

数据智能: 趋势与挑战

吴俊杰^{1,3,4}, 刘冠男¹, 王静远^{2,3}, 左源¹, 部慧¹, 林浩¹

(1. 北京航空航天大学 经济管理学院, 北京 100191; 2. 北京航空航天大学 计算机学院, 北京 100191; 3. 北京航空航天大学 大数据科学与脑机智能高精尖创新中心, 北京 100191; 4. 城市运行应急保障模拟技术北京市重点实验室, 北京 100191)

摘要 随着大数据和人工智能的兴起, 数据智能 (data intelligence) 逐渐成为学术界和产业界共同关注的焦点. 数据智能具有显著的大数据驱动和应用场景牵引两大特征. 其融合场景内外的多源异质大数据, 利用大规模数据挖掘、机器学习和深度学习等预测性分析方法和技术, 提取数据中蕴含的有价值的模式, 并用于提升复杂实践活动中的管理与决策水平. 本文指出了推动数据智能实现迭代发展的三维要素: 数据、算法和场景, 然后围绕这三大要素介绍了数据智能的前沿热点、发展趋势和存在挑战, 特别对数据智能与管理学交叉的研究与应用问题进行了较为深入的探索. 论文还尝试给出一些具有前瞻性的观点或评论, 一来希望为有兴趣进入数据智能领域的读者提供指引, 二来希望能够在管理同行中起到抛砖引玉之效.

关键词 数据智能; 管理与决策; 大数据; 人工智能; 物联网

Data intelligence: Trends and challenges

WU Junjie^{1,3,4}, LIU Guannan¹, WANG Jingyuan^{2,3}, ZUO Yuan¹, BU Hui¹, LIN Hao¹

(1. School of Economics and Management, Beihang University, Beijing 100191, China; 2. School of Computer Science, Beihang University, Beijing 100191, China; 3. Beijing Advanced Innovation Center for Big Data and Brain Computing, Beihang University, Beijing 100191, China; 4. Beijing Key Laboratory of Emergency Support Simulation Technologies for City Operations, Beijing 100191, China)

Abstract With the unprecedented development of big data and artificial intelligence, data intelligence has emerged as a focal point in both academia and industry. It features in a set of predictive data analytics methods gathered in a big-data driven and applications oriented manner, including data mining, machine learning, deep learning, etc. It aims to extract valuable patterns from big data generated inside and outside targeted application scenarios so as to enhance real-life management and decision-making levels. This paper thus focuses on introducing the recent advances in data intelligence, which is formulated as a cyclic system including three naturally integrated and mutually functional dimensions: Data, algorithms, and scenarios. We discuss the hot topics, growing trends, as well as research challenges in data intelligence, with our own comments and opinions aiming to provide guidance for entering the area of data intelligence and arouse peer discussions on this exciting field.

Keywords data intelligence; management and decision making; big data; artificial intelligence; internet of things

1 引言

近年来“大数据”和“人工智能”先后成为学术界和产业界讨论最多的话题之一. 对于大数据一般有两

收稿日期: 2020-01-09

作者简介: 通信作者: 王静远, E-mail: jywang@buaa.edu.cn.

基金项目: 国家重点研发计划重点专项 (2019YFB2101804); 国家自然科学基金 (71531001, 71725002, 71701007, 61572059, 71901012, 91846108); 国家博士后基金 (2018M640045)

Foundation item: National Key R&D Program of China (2019YFB2101804); National Natural Science Foundation of China (71531001, 71725002, 71701007, 61572059, 71901012, 91846108); Postdoctoral Research Foundation of China (2018M640045)

中文引用格式: 吴俊杰, 刘冠男, 王静远, 等. 数据智能: 趋势与挑战 [J]. 系统工程理论与实践, 2020, 40(8): 2116-2149.

英文引用格式: Wu J J, Liu G N, Wang J Y, et al. Data intelligence: Trends and challenges[J]. Systems Engineering — Theory & Practice, 2020, 40(8): 2116-2149.

个视角. 一是数据视角, 认为大数据是由列表数据、时间序列、空间序列、关联网络、文本、图像、多媒体等构成的多源异构信息, 带来体量大、模态多、速度快、杂质多等 4V (volume, variety, velocity, veracity) 技术挑战^[1], 因此大数据处理技术, 特别是 NoSQL 的存取技术^[2] 与多模态计算技术^[3], 得到了广泛的研究和产品化尝试, 而大数据分析法则在很长时期内仍停留在数据挖掘以及分析结果可视化的层面. 二是决策视角, 即所谓的大数据思维^[4], 如大数据应作为企业战略资产、大数据改变信息不对称进而驱动市场均衡发生变化、大数据催生新的产业并推动价值创造等. 特别地, 清华大学陈国青等针对大数据环境下的管理与决策问题, 指出了大数据具有粒度缩放、跨界关联、全局视图的决策特征, 并据此提出了考虑假设转变、跨域转变、流程转变的决策新范式, 以及包括范式 (paradigm)、分析 (analytics)、治理 (governance) 和使能 (enabling) 等四大维度的大数据研究 PAGE 框架, 极大地丰富了大数据的决策理论^[5].

以深度学习为标志的人工智能的兴起, 标志着大数据研究进入了以大数据分析方法为中心的时代. 各种深度神经网络结构被提出, 以解决不同领域的大数据的建模问题, 如卷积神经网络之于图像^[6]、循环神经网络之于时间序列^[7]、图神经网络之于异质网络^[8]、注意力网络之于文本^[9] 等, 不一而足. 这些方法得益于近年来非结构化大数据如文本、图像、多媒体等的广泛涌现, 可以在大规模训练中不断调优和进化, 甚至开始以提供预训练模型的方式被用于构建和使用领域通用知识^[10], 为推进科学家长期向往的“直觉 AI” 提供了思路. 然而, 随着人工智能的不断升温, 学术界和产业界又开始出现了“AI 可能昙花一现” 的担忧. 从整个 AI 领域来说, 确实经历过挫折, 如上个世纪中期开始的知识推理和知识系统研究^[11], 成果未能尽如人意. 但如果考虑到这次的 AI 热潮不仅有大数据作为支撑, 而且有丰富的应用场景作为牵引, 这种担忧似乎来得有点过早. 事实上, 现在大众眼中的 AI 是 AI 应用——如聊天机器人、脸部识别设备、无人驾驶汽车等——而非 AI 技术本身, 这充分表明这次的 AI 热潮是冲着落地应用而来, 其向各应用领域的渗透远未结束.

综上, 大数据分析与应用中已经深深嵌入了现代人工智能技术, 并成为推动其发展的核心动力, “数据智能” (data intelligence) 的概念呼之欲出. 从管理的视角出发, 可将数据智能定义为: 通过大规模数据挖掘、机器学习 and 深度学习等预测性分析技术, 对现实应用场景的内外部多源异质大数据进行处理和分析, 从中提取有价值的信息或知识, 并用于提升复杂实践活动中的管理与决策水平. 根据上述定义, 数据智能首先是面向大数据的预测性数据分析技术, 其涵盖了 20 世纪 80 年代以来基于人工特征的机器学习、90 年代以来源自数据库的数据挖掘, 以及 21 世纪初以来的深度学习, 当然也包括了传统的统计分析及可视化技术——只是需要面向大数据进行调整和创新——以及大数据获取与处理方法和技术. 其次, 数据智能还是面向应用场景的预测性数据分析技术, 目的是为复杂实践活动中的管理与决策提供技术支持, 因此一些基础性的数据分析技术将在应用场景的需求牵引下进行创新和整合, 形成综合性的数据分析技术, 如现在已在电商平台广泛使用的推荐系统就是一类典型的综合性技术. 从上述两点来看, 大数据驱动和应用场景牵引, 是数据智能的关键特征, 也是数据智能异于人工智能的关键点, 因为后者不一定数据驱动、也更强调研究通用的方法与技术. 而源自统计学领域的数据科学概念^[12], 则在方法上范围更宽 (如涵盖计量等因果分析方法)、距离具体应用场景也通常更远.

与数据智能较为相关的一个概念, 是 20 世纪 90 年代提出的商务智能^[13], 二者相似之处在于数据驱动和场景牵引, 但具体含义已有较大差别. 首先从数据看, 商务智能的输入主要是大规模结构化数据, 而数据智能则作用于真正意义上的多源异质大数据, 即以文本、网络、时间序列、空间序列等应用领域涌现的大数据为主要分析对象. 其次从场景看, 商务智能主要应用于商业领域, 而数据智能的应用领域已远远超出商业领域, 如智慧城市、智慧金融、智慧制造、智慧医疗等, 也是数据智能的典型应用场景. 最后从方法看, 商务智能主要应用数据立方体、数据仓库技术和数据挖掘技术, 基本不涉及大数据处理技术, 也不包括前沿的机器学习和深度学习方法, 而后者正是数据智能实现大数据分析的核心技术. 尽管如此, 商务智能于提出伊始就把数据、算法、场景融合一体, 为在大数据和人工智能时代提出数据智能概念, 奠定了很好的基础. 具体而言, 大数据可以推动算法、特别是深度学习算法研究, 使算法得以在实际场景中应用, 而应用将产生更多的大数据, 它们将被用以持续提升算法. 如此一来, 数据、算法、场景三者将循环迭代、螺旋上升, 成为推动数据智能长期发展的系统动力.

本文聚焦于数据智能, 尝试从数据、算法、场景三个维度, 介绍数据智能的前沿热点、发展趋势和存在挑战. 本文无意成为数据智能领域的一个完整的综述——鉴于本领域的跨学科特征和快速发展特点, 这似乎也是不可能的——而是选取数据智能的一些热点问题, 尤其是融合管理与技术视角后有价值的研究问题, 进行脉络性介绍和趋势分析, 其中也尝试给出了一些可能不太成熟的观点和评论, 仅供读者参考并一起探讨. 本文也不是一个特定领域的教程 (tutorial), 但希望能为读者特别是管理领域的读者进入数据智能这一新兴领域提供一些方向性的指引, 尽管论文对于热点的选择肯定存在偏颇之处——也许反而能激发对数据智能的热烈讨论.

本文剩余章节安排如下. 第 2 节聚焦于数据智能的数据维度, 特别是大数据的获取问题与治理问题. 第 3 节介绍算法维度的一些前沿问题, 这里既包括虽然经典但仍被广泛使用的机器学习策略, 也包括近期被广泛关注的深度学习方法, 以及异质数据融合建模、可解释预测等核心挑战. 第 4 节介绍数据智能在若干重要场景的应用现状与趋势, 其中与移动物联网相关的场景得到了特别的关注. 最后一节总结全文并尝试给出具有前瞻性的评论.

2 数据维度

大数据是数据智能的起点, 直接驱动了数据智能算法的发展以及面向各类场景的实际应用. 数据量的大小、数据质量的高低也在一定程度上决定了各类模型的训练效果. 特别是在大数据时代, 数据分布广泛、源头庞杂, 个体借助不同类型的设备、渠道在不同的时空场景中产生数据. 因此, 如何在泛在的网络空间中高效获取和感知数据, 进而全面治理和使用数据, 成为数据智能领域的研究热点. 而在大数据资产化的背景下, 人类将更多地参与大数据的获取与治理环节, 这就意味着存在更多的管理交叉研究机会有待探索.

2.1 大数据感知

近年来随着移动互联网、物联网以及移动设备的广泛使用, 大数据来源愈发分散且呈现井喷态势. 大数据感知的核心是从大规模、分布式、泛在性的大数据中获取有价值的信息. 如何从细粒度的个体层面更有效地感知数据, 并利用群体智能 (crowd intelligence) 来提升大数据的感知能力, 成为了研究热点; 特别是其中应用到的任务分配、优化、调度等思想, 是管理科学长期关注的重点, 也是在数据智能时代值得持续关注的科学问题之一.

众包 (crowdsourcing) 是指在有限预算和时间内, 公开面向互联网大众寻求服务的分布式问题解决机制, 因而被认为是一种利用群体智能实现高质量的数据感知的机制^[14]. 具体而言, 基于众包的数据感知是将感知任务拆解成简单子任务发布在互联网上, 利用网络大众以有偿或无偿的形式来完成子任务并反馈给请求人^[15]. 目前国内外常见的众包商用平台包括 Amazon Mechanical Turk (AMT)、gMission、CrowdFlower、猪八戒等. 研究者通常引入众包来完成一些复杂的数据感知和数据采集问题, 如图像打标签^[16]、实体标注^[17]等. 众包感知机制的核心问题和挑战在于如何将任务拆解分配并确保任务能够高质量和有效执行. 一方面, 可采用分治等策略分解复杂任务, 减小子任务执行难度^[18]; 另一方面, 则根据任务属性设计不同激励措施以提高任务完成人的积极性^[19,20]. 此外, 众包参与者的筛选与子任务分配也是实现成本最小和质量最优化的重要策略^[21,22]. 在参与者提交任务结果之后, 如何推断最优答案并进行任务的质量控制, 目前尚无所有任务通用的整合策略^[23,24], 因而有待进一步研究.

