

算法创新与企业全要素生产率提升

——来自专利文本的经验证据

蒋伟杰 齐晨扬 陈琦*

摘要:在数字经济加速深化的背景下,算力、算法与数据已成为支撑实体经济转型的关键资源。算法创新作为激发数据要素潜在价值的核心引擎,正逐渐成为推动企业高质量发展的新动能。本文利用2010~2021年中国上市公司数据,引入MacBERT大语言模型处理海量专利摘要文本,精准识别算法专利,并以此刻画地区算法创新水平,实证检验其对企业全要素生产率的影响。研究发现,算法创新能够显著提升企业全要素生产率。机制分析表明,算法创新主要通过提升投入端和产出端数据利用效率发挥赋能作用,一方面,算法创新能够缓解信息摩擦,纠正要素错配并提升产能利用率;另一方面,算法创新能够增强市场需求识别能力,驱动研发创新、降低销售费用并加速资金周转。异质性分析显示,在数字基础设施完善、数据开放程度高的地区,以及需求波动大的行业,算法创新的促进作用更为明显。本文为充分挖掘数据要素价值提供了微观经验证据,为构建算法创新支撑体系、优化数字化发展环境提供了重要的政策启示。

关键词:算法创新 数据要素 全要素生产率 大语言模型

一、引言

党的二十届四中全会明确提出“强化算力、算法、数据等高效供给”,体现出算

* 蒋伟杰,副教授,宁波大学商学院,电子邮箱:jiangweijie@nbu.edu.cn;齐晨扬,硕士研究生,宁波大学商学院,电子邮箱:solution51@163.com;陈琦(通讯作者),副教授,宁波大学商学院,电子邮箱:chenqi2@nbu.edu.cn。本文获得浙江省社会科学界联合会研究课题项目(2026N164)、国家社会科学基金一般项目(25GBL177)和浙江省哲社规划课题之江青年专项(24ZJQN098Y)的资助。本文入选“第十届香樟经济学年会”。本文未使用AI。感谢匿名审稿专家的宝贵意见,文责自负。

力、算法与数据已上升为支撑中国式现代化的重要战略资源。自党的十九届四中全会提出将数据列为新型生产要素以来,中国数据资源规模持续扩大。国家互联网信息办公室发布的《数字中国发展报告(2022年)》显示,2022年中国数据产量高达8.1ZB,约占全球数据总产量的10.5%,数据要素市场规模已攀升至1018.8亿元。然而,在数据要素蓬勃发展的同时,数据要素的利用效率仍待提升,需要将数据要素的快速积累充分转化为现实生产力。

事实上,数据本身并不是直接的生产力,而是一种潜在的生产力(徐翔等,2024;陈荣达等,2024)。现有研究主要从数字化转型、数字技术创新和数据要素融合协同等角度,探讨数据要素对全要素生产率的影响。在数字化转型方面,大量研究借助企业年报文本分析衡量企业数字化水平(吴非等,2021),并验证了其在提升专业化分工(袁淳等,2021)、改善投资效率(李万利等,2022)及增强跨界竞争能力(张骁等,2019)等方面的积极作用。进一步地,部分文献聚焦于数字化转型对企业生产效率的直接影响,发现其可通过提高创新能力、优化人力资本结构、推动产业融合发展及降低成本等路径提升全要素生产率(赵宸宇等,2021;刘淑春等,2021)。在数字技术创新领域,不同于企业层面的数字化应用,现有研究更关注大数据、云计算及区块链等前沿技术带来的创新驱动效应(黄先海和高亚兴,2023;Vial,2019)。学者们利用文本分析(Liu等,2023;黄勃等,2023)及专利IPC分类(陶锋等,2023b)等多种度量方式发现,数字技术创新主要通过重塑企业组织治理结构(Svahn等,2017)、提升产品与服务价值(Nambisan等,2017)以及优化创新资源配置(罗佳等,2023)等渠道,促进企业生产效率提升。

值得注意的是,数据要素具有非竞争性和非排他性等特征(蔡跃洲和马文君,2021;Jones和Tonetti,2020),导致其参与生产活动的方式,与资本、劳动力等传统生产要素有所区别。数据要素需要经过深度加工,物化为资本和劳动等传统要素的知识,进而渗透在生产过程中产生倍增效果(徐翔和赵墨非,2020)。在大数据时代,算法作为智能化处理数据的技术方法,在助推数据从潜在生产力转变为实际生产力的过程中发挥着关键桥梁和核心引擎作用(丁晓东,2020)。尤其是在数据爆发式增长时期,大量数据资源具有信噪比低、格式高度非结构化等特点(Camacho等,2016),需要借助先进的算法从海量的数据中提取出有价值的信息,进而赋能传统生产要素以激发新质生产力。作为一种现代信息技术,算法也有水平高低、效率优劣之分。先进的算法能够显著提升数据处理的速度和精度,最大限度激发数据要素利用价值。因此,不断迭代升级原有算法或创造新的算法,推动算法创新,是驱动数据要素赋能全要素生产率的关键所在。然而,现有文献缺少从算法视角探究数据要素利用和赋能的相关研究。

为此,本文尝试从算法创新的微观视角,揭示数据要素赋能企业全要素生产率的内在机制,以期为推动数字技术创新向新质生产力转化提供理论参考。首先,本

文通过机器学习的方式,训练大语言模型通过专利文本摘要识别海量专利中的算法专利。然后,以各地区新增算法专利授权数量作为算法创新的度量指标,实证检验各地区算法创新对企业全要素生产率的影响。研究发现,算法创新显著地提升了企业全要素生产率水平。机制分析表明,算法创新通过提升企业投入端与产出端的数据利用效率,促进企业全要素生产率提升。异质性分析结果显示,算法创新对全要素生产率的促进作用在数字基础设施完善、数据开放程度高的地区以及需求波动大的行业中更为明显。

本文可能的边际贡献体现在以下三个方面。

第一,构建算法创新影响企业全要素生产率的分析框架,拓展了数字技术与实体经济融合研究的分析边界。算力、算法和数据是数字经济运行的三大核心要素,三者相互依存、相互促进,共同支撑着数字经济的快速发展(马鸿佳等,2024;任保平和豆渊博,2024)。既有研究更多聚焦于算力维度,从算力部署与算力基础设施建设等方面,考察其对生产率和高质量发展的影响(许诺等,2025;杨壮和吴福象,2025)。由于企业层面直接量化数据要素利用程度存在较大困难,现有研究往往采用上市公司年报中的关键词词频来度量企业的数据利用能力,难以真实刻画企业对数据要素的真实利用水平。相较而言,算法作为连接数据与算力的关键环节,是将数据要素转化为智能决策的核心驱动力。遗憾的是,鲜有文献从算法视角,分析数据要素充分利用对企业绩效的影响。基于此,本文以算法创新为切入点,通过识别并量化地区算法创新水平,将其作为衡量企业有效利用数据与算力的重要代理指标,系统考察算法创新对企业全要素生产率的影响,从而为理解数据要素如何转化为企业生产率提供了新的研究视角。

第二,从“投入端优化”与“产出端增值”双重渠道,揭示算法创新影响企业全要素生产率的内在机制,深化对企业数据要素利用机制的理论认知。现有研究多聚焦数字技术对企业内部资源配置效率的改善,如缓解资本及劳动要素错配(Jiang和Li,2024)、降低信息不对称(吕铁和李载驰,2021)或优化供应链管理(肖红军等,2024)等,却鲜有文献系统考察其对企业市场价值创造的影响。本文突破传统单一效率分析框架,基于企业“生产—销售”闭环,首次将地区算法创新的作用分解为两个维度:在投入端,算法通过精准匹配投入要素、动态调整生产计划,缓解信息摩擦导致的劳动与资本要素错配,提升产能利用率,从而优化内部资源配置效率;在产出端,算法通过用户画像、需求预测与个性化推荐,增强企业创新能力、降低销售成本并提高资金使用效率,进而提升市场价值创造能力。上述双重路径机制的分析,为理解数字经济下全要素生产率的跃升提供了新的理论框架。

