

管理学研究中的大数据： 科学问题与未来发展

○ 黄楠 李竹伊 李雪岩 宫大庆

摘要 在数字化时代,数据已经成为一种新的生产要素,被认为是第四次技术革命的关键资源。大数据的开放、共享和流通成为万物互联的基本需求,而大数据多样性的基本特征,也使人思考在管理学研究当中需要解决的基本理论与应用问题。本文深入梳理与分析了“大数据”概念被提出以来,人们认识世界角度与方式发生的颠覆式转变,以及大数据科学与AI技术在应用中所产生的管理学问题,总结指出了大数据的科学描述、数据的联动协调、全域管理目标场景谱系、大数据共享的管理机制,以及AI技术与算法创新等关键问题对管理学在研究范式层面的挑战。面对未来发展问题与挑战,本文提出了相应理论、管理机制等方法对策,并结合相关案例研究说明其未来应用的可行性。

关键词 大数据科学;大数据管理;大数据全域联动;挑战与对策

引言

自大数据的概念被提出以来,海量数据形成的信息流不仅使人类加速步入万物互联时代,也促使人类活动环境从工业社会向数据驱动的社会系统转变。在这个数字化的复杂超巨系统中,众多原本不相关的事物通过数

据传导相互联系,引发了生产生活方式的巨大变革。特别是在工商管理领域,大数据的运用已成为提高决策效率、优化市场策略、加强客户关系管理的关键工具。^[1]在决策支持方面,企业通过分析大数据,可以具备更深入的市场洞察力,从而以数字化决策驱动管理机制变革。^[2]在运营优化方面,数字化决策帮助企业发现生产和运营中的效率提升点和成本节约机会。^[3]在风险管理方面,企业利用大数据分析预测市场趋势和管理风险,如库存和供应链管理。^[4]在客户管理方面,通过客户大数据分析,企业可以提供更个性化的服务和产品。^[5]然而,随着数据资源及其对应的技术逻辑不断演化发展,大数据在使企业发生变革的同时,也带来了许多挑战与问题,这些问题也日益成为管理学领域的关键探索内容。

国家网信办发布的《数字中国发展报告(2022年)》显示,2022年我国数据产量达8.1ZB,同比增长22.7%,全球占比达10.5%,位居世界第二位。早期国家层面已有大数据战略,如今党的二十大报告提出,^[6]构建全域联动、立体高效的国家安全防护体系,是实现国家治理体系和治理能力现代化的重要方面,并且明确作出了“加快建设数字中国”的部署安排。如今,数据早已成为新的生产要素,数据科学及其应用技术飞速发

作者简介 黄楠,中央民族大学管理学院讲师、博士,研究方向为战略管理、创新管理、资本市场;李竹伊(通讯作者),中国人民大学商学院助理教授、博士,研究方向为组织行为学、数字化与人力资源管理;李雪岩,北京联合大学管理学院副教授、博士,研究方向为复杂系统建模与仿真、计算经济;宫大庆,北京交通大学经济管理学院副教授、博士,研究方向为大数据算法分析

基金资助 本文受国家自然科学基金项目(72103019、62276020)、中国国家铁路集团重大项目(N2021S017)资助

展,^[7]大数据的开放、共享、流动已成为当今数字化万物互联时代的基本特征。

随着数据科学的发展,大数据技术的使用不仅使各部门的信息更易于获取,降低了信息不对称导致的交易成本,同时企业可以重新审视和优化其价值链,从而改变企业的边界。^[8]在大数据关键技术的支持下,对内,企业可以更准确地识别和专注于其核心能力;^[9]对外,企业可以重新定位自身在价值链中的位置,从而使价值链重构,产生新的增长点,^[10]为企业及其生态系统创造新价值。此外,数据兼有技术工具与战略工具的双重特征,对企业的边界影响深远。可见,海量数据时代,如何创新与外部数字化世界的交互,在拓展数据资源挖掘广度的基础上,运用数字化方式解构不同研究对象之间的联动性,是企业管理变革的基础。

对企业管理者而言,需要充分应用数据和技术,以使数据技术逻辑能够不断逼近企业的商业逻辑与管理机制,当前,通过数据的跨行业、跨领域、跨时空乃至跨生态的横向全域联动,挖掘表面上毫不相关、甚至相互矛盾的事物之间的内在联系,并使其产生“化学反应”,是数据赋能企业创新的重要动力。这个问题包含数据资源在横向和纵向的联动,需要系统化、全局化的知识视图。^[11]纵向联动是指在同一领域或企业内部,将不同层次或时间序列的数据整合连接,分析数据的关联和趋势。横向联动则是指在不同领域、行业或部门之间构建数据的联动空间,以挖掘跨界关联性,进行综合施策。实际上,这一过程与人脑通过对多渠道信息的广泛交互完成学习,从而实现对认知对象的高层次识别与理解,具有内在一致性。横向联动数据的汇聚,能够降低信息交易成本,带来企业大脑的知识涌现,实现企业内部深层次传导路径的挖掘。

大数据的纵向联动已经取得显著的进展,对于特定行业企业,纵向联动较为普遍,数据联动的研究主要体现在如何拟合缺失数据,^[12]突破数据采集方式的限制,展开高维度、多源异构数据的缺失填补,^[13]从而实现价值链的高度垂直。而在行业的价值链层面,高度数字化也是人才、技术、资本、管理等创新要素实现价值链联动的中枢,^[14]从产业的基础研究、产品设计再到市场拓展,横向联动的数字化生态体系(如企业统一社会信用代码)能够实现各要素、资源、需求的精准对接。例如,章琰等提出科学数据联动的产业技术联盟模式,^[15]以患者所涉及的医院、医生、护士、上下游医药设备企业、科研机构、医保、医疗部门等作为链条进行联动,提升了疾病诊疗质量。新冠疫情期间,

Jia等将人口流动数据、确诊病例数据以及地理位置数据进行联动分析,能够提前准确预判高风险地区的位置。^[16]可见,数据流的真实价值根植于科学完善的横向管理协调机制。

当前,数据资源的全域联动问题,给数字化时代的管理学研究带来了诸多挑战。

(1)大数据的纵向联动和完备化已有较为成熟的技术手段。给定一个行业,每一种纵向关联都已有较为成熟的技术方案,如知识关联、知识图谱等。^[17]但在行业纵深数字化条件下,数据在企业间乃至行业间的横向完备化,依然存在管理协调机制模糊、资源使用差异性矛盾较大、沟通成本较高等问题。

(2)企业间的数据共享机制存在缺陷。关于数据共享方案,前沿的云存储技术目前主要集中于对微观用户行为的约束和管理,^[18]这可以防止数据篡改。代理加密区块链技术则在隐私保护和数据有用性之间平衡的问题上提供了解决方案。^[19]大数据背景下,管理活动又呈现出群体决策的特性。^[20]大多数企业都将数据视为自己的战略资源和竞争优势,因此企业中存在分散的监管格局、潜在的利益冲突等系列问题。^[21]打破部门之间的数据共享障碍,建立值得信赖的共享环境,仍然面临巨大挑战。^[22]同时,数字化转型过程中,公共数据如何通过有效分享实现社会价值最大化仍有待深入研究。^[23]邓维斌等构建了政府、技术支撑企业与监管机构之间的数据共享三方博弈模型,发现数据共享的非主观化是三方协调的基础。^[24]2022年5月,欧盟发布的《数据治理法案》为“数据利他”提供了明确法律框架。在这一背景下,国际上数据空间的解决方案以数据的标准化为驱动,如面向企业的B2MML集成标准、自动化标记语言(AutomationML)等,^[25]为工业4.0相关系统的开发提供了支持。可见,数据共享机制的缺陷极有可能使较大尺度场景下管理实证问题分析的有效性降低。