广泛渗透的在线社会化媒体也提供了数据感知的平台基础. 用户在社交媒体平台上产生的内容、图片、位置等信息都可被视作数据源. 特别是用户在使用各种社交产品时产生的情境状态等相关数据 (如微博签到、打卡等), 经过融合和挖掘后能够提供重要的信息价值^[25]. 然而, 由于社交媒体的自媒体性和参与行为的异质性, 总体信息呈现杂质过多、信息密度过低的特点. 因此, 如何在分布广泛、内容质量参差不齐的数据源中实现有效感知, 也成为了一个重要的研究问题. Yuan 等^[26]提出了社会感知最大化 (social sensing maximization) 的概念, 通过在微博平台设置社会感知器 (social sensors), 实现了对热点话题、爆发话题的快速有效感知. Xie 等^[27]提出通过设置限定规模的社交网络感知器进行信息流的组织, 以作为突发信息监测的数据源, 能够减少计算复杂度和提升性能. An 和 Weber^[28]发现选择有完整背景信息的用户作为社交媒体感知器, 能够在获取更高质量感知数据的同时显著降低数据规模. 目前针对社会传感的研究仍处于起步阶段, 如

何实现传感群体的实时动态评估以面向具有不确定性特征的突发感知任务, 仍然是一个颇具挑战性的问题。

随着嵌入式设备、无线传感网络、物联网和智能移动终端等的快速发展, 大规模感知系统的部署极大增强了数据感知能力, 也拓宽了众包感知和社会感知等概念的内涵, “移动群智感知”(mobile crowd sensing)^[29]的概念应运而生, 即以普通用户移动智能终端和移动传感设备作为基本感知单元, 协调物联网/移动互联网进行任务分发、数据收集、融合、分析挖掘等, 最终完成大规模、复杂的城市与社会感知任务, 亦称为空间众包(spatial crowdsourcing)^[30]。与传统感知方法相比, 移动群智感知更加强调大量普通用户行为的广泛性、移动性和机会连接性。针对移动群智感知的特点, 目前的相关研究着重关注以下几个方面的挑战。其一, 群智任务分配感知问题。群智感知严重依赖用户参与, 需要针对感知任务的特定需求, 结合任务的时空特征和用户属性实时地将任务分配给适合的参与者执行, 并设置合理的激励措施以高效完成感知任务^[31]。其二, 群智数据质量控制问题。用户感知行为受到感知情境、设备状态、技能水平等因素影响, 造成感知数据的不完整性、冗余性和不准确性等质量问题, 如何从状态不一、良莠不齐的原始数据中清洗出优质数据, 目前仍是巨大的挑战^[32,33]。其三, 群智数据建模分析问题。群智数据蕴含丰富信息, 需要对多模态的时空感知数据进行有效的融合和建模, 以实现有价值信息的深度发掘^[34,35]。其四, 群智感知隐私保护问题。群智感知网络记录用户行为轨迹, 通过用户行为轨迹数据挖掘能够获得用户日常行为、职住位置等敏感信息, 因此很多时候需要在保护用户隐私的前提下来执行数据采集任务^[36,37]。

2.2 数据治理与数据共享

在大数据感知的基础上, 另一个重要的挑战则是实现对数据的有效治理。一方面希望通过数据共享来更好地整合多源异构数据, 实现应用创新; 另一方面, 数据安全又日益成为一个社会大众密切关注的问题。事实上, 在大数据环境下, 隐私保护已成为社会共识; 特别是在 Facebook 等社会化媒体平台出现数据泄露问题后, 如何实现数据共享与数据安全的平衡, 也日益被业界和学术界所重视。

一类研究聚焦于设计算法以对数据记录进行匿名化处理, 从而保护用户隐私。数据记录的匿名化可以通过不同的隐私保护模型来实现, 其中 k -匿名模型^[38]是最经典的模型之一, 其基本思想是通过将原始数据记录的处理(如添加数据噪音、数据粒度概化), 确保每条记录在可用于识别用户身份的属性(如年龄、性别等)上至少与其他 $k-1$ 条记录相似, 从而令攻击者无法识别某条记录的身份。在此基础上衍生出诸多扩展模型, 如 l -多样性模型^[39]、 t -贴近度切片^[40]等。此外, 面向特定数据类型和应用场景, 研究者有针对性地提出了各类隐私保护方法。如 Li 等提出了一种新方法提取、聚类和匿名化医疗文本记录^[41]; 面向交易数据的隐私保护, 则提出了一种基于最小生成树技术的聚类方法, 并利用群集级微扰方法来解决数据共享的隐私问题^[42]。这类匿名化的隐私保护方法被应用于电子健康^[43]、智能电网^[44]等领域。在保证匿名化效果的同时, 如何保障经匿名化处理后的数据集在各类数据分析应用下的服务效用(utility), 则是另一类研究重点。已有研究提出采用诸如互信息(mutual information)、KL 散度(KL-divergence)等指标对隐私保护后的数据进行可用性度量, 并设计优化算法在保证隐私保护效果的前提下, 最大化数据可用性^[45]。同时, 数据挖掘和机器学习技术的发展也进一步推动了隐私攻击方法的发展, 这使得共享和挖掘用户非隐私数据亦可能导致用户的隐私数据被准确推断(即属性推断攻击)^[45,46]。因此, 数据智能环境下的攻击方法更加难以防护, 这也使得隐私数据保护面临巨大挑战。

要更深层次地解决数据的隐私、共享、定价等治理问题, 需要从机制设计、优化决策等角度进行深入的探讨, 这本身也是管理学与数据智能的典型交叉领域问题之一。其中一个重要角度是从用户行为中识别其对于隐私性问题的认知, 从而设计合理的隐私管理机制。例如, Heimbach 和 Hinz^[47]利用随机控制实验来验证社会化媒体中的内容共享机制对共享行为的影响; 他们发现在合理的共享机制下, 内容提供者将显著地增加共享内容, 同时验证了“两次点击”的机制设计会负向影响内容共享行为。同时, 大型互联网企业对用户信息的收集也引起了用户对于隐私的担忧并唤醒了其对私人信息价值的意识, 因此对于隐私价值的研究也引起了广泛的研究兴趣^[48,49]。例如, 有研究讨论了手机 APP 的用户隐私权限与价格的交换机制, 发现价格较低的 APP 使用了更多的敏感信息, 但在相同的价格和功能的前提下, APP 使用过多的敏感信息将导致用户的需求更低^[50]。Schneider 等则设计了一种保护市场数据的方法, 以在信息损失和保护缺失之间取得较好的平衡, 该法特别适用于数据提供商能选择数据保护程度的场景^[51]。实际上, 数据金钱化策略(monetization)是实

现数据共享的一种有效方式,具体策略包括了交易费用、访问费用以及对单边的补贴策略等^[52]。在此基础上可建立新型的数据市场,将数据交易作为核心活动,或者将被交易的数据本身作为平台核心^[53]。一般而言,数据市场是多边平台,由电子平台作为中介连接数据提供者、购买者以及其他技术的提供者;但就现实而言,数据仍然很少通过多边平台进行大规模交易和共享^[54]。Koutroumpis 等提出区块链技术有潜力为未来的数据交易市场提供可执行的合同,同时还详细说明了集中多边数据交易平台的机制^[55]。

当前,越来越多应用场景的大数据呈分布式存储且囿于特定原因很难实现共享。如何为这类场景训练数据智能模型,成为一个挑战性问题。一类解决方案是在采用安全数据传输协议的基础上对数据进行汇总训练,例如采用的分布式隐私保护协议主要包括一组安全协议和基于操作的协议两种类型,前者包括遗忘的传输协议^[56]和同态加密协议^[57]等,后者包括安全总和协议、安全集合并集协议^[58]等。另一类方案则追求在数据无须汇总的情况下完成模型训练。谷歌于 2016 年率先提出联邦学习 (federated learning)^[59] 的概念,以在各终端不共享隐私数据的条件下实现协同学习。联邦学习无需汇聚各终端的数据进行集中训练,而是可以在各终端上进行加密的分布式计算,且各终端之间通过加密机制实现参数交换,从而建立一个虚拟的公有模型。激励各终端参与是联邦学习的重点,其核心机制是模型效果的提升,模型效果会通过联邦机制反馈给各节点,形成一个正反馈的闭环。根据不同终端中的数据在数据项与特征项上的重叠程度,联邦学习可以被进一步分为横向联邦学习和纵向联邦学习。这种新兴人工智能学习策略和技术有望成为未来人工智能分布式学习和协同学习的基础,尽管有研究指出联邦机制下共享梯度仍可能带来信息泄露问题^[60]。

总体而言,数据共享是数据智能取得进步的重要基础。在当前政府、企业、个人都越来越注重数据隐私和安全的前提下,如何从技术层面设计更好的算法来加强隐私保护,同时从机制角度设计合理的数据交易平台和交易机制,仍面临巨大的挑战。管理学领域的研究者可以借助优化决策、实证分析等手段对数据这一数据智能的重要维度开展更加深入的探索。

3 算法维度

数据智能中的预测性数据分析算法来自数据挖掘、机器学习、人工智能等多个领域,为阐述方便以下简称“智能算法”。智能算法是实现大数据智能的灵魂,决定了数据能够发挥的效用以及在实际场景中的应用效果,其重要性不言而喻——尽管业界也广泛存在数据提升应优于算法提升的看法,但二者也不矛盾。一般认为,智能算法建模涵盖特征学习和基于特征的模型学习两个阶段,前者的重要性在基于连接主义的深度学习出现以后得到了更为充分的认识,而端到端学习的流行则进一步将两阶段合二为一以取得更好的建模效果。由此可见,深度学习在智能算法中占据了特别重要的地位,因此是本节阐述的重点内容。尽管如此,一些经典的基于符号主义的机器学习策略如组合学习、强化学习等,仍在当前的许多应用场景中发挥着非常重要的作用,而且易于结合深度学习以达到更好的效果,因此也是本节关注的重点。此外,鉴于概率模型在管理领域的重要性——在完成预测任务的同时容易得到统计意义上的解释性——以及文本数据在管理领域的普遍性,本节还将重点关注概率隐变量模型和文本分析模型。

3.1 经典学习策略

传统的知识发现和数据挖掘过程通常涉及分类、聚类、关联等基本算法,然而这些算法应用于真实数据场景时往往面临训练数据稀疏、标签不足、噪声大等问题,因此有必要采取一些特殊的学习策略以提高基础算法的鲁棒性,进而提升学习效率和预测效果。有鉴于此,本节关注四类经典学习策略以应对不同的复杂学习场景: 1) 组合学习,主要应对单一基础学习器的过拟合、局部最优化等问题,面向多基础学习器进行集成建模; 2) 融合学习,主要应对基于单一数据源的模型预测精度不足的问题,侧重于多源数据的融合建模; 3) 迁移学习,目的在于将学到的知识从一个场景迁移到另一个场景,侧重于知识的迁移、复用和适应,以避免新场景中高代价的数据收集或标注工作; 4) 强化学习,主要解决复杂场景(如机器人行动或游戏)中数据标签的延时反馈问题,侧重于学习路径的优化。由于具有明显的实用性和有效性,以及易于容纳深度学习的适应性,可以预期这四类学习策略在相当长的时间内仍将具有非凡的活力;但面对大数据时如何提升学习效率,始终是应用这些策略时值得关注的问题。

3.1.1 组合学习

组合学习旨在生成并整合多个基础学习器(如分类或聚类模型),以得到更为稳健的组合模型并提升预测效果.组合学习能够有效解决单一模型的过拟合、局部最优化等问题,因此被广泛应用于分类^[61]、聚类^[62]、关联分析^[63]、异常检测^[64]等经典数据挖掘任务,同时也与半监督学习、主动学习、代价敏感学习等学习方法结合紧密.本节以组合分类和组合聚类为例,对组合学习的关键技术进行介绍.

在组合分类方面,传统方法主要通过模型训练过程中引入多样性来提高分类预测精度.常见方法中有的对样本点进行再采样(resampling),进而产生多个训练集,如装袋法(bagging)^[61]和提升方法(boosting)^[65];有的对特征集合进行选择,如随机森林(random forest)^[66];有的对类别标签进行处理,如错误-纠正输出编码^[67];还有的对学习算法进行调整^[68]等.除此之外,还存在一类对基础分类模型输出进行融合的方法,包括加权方法^[69]、元学习^[70]等.近年来,大规模数据中的快速组合分类成为一个热门研究主题.Chen和Guestrin提出快速梯度提升法(XGBoost),该方法使用节点的默认方向处理稀疏数据,使用合并和修剪操作处理加权数据,并且高效地枚举所有可能的拆分^[71].鉴于神经网络包含多个非线性操作层,因而能够在大规模数据集中挖掘复杂结构并学习高层次语义,一系列结合深度神经网络的组合方法逐渐涌现,如Kontschieder等提出深度神经决策树^[72],将卷积神经网络与决策树进行融合.

与组合分类相比,组合聚类的一个难点在于缺乏真实的外部类别标签信息^[62],这导致很难确立一个客观上“好”的优化目标.目前的组合聚类大致可划分为有明确全局目标函数和无明确全局目标函数的两类方法.前者通常采用一个全局效用函数来衡量所有基础聚类结果与组合聚类结果之间的相似性^[73];而后者则着力于设计能有效集成多个基础聚类结果的数据结构——如超图(hypergraph)^[74]、协联矩阵(co-association matrix)^[75],然后采用已有的方法来完成对数据的建模.近年来,对组合聚类的收敛理论和学习效率研究是该领域的研究热点.Wu等提出基于列联矩阵(contingency matrix)的大数据组合聚类方法,实现了与K-均值算法效率相媲美的快速组合聚类^[76].Liu等则进一步提出基于协联矩阵的谱组合聚类算法,用于大规模文本分类与社会化媒体用户分群,并首次从理论上严格证明了组合聚类的泛化性、鲁棒性和收敛性^[77].