第三,在核心指标构建方法上,借助大语言模型与自然语言处理技术,构建了算法专利的精准识别方法,为相关领域研究提供了更可靠的量化思路与度量指标。算法创新属于数字技术的创新应用,以往文献通常基于企业年报文本数据,

衡量企业对数字技术的利用及重视程度(吴非等,2021)。但这种衡量方式实质上更多反映的是企业“怎么说”而非“怎么做”,企业可以为了蹭热点话题而在管理层分析与讨论中策略性地增加数字技术相关内容,进而产生较大的估计偏差(黄先海等,2023)。也有研究基于专利IPC分类号识别数字技术相关专利,进而对企业的数字技术水平进行衡量,此类方法也可能存在识别不准确、类目过于宽泛等缺陷(罗佳等,2023),导致估计结果夹杂其他混淆因素。本文借鉴创新领域内前沿文献的研究方法,利用大语言模型、文本分析以及机器学习等文本数据挖掘技术,对专利摘要文本进行精确识别,首度量化各地区算法创新水平,并以此衡量企业利用数据要素提升生产效率的能力。该方法创新不仅有效缓解了传统测度方法难以准确衡量算法的问题,也为后续从算法视角探讨数字经济相关研究提供了可靠的度量指标。

二、理论分析与研究假说

进入数字经济时代,数据已经成为新的生产要素,其赋能作用本质上体现在数据要素与实体经济的协同融合(洪银兴和任保平,2023;袁淳等,2021)。数据要素能够发挥“黏合剂”作用,凭借其跨界融合性与其他传统要素结合,形成新的要素组合和要素结构,促进新质生产力的形成(李海舰和赵丽,2021;徐翔等,2024)。数据要素对传统要素的赋能会加深要素之间的关联性,促进企业改变研发决策、优化要素配置,进而提升生产效率。除数据要素与传统要素融合作用外,也有学者关注到数据要素还能通过对传统要素的替代效应,作用于全要素生产率。二者的替代关系可通过要素替代弹性表征,具体而言,当替代弹性较高时,企业通过构建智能化管理平台大幅节约人力与管理成本,将资源重新配置至研发创新环节,从而实现降本增效和资源配置优化(郭凯明等,2024;李娅和李卿芸,2025)。

然而,数据要素本身不是实际生产力,其能否产生赋能作用取决于现实中处理数据的“工具”,即算法。算法技术的研发具有较高的门槛和专业性,且先进技术的推广和使用具有显著的本地化特征(Ganguli等,2020;Balsmeier等,2023)。一个地区的算法创新水平不仅代表了该区域在人工智能与大数据技术上的原始创新能力,更意味着该区域具备了丰富的算法技术供给、成熟的数据服务市场以及活跃的技术扩散环境。对于微观企业而言,身处算法创新高地意味着能够以更低的成本获取外部先进的算法解决方案或技术服务(王雅洁和韩孟亚,2021),从而依靠先进的算法实现自身数据的高效利用,将区域层面的“技术红利”转化为企业内部的“治理红利”。考虑到企业收集、利用的数据主要来自生产活动的投入端和产出端两个渠道,本文尝试从这两个维度的数据利用视角揭示地区算法创新赋能企业全要素生产率提升的内在机制。

地区算法创新有利于促进投入端的数据利用,进而缓解信息摩擦并实现要素

优化配置。在投入端,企业面临的核心挑战是如何在信息不完全以及不对称的条件下实现生产要素的最优配置(Blouch等,2021)。传统生产模式下,由于企业内部各环节存在严重的信息壁垒与“信息孤岛”,管理者难以实时掌握设备运行状态、劳动力技能匹配度以及资本回报的动态变化,导致各种“信息摩擦”广泛存在(David等,2016)。这种摩擦直接导致了产能闲置、劳动力错配和资本低效等资源配置扭曲现象,抑制了全要素生产率的提升(王京滨等,2024)。

地区算法创新则为企业突破上述困境提供了可能。首先,区域内活跃的算法技术环境能够为企业提供更精准的人力资源匹配工具。通过引入先进的匹配算法,企业能够挖掘有效信息,精准刻画劳动力的人力资本特征与岗位需求,将合适的人才配置到最需要的岗位上,从而有效纠正劳动力错配,提高人力资源使用效率(丛屹和俞伯阳,2020)。其次,依托地区算法技术,企业能够获得更优的投资决策支持工具,基于数据分析精准识别高回报项目,引导资本流向高效率部门,从而缓解资本错配(文雯等,2025)。最后,地区算法创新水平的提升意味着区域内存在丰富的数据处理技术与智能化服务,企业能够以较低成本引入先进的智能排产与预测性算法。这些外部技术的引入能够帮助企业对生产流程中的海量数据进行实时采集与分析,动态优化设备负荷与生产安排,从而显著提升企业的产能利用率,减少由于信息滞后导致的资源闲置(陶锋等,2023a)。基于此,本文提出研究假说1。

假说1:算法创新能够通过缓解信息摩擦,纠正劳动与资本要素错配,提升产能利用率,进而从投入端促进企业全要素生产率提升。

在产出端,企业的核心目标是将生产产品高效地转化为市场价值,这一过程高度依赖对市场需求的精准识别与快速响应。然而,传统模式下,企业面临消费者需求碎片化、营销渠道分散化以及资金回笼周期长等多重挑战,导致产品与市场错配,制约了全要素生产率的提升(Eeckhout和Veldkamp,2022)。

活跃的地区算法创新为解决上述问题提供了强有力的外部工具和服务支持。首先,算法技术的本地集聚,使区域内的大数据分析服务、用户画像工具和行业知识库日益成熟。企业借助这些外部技术服务,能够深度挖掘消费者行为数据,精准识别潜在需求与细分市场,从而驱动产品研发的差异化与迭代创新(宋德勇和陈梁,2024;李宏等,2023)。其次,地区算法创新优化了区域内的营销技术环境。借助区域内普及的智能推荐算法与数字化营销平台,企业可以实现广告的精准投放与销售渠道的自动匹配(Purnomo,2023)。这种技术赋能使得企业能够大幅减少无效营销投入,优化营销资源配置,从而降低单位收入所需的销售费用,即降低了销售费用率,显著提升销售资源使用效率(Bushehr,2024)。最后,地区算法创新提升了企业的资金周转效率。在需求精准识别与营销效率提升的基础上,企业能够更快地实现存货变现,缩短从生产到销售的周期,从而提升运营资本周转率。同时,

随着销售回款能力的增强和现金流的改善,企业在供应链中的议价能力提升,能够优化付款节奏,进而推动应付账款周转率的合理调整。基于此,本文提出研究假说2。

假说2:算法创新能够通过促进市场需求精准识别,增强企业创新能力,降低销售费用率并加速资金周转,进而从产出端促进企业全要素生产率提升。

三、算法创新的定义与识别

(一)算法专利的定义和人工识别

算法创新的识别与度量是实证研究的核心难点。新专利的授权需要对现有专利进行扩展和完善,因而在本文的分析框架中,将新授权的算法专利定义为算法创新^①。新授权的算法专利越多,意味着企业利用数据要素的能力就越强。但直接采用企业层面的算法专利授权量衡量企业数据要素利用能力会存在以下问题:首先,持证企业与非持证企业间存在显著差异,易导致严重的样本选择偏误;其次,算法专利的开发方与实际使用方常出现分离,存在空间上的不匹配问题,比如平台企业往往对外提供算法服务,而非直接利用专利处理内部数据。此外,地区间的知识溢出往往受到行政边界和地方保护主义的限制,导致知识传播具有明显的本地化特征(陈志远等,2025;刘修岩和王峤,2022)。同时,语言边界通过削弱知识传播能力、增加信息沟通成本使专利引用活动局限于城市内部(Hussler,2004),进一步强化了知识溢出的本地化特征。考虑到上述问题,本文将算法创新扩展到城市层面,各城市新增算法专利授权数量越多,城市内部企业利用数据要素的能力也就越强,从而能够更大限度地发挥数据要素的赋能作用。同时,由于算法和数据之间存在高度互补关系,城市内部算法专利的增加,可以激励企业收集、利用数据要素,进而提升全要素生产率。