(3)当前以大模型为核心的人工智能生态体系正在快速迭代。自2022年8月微软推出BEiT-3模型、2023年3月OpenAI推出多模态大模型GPT-4,到2024年2月推出文生视频大模型Sora,多模态大模型进行预训练的数据需要各领域外部知识不断被打通。大模型需要具备以下条件:①是典型的多变量超参数模型,参数规模可达百亿以上;②往往涉及两种及以上的模态,如AI产生的图像、文本等,将多个模态的特征进行高效融合,需要深入研究;^[26]③模型能够产生意料之外的“化学反应”;^[27]④基于先进的计算技术(如量子计算),模型能够做到反应迅速;^[28]⑤不再局限于特

定的领域或专门的平台,而是能够处理各领域的复杂问题。可见,大模型的发展趋势,将会给复杂企业生态系统的管理带来新的挑战。^[29]例如,企业集群大规模数据交互条件下如何设计产业政策,行业内部的数据共享也需要向数据生态系统转变。^[30]

上述问题,既会带来管理实证研究的变革,同时也为AI驱动下商业逻辑的创新提供了机遇。

根据上文提到的主要问题,本文的创新工作以管理学研究的变革为视角,分为以下几方面:首先,通过分析数据在不同场景间联动的内在机理,论证数据联动与认知仿生过程的内在一致性,提炼可行的跨场景数据横向联动协调机制。其次,探讨通过目标谱系建立跨场景数据资源管理机制的可行性。在有序的目标管理基础上,从企业管理的角度出发解构数据共享机制,特别是数据共享与安全方面的挑战与对策。再次,在上述联动机制的基础上,针对企业集群与海量数据条件下产业政策的设计问题,力图突破既有的管理实证研究局限性,建立一套新的方法论,形成集成创新。最后,对数据驱动管理学研究的未来发展进行展望。

一、认知仿生过程对数据联动式管理的启示

1. 数据联动的仿生机理

在数字化环境中,终端通过整合来自多个源头的的数据,揭示隐藏的模式和联系,形成驱动企业创新的管理机制与非线性路径。这个过程的内在机理本质上类似于对人类大脑处理外界信息的仿生过程。这是由于,人类大脑对外部世界的理解和反映,并非将外部世界的对象进行简单累加,而往往是通过“联想”将来自不同源头的的数据和客观对象进行连接和加工,进而产生思维和认识的螺旋式上升。

在数字化的管理情境下,AI与数据密不可分。人脑是迄今为止最高形式的智能模式,AI对客观世界的学习,本质上就是仿生人类大脑对纷繁复杂世界的认知过程,人类从婴儿成长为成年人,大脑不断形成对外部世界各种事物的映射,进而转化为概念(如猫、狗、山川、河流等),多种概念的复杂交互,最终产生知识的涌现,帮助人类形成人生观、世界观、价值观。企业的生命周期也具有这一特征。人工智能的目的是运用“大模型”模拟并升级生理层面人脑的思维功能,而AI对外部世界映射的实现存在一定的技术缺失与数据偏见。例如,到目前为止,先进的无人驾驶汽车往往不能完全取代优秀司机的经验和应变能力,AI算法也无法完整模拟道德直觉,当遇到伦理问题、敏感争议事件的法律判决等

问题时,即便是专业领域的垂类大模型也有局限性。产生这些难题的原因在于,AI在确定性信息的处理方面具有优势,并早已超越人类,然而在人类对世界的认识过程中,还存在若干未知或难以精准量化的对象,比如情绪、人格、非理性、意识、应激反应、梦境、死亡等,而对于这些概念,AI并没有给出具备权威性的元数据。AI是针对人类理性、确定性思维的精确模拟,而对于高级的、非确定性大脑活动的模拟则前景未定。数据的全域联动性必然会衍生出新知识、新商业模式的交互式涌现,从而弥补(不是消除)这一差距。因此,数据壁垒的打破、企业的数据驱动与智能体的认知仿生具有逻辑一致性。

人工智能的发展伴随着对高级思维模式与不确定性更加细致的处理。智能体通过视觉、听觉等的输入形成接触对象的概念,若干概念通过广泛联系涌现出知识(见图1)。全域数据联动功能的实现也是通过“小数据”实证结果(概念)之间的协调,演化出更高价值密度(高级思维形式)的结果。

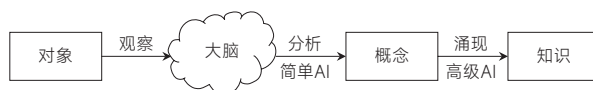


图1 大脑的认知仿生过程

2. 粒计算中的数据管理思想

与解析化的数学公式不同,人脑认识客观世界的过程乃至企业数字化的过程并不是连续的,而是具有很强的离散化特征。“信息粒”是对现实的一种模块化、抽象和分层,广泛存在于企业管理实践中。在数学领域,“粒计算”是模拟不确定性问题的方案之一,粒计算没有局限于机械的算法设计,而是将复杂的对象尽可能通过近似关系、模糊关系、功能关系、距离关系等划分出不同的“块”。这种“块”的划分,一方面符合人类学习认知事物的过程,另一方面实现了认知概念的过渡。例如,粗糙集中使用“上近似集”与“下近似集”表达不确定或不精确的知识,“三支”决策模型中除了传统决策的“接受”和“拒绝”两个选项,又引入了“不承诺”这一过渡选项。“颗粒化”的处理方式,均体现了管理实践中认识客观对象过程的仿生特点。

同时,粒计算也在一定程度上模拟了人脑内部的分类压缩功能,把人类的语言和思维过程提炼为模型式表达。目前,每一种粒计算模型都是针对认知对象某一方面的不确定性进行设计,力图用自然语言和计算机直接沟通。例如,模糊集是用来刻画认知对象中那些不能简单用“是”或“否”来判定的属性(如年轻、暖和、

优秀等概念)；Vague集又把模糊集中隶属的概念扩展为真假隶属度，能同时给出支持和反对的证据。

“举一反三”是人类大脑成长发育的重要标志，管理对象往往是动态时变的，因此，知识的迁移和联想意味着这些对象往往处于多个重叠的论域，在每个论域中，它们都会展示出相应的属性。可见，在企业数字化的横向联动过程中，“颗粒化”可以求同存异，将同一特征映射到不同场景下，实现不同场景下的差异化表达，在数据的跨论域横向联动基础上，实现不同创新过程的协调，降低沟通成本。

3. 全域集合的思想及其企业管理实践应用

此外，人脑认识外部世界的一个重要机制就是对客观对象的分类活动，集合论便是从这一过程中抽象而来。集合论从诞生至今已经有100多年的历史，起源于集合中的对象 X 与集合 S 之间的二元关系，为了统一描述各类集合在管理实践中的协调关系，用一个五元组来表示数据样本在论域空间（包含所有合作企业的全部数据资源）内的概念： $S(X) = (A, B, F, J, D)$ 。其中， A 表示某企业样本 X 数据分析业务系统的字段集合； B 为字段集合 A 的全域关联字段（其他企业中与 A 相关的数据字段）； F 表示由 A 映射到 B 所处问题场景的跨场景变换； J 表示数据分析需求与企业业务范围的范围； D 表示用于存储数据资源的表结构。 B 与 A 的关系为： $B = F(A) \vee B'$ 。例如，字段 A 表示“身份证号”， B' 则表示其他企业系统中与身份证号关联的重要字段。不同的场景具有不同的特征，因此映射 F 表示的过程不仅包括若干系统字段的简单叠加，还包括在数字化过程中高级别智慧的涌现。

两个场景的联动关系。给定数字化场景 R_1 与 R_2 ， $R_1, R_2 \in D$ ，令 X 表示样本集合变量， $X \in R_1$ 表示 X_1 是 R_1 中的样本集合， A_1 为场景 R_1 中的字段幂集， A_2 为场景 R_2 中的字段幂集 $a_1 \in A_1, a_2 \in A_2$ ，则从场景 R_1 与场景 R_2 的广义笛卡尔积中选取字段间满足一定条件的样本，可以形成两个企业主体之间的连接运算： $R_1 \triangleright \triangleleft R_2 = \{X_1 X_2 \mid X_1 \in R_1 \wedge X_2 \in R_2 \wedge V[X_1(a_1)] \leftrightarrow V[X_2(a_2)]\}$ 。其中， $\leftrightarrow$ 表示联动关系， $V(\cdot)$ 表示字段的取值，可见，当 $\leftrightarrow$ 为“等号”时， $R_1 \triangleright \triangleleft R_2$ 即为等值连接，此时，若存在 $\{X_1 \in R_1 \wedge X_2 \in R_2 \wedge V[X_1(a_1)] \leftrightarrow V[X_2(a_2)]\}$ ，则场景 R_1 与场景 R_2 形成联动关系。