时至今日,组合学习仍在众多应用场景中被广泛使用,如XGBoost甚至已经成为很多新算法的“梦魇级”基线方法;最新的研究也表明在防御对抗样本(adversarial examples)的直接攻击方面,随机森林与深度神经网络相比具有明显优势^[78].然而在大数据背景下,如何应对数据不均衡、数据特征和标签分布动态变化、数据高维稀疏等长期问题,并适应在线和离线数据的快速和海量计算需求,将持续成为组合学习的关键挑战.

3.1.2 融合学习

融合学习旨在对异质信息进行融合建模,以解决基于同质信息的模型在鲁棒性、预测力上的先天缺陷.融合学习的核心在于对异质信息进行分解、表达和重组,从而实现对异质信息之间的复杂关联性的刻画和使用.从当前的研究看,基于概率框架的概率图模型和基于深度网络的表示学习,已经成为融合学习的重要方法;传统的多视角学习(multi-view learning)也是一类融合学习策略,但经常受限于过于复杂的优化求解算法.从广义角度看,可以把上一节的组合学习策略认为是一类融合学习策略,只不过后者更强调从数据层面就开始融合,而前者则聚焦于模型层面的融合.由此不难推断,融合学习的复杂度往往会高于组合学习,因此提升融合学习效率和设计有效的融合策略一样,都是这类学习策略的核心挑战问题.下面介绍融合学习的三类重要方法:概率图模型、表征学习、基于正则项的多视角学习.

概率图模型(probabilistic graph model, PGM)是一种描述数据生成过程的统计学习方法,其通过引入隐变量来刻画数据之间的关联性.在建模同质信息时,这种关联性通常可以理解为降维,例如建模文本生成过程的隐狄利克雷分配(latent Dirichlet allocation, LDA)模型^[79],通过引入主题作为隐变量,将词聚成了语义相关的词簇.而在建模异质信息时,隐变量可以成为关联异质信息要素的桥梁,即以隐变量张成的潜在空间连接异质信息.此外,由于概率图模型可描述信息的生成过程,因此它们不仅仅能够融合异质信息做出预测,还可为融合的过程提供较好的解释性,因此是一类具有可解释预测能力的模型.基于概率图模型的融合学习模型包括:同时建模文本和数值信息的协同主题回归模型^[80,81],同时建模飞机开关控制和传感器观测数据的动态贝叶斯模型^[82],以及同时建模序列和网络的异常用户行为检测模型^[83]等.这类模型的寻参过

程通常较为复杂,收敛速度一般较慢。

表征学习 (representation learning) 是随着深度学习流行起来的一类方法,旨在应用神经网络模型从数据中自动抽取实值向量作为样本的表征,从而免去手动特征工程的操作,同时也便于在表征后的特征向量上直接进行特征融合,从而屏蔽异质的原始数据在形态上的差异性。典型的多模态数据表征方法包括变分自编码 (variational autoencoder, VAE)、网络嵌入式表示 (network embedding) 等。在对表征后的特征向量进行融合时,常见的做法包括在同一个隐空间对文本、图像、数值等多模态数据的隐含表示进行对齐^[84],利用注意力机制对文本、用户、网络等信息的嵌入式表示进行融合^[85],通过网络元路径对异质信息网络中的节点的嵌入式表示进行关联^[86],甚至直接把异质信息的嵌入式表示拼接成向量然后输入组合学习器如 XGBoost 等。凭借深度学习的优越性,基于表征学习的融合学习策略通常表现很好,但缺乏良好的可解释性成为一个不容忽视的约束。本文将在 3.2 节的深度学习模块对表征学习进行较为详细的介绍。

基于正则项 (regularization) 的多视角学习,旨在将辅助信息以正则项的方式融入到面向主体信息的模型学习过程中。此类方法将目标函数定义为多个子损失函数相加的形式,其中有一部分子损失函数是正则项,用以对一些辅助信息加上先验约束。例如, Wang 等在对城市浮动车流量进行 Tucker 张量分解基础上,引入城市兴趣点信息作为正则项,从而获得对城市时空分布规律的认识^[87]。Wang 等将知识图谱这样的外部信息作为正则项,引入到基于用户与物品交互的矩阵分解框架中,有效地提升了推荐系统的精度^[88]。Wang 等在研究医院和居民区之间的客流分布时,除了引入医院和居民区的地理位置信息作为正则项外,还把在缺失值上的校准信息作为正则项引入,在提升客流预测精度的同时解决了冷启动问题^[89]。基于正则项的多视角学习通常在预测精度和学习效率上有较好的平衡,而且具有一定的可解释性,因此在进行融合学习时被广泛使用。

融合学习作为一个前沿方向,在如下两个方面呈现出新的趋势。一方面,模型由少数模态的简单融合向复杂多模态(如网络、文本、图像等)的融合表示学习转变,出现了一些基于深度生成模型的融合表示学习方法,如基于变分自编码的多模态学习模型^[90]。这类模型不仅在多模态协同表示能力上有较大增强,同时在参数学习效果方面有较大改进。另一方面,针对传统融合模型目标函数中损失函数权重需要人工设定或者以搜索方式进行寻优的问题,已有模型开始尝试对多个组件权重进行自动学习。如 Kendall 等^[91]在基于深度学习框架对不同模态数据进行表示的同时,引入贝叶斯学习对不同模态数据的不确定性进行建模,以自动衡量融合模型中不同模态的权重。

3.1.3 迁移学习

尽管身处大数据时代,各应用领域的数据特别是人工标注信息仍有贫富之分,如英文语料的情绪标签就远比中文语料来得丰富,新闻语料的话题标签也比用户生成内容的话题标签丰富得多。如何对数据资源相对匮乏的领域进行有效建模,是一个颇具挑战性的问题。另一方面,现有的计算和存储资源仍然难以满足所有领域的计算需求,这就意味着应该在不同领域间共享一些计算结果以减少总体计算成本。为了解决这些问题,迁移学习 (transfer learning) 应运而生。迁移学习是利用数据、任务或模型之间的相似性,将应用于某领域的模型迁移应用到新领域的学习过程。

迁移学习方法大致可分为四类^[92],即基于实例的迁移学习 (instance-based transfer learning)、基于映射的迁移学习 (mapping-based transfer learning)、基于网络的迁移学习 (network-based transfer learning) 和基于对抗的迁移学习 (adversary-based transfer learning)。基于实例的迁移学习认为源域中的部分实例有助于目标域的学习。其使用特定的权重调整策略,通过比较源域与目标域的相似度为源域实例分配适当的权重,以从源域中选择部分实例补充目标域训练集来完成迁移^[93]。该类方法适用于源场景和目标场景具有相似样本分布但具体实例有较大差异的情形,例如在具有相同竞品的不同公司之间进行模型迁移。基于映射的迁移学习方法认为在精心设计的数据空间下,源域和目标域可以更相似。其通过特征变换的方式将源域和目标域中的实例映射到新的数据空间,以减少源域和目标域之间实例的差异性^[94]。该类方法适用于处理特征不同但是具有一定相似性的源场景和目标场景之间的模型迁移,例如在具有类似业务的公司之间进行模型迁移。基于网络的迁移学习认为迁移学习过程是一个迭代的、连续的抽象过程。其在源域预先训练好部分网络,并将其看作特征提取器,提取出的特征应用于所有目标域^[95]。该类方法被广泛应用于从大数据集向小数据集

的迁移, 即源场景具有丰富的数据样本、但目标数据集的样本较为稀疏, 如自然图像数据集向医学图像数据集之间的迁移. 基于对抗的迁移学习在生成对抗网络的启发下引入了对抗学习技术, 试图寻找同时适用于源域和目标域的可迁移表示^[96]. 该方法是迁移学习近两年的研究热点.

在数据智能时代, 迁移学习得到了前所未有的关注和研究, 并在城市交通管理、电子商务、医疗健康等领域得到了广泛的应用^[97]. 尽管如此, 迁移学习仍有大量开放问题尚待解决. 首先迁移学习与深度学习类似, 也存在可解释性弱的问题, 这极大地限制了迁移学习在决策敏感领域的应用. 其次, 如何将人类经验融入迁移学习的过程中, 将机器智能与人类经验结合起来, 也是迁移学习下一步的研究重点. 此外, 如何利用迁移学习完成普适知识的提取从而实现较远的领域之间的迁移, 如何将迁移学习应用于实时数据从而完成在线迁移学习等, 都亟需进一步的研究和探索.

3.1.4 强化学习

强化学习 (reinforcement learning) 是指智能主体 (intelligent agent, 如模型、算法) 在与动态环境进行交互中, 通过反复试验和试错来进行学习的方法. 相较于传统的有监督学习, 强化学习的应用场景通常具有两个特征 (通常第一个特征是必要的, 第二个特征是可选的): 首先在强化学习中, 智能主体不会被显式地告知要采取哪些动作, 而必须尝试和发现哪些动作可以获得最大的回报^[98]; 此外, 主体的行为可能会对环境产生影响, 即环境和主体之间会存在一个互相影响的交互关系. 显然, 强化学习比传统的有监督学习更接近真实的应用场景, 因此也就具备了更加广阔的应用空间.

经典的强化学习方法采用基于价值函数 (value-based) 的建模方法^[99], 通过蒙特卡罗法和时序差分法进行训练. 其中时序差分的思想更为主流, 以此为基础发展而来的经典算法包括 Q-learning 和 Sarsa^[100] 等. 基于价值函数的建模方法在面临连续空间决策时, 会陷入维度灾难的困境. 为解决该问题而设计的强化学习算法, 包括价值函数近似、策略梯度以及同时进行价值函数和策略逼近等^[99]. 目前, 上述强化学习方法已经在围棋^[101]、机器人^[102]、游戏^[103] 等领域取得了众多突破, 其中最为有名的就是 AlphaGo 围棋机器人的巨大成功.

强化学习核心是解决序列决策问题, 因此天然契合管理实践中面临的复杂动态决策问题. 随着强化学习理论和应用研究的不断深入, 强化学习在交通管理、金融决策、物流调度等领域也获得了广泛的应用. 在交通管理领域, 强化学习方法的一个重要应用就是对于交通信号灯的管理控制. 传统的交通信号灯控制方法大多基于预设的控制策略, 对于交通状况的动态性建模不足; 而使用强化学习的方法, 可以根据不同时段交通动态制定交通信号灯的控制策略^[104]. 例如, 可使用基于价值函数近似的 deep Q-learning 深度强化学习框架来优化道路交通信号灯控制^[105]. 通过将深度神经网络与交通领域知识相结合, 可设计面向复杂路况的深度强化学习模型以支持大规模交通灯管理决策^[106].

在金融投资领域, 强化学习方法可用于生成动态资产投资策略. 在传统基于有监督学习的投资策略生成方法中, 人们通常会构建预测模型对资产的未来状态进行预测, 并基于预测状态构建投资策略, 从而形成“预测 + 策略”的两步策略生成方法. 而基于强化学习的投资策略能够将预测和策略整合为一个模型整体, 因此通常可以获得更好的投资优化效率. 早期的强化学习投资策略基于 Q-learning 及其改进的 multi-agent Q-learning 算法进行金融市场交易决策^[107]. 在策略梯度框架下, 经典的循环强化学习架构及其神经网络扩展用于生成投资组合策略^[108]. 随着深度学习方法和可解释预测方法的兴起, Wang 等结合深度学习时序建模方法, 提出基于策略梯度的深度强化学习框架 AlphaStock^[109]; AlphaStock 通过优化夏普比率来构建“买赢家、卖输家”的投资组合, 不仅表现出具有很好的投资绩效, 还对投资策略的生成提供了很好的解释性, 这对于金融领域的应用尤为重要.

强化学习同样也广泛应用于资源的分配调度问题, 如供应链库存管理^[110]、资源分配^[111]、车辆调度^[112,113] 等. 近年来随着网约车应用的兴起, 出行订单的分派成为强化学习在该领域的重要研究问题. 滴滴等公司将司机和乘客匹配的订单分派问题建模为一个序列决策问题, 将融合强化学习和组合优化理论设计的智能派单算法部署于实践应用, 可在确保乘客出行体验的同时提升司机收入^[114,115].