据此,本文研究的难点转化为“算法专利”的识别。尽管IPC分类号是识别专利主题的传统工具,但其存在显著局限:一是IPC重功能分类而轻应用分类,难以精准映射算法主题;二是分类类目粗泛,同一号段下包含多种异质技术,无法实现特定领域的精确识别(刘青和肖柏高,2023)。因此,有必要重点关注专利内容本身,将“专利类型分析”回归到对专利摘要文本的处理。由于专利数量以百万计,采用人工方式对这些专利进行筛选和分类不具可行性。本文将基于MacBERT自然语言处理模型,设计专利主题分析模型,通过对专利摘要文本的分

^① 在创新经济学的研究中,通常将技术进步看成是产品种类的增加(Variety Model)或者是产品质量的提升(Quality Ladder),两者对经济增长具有等价的作用。在本文的框架下,无论是数据处理算法种类的增加还是质量的提升,均可以提高企业的数据处理能力,能够更大限度地发挥数据要素对实体经济的赋能作用,为方便表述,本文将算法专利的增加表述为“算法创新”,而不做进一步区分。

析识别算法专利。首先以自然语言定义分类规则,其次将分类规则文本与专利摘要文本通过 MacBERT 模型进行向量化处理,计算余弦相似度,最后利用阈值筛选算法专利。

算法是指一组明确定义的计算步骤,其本质是依据特定的逻辑规则和方法,对输入的数值或数据集进行处理,以生成相应的输出(Cormen 等,2022)。此外,算法不仅是执行计算任务的具体指令,更包含支撑这些指令的逻辑规则与方法。随着大数据、人工智能等技术的发展,算法在数据处理、模式识别和自动决策等方面的作用愈发突出。算法是发挥数据要素潜在价值的主要驱动力(Jones 和 Tonetti, 2020),因此本文将算法专利界定为能够对特定领域的数据进行处理、分析或利用的技术。基于上述认识,本文从以下两个维度对算法专利作进一步说明:

定义 1: 此类专利应属于数据处理、分析或利用领域的技术创新,通过对已有数据的处理与应用提升技术水平或数据利用价值,通常包含关键词“数据”“处理”。

定义 2: 此类专利的核心改进应为算法或方法层面的创新,通过优化数据处理算法或数据应用方法提升数据利用效能,通常包含关键词“算法”“方法”。

下面在表 1 中列举了一些按上述定义进行分类的具体例子,以更好地阐明这种分类方法的核心思想。在下面的例子中,专利 1 公开了一种利用车辆行驶视频的道路监控系统,其对车辆行驶时的视频数据进行分析并加以利用,但这项专利的核心是设计一套硬件设备系统来支撑系统的运行,符合上述定义 1 但不符合定义 2。专利 2 通过优化 RFID 设备传输能耗提升设备性能,符合定义 2 但不符合定义 1 中“提升了原有数据的利用价值”的算法或方法,因此也不是本文筛选的目标专利。专利 3 是一种基于数据挖掘算法分析电能检测大数据的方法,通过改进分析方法从海量数据中自动提取有价值信息,提升了数据利用价值,同时符合定义 1 和定义 2。

表 1 算法专利示例

序号	公布号	摘要	定义 1	定义 2
1	CN102750755A	本发明涉及一种交通道路监控系统,……模拟信息数字化并保存下来提供实时同步的照片或视频……包括 SOC 视频信号主处理器, GPSGPS 接收天线……GPS 导航定位模块,本发明适用于各种车辆行驶定位视频安全系统	符合	不符合
2	CN103208019A	本发明涉及无线射频识别中的标签防碰撞技术,具体地说是适用于 RFID 系统的多叉树防碰撞算法……不仅减少了数据的传输,降低了能量消耗,而且消除了空闲时隙,缩短了识别时间	不符合	符合

(续)

序号	公布号	摘要	定义1	定义2
3	CN108694175A	本发明公开一种基于数据挖掘算法分析电能检测大数据的方法……结合决策树算法与组合支持度关联算法从海量的数据库中提取有用的数据信息……从大量的数据中自动分析获得规律,使用方便	符合	符合

上述对算法专利的定义和相关例子,基于自然语言以文本的形式定义了本文期望筛选出的专利类型,符合人类直觉。但对计算机来说,要理解复杂的人类语言并非易事。因此,本文以计算机自然语言处理领域的最新技术 MacBERT 模型为基础,设计针对专利主题分析的模型,结合专利摘要文本信息对不同的专利进行识别和分类,最终寻找出符合本文定义的算法专利。下面将对 MacBERT 模型进行详细介绍,并结合其优势说明选择该模型的具体原因。

(二)构建基于 MacBERT 的专利主题分类模型

随着自然语言处理(NLP)技术进入神经网络时代,基于专利文本的自动分类成为可能。早期研究多采用 Word2vec 方法进行主题关联演化分析(徐红姣等,2018)。然而,Word2vec 本质上是基于关键词的向量化模型,在计算前需进行分词与关键词提取,这一过程极易受分词差异影响,产生累积性的语义误判,难以胜任专利摘要等复杂长文本的精确分类。BERT(双向编码器表示模型)的出现克服了上述局限,通过变换器架构实现了文本特征的全局空间向量表示,让计算机直接处理完整语义成为现实。针对中文专利文本分析需求,Cui 等(2020)基于中文语料预训练并改进了 MacBERT 模型。相较于传统 BERT,MacBERT 显著增强了对中文语义的捕捉能力,并修正了“预训练与微调阶段不一致”的问题,在下游任务中表现出更高的有效性。

已有研究基于 BERT 模型和 SoftMax 分类器模型对劳动节约型专利进行了分类和筛选(刘青和肖柏高,2023)。但在本文分类筛选算法专利的过程中,使用分类器模型会带来比较严重的问题:在使用 SoftMax 分类器模型前需要使用人工标注数据对其进行训练,使其学习分类样本之间的差异,而后基于学习得到的模型对专利摘要文本进行分类。该方法适用于不同类别的文本数量相对均衡的情形,例如劳动节约型专利在总专利数量中的占比接近 40%,正负样本数量较为均衡。而本文的研究在对专利库进行抽样标记后发现,在 3000 个专利样本中仅有 2% 左右的专利属于本文所定义的算法专利,存在严重的类别不平衡问题,这将导致分类器模型对数量较多的类别产生过度拟合,而忽略数量较少的类别。

为进一步优化专利主题识别模型,本文在 MacBERT 预训练模型的基础上引入了 CoSENT 方法进行微调,并采用 Pooling 池化策略以增强模型对于算法专利的分类

性能。CoSENT方法由苏剑林(2022)提出,专门针对“预训练和微调阶段两者之间存在差距”的问题。该方法通过引入基于余弦相似度的排序损失,使模型的训练目标与下游任务——文本相似度计算——实现了更紧密的匹配。与传统的SoftMax分类器不同,CoSENT无需依赖大量的人工标注数据,而是直接利用文本对之间的相似度关系来优化模型,从而在数据类别不均衡的条件下更具优势。

具体而言,传统的SoftMax分类器容易在样本类别分布不均衡时,对占比更高的类别产生过拟合,而CoSENT模型则通过余弦相似度计算,有效避免了模型对少数类别(如本文定义的算法专利)发生误判的情况。实验结果表明,CoSENT在收敛速度和最终效果方面均优于InferSent和Sentence-BERT等方法。此外,为进一步优化文本嵌入的降维效果,本文引入了Mean Pooling策略,以提升模型在长文本数据处理中的特征提取能力。考虑到专利摘要文本往往较为冗杂且包含大量与分类主题无关的信息,Mean Pooling策略通过对MacBERT模型所有Token的隐藏层输出取均值,从而获得文本的全局特征表示。相较于传统的CLS标记池化方法,Mean Pooling避免了过度依赖特定Token的信息,能够更全面地保留文本的深层次语义特征,有效减少了无效信息对模型性能的干扰。