全域场景联动关系。给定场景 R_i 及其对应样本 X_i ，对于界壳内包含多个场景的系统 $D = \{R_1, R_2, \dots, R_n\}$ ，所有满足条件 $\{X_i \in R_i \wedge X_j \in R_j \wedge V[X_i(a_i)] = V[X_j(a_j)]\} \neq \Phi$ 的场景集合 $U_j \in [1, n] R_j$ 为样本 X_i 的全域联动

关系， $U_j \in [1, n] A_j$ 为样本 X_i 的全域字段。

给定一个全域集合 $S(X)$ ，在 $S(X)$ 上定义若干种由 A 到 B 的映射关系，形成一套多个企业的联动系统，表示为： $\Phi = (S'(X), \mu_1, \mu_2, \dots, \mu_N)$ 。其中 $S'(X)$ 表示 $S(X)$ 的幂集， $\mu_1, \mu_2, \dots, \mu_N$ 表示定义在 $S'(X)$ 上 $F(A) \vee B'$ 的运算。显然，在对应的实证研究中，研究对象的量表构建要参考的因素集合即为 $F(A) \vee B'$ ，这个多场景联动系统同时形成一个逻辑推理系统，表示为 $\Psi = (K(X), r_1, r_2, \dots, r_N)$ ，其中， $K(X)$ 表示联动系统中所有研究问题的集合， $r_1, r_2, \dots, r_N$ 是不同场景下产生的不同字段（因素）的关联关系，界壳即对联动关系的约束，与 $\mu_1, \mu_2, \dots, \mu_N$ 一一对应。进一步，令 $\Pi(r_1, r_2, \dots, r_N)$ 表示多种字段（因素）之间关联关系的涌现结果，任意不同场景内关联关系 r_i 与 r_j 的组合可以涌现产生新的知识。

同时，也可以对同一个全域集进行 s 重自我迭代，即 $B = F_s(F_{s-1}(\dots F_2(F_1(A) \vee B') \vee B') \vee B') \vee B'$ ， F 可以用一系列分量表示为 $F = (F_1, F_2, \dots, F_s)$ ，构成联动系统 $\Phi = (S'(X), \mu_1, \mu_2, \dots, \mu_s)$ ，对应于逻辑推理系统 $\Psi = (D, r_1, r_2, \dots, r_s)$ ；则 $\Pi(r_1, r_2, \dots, r_s)$ 表示针对同一场景进行 s 重迭代， $\Pi(r_1, r_2, \dots, r_s)$ 构成的自我迭代系统可在特定的参数条件下通向混沌状态，目的是在大量无序的因素中寻找影响研究目标的序参量因素。在管理实践中，全域集系统的 s 重直积可以用于挖掘测度各类因素的指标体系。

假设多场景的分析结果表示为 $\Psi = (K(X))$ ，则 $r_1, r_2, \dots, r_N$ ， $r_1, r_2, \dots, r_N$ 分别对应于数据分析得到的不同场景下样本的规则组合。显然，不同场景下的推理结果 $r_1, r_2, \dots, r_N$ 之间不应预设权重。此处给出一种可量化测度的数据关联协调方法：将不同的结果形成的规则组合 $(r_1, r_2, \dots, r_N)_1, (r_1, r_2, \dots, r_N)_2, \dots, (r_1, r_2, \dots, r_N)_m$ 设定为基于帕累托最优解的“支配”与“非支配”关系。如果在界壳空间内存在一系列不被其他任何推理规则支配的推理规则 $(r_1^*, r_2^*, \dots, r_N^*)_1, (r_1^*, r_2^*, \dots, r_N^*)_2, \dots, (r_1^*, r_2^*, \dots, r_N^*)_k$ ，则这些规则可以形成一个联动平面，数据在不同场景之间的联动关系可以用规则空间 Ω 条件下不同规则之间的距离表示： $\hat{d}_\Omega[(r_1^*, r_2^*, \dots, r_N^*)_i, (r_1^*, r_2^*, \dots, r_N^*)_j]$ ，这些非支配规则之间在规则空间条件 Ω 下的距离即表示不同企业之间的数据联动强度。距离越近的场景，字段之间的联动性越强。这一系列不被其他任何推理规则支配的规则同时形成一条逻辑推理链条，在多重直积的条件下可以涌现出新知识。

4. 联动空间

当联动的场景数量多于两个时, 多个场景之间不同形式的协调规则空间 Ω 便同时形成了一个联动空间 (需要注意的是, 联动空间中的多个场景, 并非由主观先验概率确定其重要性, 上文提出的“支配”与“非支配”关系是一种可行的场景间协调关系)。因此, 在企业的管理实践中, 联动的思想可以集成各类业务系统的联动描述 (例如, 将交通、医疗、学校等行业的活动进行连接)。对任一界壳 J 内的 $X_i \in B$, $X_i \subset B$, 与已有 $A = \sum A_i$, 形成 X_i 的联动子空间: $X_i \in B, \subset B \Rightarrow S(X_i) = \{X_i \cap \sum A_i \neq \Phi\}$ (子空间之间的协调关系同上小节)。而 $\forall X_i / J \Rightarrow \sigma(S) = \{S(X_i)\}$ 可以从理论上称为界壳 J 内关于 X_i 的全域联动空间, $\sigma(S) / J$ 是所有潜在的联动场景分析基础。 $\sigma(S)$ 可以从理论上称为界壳内的全域联动空间 (Global Linkage Space)。

对任意行业、业务、论域的界壳 J , A_i 的结构即针对具体问题的、各种用于测度自变量、因变量或构建量表的因素集合。建立各类实证模型的依据可认为是 $S = (A, B, F, J, D)$ 中, 运用 AI 模型学习到的 B 与 $A = \sum A_i$ 的联动关系。

二、数据资源共享的目标管理：目标谱系

高级思维过程的形成不是杂乱无章的, 而是在看似无序的联动中形成有序的学习过程。“横向联动”与“举一反三”需要明确的“学习目标”引导。将这一认知过程延伸到类似人类大脑的重要决策主体, 如政府机构、企业、社会团体等, 它们均有着更为具体的层次化管理目标。海量的全域数据无法直接进行全要素基础支撑, 因此, 以目标管理为导向生成数据资源联动需求, 继而建立需求之间的关联, 是当前 AI 背景下企业数字化过程的基础。

1. 目标谱系棱镜模型

大数据环境下人类活动的各个场景要实现的目标可以视为光谱中的谱线, 各条谱线之间均存在联动关系, 海量数据通过横向联动, 可以产生隐含目标的涌现, 提升数据价值密度。将光谱的分析过程进行跨学科概念迁移可知, “光谱”系统是开放的、发展的, 如图 2 所示, 谱系的分支可以不断演化、增长, 不同的分支既有独立的标识, 又存在普遍的联系。

光谱分为可见部分 (既有的管理场景) 与不可见部分 (可预测、有待发展的业务)。以智慧城市的管理为例, 目标谱系是智慧城市中海量数据与分析测度工具 (如 AI) 之间产生化学反应的催化剂 (见图 3)。这里选取智

