中“苹果”是一个品牌。单词对应的词向量实际上是基于其所处的局部上下文共现关系学习得到，从而可以用于识别相同单词在不同文本语境下的不同含义。



图IV1 Word2Vec 语义推算示意图

注：该例子中参数设置window=2，表示在预测单词含义时考虑的上下文最大距离为 2 个单词。

表IV1展示了 GloVe 模型计算和利用单词共现概率的过程。表中第一行和第二行分别表示每个单词 k 与“ice”和“steam”的共现概率，第三行表示每个单词 k 与“ice”的共现概率相对于 k 与“steam”的共现概率比值。单词“solid”与“ice”共现的概率远高于其与“steam”共现的概率，因此“solid”一词共现概率的比值较大；而“gas”与“ice”共现的概率远低于其与“steam”共现的概率，因此“gas”一词共现概率的比值较小；“water”则与“ice”和“steam”都密切相关，其比值接近于 1；“fashion”与“ice”和“steam”均无关，其比值也接近于 1。可以看出，单词共现概率的比值随着其与共现词的相关性而表现出一定的规律，当 k 与两个词均相关或均无关时，共现概率的比值趋近于 1；而当 k 与一个词的相关性较高（或较低）时，共现概率比值出现极大（或极小）的现象。这些概率比值能比原始的共现概率更好地区分相关词（如“solid”、“gas”）和与两者都相关或不相关的词（如“water”、“fashion”），并能区分两种相关的词。GloVe 模型是通过构造一个能够近似表达该语义关系的目标函数，使得其能够计算并拟合这些有区分度的共现概率比值，揭示出词语之间隐含的语义关联和特征，使得词向量能更准确地反映单词之间的差异和相似性。

表IV1 GloVe 方法的词向量表示举例

0	Probability and Ratio	k = solid	k = gas	k = water	k = fashion
1	$P(k ice)$	1.90×10^{-4}	6.60×10^{-5}	3.00×10^{-3}	1.70×10^{-5}
2	$P(k steam)$	2.20×10^{-5}	7.80×10^{-4}	2.20×10^{-3}	1.80×10^{-5}
3	$P(k ice)/P(k steam)$	8.9	8.50×10^{-2}	1.36	0.96

GloVe 这一利用全局词共现信息进行语义分析的特点在一定程度上弥补了 Word2Vec 在捕捉长距离语义关联上的不足，提升了语义分析的全面性。仍以“苹果”一词为例，Word2Vec 基于局部上下文窗口构建“苹果”的词向量，使其能够反映与之共同出现的相邻词汇的语义特征，而 GloVe 模型则通过构建一个反映“苹果”与语料库中所有单词共现概率的矩阵，

学习出能够整合不同语境中语义联系的词向量表示。这一依赖全局共现统计的建模方式，不仅在刻画词语间整体语义结构方面具有独特优势，也使得 **GloVe** 更适用于大规模、结构较为稳定的文本语料分析场景；相较而言，**Word2Vec** 则因其对局部上下文的高敏感性，更适用于处理短文本或语境变化较大的文本数据。

附录V 大语言模型的评估方法

随着大语言模型的快速发展和迭代，以及其在科研和工业界日渐广泛的应用，模型的性能评估也愈发重要。大语言模型的评估可以被分为针对特定任务的指标评估、基准评估、自我评估以及人工评估。具体地，

第一，针对特定任务的指标评估包括用于分类任务的准确率（Accuracy）、精确度（Precision）、召回率（Recall）、F1得分（F1-Score）等指标（Minace et al., 2024），用于生成任务的BLEU（Bilingual Evaluation Understudy）、ROUGE（Recall-Oriented Understudy for Gisting Evaluation）等指标（Papineni et al., 2002; Lin, 2004）。BLEU是基于n-gram的生成文本和参考文本相似度指标，适用于精确匹配的文本翻译类任务评估。ROUGE计算生成文本与参考文本之间的重叠短语、词汇和字符比例，适用于文本摘要任务，评估生成文本中是否包含了参考文本的关键信息。

第二，基准评估通常基于标准化评估框架中的一系列任务和指标，对模型的表现进行量化打分，从而全面衡量模型的性能。Wang et al. (2019) 和 Wang et al. (2020) 分别构建GLUE（General Language Understanding Evaluation）和SuperGLUE评估框架，从情感分析、文本推断、文本相似度等九个角度评估模型的自然语言处理能力。清华大学基础模型研究中心提出SuperBench评测体系，从语义理解、代码生成、安全性等多个角度综合评估大语言模型的性能。当下流行的测试集还包括MMLU、CMMLU、LongBench等文本语义理解测试集，MATH、AIME2024等数学能力数据集，LiveCodeBench、HumanEval等代码能力测试集，C-simpleQA等中文知识测试集等。

第三，模型的自我评估则通过自监督学习或奖励机制等对生成内容的逻辑性、准确性以及任务完成度等进行自我评分。

第四，人工评估通常由领域专家对模型输出结果进行评分。这种方法可能是最为可靠的评估方式，但其成本较高且难以规模化。因此，实践中通常将人工评估与其他评估方法相结合使用以兼顾评估结果的可靠性和效率。例如，发布于HuggingFace网站上的对话机器人排行榜（Chatbot Arena Leaderboard）结合了MMLU基准评估、MT-bench基于GPT-4的模型评估和Chatbot Arena基于人工评估投票的方法对现有各类大模型的表现进行综合打分并排行。