因此,本文选用CoSENT在STS语料集上微调后的MacBERT模型作为核心神经网络模型,将句子或段落文本转化为向量,成为专利主题分析模型的核心模块。基于上述考虑,本文引入余弦相似度,计算专利摘要文本之间的相似度,并以此解决类别不平衡的问题。余弦相似度是在向量空间模型中常用的相似度度量方法,主要用于比较两个非零向量的方向或者方向上的相似性。该函数的取值范围是 $[-1, 1]$ 。给定两个向量 A 和 B ,它们的余弦相似度可以定义为:

$$\cos(A, B) = \frac{A \times B}{\|A\| \times \|B\|} \quad (1)$$

至此,只需要把专利摘要和预定主题通过MacBERT大语言模型分别计算出两者在空间中的向量表达,就可以利用余弦相似度计算公式(1)对两个向量的余弦相似度进行计算,计算结果即表示专利摘要和预定主题的文本相似程度。

综上所述,本文主要构建了3层结构来分析专利摘要文本与预定主题的相关性,具体流程如图1所示。在计算主题相似度之前,先要明确本文所预定的主题。依照前文对算法专利的定义,本文将分类模型的预定主题设置为:应用在某个领域的数据处理、分析、利用的算法或方法。计算专利摘要和预定主题的相似度,具体流程主要包括以下三个方面:首先,采用基于CoSENT方法微调的MacBERT预训练模型对专利摘要文本和预定主题文本进行文本向量化,将文本文字映射为空间向量;其次,对两个空间向量进行Pooling降维,以确保在保留更多特征的同时,得到固定长度的向量;最后,利用前序步骤得到的向量,计算出两个相同长度向量之间的余弦相似度,从而得到专利摘要文本与预定主题的相

似度。

由于本文设计的专利主题分类模型的输出结果为0~1范围内的相似度指标^①,需要选取合理的临界值作为识别算法专利的阈值。经过人工标注检验和模型筛选,本文最终选定0.61作为本文筛选算法专利的阈值。在这个阈值下,自然语言模型筛选出的专利与人工标定的专利有超过80%的匹配率,且该阈值所对应的准确率较高,比较适合作为模型判断的临界值。通过前文所述模型和方法对2010~2021年中310万条专利数据的摘要文本与预定主题的相关程度进行了计算,最终从310万条专利数据中筛选出近5万条算法专利数据。考虑到算法专利具有准公共物品的特征,同时也为了避免严重的内生性问题,本文选择将算法专利加总到城市层面,考察地区算法专利对其内部企业全要素生产率的促进作用。限于篇幅,有关识别阈值设定的过程见附录A^②。

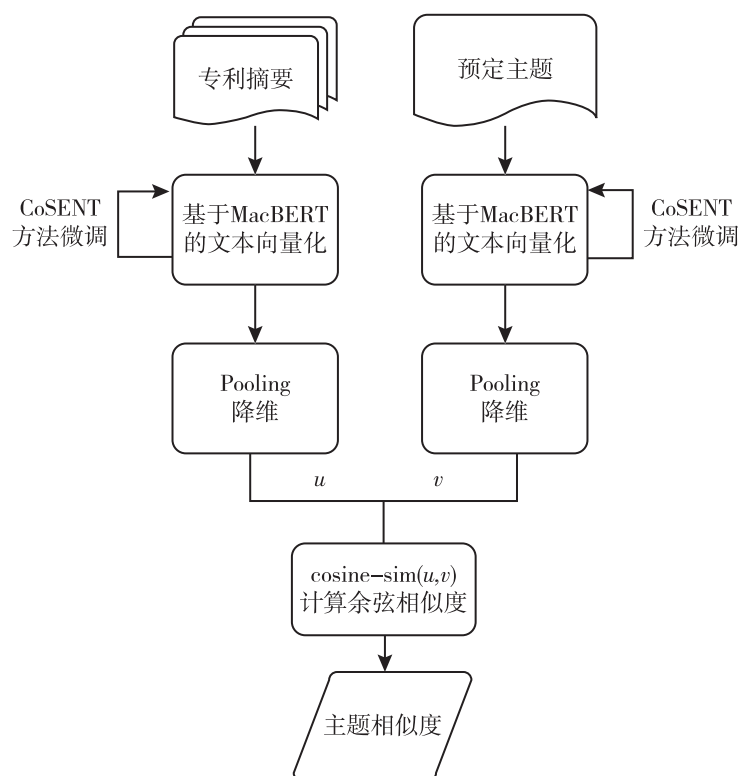


图1 专利主题分析模型结构

① 虽然余弦相似度本身范围为 $[-1, 1]$,但考虑到在文本处理中,常用的向量化方法生成的向量基于词频、权重或语义编码,这些值通常是非负的。正是如此,本文的预定主题与专利摘要文本这两个非负向量之间的余弦相似度不会出现负值。所以本文输出的余弦相似度结果在0~1范围内。一般情况下,余弦相似度的值越接近1,表示两个向量的相似度越高。

② 本文附录详见《数量经济技术经济研究》杂志网站,下同。

四、样本选择与模型构建

(一)样本选择

本文研究使用的样本数据主要来自以下两个数据库:其一,企业地理信息和财务数据来自国泰安数据库,以2010~2021年中国A股上市公司作为研究样本。其二,专利文本信息来自国家知识产权局专利数据库。本文选择了2010~2021年间由企业申请的发明专利以及实用新型专利数据,共计约310万条,具体包括专利所属公司名称、所属地区、申请文件摘要和权利要求等文本内容,能够为本文研究提供详细的专利文本信息。

为获得回归分析所需要的面板数据,本文对上述不同来源的数据进行了如下处理:首先,本文对上市公司数据做以下处理:剔除ST、ST*、PT类公司样本;剔除金融行业相关公司样本;剔除IPO未满一年的公司样本;剔除总资产小于总负债的企业样本。经过以上处理,本文得到公司一年度的样本观测值共计26832个。其次,为缓解内生性问题,本文将企业算法专利数按年份汇总到城市层面,再将该数据与上市公司的样本进行匹配,得到各企业所在城市相对应的算法专利数。之所以将算法专利数据汇总至城市层面是基于以下考虑:创新技术在实现商业化过程中,往往需要与其他创新成果和专利相互嵌套(吕铁和李载驰,2021)。在此过程中,城市内部制度壁垒相对较低,有利于知识溢出。因此,将核心解释变量算法专利数量加总到城市层面更具合理性。为缓解极端值对估计结果的影响,本文对连续型变量进行了前后各1%的缩尾处理。

(二)变量定义

为系统考察算法创新对企业全要素生产率的促进作用,本文采用企业全要素生产率作为核心被解释变量。全要素生产率是衡量企业综合生产效率的重要指标,在研究中得到了广泛的应用。估计企业全要素生产率的经典方法主要包括OP法(Olley和Pakes,1996)、LP法(Levinsohn和Petrin,2003)以及ACF法(Ackerberg等,2015)等。OP法以投资作为代理变量,且要求投资与总产出保持单调关系。而本文所采用的数据中,投资额数据缺失较为严重,这意味着无法估算样本中投资额为0的企业全要素生产率(鲁晓东和连玉君,2012)。为此,本文借鉴鲁晓东和连玉君(2012)以及Levinsohn和Petrin(2003)的研究,采用LP法估算的全要素生产率作为核心被解释变量。此外,在后续的稳健性检验部分,也采用其他多种方法计算的企业全要素生产率,对基准回归进行稳健性检验。

本文关注的核心解释变量为各城市的算法创新。考虑到年报关键词词频法易受企业主观包装干扰(黄先海等,2023),本文采用客观性更强的专利数据进行度量。参照刘青和肖柏高(2023)的研究思路,本文基于大语言模型对专利摘要进行

深度语义识别,从海量专利中精准筛选出具有数据处理与分析功能的算法专利。相比于年报词频,基于专利申请行为的度量方式更客观地反映了企业实际的技术研发与应用路径,能够更精准地刻画各地区在算法驱动数据要素利用方面的真实水平。