从现有研究中不难发现, 强化学习的发展具有两个明显的趋势: 一是与更多实际管理决策场景结合更加紧密, 其核心研究点在于寻找合适强化学习特点的应用场景, 并对场景进行模型抽象, 设计合理的强化学习

目标函数;二是将深度学习、元强化学习等复杂场景建模方法同强化学习相融合,设计能够有效建模应用场景复杂动态性的模型函数,从而提升强化学习的性能,提高模型在真实场景中的可用性^[116]。

3.2 深度学习

随着信息技术的进一步发展,其瓶颈已经逐步从数据获取与计算的能力,转变为从海量异构数据中提取抽象信息并转化为可用知识的能力^[117]。在此契机下,以深度学习为代表的人工智能技术迅速崛起。深度学习在对数据进行抽象化表征和端到端建模^[118]等方面的出色表现为上述瓶颈提供了一种可能的解决途径。但深度学习方法缺乏可解释性,这限制了其在许多领域的应用。因此,近几年对可解释深度模型的研究越来越多。本部分将分别从端到端学习、表征学习和可解释性等三个角度,对深度学习的研究进展进行简要介绍。当然,作为当前炙手可热的一个领域,深度学习的研究热点绝不仅限于此——如应包括深度神经网络模型的对抗防御(adversary defense)问题,但限于篇幅不再展开。

3.2.1 端到端学习

端到端学习(end-to-end learning),是指模型能够直接将原始输入映射为特定任务需要的最终输出。在传统基于机器学习的数据分析任务中,通常需要先对数据进行特征工程,提取数据特征,然后再选择合适的机器学习模型,模型的输入是人工设计的特征,输出是任务需求。深度学习模型将二者融合为一体,端到端地完成特征工程与机器学习的步骤。深度学习模型的这一特点,非常适合图像、语音、自然语言处理等难以进行人工特征提取的数据,因此在这些领域得到了非常广泛的应用^[119]。

端到端模型的发展历程是一个随着数据需求不断产生特殊的模型网络结构的过程。一方面,在不同类型的数据分析需求的驱动下,研究者不断提出能够适应数据特点的深度学习神经网络结构;另一方面,深度神经网络灵活多变的特点,为新型网络结构的不断涌现提供了广阔的空间。最基本的深度神经网络是全连接多层感知器(multi-layer perception, MLP),也被称为前向传播网络或深度前馈网络^[120]。目前较为流行的网络结构包括卷积神经网络、循环神经网络、注意力网络和图神经网络等。

卷积神经网络(convolutional neural network, CNN)是针对图像等具有矩阵结构的数据而设计的。CNN通过引入包含卷积和池化的操作的卷积层,可以更好地挖掘特征的局部关系;使用权重共享, CNN的参数规模也得到了极大的减少。具有代表性的CNN网络包括VGG系列网络、GoogleNet、ResNet等^[121]。这些网络大多被应用在图形图像相关的计算机视觉领域。由于围棋棋盘数据具有非常类似图像的矩阵特性,因此著名的AlphaGo围棋程序也是基于CNN进行设计的^[122]。

循环神经网络(recurrent neural networks, RNN)被设计用以处理具有序列特征的数据,如时间序列数据、文本序列数据等。RNN网络的特点是将当前时刻的数据和上一时刻网络的状态一起输入到当前时刻的网络当中,从而建模数据中的时间依赖关系。最早提出的RNN网络被称为Basic RNN,其在训练时无法很好地解决长距离依赖和梯度爆炸或梯度消失问题。LSTM(long short-term memory)^[123]在传统的RNN网络结构上加入输入门、遗忘门和输出门,通过门控结构解决长距离依赖问题。GRU(gated recurrent unit)^[124]则在简化LSTM结构的同时保持着和LSTM相同的效果。目前,LSTM和GRU是应用最为广泛的RNN模型,其主要应用场景为语音处理、文本处理、时间序列和轨迹建模等。

注意力神经网络(attention networks)常与RNN结合用于解决序列建模问题。注意力(attention)机制广泛存在于人类神经系统中:当阅读或翻译时,人们将注意力集中在关键的词句上而不是整篇文章。类似地,将注意力机制加入深度神经网络中,可以使模型更加关注输入数据中相对重要的部分。广泛的实验和应用显示,引入注意力机制的RNN网络在机器翻译等多项任务中可以取得更加优异的性能^[125]。2017年,Google提出了仅使用注意力机制的Transformer模型^[9],其使用注意力机制完全代替了RNN的网络时间序列依赖结构,并声称“Attention is All You Need”。后续,Google基于Transformer提出了具有3亿个参数的BERT模型,并提供了在大规模文本上进行预训练的模型参数^[10]。BERT的出现催生了大量基于BERT预训练参数并在特定任务上进行调优的应用研究。

除了适合CNN的矩阵型数据和适合RNN的序列型数据,现实世界中的数据有时会表现为更复杂的网络结构,如社交网络、蛋白质网络、交通结构等。为更好地建模此类数据,图神经网络(graph neural networks, GNN)应运而生^[126,127]。GNN模型在网络结构上进行信息的聚合,通过临近节点的权值来更新隐

藏节点的状态, 可视为传统随机游走模型的深度神经网络版本. 由此发展的图卷积神经网络 (graph convolutional networks, GCN) 将卷积操作引入图神经网络^[128], 挖掘图中局部感受野的邻域信息. 图注意力网络 (graph attention networks, GAT) 则将注意力机制应用于图的信息收集^[129]. GNN 及其改进模型对非结构化图数据的有效建模, 使其在交通流量预测、推荐系统、知识图谱等众多领域取得了广泛的应用^[127].

通过梳理深度学习端到端模型的发展历程可以看出, 不同数据类型的涌现驱动着神经网络从基础的 MLP 发展出了 CNN、RNN、GNN 等多个不同分支. 但管理学要研究的是非常复杂的社会经济系统, 数据不会局限于某一类型, 因此多种神经网络家族之间的交叉融合将会是深度学习技术在管理学领域的重要发展方向. 此外, BERT、ResNet 等超大规模预训练网络的出现, 给相关领域的研究创造了机遇的同时也带来了挑战. 在机遇方面, 大量面向应用的研究可以利用这些预训练模型来完成之前难以完成的任务; 在挑战方面, 很多人也在担心是否除了 Google 等行业巨头, 其他科研机构已经难以在模型设计上有所突破——没有足够的资源完成参数学习和效果验证, 因此只能进行应用上的研究创新. 如何在深度神经网络的基本原理上有所突破, 如实现真正的类脑计算, 也许是交叉领域的专家学者应该突破的重要方向.

3.2.2 表征学习

利用机器学习方法的前提是将数据表示成方法能够处理的形式. 传统的做法是在领域专家或知识的指引下, 通过特征工程手动从原始数据中抽取特征, 进而将数据表示为特征向量. 这类做法虽然行之有效, 但是通常费时费力, 且严重依赖专业知识. 随着深度学习的兴起, 表征学习 (representation learning) 得到研究者的广泛关注. 配合端到端学习, 表征学习不但能自动从数据中学习特征向量 (即表征), 还能利用这些特征完成具体任务. 这不但避免了手动抽取特征的繁杂工作, 往往还能在任务上获得较好的性能. 近年来, 表征学习在文本、网络 and 知识图谱等非结构化数据上取得了很好的效果. 以下以这三类数据为例, 介绍表征学习的研究特点.

文本的表征学习主要包括词嵌入表征和文本表征. 词嵌入表征的概念最早由 Bengio 等^[130] 在 2003 年研究神经网络语言模型时提出. Mikolov 等在 2013 年提出的 Word2Vec 算法^[131] 是首个专门研究词嵌入表征的研究, 引发了自然语言处理领域的一场变革, 即在许多方法中用词嵌入表征取代传统的 one-hot 表示. 词嵌入表征是一种稠密的低维实值向量, 避免了传统 one-hot 表示的高维稀疏的缺陷. 此外, one-hot 表示无法刻画词之间的语义关联, 而词嵌入表征可以通过表征之间的相似度来反映词之间的语义相关性. 文本表征可以简单地通过对其中的词表征进行聚合获得, 例如取文本中所有词向量的加权平均值^[132] 得到句子向量. Doc2Vec^[133] 是 Word2Vec 的拓展, 该算法用于预测一个向量来表示不同的文档, 克服了词袋模型 (bag-of-words) 缺乏语义的不足. TextCNN^[134] 则使用卷积核提取局部特征, 通过整合不同大小的 n-gram 特征作为整个句子的表示. RCNN^[135] 利用 RNN (LSTM/GRU 等) 来克服词袋向量的语序问题, 以最后一个隐状态作为整个句子的表示.

在文本表征学习的启发下, 大规模网络表征学习很快也成为了热点研究问题. Perozzi 等最早提出了 DeepWalk^[136], 利用随机游走生成大量节点序列, 然后视为文本并利用 Word2Vec 的思想学习节点表征. Node2Vec 在此基础上改进了随机游走过程^[137]. Tang 等进一步提出了 LINE, 适用于大规模的有向带权图的网络表示学习^[138]. 考虑到网络动态演化的特性, Zuo 等提出了能够刻画节点邻居变化过程的大规模动态网络表征学习模型 HTNE^[139]. 此外, 部分工作利用深度神经网络对一些中小规模的网络进行表征学习, 这类方法能够刻画节点间的非线性关系, 并且便于引入网络拓扑结构以外的信息. 例如, SDNE 使用深层神经网络对节点表示间的非线性关系进行建模, 将深层自编码器的中间层作为节点的表征^[140].

知识图谱是一种特殊的网络, 每个节点代表现实世界中的实体, 节点间的边表示实体之间的关系. 知识表征学习即面向知识库中的实体、关系进行表征学习, 以在低维空间中高效计算实体和关系的语义联系, 有效解决数据稀疏问题, 使知识获取、融合和推理的性能得到显著提升. 以 TransE^[141] 为代表的知识表示学习模型, 已经在知识图谱补全、关系抽取等任务中取得了瞩目成果, 其基于实体和关系的分布式向量表示, 将每个三元组实例 (head, relation, tail) 中的关系 relation 看做从实体 head 到 tail 的翻译. 然而, TransE 只适合处理一对一的关系, 不适合一对多或多对一的关系. 因此, 后续研究从不同角度尝试解决复杂关系建模问题, 提出了 TransH^[142]、TransR^[143] 和 KG2E^[144] 等多种模型, 以期有效提升 TransE 的性能.

表征学习还被广泛用于其他类型大数据如空间轨迹^[145]和社媒用户^[146]等,此处不再详述。从相关的最新工作来看,表征学习的研究趋势包括但不限于如下两点。一是通过学习“分离表征”(disentangled representations)——指表征的各个维度独立且具有特定含义——来增强表征的可解释性和可操作性^[147]。例如,文本的分离表征中某些维度控制情绪,而某些控制时态等语法^[148],通过修改相应维度的表征可以控制生成文本的情感或时态。二是在端到端表征学习的同时融合更复杂的任务,而不再仅仅基于表征做分类、回归或结构学习等预测任务。例如,将现有组合求解器(combinatorial solvers)融合到端到端的前馈和梯度反传中,将神经网络的表征学习和组合优化有机地融合起来^[149]。这有利于将深度学习应用到更多真实场景中。

3.2.3 可解释性

近两年,随着深度学习模型的广泛应用,其存在的种种弊端也不断地暴露出来。其中最被人诟病之处,就是模型的可解释性问题,即深度学习模型的“黑盒”问题。此问题体现在两个方面,一是对观察者而言,无论深度学习的内部结构是未知还是已知的,其运行逻辑都是人类无法理解的;二是对于模型使用者而言,模型所作出的决策依赖于哪个具体的输入变量,是很难确定的。深度学习的不可解释性在管理决策领域带来的问题尤为突出。使用人类不能真正理解的模型进行管理决策不仅会产生道德上的风险^[150,151],还会在责任^[152]、安全^[153]、信任^[154]等方面带来诸多的不确定性。这种缺乏解释的特性极大地限制了深度学习在信用评价、金融决策、军事指挥等决策敏感领域上的应用。

深度学习的可解释性研究,旨在将深度神经网络的运行逻辑以人类可以理解的术语或概念进行呈现^[155]。目前,关于深度学习可解释性的研究主要包括两大类:一是针对具体实例的模型解释,也称局部解释方法,目的在于解释模型在特定输入下给出相应输出的原因。另一种则是对深度学习模型的内部逻辑和规律进行解释,也称全局解释方法,目的是解释模型做出决策所经历的整个判定过程和判定依据。相对而言,第一种获取模型局部解释的风格更具实用特性,它可以帮助在某一特定的输入环境下为某一具体决策做出解释,而第二种获取模型全局解释的风格则更具理论特性,其侧重于研究是什么样的测量导致黑盒模型能够取得远高于人类预期的预测性能,以帮助相应领域的研究者对问题的本质进行更深入的研究。

局部解释方法是对深度学习在某一个输入样本下的输出作出解释。获取局部解释的通用思路,是设计一个可解释模型在局部上代理黑盒模型,进而得到在某个特定实例下的解释。在当前研究中,线性模型、决策树模型和基于规则的模型被公认为是天然易于人类理解的可解释模型,因此局部解释方法也多使用这些模型对深度学习进行局部代理。在该领域较为经典的方法是 Ribeiro 等提出的 LIME 模型^[156]及其衍生模型^[157],它们通过对深度学习的输入进行扰动制造数据集,再使用数据集训练可解释模型进行局部建模来获取解释。

除了采用代理方法之外,另一类局部解释方法是利用深度学习模型的一些自身性质来达到解释的目的,如注意力机制(attention)^[158]、分层相关性传播技术^[159,160]、类激活映射方法和激活最大化方法^[161,162]等。上述方法的一个共同特点,是通过分析给定输入样本之后神经网络内部结构输出的变化来进行模型解释。也有一部分工作使用上述思想构建新的具有一定可解释性的网络,如 Hinton 等提出的 Capsules 网络^[163], Wang 等提出的深度 SVM 栈式网络^[164]等。