附录VI 研究中利用 NLP 方法的研究场景

一、理解文本的应用场景^①

在实证研究中，许多关键经济或管理变量往往难以被直接观测。过去的研究通常依赖结构化数据构造代理变量来进行间接测度，但这种方法可能因测量误差导致识别偏差，从而影响研究结论的准确性。随着自然语言处理技术的发展，研究者可以通过对文本数据的识别、提取与量化，直接构造相关变量，从而缓解代理测度所引致的内生性问题，更好地回答了过去研究未能有效回答的重要问题。本附录列举了现有研究中基于自然语言处理方法构建的文本指标，及其构造方法随技术演进的演变。表VI1列示了正文及附录中梳理的基于自然语言处理方法构建的文本指标及相关文献。

1. 企业文化

企业文化一直以来被认为是影响企业发展的重要因素，然而早期研究大多停留在理论分析或案例研究层面，这主要是因为企业文化涵盖的特征范围广泛，使用传统研究方法无法对其进行准确刻画，而自然语言处理为恰当定义和准确刻画企业文化提供了一个可能的路径。Guiso et al. (2015) 基于S&P500公司的官方网站描述，采用聚类分析法将企业文化定义为五大类^②，考察不同种类企业文化的价值相关性。Li et al. (2021) 基于企业电话会议文本^③，利用Word2Vec方法扩充了Guiso et al. (2015) 定义的企业文化词汇，构建企业文化词典并据此构造了企业文化的文本度量指标，研究发现该指标能够更加准确且广泛地刻画企业文化。Li et al. (2025) 进一步利用GPT从分析师文本、电话会议文本和员工评价文本中提取和识别企业文化的类型、成因及其经济后果，从而探讨不同利益相关者对企业文化的认识差异及其经济影响。研究发现，不同利益相关者对企业文化的评价存在系统性差异，该差异与他们在经济体系中的角色和激励机制有关，同时这种差异还会影响资本市场的定价与投资决策。

2. 劳动力技能

劳动力是经济活动的核心生产要素，而人类历史上持续不断的技术变革对劳动力市场产生了巨大影响，特别是本世纪以来技术的飞速发展如何重塑劳动力市场是经济学的重要研究议题。过去研究大多通过分析劳动力雇佣中的职业份额变化来描述劳动力市场结构的变化，或者通过单一案例研究来探讨市场对劳动力技能需求的变化。事实上，岗位名称难以准确反映技能需求的变化，同一岗位的技能要求会随时代发展不断演变，而这种变化通常体现在招聘文本的描述中。自然语言处理为更准确提取劳动力技能需求提供了一种可能。Deming and Kahn (2017) 基于企业的招聘文本数据，利用关键词识别方法将上市公司的劳动力技能需求分为10类，包括认知技能、社交技能等，研究发现，公司招聘时对劳动力认知技能和社交技能要求越高，其薪水溢价越高。Turrell et al. (2019) 基于2008年至2016年发布在英国某招聘广告网站上约1500万公司招聘数据，基于词典法开发了分类转换算法，通过将企业的招聘文本数据转换为职业分类，全面刻画了英国劳动力市场的用工需求现状。类似地，Atalay et al. (2020) 收集了1950年至2000年间美国主流报纸（波士顿环球报、纽约时报以及华尔街日报）上发布的共780万招聘文本数据，利用字典法将劳动力技能进行分类并发现，美国劳动力市场从常规认知和体力工作逐渐向非日常互动和分析工作转变。

① 本部分是对正文中第四章第（一）部分第2小节的补充内容。

② 文中定义的五种企业文化：Innovation, Integrity, Quality, Respect, Teamwork。

③ 考虑到一个公司的文化主要受到其高层管理者的影响，高管在电话会议中的发言能够反映其价值观念，而且与公司网站声明或媒体新闻报道这样的“廉价交谈”相比，电话会议更能传递出真实的企业文化。因此，作者基于企业电话会议文本构建了企业文化词典。

3. 企业数字化转型

时代的革命性变革要求企业与时俱进,加速数字化转型。早期针对企业数字化转型的研究主要采用定性分析的方法,吴非等(2021)指出,数字化转型作为企业的重大发展战略,通常会体现在企业年报的文本信息披露中,因此,采用自然语言处理从文本数据中构造变量是一种可行的研究方法。从技术人才需求的视角出发,谭莹等(2025)基于300余万家企业在前程无忧平台上发布的招聘数据,将词典法与监督机器学习方法相结合,搜索、匹配与统计招聘文本中对数字技术人才的需求,从而构建了企业数字化转型程度的测度指标。余典范等(2025)则利用BERT模型构造企业招聘文本的向量化表示,通过计算该文本向量与我国职业分类大典中对职业描述的文本向量的余弦相似度,识别企业数字职能岗位的需求量,从而构造企业的数字化程度指标。

总体上,随着自然语言处理方法在经管领域中的广泛应用,简单地将文本数据作为单一分析对象或利用文本数据的优势来补充回答经典研究问题的机会正逐渐减少。通过挖掘文本内涵,打开传统研究方法无法打开的“黑箱”,从而为研究命题提供更精准的机制解释将成为未来文本理解的主要应用场景。以探索媒体如何发挥信息中介作用为例,早期研究发现媒体在资本市场上发挥着重要的信息中介作用,但是对于媒体如何发挥信息中介作用的机制一直未得到有效识别。Guest(2021)发现媒体报道与公司年报间差异越大,对市场交易量的影响也越大,从而证明了媒体报道能够通过信息创造发挥信息中介作用。再如,早期研究发现投资者交流平台如“互动易”和“e互动”,能够改善市场信息效率,但其如何发挥作用一直未得到回答, Lee and Zhong(2022)利用BERT识别投资者交流平台所解决的投资者问题的类型,发现投资者的大多数帖子是寻求对特定项目的解释,表明信息交流平台通过改善投资者的信息处理效果而提高了市场信息传递效率,从而揭示了投资者交流平台发挥作用的内在机制。