考虑到企业特征可能影响其全要素生产率,本文控制了企业层面的相关变量,以缓解遗漏变量带来的内生性问题。为缓解异方差问题,对除比例变量之外的连续变量取自然对数。参考黄勃等(2023)的研究,本文在企业层面控制以下变量:企业成长速度(*grow*),采用营业收入的年增长率进行衡量;资产负债率(*leve*),采用总负债与总资产之比进行衡量;企业盈利能力(*profit*),采用净利润与营业收入之比进行衡量;流动比率(*cratio*),采用流动资产与总资产之比进行衡量;企业规模(*lnsize*),采用企业职工人数取自然对数进行衡量;企业性质(*state*),企业为国有企业取值为1,否则为0。相关变量的描述性统计见附录B。

(三)计量模型设定

为实证检验算法创新对企业全要素生产效率的影响,本文构建如下计量模型:

$$tfp_{it} = \beta_0 + \beta_1 \ln algo_{ct} + \beta_2 X_{it} + u_i + v_c + w_t + \sigma_j + \varepsilon_{ijct} \quad (2)$$

其中,下标*i*表示企业,*t*表示年份,*tfp_{it}*表示企业全要素生产率,*lnalgo_{ct}*表示企业所在城市*c*当年新授权的算法专利数量,用以反映该地区的算法创新水平。考虑到算法具有非竞争和非排他等准公共物品的属性,本文采用各地区新增算法专利作为衡量企业数据利用能力的代理指标。*X_{it}*为企业层面相关控制变量,*u_i*、*v_c*、*w_t*和*σ_j*分别表示企业、城市、年份以及行业层面的固定效应,*ε_{ijct}*为随机扰动项。考虑到行业内部企业之间TFP的相关性大于地区内不同行业企业TFP的相关性,参考现有研究(黄先海等,2016;刘成奎和王浩,2025;卢福财等,2024),本文采用行业层面聚类稳健标准误。

五、实证结果分析

(一)基准回归结果

本文的基准回归结果如表2所示。列(1)、列(2)表明,算法创新(*lnalgo*)对企业全要素生产率的回归系数均显著为正,意味着地区算法创新能够促进企业全要素生产率提升。根据列(2)的估计结果,若地区算法专利数量翻倍,能够使得该地区代表性企业的全要素生产率提高1.7%。基准回归估计结果表明,地区算法创新水平的提升,会显著增加该地区企业获取和利用前沿算法的机会,增强对数据要素的处理及利用能力,进而提升企业全要素生产率。

表2 算法创新与全要素生产率:基准回归结果

变量	(1)	(2)
	tfp_{LP}	tfp_{LP}
$\ln algo$	0.016** (0.007)	0.017*** (0.006)
常数项	10.656*** (0.200)	10.677*** (0.196)
样本量	18820	18820
R ² 值	0.826	0.835
控制变量	是	是
企业固定效应	是	是
城市固定效应	是	是
年份固定效应	是	是
行业固定效应	否	是

注:*、**、***分别表示在10%、5%、1%的水平上显著,括号内行业层面聚类稳健标准误。

(二)稳健性检验

本文开展了一系列稳健性检验,包括替换TFP估计方法、控制行业一年份固定效应、加入城市固定效应与线性时间趋势交互项、更换聚类稳健标准误、剔除直辖市、剔除专利数量最多的五个城市样本、剔除计算机软件行业、对地区算法专利进行标准化、采用不同阈值筛选的算法专利以及直接使用企业层面算法创新对企业全要素生产率进行回归等,结果均支持基准回归的相关结论,上述稳健性检验的估计结果见本文附录C。

(三)潜在内生性问题

尽管本文在回归方程中尽可能控制了影响企业全要素生产率的相关变量,但仍可能存在反向因果或遗漏变量导致的内生性问题。为此,本文参考Acemoglu和Restrepo(2020)构建机器人渗透率的思路,采用企业所在地级市到光缆骨干节点(1998年)的最小距离与全球(剔除中国)算力规模的乘积取自然对数^①,作为算法创新的工具变量(IV)。

一方面,地理上更靠近光缆骨干节点的城市更加有利于低成本进行数据交换以及获取外部技术服务,从而更有利于算法研发与落地。与此同时,全球算力规模

^① 由于中国统计算力规模时间较晚,这里仅采用2016~2021年算力规模与最小距离相乘,全球算力规模数据来源于中国信通院《中国算力发展指数白皮书》。

反映了世界范围内可用算力资源的发展态势,当全球算力供应充足且外部算力发展快速时,即便本地算力有限,城市通过带宽、跨境数据或计算服务也更容易获得算力支持以开展算法训练,因此该工具变量具有较强的相关性。另一方面,骨干光缆节点的布局源于1998年建成的“八纵八横”光缆传输骨干网,其选址主要基于当时的地理条件和通信布局考虑,与企业后续生产率提升无关,具有历史外生性。同时,全球算力规模(剔除中国部分)反映的是国际技术环境的总体趋势,不会直接影响中国企业的生产率表现,因此满足外生性假设。

结果如表3所示,在列(1)工具变量 IV 对内生变量 $\ln algo$ 的回归中, IV 的系数在1%的水平上显著为负,意味着与光缆骨干城市地理距离越近,地区算法专利的数量越多,满足工具变量相关性的要求。由于第一阶段F统计量大于临界值,可以拒绝弱工具变量的原假设。列(2)报告了第二阶段估计结果,变量 $\ln algo$ 的回归系数为0.415,表明引入工具变量缓解内生性问题后,基准回归结果依然成立,意味着各地区算法创新对该地区内部企业全要素生产率的促进作用确实存在。

表3 算法创新与全要素生产率:TSLS

变量	(1)	(2)
	$\ln algo$	tfp_{LP}
IV	-1.531*** (0.193)	
$\ln algo$		0.415*** (0.129)
控制变量	是	是
企业固定效应	是	是
城市固定效应	是	是
年份固定效应	是	是
行业固定效应	是	是
K.P.F统计量	62.731	
样本量	4922	4922

注:同表2。

(四)潜在机制分析

上述结果表明,地区算法创新能够显著促进内部企业全要素生产率提升。进一步地,本文基于投入端与产出端数据利用的双重视角,对地区算法创新作用于企

业全要素生产率的潜在机制展开检验,以验证理论分析部分的研究假说。

1. 投入端数据利用机制

地区算法创新水平的提升能够为企业高效利用数据提供更多解决方案和专业化的服务,进而推动企业在生产经营决策过程中更广泛地利用数据分析进行智能化决策。现有研究表明,信息摩擦是影响发展中国家资源配置效率的重要因素(David等,2016)^①。而算法技术的嵌入有助于企业在生产计划、要素投入等环节充分利用现有信息开展智能决策,从而减少由于信息摩擦带来的资源错配问题,提升企业全要素生产率。为实证检验上述作用机制,本文借鉴Jiang和Li(2024)的研究,构造了劳动力错配(τl)和资本错配(τk)指数,以此衡量劳动和资本要素偏离最优投入状态的程度。该指数数值越大,表示生产要素错配程度越高。回归结果如表4列(1)、列(2)所示,算法创新 $\ln algo$ 的估计系数均显著为负,说明在算法创新更活跃的地区,企业能够更有效地利用数据要素,从而有效缓解信息摩擦带来的生产要素错配问题,提升企业全要素生产率。

此外,参考李雪松等(2017)的研究,将企业的实际产出与前沿产出的比值作为产能利用率的衡量指标,并采用随机前沿生产函数法进行估计。具体而言,利用企业的总产出、资本存量和劳动力投入构建随机前沿生产面,从而估算各企业的产能利用率水平。地区算法创新对企业产能利用率的估计结果如表4列(3)所示,结果表明算法创新能显著提升企业产能利用率,假说1得到验证。

表4 作用机制分析I——投入端数据利用

变量	(1)	(2)	(3)
	τl	τk	CU
$\ln algo$	-0.116** (0.058)	-0.111* (0.059)	0.124*** (0.045)
常数项	-2.209 (1.584)	7.197*** (1.344)	77.221*** (1.426)
样本量	13472	13472	18820
R ² 值	0.595	0.631	0.794
控制变量	是	是	是
企业固定效应	是	是	是
城市固定效应	是	是	是