全局解释是对整个深度学习模型逻辑的解释。在该方法中,全局性的可解释代理模型依然是一种重要的解决思路。例如使用基于决策树和规则集进行模型代理^[165,166],使用树集成模型进行代理^[167,168],利用广义加性模型拟合深度学习模型^[169,170]等。另外一种思路是通过输入的重要性分析来解释模型中输出对对应输入的依赖。如使用数据随机化方法对原始数据进行随机扰动,并观察扰动前后的输出变化来探测属性和属性组对预测能力的贡献^[171];通过估计特征重要性评分的方式进行模型解释,如 POIMs 方法^[172]等;使用深度学习模型输出输入的梯度积分进行全局模型解释^[173]等。此外, Horel 将显著性检验引入到神经网络当中,构造了类似梯度平方的统计量对单层神经网络进行解释^[174],迈出了基于统计方法进行深度学习模型解释的第一步,尽管针对多层神经网络的统计检验似乎仍然遥不可及。

随着深度学习解释方法的不断深入,将可解释深度学习模型应用于管理实践也成为一种趋势。Wang 等将解释算法应用于医学领域,利用重要性分析方法对多层小波分解模型进行解释分析,在 ECG 数据预测过程中发现了 T-wave 波段对结果的显著影响^[173]; Yang 等使用注意力机制对 RNN 模型进行解释,用于分析 ICU 患者的死亡率与医疗条件之间的关系^[175];在前文提到的 AlphaStock 中,可解释深度学习方法也被应用

于量化策略构建中以识别“赢家股票”的基本特征^[109]。有趣的是,目前也有研究在尝试将可解释的因果关系与基础的机器学习模型相结合,以提高模型的预测性能。例如 Wager 和 Athey 提出了名为因果森林 (causal forest) 的方法,该法利用随机森林来估计异质处理效应^[176]。Johansson 等提出了一种具有领域适应性的表征学习方法,从而提高模型的反事实推断 (counterfactual inference) 的能力^[177]。尽管如此,文献中更多的还是将预测方法应用于因果模型的工作,应用方式一般包括外部嵌入 (引入变量)、技术增强 (引入模型)、场景应用 (引入应用) 三类,陈国青等对此有较为深入的论述^[5]。本文聚焦于预测性分析方法,不再赘述。

可解释的深度学习方法依然是一个正在发展的领域,现阶段学术界对其定义还没有达成一致,对解释需要保证的性质——如可靠性、完整性、可理解性等——和定量描述还没有统一的标准。除此之外,如何利用得到的解释完善深度学习模型、模型的可解释性是否可以有效抑制对抗样本、当存在潜在的和未观测到的特征时如何基于模型做出决策等,仍属高度开放问题,亟需进一步探索和完善。

3.3 概率隐变量模型

概率隐变量模型是一类结合了图论和概率论知识的概率建模方法,鉴于其可刻画数据潜在维度、可建模不确定性、可解释性高等特点,已成为数十年来概率建模方法中的研究热点。从应用层面而言,概率隐变量模型在文本分析、用户行为分析等领域得到了广泛的应用。从方法层面而言,随着深度学习技术的发展,传统的概率隐变量模型与新兴的深度学习技术相互融合,衍生出贝叶斯深度学习,正成为当下机器学习社区最热门的研究主题之一。

在文本分析方面,概率隐变量模型主要通过隐变量表示文本中的主题信息,进而产生众多概率主题模型,其中的代表性模型包括潜在语义分析 (latent semantic analysis, LSA)^[178]、概率潜在语义分析 (probabilistic latent semantic analysis, PLSA)^[179]、隐含狄利克雷分配 (latent dirichlet allocation, LDA)^[79] 等,其中以 LDA 最为有名。LDA 将文档中主题的混合权重看作随机参数,并为其引入先验分布,实现了高质量的文档语义分析。近年来,随着 Twitter、微博等在线社交平台的兴起,概率主题模型的前沿研究开始由通用的概率主题模型,逐步转向面向短文本的概率主题模型,重点在于解决传统的概率主题模型在对短文本进行文档生成建模时面临的文档特征稀疏的问题。其解决途径主要包含两种,其一是利用辅助信息,包括: Wikipedia、WordNet 等外部信息,主流新闻等长文信息,以及文档的用户、Hashtag、时间戳、短链接等上下文信息^[180,181] 等;其二是利用内部信息,打破短文本的文档边界,如自聚合主题模型^[182]、伪文档主题模型^[183] 等。

在用户行为分析方面,概率隐变量模型主要利用隐变量刻画用户的内容偏好、环境、社交、用户意图等多种要素。主流的基于概率隐变量模型的用户行为分析方法,对接的下游应用是推荐系统,如音乐推荐^[184]、兴趣点推荐^[185]、商品推荐^[186]、社会化推荐^[187] 等。随着物联网和移动电子商务技术的发展,移动设备中的日志信息挖掘^[188]、移动 APP 中用户行为分析^[189] 等成为用户行为分析研究的新的增长点。

近年来,传统的贝叶斯学派通过使用更灵活的函数形式 (如深度神经网络),对部分随机变量之间的映射关系进行刻画,用以弥补传统概率结构在高维复杂数据表示上的不足,同时保留了贝叶斯学习在不确定性建模和小样本学习中的优势。这也成为了概率隐变量模型相关研究的重要发展趋势之一。其中,深度生成模型是贝叶斯深度学习研究中最值得关注的部分,其代表性模型包括变分自编码 (VAE)^[190] 和生成对抗网络 (generative adversarial networks, GAN)^[191]。VAE 开启了推断网络和生成网络并存的网络结构设计,而 GAN 则引领了判别器和生成器对抗式训练的全新学习模式。而如何对模型判别能力与生成能力进行权衡^[192],以及深度生成模型与传统学习策略的结合如融合学习^[90]、迁移学习^[193] 等,则成为了前沿研究方向。此外,深度生成模型在图数据的生成式建模上也受到了极大的关注^[194]。

3.4 文本分析

在大数据时代,互联网无时无刻不在产生诸如新闻、在线评论、社交媒体帖子等不同类型的海量文本。文本中蕴含丰富的信息,可用于股票指数预测^[195,196]、工作满意度因素调查^[197]、垃圾评论识别^[198]、消费者与在线零售品牌^[199] 等研究,对管理学领域有重要价值。随着大数据和深度学习技术的发展,文本分析也得到了长足的发展。下面将从特征、模型、应用和学习方式四个方面,介绍文本分析技术的发展趋势与研究挑战。

随着深度学习进入文本分析领域,文本表示由词袋模型转向分布式表征。相应地,文本表示的特征由词或 n-gram^[200] 转向基于神经网络的表征学习。具体而言,随着 Mikolov 等人提出词向量模型^[131] 并开源

Word2Vec, 文本分析领域开始大量采用词向量取代词袋模型中的 one-hot 向量. 词向量蕴含词语的语义信息, 有助于提升各类文本分析任务的效果. 受 Word2Vec 等研究的启发, Le 等提出针对文本的表征学习模型 Doc2Vec^[133]. 此外, 不少研究工作在 Word2Vec 模型基础上, 引入神经网络实现文本语义的深度表征, 进一步提升了各类文本分析任务的效果. 例如, Severyn 和 Moschitti 利用深度卷积神经网络进行 Twitter 文本的深度语义表征, 提升了 Twitter 文本情感分类的效果^[201]; Chiu 和 Nichols 提出双向长短期记忆网络和卷积神经网络联合模型, 提升了命名实体抽取的效果^[202].

在模型方面, 深度学习模型逐步成为文本挖掘的主流方法. 早期的文本分析以浅层学习方法为主, 代表性的模型包括隐狄利克雷分配模型^[203]、矩阵分解^[204]等. 这类方法具有实现简单、可解释性强等优点, 但是依赖高质量的特征, 而高质量的特征需要业务专家或领域知识的指引. 随着深度学习的发展, 文本分析方法由浅入深, 开始转向变分自编码器^[205]、循环神经网络^[135]、卷积神经网络^[206]和完全基于注意力机制的 Transformer^[9]等深度学习方法. 这类方法能够实现特征自动学习与文本语义的深度刻画, 但其缺陷是缺少可解释性.

在应用方面, 文本生成、自动对话和自动问答等依赖文本理解的应用受到了更多的关注. 深度学习方法的发展实现了对文本语义的深度表征, 这一方面提升了经典文本分析任务的效果, 如情感分析^[207]、信息抽取^[208]和知识图谱^[209]等; 另一方面也推动了依赖文本理解的任务的发展, 使得包括文本生成^[210]、自动问答^[211]和阅读理解^[212]等应用的研究越来越受到研究者的关注.

在学习方式方面, 文本分析由学习特定任务转向不面向特定任务的预训练. 最早使用预训练思想的是 ELMO 模型^[213], 其训练过程包括两个阶段, 第一阶段利用双向长短期记忆网络进行预训练, 第二阶段针对不同的下游任务, 从预训练的网络中提取对应单词的语义向量作为新特征补充到下游任务中. 类似 ELMO 模型, Radford 等提出了 GPT 模型^[214], 该模型将 ELMO 模型预训练阶段的双向长短期记忆网络改成了 Transformer^[9], 并利用 fine-tuning 的方式进行下游任务的学习, 获得了比 ELMO 模型更好的效果. 考虑到 GPT 模型是单向语言模型, Devlin 等提出预训练阶段是双向语言模型的 BERT 模型, 并在 11 类文本分析任务中得到了当时最好的效果^[215]. 后来, Yang 等^[216]和 Lan 等^[217]分别提出了 BERT 模型的改进版: XLNET 模型和 ALBERT 模型. XLNET 模型采用自回归语言模型替代 BERT 的自编码语言模型, 解决了预训练过程和微调过程的不一致性问题. ALBERT 模型提出了分解的嵌入表示和跨层的参数共享两种减少内存的方法, 提升了训练速度. 因此, 预训练模型的快速发展为各类文本分析任务搭建了一个通用基础平台, 各任务可基于预训练模型进行参数微调, 无需从头开始构建模型, 从而可以提升任务效果并节省大量的时间和计算资源. 从某种意义上说, 预训练模型为实现人工智能最初的理想——知识描述、推理和应用——提供了一个异于“笨重”的知识图谱的极佳选择 (其实如前文所述, 知识图谱也已成为表征学习的建模对象), 也可能将为自动驾驶等复杂场景急需的直觉 AI 技术研究提供有益思路.

4 场景维度

数据智能的最终目标是将大数据和智能算法应用于各类实际场景并创造价值. 其推动了计算机科学与管理学、社会学、经济学等不同学科的有机融合, 拓展了原有学科领域的边界, 在学术界和工业界都产生了诸多创新的应用成果, 也带来了许多研究机会与挑战. 本节以商务、金融、社会、移动物联网等四个领域为例, 对数据智能在相关领域应用的现状、趋势与挑战进行分析和讨论.

4.1 数据智能与商务场景

如引言所述, 数据智能可以被认为是商务智能的深化与扩展, 是对企业全量全程商务数据的重新审视, 以及对企业内外部大数据的有效关联和融合, 目的是使企业的商务决策活动更加智能化. 其中的一个经典应用, 就是实现面向用户偏好的个性化推荐. 经过二十多年的发展, 个性化推荐逐渐成为了数字化商务平台的一种必备工具, 为消费者提供了极大的便利, 也发展出了类似字节跳动 (ByteDance) 这样的新型信息消费服务类企业, 呈现出对传统搜索框的替代价值. 在大数据和智能算法推动下, 个性化推荐方法也在不断发展, 并被应用于各类丰富的商务场景中, 以及整个商务链条的不同环节. 本节将主要围绕数据智能的个性化推荐应用展开分析和讨论.

一般而言, 个性化推荐方法分为基于内容的推荐算法和基于协同过滤的推荐算法. 其中基于内容的推荐算法^[218] 核心在于利用特征信息来构建用户和物品的画像, 再通过画像的匹配算法来实现个性化推荐; 而协同过滤算法^[219] 的核心假设是根据相关群体的喜好来推测个人用户的喜好. 为解决用户购买或评价产品所面临的维度灾难及稀疏性问题, 面向矩阵分解的各类方法被广泛应用到推荐系统中. 其中隐因子分解模型 (latent factor model)^[220] 及概率矩阵分解模型 (probabilistic matrix factorization)^[221] 等极大地提高了推荐的精度, 也影响了后续相关算法的发展. 此外, 推荐系统的多角度评价^[222]、长尾效应^[223] 等长期问题, 也得到了大量关注. 随着深度学习的发展, 研究者利用深度神经网络结构来表征用户和产品相关的异质数据, 以提升推荐效果. 例如, NeuMF^[224] 通过联合学习矩阵分解和多层前馈神经网络来解决隐式反馈推荐问题, 而 Wide&Deep^[225]、Deep&Cross^[226] 和 DIN^[227] 等模型则直接利用深度神经网络结构来学习产品特征的交互.