4. 经济政策不确定性

早期文献大多基于市场变量构建经济政策不确定性的衡量指标,但这些指标存在低频、预测周期短、可比性差等缺陷。借助于自然语言处理方法, Baker et al.(2016)基于来自11个主要经济体的海量媒体新闻数据,采用词典法构造了月度频率的EPU指数^①,以更好地衡量经济政策不确定性。聂辉华等(2020)利用词典法基于对企业年报中管理层讨论与分析文本的分析,构造了企业层面的政策不确定性感知指标,研究发现当感知政策不确定性升高时,企业会减少实业投资,同时增加金融资产配置。Caldara et al.(2020)还利用词典法,基于公司盈余电话会议文本构造了公司层面的贸易不确定性指数。Hanley and Hoberg(2019)基于对银行年报(10-K)文本的主题分析,以及由日股票收益数据构造的股价变动协方差矩阵,构建了针对系统性金融风险的动态预测模型,利用该方法能够实时预测系统性金融风险并识别出风险可能发生的领域,从而推动了金融风险预测方面的研究进展。Hassan et al.(2019)构建了公司层面的政治不确定性指标。具体地,他们首先基于政治教科书的内容特征将训练数据标注为与政治性质相关的文本、与特定政治主题相关的文本以及与政治无关的文本,然后利用监督机器学习算法训练管理层电话会议文本并得到政治不确定性内容的占比,从而构造了企业政治不确定性指标,研究发现,更高的政治风险会使企业削减投资和劳动力雇佣,并增加游说和政治捐赠活动。杨国超等(2023)利用Word2Vec方法对MD&A文本进行训练,分别构建了表示宏观、行业和企业个体的词表,通过筛选MD&A部分具有“不确定性”含义的句子,将这些句子分类为宏观、行业和企业三个层面的不确定性表达,从而构造了企业面临的宏观、行业和个体异质性不确定性指标。

① 目前,他们已经针对22个国家和地区构建了相应EPU指数,详见<https://www.policyuncertainty.com>。

5. 行业分类

正确的行业分类对理解经济结构变化至关重要。然而，早期文献大多基于固有国民经济行业分类开展宏观研究（Campello, 2006；伊志宏等，2010）。这种衡量方式存在一个关键缺陷，即企业的实际商业活动并非一成不变，固有行业分类对企业所处竞争市场的认定过于僵化，且同一行业的不同公司所面临的产品市场竞争程度也不尽相同。对此，不少研究尝试从文本的视角入手，刻画企业所面临的竞争态势，从而重新构建行业分类。Hoberg and Phillips（2016）基于企业在年报（10-K）中的产品描述，利用文本相似度和聚类算法，构建了基于产品市场的动态行业分类方法，该方法通过直接度量企业间的产品相似度来划分企业所处行业，并利用聚类算法识别企业真正的竞争对手，从而能够准确度量企业面临的产品市场竞争程度并据此确认其所处行业。许多研究借鉴此类方法重新衡量企业的竞争程度。例如，任宏达和王琨（2019）从中国上市公司年报中提取企业对产品业务的描述内容，借鉴上述方法构建了中国上市公司的产品市场竞争指标。田高良等（2019）基于上市公司财务报告文本附注中披露的经营范围信息，借鉴了类似的方法构建公司的市场中心度指标，用于衡量公司经营业务的竞争强度。Eisdorfer et al.（2021）则通过识别其他企业提及某企业的次数，即“竞争提及”，来衡量一家公司所面临的竞争程度，研究发现被其他公司频繁提及的公司具有更高的回报率，表明“竞争提及”这一基于文本构建的指标能够准确识别出更具竞争力的企业。

6. 企业风险

早期文献主要采用盈余波动性、股票回报波动性、负债比等财务或市场指标（Coles et al., 2006；John et al., 2008；Faccio et al., 2011）间接衡量企业风险，而这些衡量指标仅能片面反映企业在投融资方面的风险，难以全面捕捉到其他来源的风险。基于2005年SEC强制要求上市公司披露风险信息这一背景，Campbell et al.（2014）巧妙地利用自然语言处理方法更为直接地构造了企业风险的度量指标。具体地，他们基于企业年报（10-K）的风险信息披露，采用词典法构造企业风险的文本衡量指标，并根据关键词的类别将企业风险划分为与财务、税务、法律、系统性因素及个体特质等相关的五类风险，发现利用词典法构造的企业风险衡量指标可以更准确地刻画企业所面临的风险水平和风险类别。Campbell et al.（2019）进一步考察公司的风险信息披露能否反映到未来的现金流水平中，以及投资者能否识别这些信息。具体地，他们以仅影响单一现金流科目的税务风险为研究对象，采用词典法构造了公司税务风险的文本指标，发现公司的税务风险信息披露与未来现金流显著正相关，而且这些信息能够被投资者识别并最终反映到股价中。