① 该研究表明,信息摩擦将导致中国的全要素生产率和产出分别损失7%和10%。由此可见,充分利用可获得的信息,是缓解资源错配的重要抓手。

(续)

变量	(1)	(2)	(3)
	τl	τk	CU
年份固定效应	是	是	是
行业固定效应	是	是	是

注:同表2。

2. 产出端数据利用机制

在产出端,地区算法创新通过增强企业对市场信息与消费者需求的识别能力,有效提升其创新能力,研发出更加符合消费者需求的产品,从而降低销售费用并提高资金周转效率。算法技术使企业能够更精准地刻画消费者偏好、优化产品研发策略并细化市场定位,从而强化市场竞争优势。为验证上述作用机制,分别采用企业当年专利申请数量(*Patents*)及发明专利申请数量(*Invention*)衡量创新能力;采用销售费用率(*SaleExpRatio*)衡量企业的销售效率,具体以上市企业销售费用与营业收入之比进行衡量;采用营运资本周转率(*WCTurn*)以及应付账款周转率(*PayableTurn*)作为资金使用效率的衡量指标,估计结果如表5所示。表5列(1)及列(2)的结果表明,变量 $\ln algo$ 对企业创新具有显著的促进作用。列(3)结果显示销售费用率的回归系数显著为负,表明算法创新显著降低了企业的销售费用率。列(4)及列(5)的结果表明,算法创新能够促进资金周转率提升。由以上结果可知,算法创新通过提高企业创新能力、销售效率以及资金使用效率,进而提升企业全要素生产率,假说2得到验证。

表5 作用机制分析 II——产出端数据利用

变量	(1)	(2)	(3)	(4)	(5)
	<i>Patents</i>	<i>Invention</i>	<i>SaleExpRatio</i>	<i>WCTurn</i>	<i>PayableTurn</i>
$\ln algo$	0.052** (0.023)	0.067* (0.039)	-0.098* (0.054)	0.358* (0.186)	0.161* (0.088)
常数项	1.129 (1.077)	0.145 (0.878)	7.341*** (0.981)	3.978 (2.911)	5.549** (2.181)
样本量	12163	12050	18820	15779	18780
R ² 值			0.909	0.589	0.810
控制变量	是	是	是	是	是
企业固定效应	是	是	是	是	是
城市固定效应	是	是	是	是	是
年份固定效应	是	是	是	是	是
行业固定效应	是	是	是	是	是

注:前两列采用高维固定效应面板泊松模型进行估计。其余同表2。

(五) 异质性分析

考虑到企业所处地区、行业差异,可能导致算法创新对企业全要素生产率的影响存在异质性,本文分别从地区基础设施水平、政府数据开放水平以及行业需求波动三个方面进行异质性分析,结果见附录D。

第一,数字基础设施是数据要素流通与应用的载体。基础设施越完善,数据传输与处理效率越高,越有助于企业将算法创新转化为生产力。本文借鉴戴魁早等(2024)的研究,采用城市人均互联网用户数衡量数字基础设施水平,并以样本均值为界,将城市划分为“数字基础设施较强”与“数字基础设施较弱”两组。回归结果显示,在数字基础设施强的城市,算法创新的估计系数显著为正,且数值更大;而在数字基础设施弱的城市中,该系数不显著。组间差异检验结果进一步验证了两组之间的显著差异,这说明算法创新在数字基础设施完善的环境中能够更有效地提升企业生产率。

第二,数据开放有助于推动数据要素的共享与交易,使得企业在算法创新的基础上能够更充分利用数据要素,进而提高全要素生产率。为此,本文依据复旦大学与国家信息中心数字中国研究院发布的“中国开放数林指数”,参考戴魁早等(2024)对城市数据开放程度分类的界定方式,将指数排名前50的城市定义为“高数据开放”城市,其余城市定义为“低数据开放”城市。回归结果显示,算法创新的促进作用仅在数据开放程度高的城市显著,且通过了组间差异检验,说明数据开放是算法创新效应得以充分释放的重要条件。

第三,从市场环境来看,算法创新通过提升数据处理能力,帮助企业在不确定性环境下做出精准决策。在需求波动较大的行业,市场信号复杂多变,算法创新对识别市场趋势、维持资源配置效率的作用尤为关键。本文参考齐鹰飞和张瑞(2015)的研究,以企业营业收入加总至行业层面后计算得到的变异系数来衡量行业需求波动性,并依据中位数将企业划分为高需求波动组和低需求波动组。回归结果显示,在行业需求波动较大的组别中,算法创新对企业全要素生产率的影响作用更大,且通过了组间差异检验。

六、研究结论与政策启示

算法创新已经成为企业充分利用数据要素、实现高质量发展的新动力。凭借算法创新,企业能够高效、充分利用海量数据信息,准确发现多样化需求,实现供需双方快速匹配,并显著降低交易成本,从而获得全方面多维度的竞争优势。这些优势为市场和消费者带来了更多创新性的产品和服务,并在提高企业全要素生产率、培育经济发展新动力等方面发挥着越来越重要的作用。本文利用2010~2021年中国上市公司数据,结合MacBERT大语言模型和专利摘要文本,识别出算法专利以度量地区算法创新水平。实证检验算法创新对企业全要素生产率的影响

响,并探索潜在作用机制与异质性作用效果。研究表明:算法创新能够提升企业全要素生产率,该结论在经过一系列稳健性检验后仍然成立。机制分析表明,算法创新能够提高企业在投入端与产出端的数据利用效率,进而提升企业全要素生产率。异质性分析结果表明,算法创新对数字基础设施完善、数据开放程度高的地区以及需求波动大的行业内部企业全要素生产率提升效果更为明显。本文基于专利文本信息构建了地区算法创新水平的度量指标,为算法创新的影响提供了新的经验证据,表明算法创新已成为推动企业形成新质生产力的关键因素,能够有效破解数字经济时代的“索罗悖论”。

基于上述研究结论,可以得到以下政策启示。

第一,构建梯度化算法创新支撑体系,激活算法技术创新活力。算法创新已经成为驱动数据要素赋能企业全要素生产率的核心引擎,未来应加快构建与算法技术属性特征相匹配的政策支持体系。具体而言,一方面,政府应加快构建与算法技术属性相匹配的政策支持机制,通过设立财政专项基金、完善研发费用加计扣除与研发补贴等激励措施,降低企业算法研发成本,重点支持人工智能、大数据分析、机器学习等关键领域的基础研究与产业化落地。财政专项基金可重点投向算法基础理论与共性技术研究,夯实创新底座。针对研发投入达到一定比例的企业制定算法研发费用专项扣除政策,并提供算法研发补贴激励,逐步构建从基础研究到应用开发再到场景落地的全链条政策扶持机制。另一方面,完善算法知识产权保护与市场化转化机制。考虑到算法专利存在的价值评估难、交易成本高等问题,可考虑推动建立算法专利池、技术评估与许可制度,通过集中许可与交叉授权缓解交易摩擦,降低企业获取核心算法技术的交易成本,促进全链条扶持体系的形成。

第二,推动算法深度嵌入企业生产经营决策,提升要素配置效率与产品市场价值。本文研究发现,算法创新通过提高产能利用率、减少劳动与资本错配、增强企业创新能力、加速资金周转以及降低销售费用率来提升企业全要素生产率。这一结论表明,算法技术的经济效应关键在于深度嵌入企业的生产组织与市场决策过程。为此,应引导企业将算法系统性应用于生产计划制定、设备调度与供应链协同等核心环节,推动形成数据驱动的智能决策体系,提升资源配置的精准性与动态调整能力,减少要素闲置与结构性错配。同时,应推动算法技术在研发创新与营销管理中的规范化与高效化应用,支持企业利用算法精准识别市场需求以驱动创新,提高营销资源配置效率,降低无效销售投入并提升资金使用效率,从而改善企业经营绩效。通过构建算法创新与企业运营深度融合的应用生态,可进一步释放算法技术在提升企业效率与推动高质量发展中的潜在效应。