近年来, 如何更加精细化的刻画用户的行为与偏好, 从而提高推荐系统的效果, 甚至从产品链的顶端修正产品设计, 也成为了管理学领域的前沿问题. 一种对用户行为进行精细化刻画的方式即为融合多源异构的行为数据. 例如, Lu 等研究了基于消费者视频的衣物推荐系统^[228], 该系统根据消费者在穿衣镜前的表情和对衣物的关注点等行为特征来推测其喜好, 并根据历史上的相似客户为其进行衣物推荐. Ansari 等基于用户发布的评论文本数据, 提出了概率图模型来洞察用户偏好^[229]. Zhang 等融合了图片、文本、网络结构等多类型数据进行推荐^[230]. Liu 等通过构建产品网络, 将单一产品的推荐扩展到对于一个产品捆绑的推荐与设计^[231]. Zhang 等则进一步提出了对产品关系进行个性化表征学习的方法, 将用户的偏好视为对应产品关系的汇聚^[232]. 此外, 随着移动商务这种新兴业态的出现, 用户行为也随之发生变化. He 等基于涉入理论, 提出了融合用户移动 APP 的下载和浏览行为的推荐方法^[233]. 同时, 由于用户可以通过移动互联网产生位置信息, 因而也有大量研究探讨基于地理位置的推荐问题^[234,235]. 由此可见, 面向不同的商务应用领域, 推荐系统正在朝着用户偏好精细化挖掘的方向发展. 特别是随着新零售的兴起, 用户在各类场景中的行为数据都能被记录下来, 为零售过程中各个环节都提供了深入洞察用户偏好的机会.

此外, 随着个性化推荐技术的不断成熟, 相关智能方法已经逐渐成为了新型商务活动中不可替代的工具, 从而影响和改变了用户的活动与行为特征. 有鉴于此, 推荐系统对用户行为的影响也引起了管理学、经济学等领域广泛的讨论, 而相关的发现也有助于推荐系统的设计和效果的提升. 有研究引入了更加丰富的情境因素与产品要素来洞察用户与推荐系统的交互行为. 例如, Kawaguchi 等通过在零售点安装视频设备来测量时间、周边大众等情境因素, 从而发现了时间的压力降低了推荐的效果, 但周边大众的存在却能提高推荐效果^[236]. 还有研究发现用户的偏好可以随时间演进, 并能逐渐学到对产品不同特征的偏好权重的变化^[237]. 此外, Oestreicher-Singer 等分析了电子商务网站中由推荐链接形成的产品网络, 发现了产品销量的溢出效应, 并提出了衡量产品网络经济价值的方法^[223,238]. Lin 等在此基础上进一步地考虑了网络中边的方向性对产品需求量的影响^[239]. Liang 等发现移动 APP 市场中的推荐产品对于其他产品有显著的溢出效应, 而这种溢出效应取决于产品之间不同的关系类型^[240].

然而, 推荐系统在影响用户行为的同时, 也可能产生一些负面的影响. 一种担心认为推荐系统倾向于推荐较为热门或用户感兴趣的部分内容和产品, 使得用户无法突破其固有认知. 因此, 有研究针对推荐系统带来的用户认知偏差进行了讨论. 其中, Adomavicius 等发现用户对于被推荐的产品支付意愿更强, 即便是针对随机产生的推荐产品^[241]. 也有研究者探究了推荐系统的“操纵”效应, 他们发现传统的 Top-N 的推荐机制放大了不同产品、新闻受欢迎程度, 而引入概率机制的推荐系统具有较强的鲁棒性^[242]. Song 等提出了一种多类型的效用模型, 来建模用户对内容的偏好和希望获取新内容的时间^[243]. 有研究进一步考虑了推荐产品的多样性^[244,245]、公平性^[246] 等问题, 并结合用户的信任机制^[247] 以降低推荐系统带来的偏差, 而这也势必成为推荐系统的前沿研究方向.

总体而言, 数据智能对商务领域最显著的意义就在于利用用户数据实现对用户偏好与需求的精准刻画, 进而帮助企业实现精准营销等目的. 而作为数据智能在商务领域最重要的应用之一, 个性化推荐方法近年来取得了长足的进步. 随着物联网、移动支付、图像识别、虚拟现实等技术的发展, 用户数据不断积累, 诸多企业已经具备了更加精细化洞察用户需求、偏好的基础条件. 因此, 着眼于个性化推荐的发展趋势, 学术界和

工业界势必将深入整合多源、异构、跨场景的数据,从而为用户提供更加精准且具有一定解释能力的推荐结果——这对于基于深度学习的个性化推荐系统而言是一个不小的挑战。与此同时,探究相关数据智能技术对用户行为、商务活动的影响,特别是带来的安全性、公平性等问题,也将成为数据智能在商务领域的研究热点。

4.2 数据智能与金融场景

数据智能方法在金融领域最直接的应用,就是对金融时间序列的预测,包括股票价格、期货价格、汇率等,这也是学术文献较为密集的方向^[248]。从 20 世纪 90 年代至 21 世纪初,金融领域应用最多的预测方法主要以人工神经网络^[249]和支持向量机为代表^[250,251];此后各类神经网络模型如卷积神经网络、循环神经网络、多种 RNN 的变体模型以及注意力机制等均被用于金融时间序列预测。近年来,以长短期记忆模型(LSTM)为代表的 RNN 类模型在金融时间序列预测方面最为常用^[252]。LSTM 模型可刻画数据长期相依关系,从而有效捕捉具有高波动、高噪音属性的金融时间序列中重要的非线性和时间相依特征。除了模型的演进外,预测方法也逐渐从单一模型发展到混合模型或组合预测以提升预测效果,如将数据挖掘和人工智能方法混合^[253],将人工智能方法与时间序列模型如 ARIMA 混合^[254],将数据分解方法如小波分析、谱分析、经验模态分解(empirical mode decomposition, EMD)与人工智能方法或时间序列模型等进行混合^[255,256]。此外,专家预测也被进一步引入混合模型中,从而发展出各类综合集成预测方法以及集成学习方法^[257-259]等。

数据智能方法在解决金融问题时如何与金融领域知识融合,可以从预测模型输入变量这个侧面有所反映。早期预测模型的输入变量都是最简单的股票、股指、期货、汇率等的价格、交易量等数据,以及基于这些数据计算的统计性指标如波动性和金融市场的各种技术分析指标^[249-251]。此后的研究逐渐将与金融领域知识相关的基本面因素^[260-262],以及对资产价格产生影响的其他因素如市场预期、投资者情绪、媒体情绪、交易行为等^[196,252],纳入到预测模型的输入变量中。自然语言处理和文本挖掘方法的发展则使研究人员能够从新闻^[261,262]、论坛^[196]、社交媒体等各类文本信息中挖掘有效信息,从而通过改善模型输入的信息量来提升预测效力。但同时应注意,一些研究工作在挖掘文本信息时仍然只关注了一般意义上的知识,而不是金融领域的特定知识或有金融理论支撑的关键信息,这使得构建的模型通常不具备鲁棒性——如只适应于某种特定市场状况——和可解释性,从而很难被金融机构所采纳。因此,如何将数据智能方法与金融理论和金融领域知识进行深度融合,仍然是需要重点关注的问题。

尽管上述研究以外部嵌入或技术增强等方式引入了数据智能方法,但大部分研究对预测的评价仍然集中于对方法的评价,即利用基于预测误差的各种统计指标来评价预测方法,而很少采用面向预测模型产出结果的评价。正如 Armstrong 在其研究中提及,理想的情况是将预测方法放在其使用的环境里去评价^[263]。金融领域更关心在不同场景下预测模型是否发挥作用,以及预测结果如何使用等问题。评价预测模型是否发挥作用通常有两方面,其一是预测方法评估,其二是根据现实情况进行决策是否采纳预测分析结果。因此,将预测模型应用到金融场景中解决问题时,需要关注对预测结果的评价,以及如何将预测和决策问题相联系从而改善预测方法。如 Chen 等将预测和投资交易问题相结合,提出了基于强化学习的预测模型,该模型融合了具有向前展望能力的结果评价,以期通过改善预测结果的应用效果来改善预测模型^[264]。Wang 等也将强化学习引入到股市预测中,以多期投资的夏普比率作为报酬,构建了买赢卖输的投资组合并跑赢了众多优秀的策略^[109]。这些研究充分说明,以金融决策的收益为导向,也许可以构建更为有效的金融预测模型。

数据智能方法应用的另一个优势领域是信用评价和信用风险管理。其涉及诸如公司信用风险评估、信用卡欺诈检测、个人信贷风险评估以及国家主权信用评级等核心问题,金融业对此有迫切的需求,从而令其成为方法和技术落地的最重要的场景之一。与金融预测研究的发展类似,信用评价研究的发展一方面依靠数据智能方法的演进以不断提升信用评价和信用风险评估的精度,另一方面模型的输入也逐渐从财务报表信息扩展到更丰富的多元异构数据。例如,除了可用于刻画宏观环境、公司经营、个人信用等信息的各种变量之外,新闻、各类披露的报告等文本信息^[265,266],被评价目标与其他主体的关联关系——如社交媒体上的关注或互动等形成的关联关系^[267,268]、通话关系网络^[269],以及电商平台上累积下来的消费和支付的行为数据,通过智能终端应用获取的电子足迹数据^[270]等,均被纳入评估中并被深入挖掘价值。这其中的重要问题仍然是如何结合信用风险管理中的具体场景和问题,更好地令智能方法与金融理论和领域特定知识相融合,从而

更好地挖掘各类数据中的有效信息。不仅如此, 面对有明确业务场景的问题, 金融实践中更重视模型的可解释性以获得对模型的控制, 因此如何更好地利用机器学习模型提升精度的同时构建可解释的信用风险评估模型, 也是长期难点问题。目前已有一些探索性工作, 但主要以规则形式固化知识^[271-273], 这距离可解释预测模型仍有较大差距, 值得深入研究。

从业界视角看, 在大数据、人工智能、云计算、区块链等新兴前沿技术的带动下, 金融科技发展迅猛并受到各方重视。金融业的模式创新、业务创新、产品创新不断涌现, 从依托互联网和支付技术的货币基金销售模式的变化, 到 P2P 的业务创新和发展, 到智能客服、无人银行的出现, 以及智能投顾、智能投研、财务机器人等的出现, 不一而足。中国人民银行在《金融科技 (FinTech) 发展规划 (2019-2021 年)》中指出, 金融科技是技术驱动的金融创新; 该规划提出到 2021 年我国要进一步增强金融业科技应用能力, 要实现金融与科技深度融合、协调发展等一系列目标。技术与金融业务相融合可以快速推动金融创新, 助力业务流程优化、提升效率、节省成本; 同时以技术赋能金融机构, 全面满足客户对金融服务流程的移动化、便捷化和个性化需求, 并增强智慧风控的能力。我国的互联网巨头也早已悄然布局受监管最严的银行等传统金融业务, 凭借其累积的数据和技术资源快速触达客户, 并对客户进行信用评价从而掌握良好的风控能力。从这个角度看, 金融科技在金融业的发展快于研究推进的速度, 也对科研提出了更高的要求。

目前, 面向市场风险、信用风险、操作风险以及系统风险 (systemic risk) 的智慧风控, 仍是亟待利用金融科技提升的重要问题。如何通过融合多源异构数据, 利用数据智能方法改善对各类风险的评价, 以打造智慧风控体系, 从而支撑信贷、信用卡、资产管理、投资理财等产品和服务, 并助力智能决策, 也是重要的研究问题。此外, 金融创新和金融监管总是相互促进、相辅相成的, 如何全面推进监管科技 (RegTech) 从而实现金融创新的有效监管, 亦是当前亟待解决的国家重大需求问题。

4.3 数据智能与社会计算场景

社会计算是大数据时代的标志性领域之一。其收集具有前所未有的广度、深度和规模的网络空间大数据, 通过数据智能方法建模, 以揭示个体或群体的行为模式^[274], 进而分析复杂且相互作用的社会经济系统的发展规律^[275]。社会计算被认为是社会科学和计算科学的交叉点, 它将社会现象视为嵌入复杂社会系统中的个体之间的社会交互作用的结果^[276]。该领域的发展为社会科学的诸多问题研究提供了新的机会, 其中社交媒体分析和社会经济系统分析等成为了研究的热点。

在社交媒体分析方面, Web 2.0 技术催生了各类社交媒体平台如 Facebook、Twitter、新浪微博等。随着社交媒体用户规模的日益增大, 其即时性、用户高交互性使得网络舆情的扩散极具影响力、复杂性和突发性。因此大量的社会计算研究聚焦于社交媒体分析, 并在社交结构特性和传播演化机制^[277-279]、在线情感表达^[280-282]、人格特质分析^[283]等方面取得了诸多成果。近年来, 由于在线社交网络在社会信息传播中的作用越来越大, 社会计算成为维护社会舆论健康的重要手段。因此, 面向网络舆情的社会感知^[26]、话题突发检测^[284,285]、谣言判别^[286-288]等成为了学术界和政企界关注的热点问题; 而深度学习及表征学习的兴起, 则为进一步提升网络舆情分析的准确性提供了支持^[289]。此外, 社会计算也被证明在地震预警^[290]、流行病预测^[291-293]等公共管理问题上卓有成效。