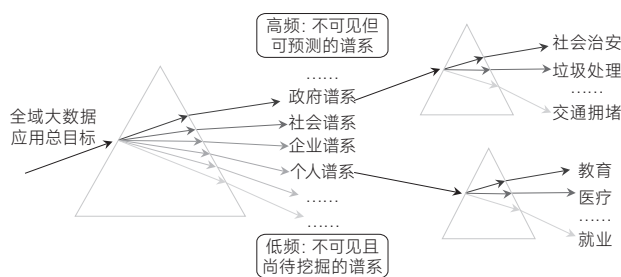


图2 目标谱系概念模型

慧城市作为目标谱系使用场景案例的原因是, 在一个大型城市的规划、设计、建设、管理与运营等领域中, 智慧城市平台为城市管理中的各类场景 (教育、医疗、交通、公共事业等) 提供高效的横向互联管理机制, 是海量数据汇聚与数字化管理对象横向联动的典型高级形态, 具有足够的数据承载力和跨领域特征。不同的谱系还可以依托各类数据资源, 借助 AI 的适应性学习, 不断生长延伸 (例如, 运用 AI 对变量的测度范围进行拓展), 产生更为具体的场景式需求。值得注意的是, “可见光”与“不可见光”部分不是相互割裂的关系, 而是目标谱系之间的隐含联动关系。

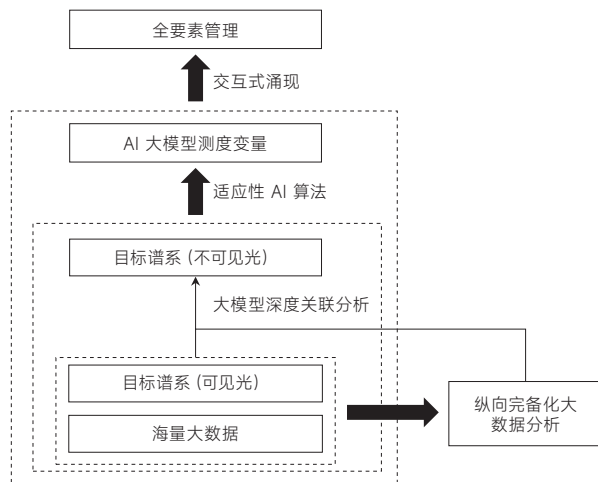


图3 智慧城市目标谱系的层次

大型智慧城市的建设本质上可以抽象为一个“城市大脑”认知客观世界的过程, 大脑在接收外界信息后, 会对信息进行分类和压缩, 构建对象与概念之间的隶属关系, 通过交互和联动产生更高级别的知识。可见, 目标谱系实际上提供了一种可行的、复杂压缩映射关系的表述方法, 更是一种数据资源管理规则建立的可行依据。无论是人脑还是 AI, 知识的形成过程都不是孤立的, 而是一个由识别“可见光”到识别“不可见光”的过程, 从管理实证研究的角度而言, 谱系方法论的引入, 可以使复杂、模糊的管理目标和影响机理得到可量化的合理

解释,为实证研究中因变量组合以及各类影响路径的构建提供更加可靠的理论依据,有利于产生不同认知交互涌现的结果,从而进一步提升实证研究的价值密度。

2. 案例分析

以大型城市的管理活动为例,将谱系思想引入类似北京这样大规模城市的管理中,通过自下而上、由微观到宏观的知识涌现,实现全域全要素精细化社会治理。在光谱系统基础上,进一步衍生出目标谱系的枝叶系统模型,为各类实证研究的因变量标签体系建立全域索引(见图4)。



图4 智慧城市(北京)目标谱系概念模型

枝叶系统能够通过树形结构实现对管理目标的结构化表达，枝叶系统包含干枝、分枝以及谱叶，干枝表示全域目标谱系，分枝是在干枝基础上生长出的智慧城市服务功能。干枝、分枝表示固定的论域、边界及业务系统，谱叶则具有动态演化特征，如同每年树叶的掉落与生长，表示在多级目标下各类管理目标的动态变化，可伴随场景的时间变化延伸出新的目标。显然，在数字化背景下的管理实证研究中，单一的因变量体系无法产生能够适应各类管理场景的“谱叶”，而大型城市中的各类管理活动普遍横向联动，产生的海量数据往往缺乏多元化的因变量评估体系，既有的因变量体系大多依赖于微观层面的研究目标（例如：数字化背景下，大数据驱动某一特定行业的创新路径）。树形目标谱系的建立能够使海量数据产生的因变量体系为局部“谱叶”的生长提供更多数据支撑，使智慧城市中各类影响因素和系统的匹配关系更加清晰，让原本互相独立的管理活动高效地共享数据，提升实证研究的准确性。

三、“中心化”数据的非中心共享机制

随着海量数据迅速渗透到企业管理的各个场景，企业信息系统的建设水平、逻辑完备性得到提升，然而，企业价值链的升级需要数据运营、处理方式的转变，从传统的内部集中、统一协调式处理，转化为多维度向外

拓展的联动式处理，方可实现信息系统的数据赋能，这也是各个业务环节数据化的趋势。

近年来,一系列新的数据协议、数据载体纷纷出现,同时成为数据共享管理中的关键科学问题。本章节以目前常见的管理海量数据的典型载体为出发点,探讨数据联动在跨企业数据调度、共享中的应用,提出一种新型的、能够服务于管理实证研究的数据共享技术机制。

1. 区块链在企业管理实践中的缺陷分析

区块链技术是一种具有分布式特征的数据存储方式，数据的使用可以不再通过中心机构的处理，表现为“去中心化”特征。

（1）区块链的场景适用性局限

区块链技术机制的本质，是致力于解决复杂业务主体之间信息的真实性问题，而非数据的共享。区块链中的共识机制可以降低交易成本，填补企业之间业务往来中的信任缺口。例如，中小企业的融资活动可通过链上的各类电子凭证直接生成信用保证；信息上链后，用户只需要到办事机构跑一次就可以实现跨部门协同办理。然而，信息整合后如何进行联动分析，区块链并未提供相应的技术机制。

(2) 区块链技术机制在数据联动共享中的缺陷

区块链中的数据传递基础，是对应的算法规则（例如共识机制与加密算法等），为保证数据传导的安全性，区块链网络中的节点状态应保持一致，在此框架下，企业需要维持不同节点算力的平衡性，防止某些企业的算力增强后，对数据信息进行篡改。显然，随着区块链的节点数量不断增加，达成共识所需的时间将会同步增加，数据传递与验证的效率将会十分低下。

2. 基于全域联动空间的共享技术机制

针对可能的数据风险及数据壁垒问题,本文提出一种“中心化的系统数据非中心化共享”机制(见图5),对企业数据应用平台进行层次划分。底层平台为基层分析应用综合平台,不同的并行子业务系统均有各自的内部数据、具体的应用场景及对应的数据分析处理算法,同时,底层平台需要完成数据的清洗、匿名化、脱敏等原始处理。当数据分析或实证模型需要外部数据关联时,则基于具体业务向上提出数据联动请求。

中层平台为中心库集群应用平台,通过引入既有共享技术机制中的访问控制技术,对数据访问进行授权和限制,允许授权用户访问和使用数据,例如针对底层平台进行身份认证、访问授权、访问审计和角色管理。进一步,针对底层平台的联动请求进行分析反馈。例如,依据本文第二节的跨场景分析结果,得到多场景之间的

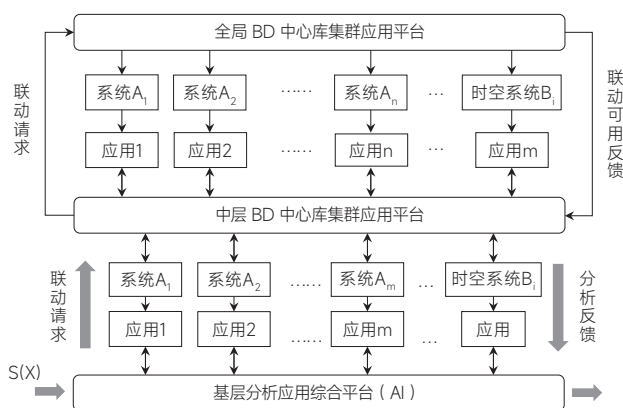


图5 BD (Big Data) 共享技术机制

非主观先验关系, 构建由多个场景需求形成的联动空间, 场景与场景之间对数据的使用遵循帕累托最优分布形成支配关系, 进一步, 依据非主观的联动关系, 制定数据共享协议、数据使用安全协议等。