7. 融资约束

早期文献大多基于公司的投资-现金流敏感度，或基于系列财务指标加权计算KZ指数，或基于动态结构估计方法计算WW指数等来间接衡量企业面临的融资约束（Lamont et al., 2001；Whited and Wu, 2006；Hadlock and Pierce, 2010；邓可斌和曾海舰, 2014）。Farre-Mensa and Ljungqvist（2016）利用三个外生冲击^①检验上述间接测量指标的有效性，发现这些早期研究中常用的间接测量指标无法准确度量企业面临的融资约束。自然语言处理方法为直接度量企业面临的融资约束提供了一种可能的路径。Hoberg and Maksimovic（2015）基于企业年报中披露的企业融资需求及融资计划信息，利用词典法构造融资约束的文本指标，为准确度量企业面临的融资约束提供了新的视角。Buehlmaier and Whited（2018）采用朴素贝叶斯法，通过对企业年报（10-K）中管理层讨论与分析（MD&A）部分的融资信息进行分类识别，构建了总体层面、股权层面和债务层面的三个融资约束指标，实证检验发现与以往基于财务数据构建的指标相比，文本指标能够更准确地衡量企业面临的融资约束水平，并且将基于该

① 分别为企业面临的信贷需求增加、银行贷款供应的变化以及企业通过发行股票回收资金的行为。

指标构建的因子加入Fama-French因子模型后，企业的年化收益率提高约7.2%。

8. 会计错报预测

早期研究通常采用定量财务数据或市场变量，如应计科目、财务业绩、股票回报率等预测会计错报（Brazel et al., 2009; Dechow et al., 2011）。然而，这种方法无法预测由连续多期操纵造成的会计错报，这是因为当公司连续多期操纵会计交易和业绩数据时，用于预测会计错报的变量本身就可能存在错报，导致预测结果失效。社会心理学研究指出，人类的欺骗行为会通过语言反映出来（Zuckerman et al., 1981），但是由于过去无法对海量的语言数据进行批量分析，因此基于语言的错报识别仅停留在实验研究层面（Newman et al., 2003）。Hobson et al.（2012）借助自然语言处理方法，通过识别管理层在电话会议中的语调不一致现象预测公司会计错报^①。Larcker and Zakolyukina（2012）对发生会计错报的公司年报进行语言特征分析，发现这些公司所披露信息中具有更多常识性内容、使用更多极端积极语调的词汇，并且与股东价值相关的信息相对更少。Brown et al.（2020）采用LDA主题分析法对企业年报（10-K）进行主题分类，将文本主题特征添加到基于财务数据和语言特征的会计错报预测模型中，研究发现该模型的预测准确率显著高于过去基于财务数据的错报预测模型，特别是与收入确认和费用相关的会计科目错报。

9. 气候风险

随着全球气候变化加剧，企业面临的气候风险逐渐成为资本市场关注的关键信息。因此，如何有效量化企业面临的气候风险，进而深入探讨气候变化对经济发展的影响成为当前研究的热点和难点。过去研究基于极端天气事件作为外生冲击，考察气候变化的经济后果（Addoum et al., 2020），而这种方法难以直接衡量企业面临的气候风险。Sautner et al.（2023a）利用自然语言处理方法，通过提取企业管理层电话会议文本中关于气候变化相关话题的内容，构建了随时间变化的企业气候风险文本指标。研究发现，该指标能够预测绿色技术增长，绿色专利等重要的绿色成果，有助于推动净零转型的实现。Sautner et al.（2023b）进一步估算了公司层面的气候风险溢价及其随时间的动态变化，研究发现，气候风险溢价具有重要的资产定价作用。

表 VI1 现有研究中运用自然语言处理方法构造的文本指标

研究话题	研究方法	代表文献
企业文化	词典法	Guiso et al.（2015）；潘健平等（2019）
	Word2Vec	Li et al.（2021）
	GPT	Li et al.（2025）
劳动力技能	词典法	Deming and Kahn（2017）；Turrell et al.（2019）；Atalay et al.（2020）
企业数字化转型	词典法	吴非等（2021）；谢绚丽和王诗卉（2022）；彭俞超等（2023）；谭莹等（2025）
	BERT	余典范等（2025）
经济政策不确定性	词典法	Baker et al.（2016）；聂辉华等（2020）；Caldara et al.（2020）；Benguria et al.（2022）
	Word2Vec	杨国超等（2023）
	监督机器学习法	Hassan et al.（2019）
行业分类	词典法	Li et al.（2013）；Bushman et al.（2016）；Eisdorfer et al.

① 语调不一致现象是指管理层在电话会议中发言语调的不一致，作者采用能够检测到语音音调不一致的程序 Ex-Sense Pro R software 识别这一现象。

		(2021)
	文本相似度	Hoberg and Phillips (2016) ; 任宏达和王琨 (2019) ; 田高良等 (2019) ; 戚聿东等 (2021)
企业风险	词典法	Campbell et al.(2014); 郝项超和苏之翔(2014); Campbell et al. (2019) ; 何捷和陆正飞 (2020)
融资约束	词典法	Hoberg and Maksimovic (2015)
	朴素贝叶斯法	Buehlmaier and Whited (2018)
会计错报预测	词典法	Hobson et al. (2012) ; Larcker and Zakolyukina (2012) ; Purda and Skillicorn (2015)
	LDA	Brown et al. (2020)
气候风险	词典法	Sautner et al. (2023a, b)

二、利用文本理解方法打开理论黑箱

事实上,随着自然语言处理方法在经管领域中的广泛应用,简单地将文本数据作为单一分析对象或利用文本数据的优势来补充回答经典研究问题的机会正逐渐减少。通过挖掘文本内涵,打开传统研究方法无法打开的“黑箱”,从而为研究命题提供更精准的机制解释将成为未来文本理解的主要应用场景。