与此同时, 社会计算也为传统社会经济问题的研究带来新的机会。首先, 海量用户行为数据被记录以为研究更大规模的社会经济问题提供了数据基础。因此大量研究聚焦于探讨宏观社会经济结构, 如基于用户的通话记录数据讨论社交关系多样性与地区经济发展间的关系^[294], 采用卷积神经网络从卫星图像中预测非洲国家区域经济发展水平^[295]等。除了宏观经济水平的预测, 区域性的失业率预测^[296]和区域发展模式^[297]等社会层面问题也开始基于大数据和社会计算方法得到讨论。其次, 用户多源异质信息得到了更加充分的记录, 因此相关研究也呈现出细粒度分析和预测个体/社会的经济水平的趋势, 如 Blumenstock 等利用通话记录中个体详尽的行为特征记录来预测卢旺达国民的贫富水平^[298], Luo 等则基于社交网络中用户的社交结构信息预测个人财富水平^[299]。此外, 相关研究也呈现出领域知识融入的趋势, 从而对传统理论进行大规模数据验证和探索性创新研究。如 Berkman 等于 1979 年发现了社会关系更强的人寿命也更长的理论^[300], 但 Hobbs 等基于 1200 万的 Facebook 用户数据开展研究发现, 该理论适用于接受友谊的社会关系, 而不是主动建立友谊的社会关系^[301]。尽管社会计算已为现代社会经济问题研究开辟了蹊径, 但在如何利用相关发现来

指导社会 and 经济发展方面, 仍然有很大的研究空间, 也将会成为持续的研究热点。

近年来, 社会计算也开始应用于政治活动预测。如奥巴马团队在 2012 年总统竞选中利用了一系列大数据分析方法来寻求政治支持^[302], 这也掀起了大数据在政治活动中的作用和收益的热烈讨论^[303]。此后预测技术也结合不同地区的政治背景信息, 在中国台湾省^[304]、英国^[305]等的政治选举预测中取得了突出结果。Cambridge Analytica 公司则被认为通过分析在线言论获得了选民的人格特质, 并据此策划了政治营销策略, 最终成功影响了美国大选和英国脱欧等政治投票事件^[306]。除了选情预测, 大数据提供的丰富个体数据, 为一些过去难以研究的问题提供了数据基础。如 Leetaru 等研究了长达三十年的世界新闻数据, 采用情感、文本挖掘技术预测分析了埃及、突尼斯和利比亚等阿拉伯国家的内战冲突^[307]。Chadefaux 等则分析了 166 个国家的报纸文章, 验证了同时间段内超过 200 次的战争, 从而构建了模型预测下一年是否会发生战争^[308]。目前社会计算在政治领域的应用仍处于蓬勃发展阶段, 如何更好地融合政治活动相关多源异质大数据、如何结合政治学相关理论提升预测精度, 仍然是需求迫切的开放问题。

总体而言, 历经多年的研究和发 展, 基于数据智能的社会计算取得了长足的进步, 其积极作用已经在政企学各界达成了共识。但如何更有效地从网络空间大数据中精准感知数据、如何提升复杂业务场景下的决策水平、如何实现人机合一的混合智能, 仍然是普遍关注的核心问题。可以预期, 社会计算将在政治安全、社会安全、城市安全等领域得到爆发式的应用, 网络空间的内容对抗将和网络底层的流量对抗一起成为常态话题。而文本语义理解技术的进步和超大规模异质网络分析技术的进步, 将对提升社会计算水平起到决定性的作用。

4.4 数据智能与移动物联网应用场景

在移动通信和智能终端应用的共同推动下, 移动物联网 (Mobile Internet of Things, MIoT) 成为计算与网络通信技术发展的新浪潮。在这一浪潮的驱动下, 数据智能技术在交通、城市、工业、医疗等领域大放异彩。移动物联网与数据智能的深度融合是一个相得益彰的过程。一方面, 移动物联网无处不在的数据获取感知能力, 为数据智能技术提供了丰富的应用场景和海量的数据分析原料; 另一方面, 借助移动网络和智能终端的连接能力, 数据智能技术强大的分析、预测、优化能力被有效地投送到目标应用场景, 充分体现了数据智能赋能多样化的管理应用场景之能力。本节将以智能交通、城市计算、众包服务等三类典型场景为例, 介绍数据智能在移动物联网领域的应用, 并对未来具有潜力的场景进行展望。

4.4.1 智能交通

智能交通是数据智能和物联网融合应用最为广泛、成果最为丰富的一个领域。其原因一方面在于 GPS、北斗等卫星定位技术的普及, 使得全量交通数据的收集成为可能; 另一方面也在于交通作为一个管理体系十分完善的社会系统, 其场景、任务、服务对象都十分明确, 因此可以称之为数据智能向物联网领域“低垂的果实”。针对智能交通场景, 数据智能的主要研究工作集中在交通预测领域, 包括个体出行预测、路段交通预测以及区域需求预测等三个方面。每个方面都经历了由经典模型到深度学习模型的发展历程。

个体出行预测研究包括出行路径推荐和出行时间预测。在出行路径推荐方面, 传统工作通常使用 A*、Dijkstra 等启发式搜索算法, 进行最短或最速路径搜索与推荐。在近期研究中, Wu 等将 A* 算法同深度学习相结合, 提出了神经网络 A* 搜索算法, 实现了个性化的出行路径推荐^[309,310]。在出行时间预测方面, 经典算法采用马尔可夫链对交通出行时间进行建模预测。基于深度方法的出行预测可分为分段预测和端到端预测两类。其中分段预测更加关注对于出行链影响要素的建模^[311], 具有较好的可解释性, 但是对未纳入模型的因素的影响较为敏感; 端到端预测则着眼于构建从能够建模整个出行行为的一体化网络模型^[312], 其模型的预测准确性较高但缺乏解释性。

路段交通预测聚焦于流量和速度的预测^[313,314]。经典的交通流量和速度预测采用 ARIMA 等时间序列预测模型, 难以对交通状况的时空上下文信息进行建模。近两年, 深度学习方法逐步被应用于路段交通预测, 例如使用 LSTM 等递归神经网络模型的混合模型对交通状况的时间特性进行建模^[315], 使用 CNN 对道路交通的时空局部性进行建模^[316], 使用图卷积网络对道路的网络结构信息进行建模^[317]等。

区域交通需求预测的主要目的是预测网约车、出租车、共享单车在某一个区域的用户需求。经典方法使用时间序列预测的方法, 例如 ARIMA 模型及其变体被广泛应用于交通需求量预测。在基于深度学习的研究

方面, 更多的外部因素被考虑了进来, 如使用 CNN、LSTM 等复杂的神经网络结构对时序因素、空间距离因素以及天气等外部因素进行建模^[318,319], 使用多模态网络将出租车、自行车等多种交通需求因素进行融合预测^[320]等。然而, 无论是经典方法还是深度学习方法, 现有工作大多是将区域交通需求当作简单的时间序列进行预测, 对于交通需求背后的供需机理分析不足, 因此 Guo 等的一系列工作从供需分析的角度研究了出租车和网约车的需求与价格关系^[321,322]。

除交通预测之外, 救护车、物流车辆、网约车的智能交通调度也是近两年智能交通关注的领域, 主要的研究趋势是使用真实的数据代替仿真的应用场景^[323], 以及使用深度学习和强化学习融合的方法进行调度问题优化^[113,324]。相比起传统的运筹规划方法, 基于数据智能的方法在仿真环境的真实性和场景复杂性建模方面均有一定的优势。

4.4.2 城市计算

城市计算是数据智能技术与城市科学的交叉融合^[325]。由于城市场景天然的时空动态特性, 卫星定位、手机信令、IC 卡、社交媒体信息等移动物联网数据成为了城市计算研究的主要数据来源。由于城市场景内涵丰富、应用多样, 这就导致城市计算的相关研究很少聚焦于单一主题。从横向看, 城市计算的研究内容学科跨度非常大, 覆盖了环境、交通、人口、能源、安全等多个领域, 是多学科交叉的汇聚点; 从纵向看, 城市计算的研究工作可以从多源数据感知、城市动态特征建模、数据驱动的应用领域三个方面进行梳理。

首先关注多源数据的城市感知研究。传统研究多使用地图、卫星等遥感测绘数据作为城市场景的感知手段, 同时配合问卷调查、人口普查等数据对城市中人的活动进行收集。近两年, 人们提出了“社会感知”的概念, 即使用装备了移动物联网终端的人、车、物作为传感器, 对城市中的人类社会活动进行感知, 对传统“目中无人”的卫星遥感数据进行补充。常用的物联网数据包括手机信令数据^[326]、GPS 轨迹数据^[327]、公交车 IC 卡数据等^[328]、室内 Wifi 定位数据^[329,330]等。与智能交通领域对这些数据的使用不同, 城市计算更关注物联网数据中包含的人类社会活动信息及其对城市各个层面的影响。随着智能手机的普及, 移动社交媒体的用户生成内容也开始被应用于城市的动态感知, 例如位置微博和签到数据等^[331]。与 GPS 等数据相比, 移动社交媒体的数据包含语义更加丰富, 但是数据在地理分布上相对稀疏。可以说, 数据来源是主导城市计算研究内容的基础环节, 而城市计算的研究也往往是使用物联网数据类型最多的一类研究。

城市动态特征建模研究主要关注城市环境中人类活动的内在模式, 如热点、社区、异常、潮汐规律等, 其目的是理解市民活动行为, 从而进行合理的管理和服务决策。在该领域, 现有研究工作可以分为空间模式建模、时间模式建模和时空模式建模三个方面。在空间模式建模方面, 相关研究聚焦于城市空间中的热点模式挖掘^[332]、城市区域的功能区划分等特点模式^[333]。在时间模式建模方面, 相关研究工作关注城市中的人流、交通流的潮汐模式和异常模式^[334]。在时空模式建模方面, 研究人员提出了城市光谱 (CitySpectrum)^[335] 的概念, 从时间模式、空间模式、时空关联模式等多极视角对城市中的人类活动模式进行挖掘展示^[336]。这类研究着眼于揭示城市中人类时空行为的基础规律, 因此能够为各类上层应用提供基础支撑。

在数据驱动的城市应用领域, 城市计算更多关注面向公共福祉的城市社会管理难题, 如城市资源管理^[89,326]、城市规划建设^[87]、城市经济发展^[297,337]、城市环境保护^[338]、城市公共安全^[293,339]等问题。将数据智能技术与更多的社会和公共管理问题相结合, 是城市计算在该领域的重要发展方向。此外, 一些在城市场景中的个体服务应用, 如个性化位置推荐^[340]、市民出行轨迹推荐^[309,310]、用户兴趣点访问预测^[341]等问题, 也经常被广义地归入城市计算的范畴。

4.4.3 众包服务

近两年, 随着 O2O 服务应用的不断普及, 移动群智计算成为数据智能与移动物联网技术相结合的新兴领域。基于移动群智计算的时空众包服务, 如网约车、外卖、快递、跑腿等业务, 成为在新兴服务业的典型应用。在这些应用中, 移动物联网终端通过无处不在的触达能力, 将大规模人群的服务供给和服务需求进行连接, 而数据智能技术则将服务供给和服务需求进行匹配, 实现对大规模人群的供需关系进行组织。基于移动群智计算的众包服务模式, 正在成为未来服务业的新型范式。以 2020 年初的新冠肺炎疫情为例, 正是在移动群智计算的支撑下, 快递、外卖、网约车、新零售等新兴服务方式保证了人们在居家隔离防疫期间的正常生

活, 减少了不必要外出的风险, 同时大大减小了疫情对第三产业的冲击, 为中国经济的快速恢复创造了条件。

在时空众包服务中, 大数据智能技术要解决的核心问题就是众包服务的任务分配问题。而从任务分配问题所依托的底层管理学逻辑的角度看, 时空众包任务的分配的技术路径又可以进一步划分为以平台为中心的优化算法设计和以用户为中心的激励机制设计^[342]。以平台为中心的任务分配方法从全局的视角入手, 以一种“计划经济”的方式通过全局性或者接近全局性的优化调度, 实现众包任务在供给和需求之间的匹配。以平台为中心的优化算法设计可分为常规优化算法和智能优化算法, 前者主要针对一些简单的问题进行计算, 后者则是针对复杂感知系统。常规算法包括贪心算法、动态规划算法等, 这类算法对于早期的小规模众包问题有较好的性能^[343,344]。随着时空众包平台规模的不断扩大, 任务分配问题逐步变为 NP-hard 问题, 使用常规的算法或者无法在多项式时间内求解, 或者难以获得令人满意的性能。为此, 研究人员将遗传算法、模拟退火、蜂群算法等智能算法引入其中, 从而求得 NP-hard 问题的次优解^[345,346]。此外, 在时空众包问题中, 众包任务必须在用户非常严格的等待容忍和给定的地理空间中完成, 因此, 时间等待和地理空间约束也被引入到了众包任务的分配当中。近两年, 随着深度学习和强化学习的不断兴起, 研究人员也在尝试使用深度强化学习算法实现更加高效和精准的众包任务分配^[347]。