第三,构建差异化的算法应用支持体系,优化数字化发展环境。异质性分析表明,在数字基础设施完善、数据开放程度较高的地区,算法创新对企业全要素生产

率的促进作用更为显著。因此,有必要针对不同区域的数字化发展条件,建立差异化的支持政策。一方面,对于数字基础设施较弱的地区,应加快5G、工业互联网、云计算等新型基础设施建设,推动重点行业和产业园区率先实现网络化、智能化改造,提升企业算法应用的承载能力。对于数据开放程度较低的地区,则应着力破除部门与行业壁垒,完善公共数据开放平台,推动政府数据与产业数据有序流通,形成共享机制。另一方面,在数字基础设施和数据开放程度较高的地区,应进一步强化政策引导,鼓励企业拓展多元化的算法应用场景,建设跨行业的算法应用示范平台,推动企业经验推广与资源共享。此外,鉴于算法在需求波动大的行业中发挥着更为突出的作用,未来应重点引导和支持此类行业加速算法赋能,利用先进算法模型增强市场预测与动态响应能力,应对市场波动风险,进而提升产业链韧性与资源配置效率。

参考文献

- [1]蔡跃洲,马文君.数据要素对高质量发展影响与数据流动制约[J].数量经济技术经济研究,2021,(3):64~83.
- [2]陈荣达,林祺,金骋路,史守学.数据资产估值定价与新质生产力发展:演进逻辑与主要挑战[J].财贸经济,2024,(8):33~51.
- [3]陈志远,于皓,张杰.中国知识溢出本地化的行政边界效应研究[J].中国工业经济,2025,(1):81~99.
- [4]丛屹,俞伯阳.数字经济对中国劳动力资源配置效率的影响[J].财经理论与实践,2020,(2):108~114.
- [5]戴魁早,黄姿,梁银笛.数据要素与服务型制造发展[J].经济研究,2024,(12):95~112.
- [6]丁晓东.论算法的法律规制[J].中国社会科学,2020,(12):138~159+203.
- [7]郭凯明,王钰冰,杭静.数据要素规模效应、产业结构转型与生产率提升[J].中国工业经济,2024,(8):5~23.
- [8]洪银兴,任保平.数字经济与实体经济深度融合的内涵和途径[J].中国工业经济,2023,(2):5~16.
- [9]黄勃,李海彤,刘俊岐,雷敬华.数字技术创新与中国企业高质量发展——来自企业数字专利的证据[J].经济研究,2023,(3):97~115.
- [10]黄先海,高亚兴.数实产业技术融合与企业全要素生产率——基于中国企业专利信息的研究[J].中国工业经济,2023,(11):118~136.
- [11]黄先海,王瀚迪,孙涌铭,虞柳明.数字技术与企业出口质量升级——来自专利文本机器学习证据[J].数量经济技术经济研究,2023,(12):69~89.
- [12]黄先海,诸竹君,宋学印.中国出口企业阶段性低加成率陷阱[J].世界经济,2016,(3):95~117.

- [13]李海舰,赵丽.数据成为生产要素:特征、机制与价值形态演进[J].上海经济研究,2021,(8):48~59.
- [14]李宏,梁宝琦,曹知修.数字化转型对企业成本加成率的影响研究[J].湖南大学学报(社会科学版),2023,(4):65~76.
- [15]李万利,潘文东,袁凯彬.企业数字化转型与中国实体经济发展[J].数量经济技术经济研究,2022,(9):5~25.
- [16]李雪松,赵宸宇,聂菁.对外投资与企业异质性能利用率[J].世界经济,2017,(5):73~97.
- [17]李娅,李卿芸.数据要素价值研究进展[J].经济学动态,2025,(12):189~204.
- [18]刘成奎,王浩.市场竞争与企业内共同富裕——来自《反垄断法》实施的准自然实验[J].数量经济技术经济研究,2025,(3):177~198.
- [19]刘青,肖柏高.劳动力成本与劳动节约型技术创新——来自AI语言模型和专利文本的证据[J].经济研究,2023,(2):74~90.
- [20]刘淑春,闫津臣,张思雪,林汉川.企业管理数字化变革能提升投入产出效率吗[J].管理世界,2021,(5):170~190+13.
- [21]刘修岩,王峤.知识溢出的边界效应——来自专利引用数据的证据[J].经济研究,2022,(11):84~101.
- [22]卢福财,王雨晨,徐远彬.头部企业在数字化转型中的作用[J].数量经济技术经济研究,2024,(5):92~112.
- [23]鲁晓东,连玉君.中国工业企业全要素生产率估计:1999—2007[J].经济学(季刊),2012,(2):541~558.
- [24]吕铁,李载驰.数字技术赋能制造业高质量发展——基于价值创造和价值获取的视角[J].学术月刊,2021,(4):56~65+80.
- [25]罗佳,张蛟蛟,李科.数字技术创新如何驱动制造业企业全要素生产率?——来自上市公司专利数据的证据[J].财经研究,2023,(2):95~109+124.
- [26]马鸿佳,王春蕾,李卅立,肖彬.数据驱动如何提升数字平台生态系统竞争优势?——基于数据网络效应视角[J].管理世界,2024,(12):170~185.
- [27]齐鹰飞,张瑞.市场集中度与产能过剩[J].财经问题研究,2015,(10):24~30.
- [28]任保平,豆渊博.数据、算力和算法结合反映新质生产力的数字化发展水准[J].浙江工商大学学报,2024,(3):91~100.
- [29]宋德勇,陈梁.数字技术应用对企业全要素生产率的影响效应研究——兼论破解新“索洛悖论”[J].科研管理,2024,(9):33~42.
- [30]苏剑林.CoSENT(一):比Sentence-BERT更有效的句向量方案[EB/OL].[2022-01-06].<https://kexue.fm/archives/8847>.
- [31]陶锋,王欣然,徐扬,朱盼.数字化转型、产业链供应链韧性与企业生产率[J].中国工业经济,2023a,(5):118~136.
- [32]陶锋,朱盼,邱楚芝,王欣然.数字技术创新对企业市场价值的影响研究[J].数量经济

技术经济研究, 2023b, (5): 68~91.

[33]王京滨,刘赵宁,刘新民.数字化转型与企业全要素生产率——基于资源配置效率的机制检验[J].科技进步与对策, 2024, (3): 23~33.

[34]王雅洁,韩孟亚.高技术产业集聚与创新绩效的交互影响及空间溢出效应——创新价值链视角下的空间联立方程研究[J].科技进步与对策, 2021, (12): 59~68.

[35]文雯,肖瑶,牛煜皓.人工智能技术水平与企业投资效率[J].经济管理, 2025, (6): 103~122.

[36]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[J].管理世界, 2021, (7): 130~144+10.

[37]肖红军,沈洪涛,周艳坤.客户企业数字化、供应商企业 ESG 表现与供应链可持续发展[J].经济研究, 2024, (3): 54~73.

[38]徐红姣,曾文,张运良.基于 Word2vec 的论文和专利主题关联演化分析方法研究[J].情报杂志, 2018, (12): 36~42.

[39]徐翔,李帅臻,李涛.数据要素赋能新质生产力形成:逻辑、机理与路径[J].经济社会体制比较, 2024, (6): 41~50.

[40]徐翔,赵墨非.数据资本与经济增长路径[J].经济研究, 2020, (10): 38~54.

[41]许诺,毛聚,毛新述,王彦超.算力部署、数据跨域流动与企业全要素生产率——来自智算中心的证据[J].中国工业经济, 2025, (4): 61~79.

[42]杨壮,吴福象.算力基础设施建设能破解“生产率悖论”吗——基于合作创新和劳动力技能结构的视角[J].经济学家, 2025, (1): 56~65.

[43]袁淳,肖土盛,耿春晓,盛誉.数字化转型与企业分工:专业化还是纵向一体化[J].中国工业经济, 2021, (9): 137~155.

[44]张骁,吴琴,余欣.互联网时代企业跨界颠覆式创新的逻辑[J].中国工业经济, 2019, (3): 156~174.

[45]赵宸宇,王文春,李雪松.数字化转型如何影响企业全要素生产率[J].财贸经济, 2021, (7): 114~129.