上层平台为全行业中心库集群应用平台, 针对中层平台的联动请求进行响应, 确定数据的权利属性, 机制如下: (1) 追溯数据的具体来源与类型, 包括数据的获取方式、时间、地点等, 能够为企业应用目标谱系的建立提供依据。(2) 数据的分析与评估。基于前述跨场景分析结果与联动空间方法, 提供进行联动数据的分配方案与潜在风险, 响应中层平台的联动请求。(3) 确认数据归属主体与数据资源分配。确认内容包括数据的所有者、使用权、保密性等。(4) 对数据进行管理和维护。确认数据资源归属主体后, 对数据权限进行管理维护。

3. 多模态大模型在企业管理中运用的关键问题

“中心化”数据的非中心共享机制为海量数据的横向汇聚提供了一条可行的技术机制和流程, 针对包含异常复杂模态的海量数据, 与之对应, 则需要具有超大参数规模、自适应性、精细化的“大实证模型”用于挖掘海量联动数据内所隐含的管理规律, 即使是专业领域内单项技能优先的数据纵向完备化, 也需要匹配高质量垂类大模型的学习。

模态 (Modality) 是指每一种信息的来源或者形式, 例如以图像、文本、视频、图形、声音等不同的媒介形式传递信息。^[31]而在当下较为热门的研究领域中, 全域数据融合包含的数据模态主要有图片、文字、音频、视频等, 这些都是在企业运营过程中产生的常见数据。如果对隐藏在多模态和跨模态中的融合表示充分建模, 可以进一步提高各种多模态应用的性能。^[32]

当前企业级别的多模态大数据处理存在一系列的困难与挑战: (1) 由于数据繁多、计算量大、耗时长, 对

设备的计算能力有很高的要求。(2) 多模态大数据包含许多跨模态信息, 捕获多模态数据的完整语义知识难度较大。(3) 多模态数据的分布是持续变化的。^[32]

上述困难导致深度学习模型在处理多模态大数据时有很大的局限性, 因此, 大模型应运而生。针对管理学研究, 大模型技术突破的意义在于: (1) 既有神经网络技术趋于成熟, 导致精度提升平缓, 自变量与因变量之间的可解释性不足。(2) 自监督学习能减少数据样本标注工作量, 使得大模型在传统的实证研究中更加高效, 减少了运算成本。(3) 针对企业所涉及的行业、工作种类繁多问题, 大模型包含的“预训练”与“下游任务微调”能够有效提升模型的泛化性, 从而应对各类企业不同的实证研究需求。^[33]

多模态大模型的组成如图6所示, 多模态大模型的核心是多模态算法群,^[34]算法群主要包括以下几种方法: 用于特征嵌入的 Embedding 技术, 用于特征提取的基于 Transformer 的模型, 用于用户生成内容的基于 GAN 的模型和交互式 AI 模型。其中 Embedding 技术是指将高维数据映射到低维空间的过程, 可用于关键因素或自变量的挖掘。Transformer 是第一个完全基于注意的序列转导模型, 用多头自注意将单个序列的不同位置联系起来。^[35]生成式对抗网络可通过神经网络之间的相互博弈, 在实证研究中提升非线性变量的模拟精度。交互式 AI (人工智能) 即能与人进行交互的智能行为技术, 通过机器学习、深度学习等功能, 实现复杂逻辑推理等行为, 可用于从案例中提取有效逻辑路径, 在管理活动中梳理复杂的影响因素传导机制等过程。

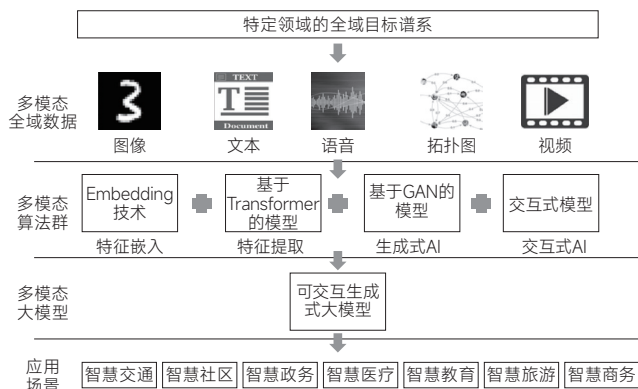


图6 多模态大模型的组成

模型体量和数据规模的增加, 一方面有助于增加实证模型的精度, 另一方面经过不同模态的数据训练后同一个模型能适用于不同的企业场景, 泛化性会提高, 有助于从多个角度挖掘企业成长的内在路径。

四、数据智能背景下企业生态系统传统研究范式的变革

通过管理目标谱系的建立与“中心化”数据的非中心共享机制,借助大模型,企业生态系统内各要素之间的非线性关联关系及其结构和功能将实现数字化。泛在的、网络化的横向联动,会使针对企业生态系统的研究彻底摆脱传统研究范式中假设驱动的限制性。可见,企业生态系统(或广义的、由管理对象构成的生态系统)的研究范式将出现重要变革。

1. 传统实证研究范式的局限性

经典的、假设驱动模型,大多运用统计分析手段来强调主要规律和主要趋势的影响,将多个企业构成的整体视为若干个体的简单相加,产生了许多归纳法和演绎法难以解决的科学问题。^[36]针对这些问题,基于假设驱动的实证方法应运而生,假设驱动的方法,虽然对企业管理活动中各类影响机理传导路径的刻画方面进行了较多的研究,但大多停留于“线性描述”上,即依托主观的量表与预设的假设,通过回归将主要矛盾呈现出来。然而,在海量数据条件下,研究对象各类变量之间具有易变性和涌现性,小样本难以覆盖,预设假设以外的自变量与因变量之间的深层次关联,传统研究则解释不足。因此,传统实证研究中仍需要解决一些方法使用局限的问题:

(1) 参数选择的不确定性。在大量企业生态系统或集群的相关研究中,一些关键参数的取值大多采用传统的实证方法,或者取值与既有研究经验值的数量级接近,尤其是在企业主体数量较多的研究中,例如产业集群、合作创新演化、供应链金融等,在保障企业主体隐私的条件下,不同主体的行为机制可由数据的横向联动分析得到(如消费习惯、风险态度、竞争合作等)。数据的缺失往往使与个体企业行为规则密切相关的参数取值偏离实际,从而影响实证结果对现实决策的参考价值,近年来,“自动化机器学习”被列为数据科学领域的前沿热点问题,正在力图实现数据密集型的模型特征、超参数选择的自动化,该方法可进一步迁移运用到企业生态系统研究中。

(2) 实证模型将企业系统抽象化引发的企业生态系统的跨场景可复现性问题。以典型的企业管理者或个体员工的行为(激励机制、创造力涌现)挖掘为例,伴随数据横向联动的实现,广义层面“个体”活动轨迹的数字化已成为现实,可以采用深度学习方法分析不同场景下的个体行为规律,通过海量数据的训练不断提高实证

模型的泛化能力,即使个体活动的时空场景发生变化,模型依然可以具备较高的准确性,缺点是模型的可解释性较差,难以解释因变量与输入自变量之间的内在逻辑关系。而采用传统实证方法,则是先通过解析模型构建特定假设(多为强假设)下个体的行为动机,因变量与自变量之间具有较好的可解释性。例如,传统的Logit模型或Probit模型就经常在研究中被用来刻画个体选择行为的变化,受数据资源的限制,在较小样本或较小环境范围内,模型的测算结果符合现实情形。因此,在大量的实证研究中,针对研究对象选择行为(如员工个人对效用的判断)的影响机制,该系列模型得到了广泛运用。然而,当研究对象所处的时空尺度发生变化,个体需要和环境产生可被记录的数据交互时,各类指标或因素影响的作用机理就会产生变化。比如,个人在不同情境下的选择行为往往是存在差异的,这种差异是由模型本身在刻画环境交互(数据横向联动)功能方面的缺失带来的,这样模型的适用性便难以保证。因此,实证结果的跨场景可复现性受到影响。