1. 媒体如何发挥信息中介作用

已有研究发现,媒体在资本市场上发挥着重要的信息中介作用 (Miller and Skinner, 2015),但是对于媒体如何发挥信息中介作用的机制一直未得到有效识别,其关键原因在于传统研究方法难以对海量媒体报道内容进行批量分析,借助自然语言处理方法则为进一步理解媒体如何发挥作用提供了新的思路。例如, Guest (2021) 和杨国超和张李娜 (2021) 通过考察媒体报道的文本相似度、文本语调等多个文本特征,探究媒体报道发挥作用的路径。Guest (2021) 发现媒体报道与公司年报间差异越大,对市场交易量的影响也越大,从而证明了媒体报道能够通过信息创造发挥信息中介作用。杨国超和张李娜 (2021) 研究发现,相较于正面、转载以及非深度报道,负面、原创以及深度报道才能对高新技术企业的研发操纵行为发挥显著的抑制作用。

2. 会计信息有用性的检验

基于有效市场假说,早期研究普遍认为公司所披露的信息会完全反映到公司的市场价格中去。然而不少实证研究发现这种公告效应似乎变得不再明显 (Brown and Tucker, 2011)。一些研究认为,造成这一现象的原因是企业信息披露的信息含量下降 (SEC, 2013; KPMG, 2011),但另一种观点则认为,这是由于投资者未能识别出信息披露公告中的信息含量所致。然而,倘若想准确回答这一问题,关键在于如何识别企业信息披露的信息含量的变化程度,早期研究以资本市场在公告后的市场反应来度量投资者对增量信息的态度,然而这无法区分市场反应的差异到底是由于信息本身的差异所致,还是由投资者的差异所致,而自然语言处理方法则能够直接测量文本信息披露的异同,从而为解决这一问题提供了可行路径。Cohen et al. (2020) 通过考察企业年报 (10-K) 文本纵向文本相似度^①与公司股价变动之间的关系巧妙地回答了这一问题。研究发现,公告后市场反应滞后现象并非是由于年报缺乏信息含量,而是由于投资者未能及时识别出年报中的细微但重要的信号所致。换言之,公司年报的价值相关性并未随着年报披露长度的增加而减弱,公司披露的基本面信息仍然具有预测股价的能力,但是年报信息披露长度和复杂度的增加导致资本市场的解读和反应出现滞后现象。此外,审计研究中也利用自然语言处理方法探究审计信息披露的沟通价值,王艳艳等 (2018) 通过

① 纵向文本相似度指当年度企业年报与其上一年年报的文本相似度。

分析关键审计事项的文本可读性、语调及“具体性”信息占比，考察了关键审计事项段提高审计报告沟通价值的作用机制。An et al. (2023) 根据是否包含存货相关信息将关键事项段分为与存货相关的关键事项段和与存货无关的关键审计事项段，通过考察存货相关关键审计事项段与公司存货管理的关系探究了关键审计事项段披露影响公司实际经营决策的作用路径。

3. 资本市场制度安排如何发挥作用

已有文献证明交易所问询函制度能够发挥监督作用（陈运森等，2019），但问询函制度如何发挥监督作用的机理尚未得到充分研究。事实上，问询函及其回函是典型的以文本为主要载体的披露形式，因此从文本角度入手可能为打开问询函监督作用的机制提供有效途径。李晓溪等（2019）和薛爽和王禹（2022）通过分析文本信息打开了监管问询的“黑箱”。李晓溪等（2019）通过计算收到问询函前后企业并购重组报告书的文本相似度及文本“具体性”程度发现，与修订前的并购重组报告书相比，修订后的报告书中披露了更多的标的公司历史信息和前瞻信息，且细节内容更多，表明问询函通过改善信息披露缓解了并购交易的信息不对称问题，从而对企业的并购重组发挥了监管作用。薛爽和王禹（2022）基于科创板IPO公司多轮审核问询回复函文本，从文本信息量、信息可视化程度、会计术语和逆接成分密度等角度系统评估回函信息质量，发现回函信息质量越高，IPO首发抑价程度越低。

4. 高管过往经历如何影响企业决策

已有研究表明高管经历会对企业决策产生较大影响。然而，过去研究仅能够基于高阶理论和印象管理理论对这些发现提供逻辑上的解释，未能提供实证证据。Yang and Hu (2022) 通过量化生产制造背景CEO对企业成本管理决策的影响，探究了高管经历影响企业行为决策的内在机制。理论上，生产制造背景CEO在成本管理方面具有更丰富的工作经验，因此更应强调企业的成本管理，从而有助于降低企业的成本粘性。验证这一作用机制的关键在于如何识别生产制造背景CEO对成本管理的重视程度，为此，他们基于企业年报的管理层讨论与分析文本，以“生产成本”和“制造成本”为种子词，采用Word2Vec方法构建了与成本管理相关的关键词表，并计算关键词在年报管理层讨论与分析中的占比衡量企业CEO对成本管理的重视程度。研究发现，拥有生产制造背景CEO的企业在年报中披露了更多与成本管理相关的内容，从而发现了高管经历影响企业决策的内在原因。