[46]Acemoglu D., Restrepo P., 2020, *Robots and Jobs: Evidence from US Labor Markets* [J], *Journal of Political Economy*, 128(6), 2188~2244.

[47]Akerberg D. A., Caves K., Frazer G., 2015, *Identification Properties of Recent Production Function Estimators* [J], *Econometrica*, 83(6), 2411~2451.

[48]Balsmeier B., Fleming L., Lück S., 2023, *Isolating Personal Knowledge Spillovers: Coinventor Deaths and Spatial Citation Differentials* [J], *American Economic Review: Insights*, 5(1), 21~34.

[49]Blouch R., Khan M. M., Shakeel W., 2021, *A Bottom-up Role of Information Asymmetry: Opening the Black-box of Firms' Resource Allocation Mechanism* [J], *Global Knowledge, Memory and Communication*, 72(12), 210~230.

[50]Bushehr I., 2024, *Investigating the Effects of Artificial Intelligence on the Optimization of*

Marketing Processes [J], Quarterly Journal of Advertising and Sales Management, 5(4), 39~58.

[51] Camacho J., Magún-Carrión R., García-Teodoro P., Treinen J. J., 2016, *Networkmetrics: Multivariate Big Data Analysis in the Context of the Internet* [J], Journal of Chemometrics, 30(9), 488~505.

[52] Cormen T. H., Leiserson C., Rivest R., Stein C., 2022, *Introduction to Algorithms, Fourth Edition* [M], Cambridge, MA: MIT press.

[53] Cui Y., Che W., Liu T., et al., 2020, *Revisiting Pre-trained Models for Chinese Natural Language Processing* [J], Association for Computational Linguistics, 657~668.

[54] David J. M., Hopenhayn H. A., Venkateswaran V., 2016, *Information, Misallocation, and Aggregate Productivity* [J], The Quarterly Journal of Economics, 131(2), 943~1005.

[55] Eeckhout J., Veldkamp L., 2022, *Data and Market Power* [R], NBER Working Paper, No.30022.

[56] Ganguli I., Lin J., Reynolds N., 2020, *The Paper Trail of Knowledge Spillovers: Evidence from Patent Interferences* [J], American Economic Journal: Applied Economics, 12(2), 278~302.

[57] Hussler C., 2004, *Culture and Knowledge Spillovers in Europe: New Perspectives for Innovation and Convergence Policies?* [J], Economics of Innovation and New Technology, 13(6), 523~541.

[58] Jiang W., Li J., 2024, *Digital Transformation and Its Effect on Resource Allocation Efficiency and Productivity in Chinese Corporations* [J], Technology in Society, 78, 102638.

[59] Jones C. I., Tonetti C., 2020, *Nonrivalry and the Economics of Data* [J], American Economic Review, 110(9), 2819~2858.

[60] Levinsohn J., Petrin A., 2003, *Estimating Production Functions Using Inputs to Control for Unobservables* [J], The Review of Economic Studies, 70(2), 317~341.

[61] Liu Y., Dong J., Mei L., Shen R., 2023, *Digital Innovation and Performance of Manufacturing Firms: An Affordance Perspective* [J], Technovation, 119, 102458.

[62] Nambisan S., Lyytinen K., Majchrzak A., Song M., 2017, *Digital Innovation Management: Reinventing Innovation Management Research in a Digital World* [J], MIS Quarterly, 41(1), 223~238.

[63] Olley G. S., Pakes A., 1996, *The Dynamics of Productivity in the Telecommunications Equipment Industry* [J], Econometrica, 64(6), 1263~1297.

[64] Purnomo Y. J., 2023, *Digital Marketing Strategy to Increase Sales Conversion on E-commerce Platforms* [J], Journal of Contemporary Administration and Management (ADMAN), 1(2), 54~62.

[65] Svahn F., Mathiassen L., Lindgren R., 2017, *Embracing Digital Innovation in Incumbent Firms: How Volvo Cars Managed Competing Concerns* [J], MIS Quarterly, 41(1), 239~254.

[66] Vial G., 2019, *Understanding Digital Transformation: A Review and a Research Agenda* [J], The Journal of Strategic Information Systems, 28(2), 118~144.

Algorithm Innovation and Promotion of Enterprise Total Factor Productivity: Empirical Evidence from Patent Texts

JIANG Weijie QI Chenyang CHEN Qi

(School of Business, Ningbo University)

Summary: Due to the accelerating and deepening digital economy, computing power, algorithms, and data are critical resources underpinning the digital transformation of the real economy. Among these, algorithm innovation, as the core engine that unleashes the latent value of data elements, is progressively becoming a new kinetic energy driving the high-quality development of enterprises.

Despite its theoretical importance, quantifying algorithm innovation at the micro-level is a significant empirical challenge in the existing economic literature. To bridge this gap, this study utilizes a comprehensive panel dataset of Chinese A-share listed companies from 2010 to 2021. Methodologically, we introduce a novel approach by employing the MacBERT large language model—a cutting-edge natural language processing technique—to process and analyze massive volumes of patent abstract texts. By fine-tuning the model for text classification and semantic similarity analysis, we accurately identify algorithm-specific patents from millions of general innovation patents. Aggregating these identified patents, we construct a robust index to delineate and portray the regional level of algorithm innovation. Utilizing this index, we empirically investigate the causal impact of regional algorithm innovation on firm-level total factor productivity (TFP).

Our baseline empirical findings reveal a robust and positive relationship: algorithm innovation significantly enhances the TFP of enterprises. To ensure the reliability of our estimations and address potential endogeneity issues arising from reverse causality or omitted variable bias, we employ an instrumental variable approach, leveraging historical and geographical technological infrastructure as exogenous shocks. The core conclusion remains highly significant and robust after applying this two-stage least squares estimation, as well as a series of rigorous robustness checks, including alternative TFP measurement methods, adjusting the sample periods, and incorporating high-dimensional fixed effects to control for unobservable confounding factors.

Through a detailed mechanism analysis, we demonstrate that algorithm innovation exerts its empowering effect primarily by substantially improving the efficiency of data utilization on both the input and output sides of the production process. On the input side, algorithm innovation significantly mitigates information friction between a firm and factor markets. By

utilizing advanced matching algorithms and intelligent production scheduling, enterprises can dynamically optimize their resource deployment, effectively correct the severe misallocation of labor and capital factors, and thus maximize capacity utilization rates. On the output side, algorithmic advancements empower enterprises with enhanced capabilities to accurately identify and predict market demand and consumer preferences. This deep market insight actively promotes targeted research and development and product innovation. Furthermore, precision marketing algorithms drastically reduce unit sales expenses and marketing costs, while intelligent supply chain management accelerates inventory monetization and overall capital turnover, thereby realizing a closed-loop enhancement of corporate value creation.

Our heterogeneity analysis uncovers substantial asymmetric effects across different external environments and industrial characteristics. Specifically, the promotional effect of algorithm innovation on TFP is remarkably more pronounced in regions equipped with highly developed digital infrastructure and a higher degree of government data openness. These favorable regional conditions provide the necessary hardware support and data liquidity essential for algorithms to function optimally. Additionally, the productivity-enhancing effect is significantly stronger for enterprises operating in industries characterized by high demand volatility. In such highly uncertain market environments, the predictive power and risk-mitigation capabilities of algorithms are particularly valuable, offering firms a stronger competitive advantage in dynamic resource allocation.

In conclusion, this study not only provides solid micro-level empirical evidence for fully excavating and unleashing the economic value of data elements but also deepens our theoretical understanding of how digital technologies integrate with the real economy. The findings offer profound policy implications for government authorities and corporate decision-makers. To foster high-quality economic growth, it is imperative to establish a comprehensive support system for algorithmic innovation, including subsidizing core research and safeguarding intellectual property. Furthermore, policymakers should strive to optimize the broader digital development environment by accelerating the construction of new digital infrastructure, dismantling data silos to facilitate public data sharing, and formulating differentiated, region-specific support policies to maximize the transformative potential of the digital economy.

Keywords: Algorithm Innovation; Data Factor; Total Factor Productivity; Large Language Model

JEL Classification: O33; D24; L25

(责任编辑:李兆辰;数据编辑:清扬)