可见,数据智能时代的来临,使企业生态系统的研究面临着范式转型。^[37]首先,数字化时代使依照横向联动数据训练各类影响因素的作用机理及不同时空尺度下的超参数具有了可行性,从而不断完善传统实证研究。其次,数字化社会的转型,衍生出大量新型业态与人际关系。传统企业生态系统内主体间的交互关系本身也在发生变革,如信息扩散加速带来的交互关系新形态。

2. 数据智能背景下企业集群交互的新形态——双层陀螺模型及其建构思想

数据智能背景下,企业生态系统内交互关系的测度与真实还原需要新的研究方法。早在2020年,DeepMind团队便提出了一种策略空间陀螺几何体理论,^[38]该理论分析了在当前海量数据背景下,若干具有强化学习机制的决策主体在进行充分的竞争、博弈、合作后产生的策略空间形态。该思想体系可在多个AI应用的大场景下产生应用价值(如企业集群管理、安全风险防控、交通秩序管理等),为新研究范式的建立提供了可行的理论框架。而在有关集群活动的刻画中,往往会形成具有两层递阶结构的、非合作的交互机制问题:一方的行为会影响另一方的选择,且任何一方都不能完全控制对方的行为(符合市场的状态)。政策制定者、管理者与企业集群的双向互动,就构成了具有相互反馈机制的双层陀螺结构:

在传统研究范式下,由于无法利用海量数据的联动对同一层次企业集群内个体之间的差异性进行校准,^①

再结合广为采用的强假设, 大多实证模型的测算范围是图7中的“行动策略1+演化规则1”, 只能得到顶端优势均衡下的结果。当企业的智能决策程度较高时(以数据信息的充分横向联动为基础), 企业生态系统中各决策主体之间的影响关系显然并不会都聚焦于稳定的均衡, 而是存在普遍的均衡转移。对于上层模型, 从行动策略1到行动策略 m , 不同管理者之间的影响机制可能产生差异化的均衡结果。对于下层模型, 从演化规则1到演化规则 n , 随着影响机制非线性程度的增加, 企业生态系统的局部嵌套式相互影响愈发容易形成策略传递压制关系, 优秀均衡结果在企业集群中的传递性受到压制, 可能会导致实证结果与实际现象之间的差异增大。整体模型转化为图7中的“行动策略 m +演化规则 n ”, 一般而言, 虽然上层模型的行动策略数量 m 小于下层模型的演化规则数量 n , 但由于多种策略与多种影响机制的差异性, 由实证得到的影响机制大多处于非传递性较强的稀释状态, 对局部类型企业之间的影响机制具有较好的可解释性。因此, 在企业生态系统研究中要解决的科学问题是如何实施合理的上层决策, 探寻在数据驱动的影响机制下, 顶端优势均衡的形成路径。

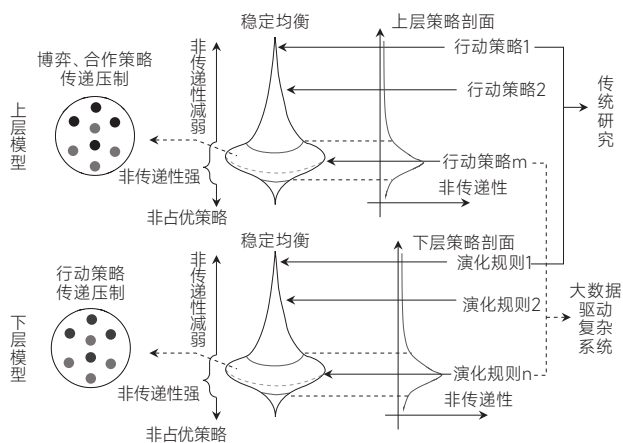


图7 双层陀螺模型思想

3. 数据驱动下企业生态系统的分析工具

数字化之后产生的数据联动复杂性给传统的企业管理模型方法论带来了新的挑战。AI方法在企业管理研究、政策研究中的模式需要创新。本文提出的数据智能背景下企业生态系统的研究思路如下:

(1) 针对内部存在海量主体网络化泛在关联的企业生态系统, 分析并测度系统内元组或个体之间的相互作用关系、偏好、价值、风险态度等差异, 刻画影响链条与相互作用规律。基于影响链条, 建立具备制衡性的相互作用关系, 同时, 保留系统中所隐含的自由特征, 继

而提出合理的相互作用模型。

(2) 收集企业生态系统的真实历史状态数据, 对主体之间的相互作用进行校准, 此时, 可引入生成式对抗网络(Generative Adversarial Networks, GAN)的思想: ①运用合理的相互作用模型产生变量的时间序列样本 $G(z)$, 数学分布为 P_g , 运用真实的变量时间序列数据 x (数学分布为 P_r)构建判别器 D , 建立合理实证模型的目标是通过模型参数的变化, 让判别器无法区分生态系统模型产生的数据与真实数据的分布差异, 而判别器的目标则是要尽可能区分生态系统模型产生的数据与真实数据的分布差异。②对 D 和 G 进行交互迭代, 直到过程收敛, 此时, 合理的企业生态系统实证模型能够在准确测度内部相互作用的同时逼近真实系统状态数据。③考虑到海量数据条件下的大规模运算代价问题, 基于经过真实数据校准后的实证模型, 可采用生成器模拟时间序列 G' (数学分布为 P'_g), 然后采用判别器 D' 对模拟出的时间序列与合理实证模型产生的实际序列(数学分布为 P_m)进行区分。④对 G' 与 D' 进行交互迭代, 直到过程收敛。此时, 生成器既能够反映真实的系统状态变化, 又能够产生与合理实证模型相同的结果, 可实现对真实企业生态系统状态变化的复现性。

(3) 引入AI中的增强学习使企业生态系统的刻画克服传统研究中“唯象描述”的局限性。依据增强学习的思想, 政策制定者或管理者对企业生态系统结构进行深层次调整的过程, 可视为一个人工智能体在复杂经济系统中不断探索寻找合理政策方案的过程, 全域联动数据与面向企业生态系统的合理政策方案(对产业结构、企业间合作关系、资源流动管理等诸多方面)之间的关联关系, 可由AI的增强学习训练过程实现。①将政策制定者对企业生态系统的反馈、调整与改造过程构造为“动作变量 A ”, 将联动数据校准的合理实证模型产生的因变量结果构造为“状态变量 S ”, 状态变量的各个维度之间均具有联动关系。②将系统运行具备稳定的规则与管理者的预期进行结合, 构建实施动作变量后的“奖励值”, “动作”的机制要立足于“简单”, 为企业生态系统复杂性“混沌边缘”状态的发生留下一定空间, 保障反馈、调整与改造过程应具有一定的韧性。③运用具有较强非线性特征描述能力的深度神经网络拟合 $\hat{Q}$ “动作变量”与“状态变量”之间的反应函数, 经过横向全域数据迭代后, 神经网络收敛并输出稳定的反馈、调整与改造方案。

上述数据智能背景下企业生态系统的建模改进思路如图8所示:

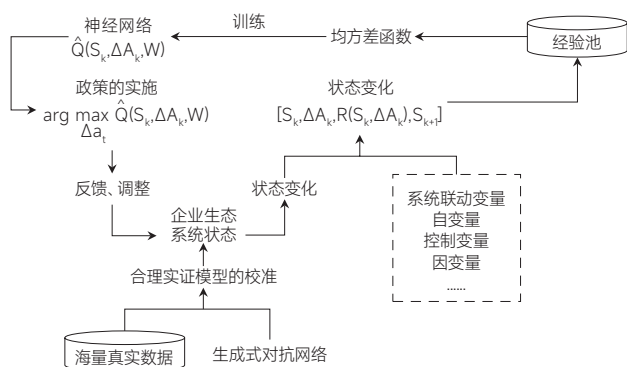


图8 数据智能背景下实证研究的改进思路

4. 数据驱动下的企业管理研究需要突破的局限与方法论创新

(1) 通用人工智能的发展为既有研究方法论的深入提供了基础, AI 算法框架是复杂模型的基础, 要打造完善的 AI 框架生态, 需将多种模型进行有机整合。

(2) 针对单一企业机构的研究, 往往会建立独立的模型, 且不同部门之间存在数据壁垒, 跨企业数据的算力协同依然面临体制机制的挑战, 可供校正传统实证模型的数据还需要大量挖掘。