5. 投资者交流平台如何发挥沟通作用

已有研究发现投资者交流平台如“互动易”和“e互动”能够改善市场信息效率（孟庆斌等，2020；谭松涛等，2016）。然而，其如何发挥作用一直未得到回答，Lee and Zhong (2022) 基于中国投资者交流平台上的投资者与上市公司问答文本，利用BERT根据文本的性质和内容对投资者发帖进行分类，通过检验投资者所关注的问题所属类别探究交流平台这种信息沟通方式如何发挥作用。研究发现，投资者的大多数帖子是寻求对特定项目的解释，表明信息交流平台通过改善投资者的信息处理效果而提高了市场信息传递效率，从而揭示了投资者交流平台发挥作用的内在机制。

参考文献

- [1] Addoum, J. M., D. T. Ng, and A. Ortiz-Bobea, “Temperature Shocks and Establishment Sales” , *The Review of Financial Studies*, 2020, 33(3), 1331–1366.
- [2] An, R., W. Li, D. Wang, Y. Wang, and L. Yu, “Do Key Audit Matters Affect Operating Activities? Evidence From Inventory Management” , *Abacus*, 2023, 59(1), 300–339.
- [3] Antweiler, W., and M. Z. Frank, “Is All that Talk Just Noise? The Information Content of Internet Stock Message Boards” , *The Journal of Finance*, 2004, 59(3), 1259–1294.
- [4] Atalay, E., P. Phongthientham, S. Sotelo, and D. Tannenbaum, “The Evolution of Work in the United States” , *American Economic Journal: Applied Economics*, 2020, 12(2), 1–34.
- [5] Baker, S. R., N. Bloom, and S. J. Davis, “Measuring Economic Policy Uncertainty” , *The Quarterly Journal of Economics*, 2016, 131(4), 1593–1636.
- [6] Benguria, F., J. Choi, D. L. Swenson, and M. Z. Xu, “Anxiety or Pain? The Impact of Tariffs and Uncertainty on Chinese Firms in the Trade War” , *Journal of International Economics*, 2022, 137:103608.
- [7] Brazel, J. F., K. L. Jones, and M. F. Zimbelman, “Using Nonfinancial Measures to Assess Fraud Risk” , *Journal of Accounting Research*, 2009, 47(5), 1135–1166.
- [8] Brown, N. C., R. M. Crowley, and W. B. Elliott, “What are You Saying? Using Topic to Detect Financial Misreporting” , *Journal of Accounting Research*, 2020, 58(1), 237–291.
- [9] Brown, S. V., and J. W. Tucker, “Large-Sample Evidence on Firms' Year-Over-Year MD&A Modifications” , *Journal of Accounting Research*, 2011, 49(2), 309–346.
- [10] Buehlmaier, M. M. M., and T. M. Whited, “Are Financial Constraints Priced? Evidence From Textual Analysis” , *The Review of Financial Studies*, 2018, 31(7), 2693–2728.
- [11] Bushman, R. M., B. E. Hendricks, and C. D. Williams, “Bank Competition: Measurement, Decision-Making, and Risk-Taking” , *Journal of Accounting Research*, 2016, 54(3), 777–826.
- [12] Caldara, D., M. Iacoviello, P. Molliogo, A. Prestipino, and A. Raffo, “The Economic Effects of Trade Policy Uncertainty” , *Journal of Monetary Economics*, 2020, 109, 38–59.
- [13] Campbell, J. L., H. Chen, D. S. Dhaliwal, H. Lu, and L. B. Steele, “The Information Content of Mandatory Risk Factor Disclosures in Corporate Filings” , *Review of Accounting Studies*, 2014, 19(1), 396–455.
- [14] Campbell, J. L., M. Cecchini, A. M. Cianci, A. C. Ehinger, and E. M. Werner, “Tax-Related Mandatory Risk Factor Disclosures, Future Profitability, and Stock Returns” , *Review of Accounting Studies*, 2019, 24(1), 264–308.
- [15] Campello, M., “Debt Financing: Does It Boost or Hurt Firm Performance in Product Markets?” , *Journal of Financial Economics*, 2006, 82(1), 135–172.
- [16] Cohen, L., C. Malloy, and Q. Nguyen, “Lazy Prices” , *The Journal of Finance*, 2020, 75(3), 1371–1415.
- [17] Coles, J. L., N. D. Daniel, and L. Naveen, “Managerial Incentives and Risk-Taking” , *Journal of Financial Economics*, 2006, 79(2), 431–468.
- [18] Dechow, P. M., W. Ge, C. R. Larson, and R. G. Sloan, “Predicting Material Accounting