数据驱动的企业生态系统模型, 其内部各类影响机制与传导路径的可解释性依然没有被完全掌握, 这就导致生成式的管理决策与政策制定存在一定的伦理失范风险。因此, 针对改进后的实证模型, 需要在实施重要因变量输出时, 完善推理过程的可追溯性, 不断形成数据驱动的规范化研究范式。

上述研究思路将为建立大数据科学与企业管理分析 (见图 9) 奠定方法论基础。数据驱动的理论模型能够形成一个改进后的实证研究体系, 形成特定行业“纵到底, 横全域”的分析平台, 不断拓展实证模型的功能与数据驱动下的企业管理研究体系。

五、总结及研究方向应用展望

大数据之大, 在于全域联动的内在规律。习近平总书记在二十大报告关于《推进国家安全体系和能力现代化, 坚决维护国家安全和社会稳定》的阐述中, 明确了完善风险监测预警体系, 构建全域联动、立体高效的国家安全防护体系, 为全域联动的理论发展指明了方向。下文将探讨不同的研究方向和应用展望:

(1) 大数据应用的核心是数据的全域联动。数据从封闭、孤立走向开放, 从训练“小模型”走向训练“大模型”面临着横向完备化的挑战。^[39] 全域联动空间的初步概念, 为多种企业、机构之间的数据横向共享提供了

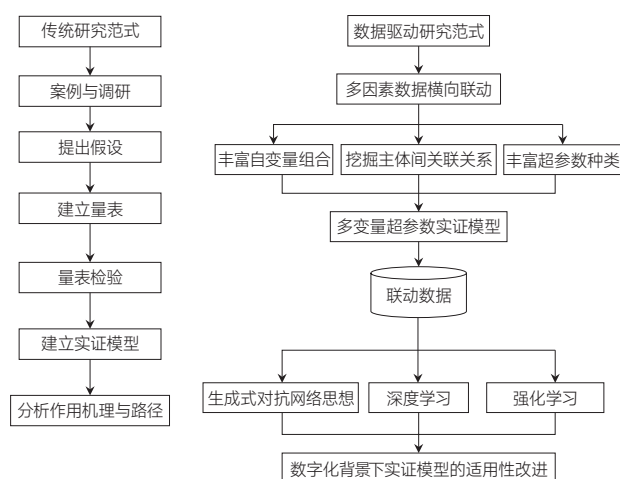


图9 企业管理研究模式创新方法论框架

一种可行的技术机制。

(2) 通过数据与算法驱动, 界定有效的界壳范围与目标谱系是未来研究的重要方向。数据的全域联动可以抽象为一个基于粒计算的大规模集成计算系统, 构成 AI 大模型的训练基础。界壳对应于目标研究问题的论域, 不但决定解决问题的目标谱问题, 同时, 更涉及 AI 算法模式的创新。因此, 运用 AI 技术与海量数据训练并学习数据分析目标的谱系, 可以在联动空间的基础上, 进一步解决企业间数据分享协调困难、数据管理秩序失调等问题。

(3) 多主体 AI 的“左右互搏”使系统的“演化”变得更加多维立体, 而不再局限于虚拟参数模型、计量实证模型和计算模拟等方法。这一新的方法论为复杂社会现象和行为机理的研究提供了崭新的视角。因此, 多模态大模型的模式创新和算法体系架构研究具有重要意义。以此为基础, 提出两个管理学领域前瞻性研究方向: 一是运用大模型感知和反馈复杂社会经济系统的涌现现象, 为企业生态系统的管理提供可操作的决策支持; 二是通过有机组合 AI 算法, 挖掘并提升算法的可解释性, 深入研究管理问题的微观决策路径。

(4) 在日常管理实践中, 有以下几个重要方面值得进一步思考: 首先, 管理决策中全域目标场景谱的设定具有学术重要性。这要求企业明确管理目标并建立相应的场景谱, 指导管理实践中的决策制定和执行。其次, 大数据共享的机制和技术逻辑也是一个需要深入研究的领域。管理者需要思考如何丰富数据共享机制, 促进组织内外部的数据共享, 从而极大地丰富应用场景。通过深入研究和思考上述问题, 管理学者和从业者可以更好地应用大数据科学方法, 提高管理决策的学术水平和