- Misstatements” , *Contemporary Accounting Research*, 2011, 28(1), 17–82.
- [19] Deming, D. , and L. B. Kahn , “ Skill Requirements Across Firms and Labor Markets: Evidence From Job Postings for Professionals” , *Journal of Labor Economics*, 2017, 36(S1), S337–S369.
- [20] Eisdorfer, A., K. Froot, G. Ozik, and R. Sadka, “Competition Links and Stock Returns” , *The Review of Financial Studies*, 2021, 9(35), 4300–4340.
- [21] Faccio, M., M. Marchica, and R. Mura, “Large Shareholder Diversification and Corporate Risk-Taking” , *The Review of Financial Studies*, 2011, 24(11), 3601–3641.
- [22] Farre-Mensa, J. , and A. Ljungqvist , “ Do Measures of Financial Constraints Measure Financial Constraints?” , *The Review of Financial Studies*, 2016, 29(2), 271–308.
- [23] Guest, N. M. , “ The Information Role of the Media in Earnings News ” , *Journal of Accounting Research*, 2021, 59(3), 1021–1076.
- [24] Guiso, L., P. Sapienza, and L. Zingales, “The Value of Corporate Culture” , *Journal of Financial Economics*, 2015, 117(1), 60–76.
- [25] Hadlock, C. J., and J. R. Pierce, “New Evidence on Measuring Financial Constraints: Moving Beyond the KZ Index” , *The Review of Financial Studies*, 2010, 23(5), 1909–1940.
- [26] Hanley, K. W., and G. Hoberg, “Dynamic Interpretation of Emerging Risks in the Financial Sector” , *The Review of Financial Studies*, 2019, 32(12), 4543–4603.
- [27] Hassan, T. A., S. Hollander, L. van Lent, and A. Tahoun, “Firm-Level Political Risk: Measurement and Effects” , *Quarterly Journal of Economics*, 2019, 134(4), 2135–2202.
- [28] Hoberg, G., and G. Phillips, “Text-Based Network Industries and Endogenous Product Differentiation” , *Journal of Political Economy*, 2016, 124(5), 1423–1465.
- [29] Hoberg, G., and V. Maksimovic, “Redefining Financial Constraints: A Text-Based Analysis” , *The Review of Financial Studies*, 2015, 28(5), 1312–1352.
- [30] Hobson, J. L., W. J. Mayew, and M. Venkatachalam, “Analyzing Speech to Detect Financial Misreporting” , *Journal of Accounting Research*, 2012, 50(2), 349–392.
- [31] John, K., L. Litov, and B. Yeung, “Corporate Governance and Risk - Taking” , *The Journal of Finance*, 2008, 63(4), 1679–1728.
- [32] KPMG , “ Disclosure Overload and Complexity: Hidden in Plain Sight ”
<http://www.kpmg.com/US/en/IssuesAndInsights/ArticlesPublications/Documents/disclosure-overload-complexity.pdf>.
- [33] Lamont, O., C. Polk, and J. Saá-Requejo, “Financial Constraints and Stock Returns” , *The Review of Financial Studies*, 2001, 14(2), 529–554.
- [34] Larcker, D. F., and A. A. Zakolyukina, “Detecting Deceptive Discussions in Conference Calls” , *Journal of Accounting Research*, 2012, 50(2), 495–540.
- [35] Lee, C. M. C., and Q. Zhong, “Shall we Talk? The Role of Interactive Investor Platforms in Corporate Communication” , *Journal of Accounting and Economics*, 2022, 74(2), 101524.
- [36] Li, F., “The Information Content of Forward-Looking Statements in Corporate Filings—a Naïve Bayesian Machine Learning Approach” , *Journal of Accounting Research*, 2010, 48(5), 1049–1102.
- [37] Li, F., R. Lundholm, and M. Minnis, “A Measure of Competition Based on 10-K Filings” , *Journal of Accounting Research*, 2013, 51(2), 399–436.
- [38] Li, K., F. Mai, R. Shen, C. Yang, and T. Zhang, “Dissecting Corporate Culture Using

- Generative AI” , *The Review of Financial Studies*, 2025.
- [39] Li, K., F. Mai, R. Shen, and X. Yan, “Measuring Corporate Culture Using Machine Learning” , *The Review of Financial Studies*, 2021, 34(7), 3265—3315.
- [40] Lin, C. , “ROUGE: A Package for Automatic Evaluation of Summaries” , 2004.
- [41] Miller, G. S., and D. J. Skinner, “The Evolving Disclosure Landscape: How Changes in Technology, the Media, and Capital Markets are Affecting Disclosure” , *Journal of Accounting Research*, 2015, 53(2), 221—239.
- [42] Minaee, S., T. Mikolov, N. Nikzad, M. Chenaghlu, R. Socher, X. Amatriain, and J. Gao , “Large Language Models: A Survey” , 2024, <https://arxiv.org/abs/2402.06196>.
- [43] Newman, M. L., J. W. Pennebaker, D. S. Berry, and J. M. Richards, “Lying Words: Predicting Deception From Linguistic Styles” , *Personality and Social Psychology Bulletin*, 2003, 29(5), 665—675.
- [44] Papineni, K., S. Roukos, T. Ward, and W. Zhu, “BLEU: A Method for Automatic Evaluation of Machine Translation” , *Annual Meeting of the Association for Computational Linguistics* 2002.
- [45] Purda, L., and D. Skillicorn, “Accounting Variables, Deception, and a Bag of Words: Assessing the Tools of Fraud Detection” , *Contemporary Accounting Research*, 2015, 32(3), 1193—1223.
- [46] Sautner, Z., L. VAN Lent, G. Vilkov, and R. Zhang, “Firm-Level Climate Change Exposure” , *The Journal of Finance*, 2023a, 78(3), 1449—1498.
- [47] Sautner, Z., L. van Lent, G. Vilkov, and R. Zhang, “Pricing Climate Change Exposure” , *Management Science*, 2023b, 69(12), 7540—7561.
- [48] SEC , “Report on Review of Disclosure Requirements in Regulation S-K ” , <http://www.sec.gov/news/studies/2013/reg-sk-disclosure-requirements-review.pdf>.
- [49] Turrell, A., B. J. Speigner, J. Djumalieva, D. Copple, and J. Thurgood, “Transforming Naturally Occurring Text Data Into Economic Statistics: The Case of Online Job Vacancy Postings” , *National Bureau of Economic Research Working Paper Series*, 2019, No. 25837.
- [50] Wang, A., A. Singh, J. Michael, F. Hill, O. Levy, and S. R. Bowman, “GLUE: A Multi-Task Benchmark and Analysis Platform for Natural Language Understanding ” , 2019 , <https://doi.org/10.48550/arXiv.1804.07461>.
- [51] Wang, A., Y. Pruksachatkun, N. Nangia, A. Singh, J. Michael, F. Hill, O. Levy, and S. R. Bowman , “ SuperGLUE: A Stickier Benchmark for General-Purpose Language Understanding Systems” , 2020, <https://doi.org/10.48550/arXiv.1905.00537>.
- [52] Whited, T. M., and G. Wu, “Financial Constraints Risk” , *The Review of Financial Studies*, 2006, 19(2), 531—559.
- [53] Yang, G., and Y. Hu, “Manufacturing Background CEOs and Corporate Cost Management: Evidence From Listed Manufacturing Firms” , *China Journal of Accounting Studies*, 2022, 10(2), 274—300.
- [54] Zuckerman, M., B. M. Depaulo, and R. Rosenthal, “Verbal and Nonverbal Communication of Deception” , *Advances in Experimental Social Psychology*, 1981, 14, 1—59.
- [55] 陈运森、邓祎璐、李哲, “证券交易所一线监管的有效性研究:基于财务报告问询函的证据” , 《管理世界》, 2019 年第 3 期第 35 卷, 第 169—185 页。
- [56] 邓可斌、曾海舰, “中国企业的融资约束:特征现象与成因检验” , 《经济研究》, 2014

年第2期第49卷,第47—60页。

- [57] 郝项超、苏之翔,“重大风险提示可以降低抑价吗?——基于文本分析法的经验证据”,《财经研究》,2014年第5期第40卷,第42—53页。
- [58] 何捷、陆正飞,“定性的未来供应链风险披露与分析师关注行为研究”,《会计研究》,2020年第6期,第36—48页。
- [59] 李晓溪、杨国超、饶品贵,“交易所问询函有监管作用吗?——基于并购重组报告书的文本分析”,《经济研究》,2019年第5期第54卷,第181—198页。
- [60] 马黎琚、伊志宏、张澈,“廉价交谈还是言之有据?——分析师报告文本的信息含量研究”,《管理世界》,2019年第7期第35卷,第182—200页。
- [61] 孟庆斌、黄清华、张劲帆,王松,“上市公司与投资者的互联网沟通具有信息含量吗?——基于深交所“互动易”的研究”,《经济学(季刊)》,2020年第2期第19卷,第637—662页。
- [62] 聂辉华、阮睿、沈吉,“企业不确定性感知、投资决策和金融资产配置”,《世界经济》,2020年第6期第43卷,第77—98页。
- [63] 潘健平、潘越、马奕涵,“以“合”为贵?合作文化与企业创新”,《金融研究》,2019年第1期,第148—167页。
- [64] 彭俞超、王南萱、顾雷雷,“企业数字化转型、预判性信息披露与股价暴跌风险”,《财贸经济》,2023年第44卷第5期,第79—90页。
- [65] 戚聿东、孙昌玲、王化成,“企业核心竞争力能够降低权益资本成本吗?——基于文本分析的经验证据”,《会计研究》,2021年第2期,第94—106页。
- [66] 任宏达、王琨,“产品市场竞争与信息披露质量——基于上市公司年报文本分析的新证据”,《会计研究》,2019年第3期,第32—39页。
- [67] 谭松涛、阚铎、崔小勇,“互联网沟通能够改善市场信息效率吗?——基于深交所“互动易”网络平台的研究”,《金融研究》,2016年第3期,第174—188页。
- [68] 谭莹、刘颖杰、张勋,“企业数字化转型程度的统计测度:以数字技术人才需求为视角”,《经济研究》,2025年第4期第60卷,第122—138页。
- [69] 田高良、田皓文、吴璇,吴其霖,“经营业务竞争与股票收益——基于财务报告文本附注的分析”,《会计研究》,2019年第10期,第78—84页。
- [70] 王艳艳、许锐、王成龙,于李胜,“关键审计事项段能够提高审计报告的沟通价值吗?”,《会计研究》,2018年第6期,第86—93页。
- [71] 吴非、胡慧芷、林慧妍,任晓怡,“企业数字化转型与资本市场表现——来自股票流动性的经验证据”,《管理世界》,2021年第7期第37卷,第130—144页。
- [72] 谢绚丽、王诗卉,“中国商业银行数字化转型:测度、进程及影响”,《经济学(季刊)》,2022年第6期第22卷,第1937—1956页。
- [73] 薛爽、王禹,“科创板IPO审核问询回复函与首发抑价”,《管理世界》,2022年第4期第38卷,第185—203页。
- [74] 杨国超、魏爽、阮茜,龚强,“企业为何选择劳务外包——基于经济政策不确定性的解释”,《中国工业经济》,2023年第9期,第136—154页。
- [75] 杨国超、张李娜,“产业政策何以更有效?——基于海量媒体报道数据与研发操纵现象的证据”,《经济学(季刊)》,2021年第6期第21卷,第2173—2194页。
- [76] 伊志宏、姜付秀、秦义虎,“产品市场竞争、公司治理与信息披露质量”,《管理世界》,2010年第1期,第133—141页。
- [77] 余典范、张宇、宋晴,“数字人力投入与企业价格加成——来自招聘大数据的证据”,

《中国工业经济》，2025 年第 7 期，第 62—80 页。

注：该附录是期刊所发表论文的组成部分，同样视为作者公开发表的内容。如研究中使用该附录中的内容，请务必在研究成果上注明附录下载出处。