实际效果。

参考文献

- [1] Simsek, Z., Vaara, E., Paruchuri, S., Nedkarni, S., Shaw, J. D.. New Ways of Seeing Big Data. *Academy of Management Journal*, 2019, 62(4): 971-978.
- [2] McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., Barton, D.. *Big Data: The Management Revolution*. Harvard Business Review, 2012, 90(10): 60-68.
- [3] Davenport, T. H.. *Big Data at Work: Dispelling the Myths Uncovering the Opportunities*. Harvard Business Review Press, 2014.
- [4] Glass, R., Callahan, S.. *The Big Data-driven Business: How to Use Big Data to Win Customers, Beat Competitors, and Boost Profits*. John Wiley & Sons, 2014.
- [5] Peppard, J., Ward, J.. *The Strategic Management of Information Systems: Building a Digital Strategy*. John Wiley & Sons, 2016.
- [6] 冯芷艳, 郭迅华, 曾大军等. 大数据背景下商务管理研究若干前沿课题. *管理科学学报*, 2013, 16(1): 1-9.
- [7] Chen, H., Chiang, R. H. L., Storey, V. C.. *Business Intelligence and Analytics: From Big Data to Big Impact*. *MIS Quarterly*, 2012: 1165-1188.
- [8] Williamson, O. E.. *The Economics of Organization: The Transaction Cost Approach*. *American Journal of Sociology*, 1981, 87(3): 548-577.
- [9] Prahalad, C. K., Hamel, G.. *The Core Competence of the Corporation*// Zack, H. H.. *Knowledge and Strategy*. Routledge, 2009.
- [10] Kaplan, A. M., Haenlein, M.. Users of the World, Unite! The Challenges and Opportunities of Social Media. *Business Horizons*, 2010, 53(1): 59-68.
- [11] 陈国青, 吴刚, 顾远东, 陆本江, 卫强. 管理决策情境下大数据驱动的研究和应用挑战——范式转变与研究方向. *管理科学学报*, 2018, 21(7): 1-10.
- [12] 汪勇, 李好, 王静. 考虑数据分布特征的多属性数据完备化方法研究. *统计与决策*, 2020, 36(24): 15-19.
- [13] 陈松蹊, 毛晓军, 王聪. 大数据情境下的数据完备化: 挑战与对策. *管理世界*, 2022, 38(1): 196-207.
- [14] 王建冬, 童楠楠. 数字经济背景下数据与其他生产要素的协同联动机制研究. *电子政务*, 2020, (3): 22-31.
- [15] 章琰, 杨一图, 吴健, 张辉. 我国科学数据共享运行机制模式创新探讨——以产业技术联盟为例. *科学学研究*, 2021, 39(11): 1970-1979.
- [16] Jia, J. S., Lu, X., Yuan, Y., Xu, G., Jia, J., Christakis, N. A.. Population Flow Drives Spatio-temporal Distribution of COVID-19 in China. *Nature*, 2020, 582(7812): 389-394.
- [17] 余艳, 张文, 熊飞宇, 孟小峰, 刘湘雯, 陈华钧. 融合知识图谱与神经网络赋能数智化管理决策. *管理科学学报*, 2023, 26(5): 231-247.
- [18] Tian, Y., Tan, H., Shen, J., Pandi, V., Gupta, B. B., Arya, V.. Efficient Identity-based Multi-copy data Sharing Auditing Scheme with Decentralized Trust Management. *Information Sciences*, 2023, 644: 119255.
- [19] Keshta, I., Aoudni, Y., Sandhu, M., Singh, A., Xalikovich, P. A., Rizwan, A., Soni, M., Lalar, S.. Blockchain Aware Proxy Re-encryption Algorithm-based Data Sharing Scheme. *Physical Communication*, 2023, 58: 102048.
- [20] 徐宗本, 冯芷艳, 郭迅华, 曾大军, 陈国青. 大数据驱动的管理与决策前沿课题. *管理世界*, 2014, (11): 158-163.
- [21] Susha, I., Rukanova, B., Zuiderwijk, A., Gil-García, J. R., Gasco-Hernández, M.. Achieving Voluntary Data Sharing in Cross Sector Partnerships: Three Partnership Models. *Information and Organization*, 2023, 33(1): 100448.
- [22] Wang, J., Gao, F., Zhou, Y., Guo, Q., Tan, C. W., Song, J., Wang, Y.. Data Sharing in Energy Systems. *Advances in Applied Energy*, 2023, 10: 100132.
- [23] 周林兴, 林凯. 城市数字化转型视域下公共数据资源开放共享机制研究. *现代情报*, 2022, 42(9): 16-28.
- [24] 邓维斌, 戴丽, 张志程, 樊自甫. 基于演化博弈的政府数据共享协调机制研究. *情报科学*, 2023, 41(2): 10-19.
- [25] Czvetkó, T., Abonyi, J.. Data Sharing in Industry 4.0-Automation ML, B2MML and International Data Spaces-based Solutions. *Journal of Industrial Information Integration*, 2023: 100438.
- [26] Feng, D., Christian, H., Rosenbaum, L., Hertlein, H., Dietmayer, K.. Deep Multi-modal Object Detection and Semantic Segmentation for Autonomous Driving: Datasets, Methods, and Challenges. *IEEE Transactions on Intelligent Transportation Systems*, 2020, 22(3): 1341-1360.
- [27] 陈剑, 刘运辉. 数智化使能运营管理变革: 从供应链到供应链生态系统. *管理世界*, 2021, 37(11): 227-240.
- [28] Pan, X., Lu, Z., Wang, W., Hua, Z. Y., Xu, Y. F., Li, W. K., Cai, W. Z., Li, X. G., Wang, H. Y., Song, Y. P., Zou, C. L., Deng, D. L., Sun, L. Y.. Deep Quantum Neural Networks on a Superconducting Processor. *Nature Communications*, 2023, 14(1): 4006.
- [29] 杨晓光, 高自友, 盛昭瀚, 黄丽华, 曾大军, 陈晓松, 杨克巍, 黄伟, 丁荣余, 郑新华, 姚轶琳. 复杂系统管理是中国特色管理学体系的重要组成部分. *管理世界*, 2022, 38(10): 1-24.
- [30] Lecun, Y., Bengio, Y., Hinton, G.. Deep Learning. *Nature*, 2015, 521(7553): 436-444.
- [31] Summaira, J., Li, X., Shoib, A. M., Li, S., Abdul, J.. Recent Advances and Trends in Multimodal Deep Learning: A Review. *ArXiv Preprint ArXiv: 2105.11087*, 2021.
- [32] Gao, J., Li, P., Chen, Z., Zhang, J.. A Survey on Deep Learning for Multimodal Data Fusion. *Neural Computation*, 2020, 32(5): 829-864.
- [33] Bommasani, R., Hudson, D. A., Adeli, E., et al. On the Opportunities and Risks of Foundation Models. *ArXiv Preprint ArXiv: 2108.07258*, 2021.
- [34] Fei, N., Lu, Z., Gao, Y., Yang, G., Huo, Y., Wen, J., Lu, H., Song, R., Gao, X., Xiang, T., Sun, H., Wen, J.. Towards Artificial General Intelligence Via a Multimodal Foundation Model. *Nature Communications*, 2022, 13(1): 3094.
- [35] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., Polosukhin, I.. Attention Is All You

Need. *Advances in Neural Information Processing Systems*, 2017: 5998-6008.

- [36] 黄晓, 吴江, 贺超城等. 仿真模拟方法: 大数据时代图情学科复杂场景的系统建模. *情报学报*, 2023, 42(2): 136-149.
- [37] 陈国青, 曾大军, 卫强等. 大数据环境下的决策范式转变与使能创新. *管理世界*, 2020, 36(2): 95-105, 220.
- [38] Czarnecki, W. M., Gidel, G., Tracey, B., Tuyls, K., Omidshafiei, S., Balduzzi, D., Jaderberg, M.. Real World Games Look Like Spinning Tops. *Advances in Neural Information Processing Systems*, 2020, 33: 17443-17454.
- [39] 冯媛. 科学数据开放共享的价值共创模型及运行机制研究. *图书馆*, 2022, (9): 29-37.

注释

- ① 典型的例子比如: 虽然我们都知道金融市场中投资者的同质性假设是不合理的, 但在大数据缺失的情况下, 行为轨迹实证模拟的准确性受到影响。

Big Data in Management Research: Scientific Issues and Future Development

Huang Nan¹, Li Zhu², Li Xueyan³, Gong Daqing⁴

1. School of Management, Minzu University of China; 2. School of Business, Renmin University of China; 3. School of Management, Beijing Union University; 4. School of Economics and Management, Beijing Jiaotong University

Abstract In the digital age, data has become a new factor of production and is considered a key resource of the fourth technological revolution. The openness, sharing, and circulation of big data have become the basic needs for the interconnection of all things. How to excavate the intrinsic connections between seemingly unrelated and even contradictory things through the horizontal full-domain linkage of data across industries, fields, time and space, and even ecosystems, and to make them produce chemical reactions, is an important driving force for data to empower corporate innovation. Therefore, the issue of the full-domain linkage of data resources brings many challenges to the study of management in the digital age.

The vertical linkage and completeness of big data have relatively mature technical means. Given an industry, each vertical association has a relatively mature technical solution. However, under the conditions of deep digitalization across industries, the horizontal completeness of data between enterprises and even industries still has many problems, such as incomplete horizontal completeness, defects in enterprise data sharing mechanisms, and the rapid iteration of artificial intelligence ecosystems posing challenges to the development of industrial policies.

In response to these issues, this paper proposes improvements to the research paradigm of management studies from the perspective of changes in management research. First, the horizontal linkage of

data is consistent with the human brain's learning process through extensive interaction with multi-channel information, thereby achieving high-level recognition and understanding of cognitive objects. Based on this, granular computing ideas and generalized set theory are organically combined to propose a quantifiable linkage space for horizontal linkage of data across scenarios. Second, for target management, drawing on the prism model in physics, a concept model of target spectrum is proposed, viewing the process of knowledge formation as a process from recognizing "visible light" to recognizing "invisible light," striving to provide a quantifiable and reasonable explanation for complex and vague management objectives and influencing mechanisms. Third, in response to the shortcomings of blockchain technology in management practice, starting from the typical carriers of managing massive data in management, a "centralized system data non-centralized sharing" mechanism is proposed, further clarifying the rights attributes of data to serve management empirical research. Based on this, an application plan for this sharing mechanism and multi-modal large models in management research is given. Fourth, the proposal of the horizontal linkage technical mechanism of data has completely freed the study of corporate ecosystems from the limitations of hypothesis-driven in traditional research paradigms. For the problems of parameter selection, cross-scenario reproducibility, and strong assumptions in traditional empirical studies, this paper proposes a two-way interactive double gyroscope model between policymakers, managers, and cluster decision-making. Furthermore, due to the complexity of data linkage after digitization, the mode of AI methods in corporate management research and policy research needs innovation. This paper organically combines generative adversarial networks with deep reinforcement learning models, proposing a methodology for the study of corporate ecosystems under the background of data intelligence, and summarizes a methodological framework for innovation in corporate management research patterns.

Finally, guided by the concept of full-domain linkage, this paper provides different research directions and application prospects, including the continued improvement of horizontal linkage technical mechanisms, data management order, large model systems for management empirical research, and full-domain scenario target spectra.

Key Words Big Data Science, Big Data Management, Big Data Global Linkage, Challenges A and Countermeasures