以用户为中心的激励机制设计更倾向于使用“市场经济”的方式, 其核心是通过给予服务提供方合理的报酬、奖励等激励机制来促使用户自身发挥主观能动性, 实现任务的优化分配^[348]。传统的众包服务方法, 会通过报酬支付、娱乐游戏、社交关系、虚拟积分等方式实现用户服务的激励。激励机制的实现方法主要包括基于博弈论的拍卖方法^[349]以及面向质量控制的优化方法^[350]。该类研究面临的一个核心挑战, 是如何对用户行为进行合理建模。拍卖方法更多利用理性人假设, 假定用户会理性地做出最大化自身收益的决策, 而质量控制方法则会通过用户的参与水平、任务完成质量、能耗效率等特征对用户进行描述建模, 通过合理激励选择出最有效的用户完成任务。在时空众包中, 由于服务场景较为复杂, 更多采用面向质量控制的优化方法, 用户的服务完成时间和地理空间位置也会被引入到用户建模描述当中^[351]。面向质量控制的优化方法的一个好处, 是将众包服务的质量控制和任务分配相结合, 能够在最优化平台收益函数的同时, 激励和改善服务提供者的服务水平。而在实际生产环境的任务分配中, 往往将以平台为中心的优化算法设计和以用户为中心的激励机制设计相结合。

4.4.4 具有潜力的新兴领域

在数据智能技术的物联网应用中, 可穿戴计算、智能制造是两个非常具有潜力的新兴领域。随着智能手机、智能手环等设备的兴起, 可穿戴移动设备逐渐成为了人们生活中经常使用的装备。可穿戴移动设备可以监测到人们的生理状况, 并通过基于云计算的服务和智能手机 APP 汇集数据, 实现对个体和群体生理行为的长期跟踪。使用可穿戴设备收集的用户生理数据进行健康管理和生活管理, 是大数据智能技术的重要应用场景。与交通、城市计算等场景不同, 可穿戴设备所收集的数据更具隐私性, 因此保障隐私安全的数据分析技术将是该领域研究需要关注的重点问题^[352]。智能制造是指将制造过程中的各个环节与新一代信息技术, 如物联网、大数据、云计算、人工智能等技术进行融合, 通过人与智能机器的合作共事, 扩大、延伸和部分取代人类劳动的新兴应用场景。新一代的智能制造业相比传统的工业制造业, 在智能感知控制、深度数据应用、开放式服务等方面有着巨大的革新, 这些革新都给大数据智能技术的应用创造了机会。不过, 当前智能制造的研究重点还更多的关注智能控制、安全协议、智能传感器等底层软硬件技术与标准的研发, 对大数据的分析和数据智能的应用还很初步, 因此也具有较大的发展空间。

综上所述, 随着 5G 技术的逐步成熟和落地, 新一代的物联网技术必定会更加全面而深入地融入社会生产生活的方方面面。数据智能技术和物联网场景先天的契合性, 也决定了其在物联网应用中必定会迎来一个新的发展浪潮。在拥抱这一浪潮的同时, 我们也要清楚地看到, 当前的数据智能技术在隐私保护、可解释性、不确定性、公平性等方面还面临诸多挑战, 通过引入管理学思维和交叉学科的技术, 研制下一代可信赖、伦理友好的智能技术, 将是物联网场景引领下数据智能的一个重要发展机遇。

5 结论

数据智能是在大数据驱动与应用场景牵引下, 融合来自多学科的数据获取、处理、分析和可视化技术, 为

现实世界的复杂管理决策实践提供可行动准则 (actionable rules) 的一个新兴学科领域。尽管有来自统计学、机器学习、数据库、数据挖掘等领域的许多专家学者, 一直致力于沿着各自领域的问题设定发展更为有效的预测性数据分析技术和方法, 但面对众多应用领域涌现的大数据分析需求, 大家始终缺乏通用且有效的解决方案, 这导致产生了大量有意义但碎片化的应用工作。这一现象直到深度学习的出现才逐渐得到了改变——深度学习及其衍生的表征学习等方法, 为连接大数据“原料”与智能应用“产出”提供了至关重要的一块拼图。自此数据、算法、场景三个维度得以真正打通, 形成有效的生长循环, 驱动商务智能进化为数据智能, 并且在三维螺旋中不断迭代成长。从这个意义上说, 数据智能虽然是方法和技术的集合, 但其应用导向特征才是其区别于数据科学、人工智能等热点概念的关键特征。它必将推动众多领域从不同层次的数字化逐渐走向智能化——这一过程从目前来看仍将持续很长的时间——并在智能化进程中发展出新的技术与方法。这意味着数据智能绝不会昙花一现。

深度学习毫无疑问将在很长一段时间内占据数据分析技术的主导地位。无论是端到端学习还是表征学习, 都非常容易适应多变的实践应用并且能够提供很好的预测力。如何进一步设计更适于某些数据特点的深层网络结构以提升其预测精度, 将会是一个重要的研究方向, 但目前来看大型或超大型 IT 公司凭借丰富的训练资源在该工作中占据了优势地位。一个在某种程度上有些“相反”的研究方向也将吸引不少学者的注意力, 即追求深层网络的压缩效率, 从而将高质量的深度学习模型更多地应用到嵌入式设备中——如手机和传感器, 这对于进一步推动物联网智能应用尤为重要。此外, 近年来的研究表明, 科学家们还将在高精度预测的基础上, 追求深层网络的知识化、鲁棒性和可解释性。“知识化”源于对近年来出现的图像、文本等的预训练模型的认识, 深层网络作为工具把某个领域的海量原始数据转化为了该领域的共性知识以供使用, 这或许将为人工智能诞生之初的知识推理理想提供落地思路, 尽管知识化的效率和公信力仍然有较大提升空间。“鲁棒性”担忧源于近年来发现的深度学习在对付对抗样本攻击时经常显示出的脆弱性, 尽管有不少防御策略和算法被提出以对抗各类已知攻击, 但脆弱性的根源是否在于深度学习的梯度计算, 仍未有确切的理论分析结论, 如何有效对抗未知类型攻击及实现对对抗攻击算法的有效保护, 仍然是值得深入研究的问题。“可解释性”是深度学习在金融、医疗、军事等敏感领域应用的共性需求, 但无论局部解释还是全局解释效果仍不尽如人意, 特别从统计推断角度来看, 构造有效的检验统计量仍然没有太大进展, 尽管梯度似乎在其中扮演了一个重要角色。最后, 将深度学习与现有浅层模型结合, 提升浅层模型在其传统应用领域的预测力, 也是一个重要的研究方向, 这里一个非常核心的问题是如何实现传统模型的“梯度化”近似求解。对上述诸多问题的解答, 将推动深度学习的研究和应用进入下一个阶段。

数据智能的核心特征是应用导向, 但采用“智能+”还是“+智能”的策略却应该因地制宜。对于互联网应用和工业工程领域, “智能+”策略往往是合理的, 因为这些领域对用户决策的依赖程度并不高, 数据智能可以成为价值创造的主要动力并推动应用创新和发展。但在一些对用户决策高度敏感的应用领域, 如金融、医疗、军事等, “+智能”是一个更为合理的策略, 有效、可信、可控的智能技术才会被引入到特定情境中, 辅助用户做出重要的决策选择, 如投资决策、诊疗决策与军事行动决策等。因此, 数据智能如何与这些领域的领域知识更好地融合, 如何提高其在这些领域应用时的鲁棒性和可解释性, 是否应该从业务绩效而非预测精度的角度来评价应用效果, 能否更好地借助业务人员经验以实现人机混合智能等, 都是值得深入研究的重要问题。从业界实践来看, 金融、医疗场景已经成为数据智能应用的热点领域, 金融科技 (FinTech) 概念和精准医疗 (precision medicine) 概念深入人心并开始产生市场价值。但从文献情况来看, 数据智能在这两个领域的应用研究仍远未成熟, 大部分成果发表来自技术领域, 这意味着对两个领域的核心问题——如金融的系统性风险识别和传染问题——涉入不深, 也为未来研究提供了充足的空间。此外, 一些令人兴奋的趋势也值得高度关注。例如, 5G 的兴起和移动终端的广泛渗透, 将为数据智能带来最为广阔的移动物联网应用场景; 这一场景将催生新的应用, 也能与传统的金融、医疗、商务等场景相互融合, 进而带动数据智能实现技术创新。又如, 在经历了图像、语音领域的突破性发展之后, 能与人类能力相媲美的自然语言理解与生成技术, 将在各类应用场景的共同需求的牵引下, 成为数据智能亟待突破的关键核心挑战。

作为数据智能的关键“燃料”, 大数据的价值正不断得到认识和发掘。这一部分得益于数据处理技术的

不断成熟, 数据处理平台“云”化和“中台”(middle-end)化趋势明显, 事实上的数据标准与可视化标准在越来越多的领域开始出现并得到认可. 在巨大价值的驱动下, 数据资产化步伐正在加快, 企业开始围绕其拥有的大数据来打造新的业务——这类业务甚至可以简单成仅提供数据查询却依然能赚得盆满钵满, 而共享数据的理想仅能靠政府来艰难推进. 然而企业私有化的、有价值的数据, 往往是所谓的“用户生成数据”(user generated content), 企业除了需要拿到用户的使用授权(也许在不完全知情状态下), 如何保障数据安全也是其面临的巨大挑战. 近年来, 许多研究人员对数据安全问题展开了深入研究. 这些研究通常将数据安全置于某类数据应用情境之下来考虑——如数据建模、数据发布、数据共享等, 联邦学习、区块链等前沿技术的潜力也得到了不少的讨论, 但距离真正解决问题还有相当的距离. 数据安全正日益成为阻碍大数据价值发现的关键因素, 值得研究者深入研究和探讨.

数据智能方兴未艾, 也给管理领域带来了丰富的研究机会. 从技术层面来看, 信息系统方向学者在应用场景理解上与纯技术方向学者相比应具有优势, 但应注意除了“用”技术也要“造”技术, 才能真正站在数据智能的风口浪尖上. 事实上, 深度学习的出现不是提高了而是降低了造技术的门槛, 它使得面向应用需求快速研发“不错的”技术方案成为可能. 在这个基础上融合领域知识和管理理念, 容易得到面向决策的可行动准则. 从管理层面来看, 数据智能整个研究链条非常长, 非技术方向的管理学者也能找到丰富的参与点, 特别是一些非常热门的学科交叉点, 如数据治理机制设计、基于预测的随机优化、基于预测的因果分析、可解释预测的统计理论等, 这些科学问题近两年在管理学顶刊催生了大量的专刊征文. 未来如果有更多数据智能研究相关的论文——不仅仅是实证范式还有预测范式——发表在管理学顶刊上, 我们不应感到惊奇.

参考文献

- [1] Bizer C, Boncz P, Brodie M L, et al. The meaningful use of big data: Four perspectives-four challenges[J]. ACM Sigmod Record, 2012, 40(4): 56–60.
- [2] Cattell R. Scalable sql and nosql data stores[J]. ACM Sigmod Record, 2011, 39(4): 12–27.
- [3] Ngiam J, Khosla A, Kim M, et al. Multimodal deep learning[C]// Proceedings of the 28th International Conference on Machine Learning (ICML-11), 2011: 689–696.
- [4] Provost F, Fawcett T. Data science and its relationship to big data and data-driven decision making[J]. Big Data, 2013, 1(1): 51–59.
- [5] 陈国青, 吴刚, 顾远东, 等. 管理决策情境下大数据驱动的研究和应用挑战——范式转变与研究方向[J]. 管理科学学报, 2018, 21(7): 1–10.
Chen G Q, Wu G, Gu Y D, et al. The challenges for big data driven research and applications in the context of managerial decision-making: Paradigm shift and research directions [J]. Journal of Management Sciences in China, 2018, 21(7): 1–10.
- [6] Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278–2324.
- [7] Elman J L. Finding structure in time[J]. Cognitive Science, 1990, 14(2): 179–211.
- [8] Scarselli F, Gori M, Tsoi A C, et al. The graph neural network model[J]. IEEE Transactions on Neural Networks, 2008, 20(1): 61–80.
- [9] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]// NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017: 6000–6010.
- [10] Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv preprint arXiv: 1810.04805, 2018.
- [11] Williams M A. Transmutations of knowledge systems[C]// Principles of Knowledge Representation and Reasoning. Elsevier, 1994: 619–629.
- [12] Van Der Aalst W. Data science in action[M]// Process Mining. Springer, 2016: 3–23.
- [13] Negash S, Gray P. Business intelligence[M]// Handbook on Decision Support Systems 2. Springer, 2008: 175–193.
- [14] Atanasov P, Rescober P, Stone E, et al. Distilling the wisdom of crowds: Prediction markets vs. prediction polls[J]. Management Science, 2016, 63(3): 691–706.
- [15] Chittilappilly A I, Chen L, Amer-Yahia S. A survey of general-purpose crowdsourcing techniques[J]. IEEE Transactions on Knowledge and Data Engineering, 2016, 28(9): 2246–2266.
- [16] Von Ahn L, Dabbish L. Designing games with a purpose[J]. Commun ACM, 2008, 51(8): 58–67.

- [17] Wang J, Kraska T, Franklin M J, et al. Crowder: Crowdsourcing entity resolution[J]. Proceedings of the VLDB Endowment, 2012, 5(11): 1483–1494.
- [18] Tong Y, Chen L, Zhou Z, et al. Slade: A smart large-scale task decomposer in crowdsourcing[J]. IEEE Transactions on Knowledge and Data Engineering, 2018, 30(8): 1588–1601.
- [19] Zhang X, Xue G, Yu R, et al. Truthful incentive mechanisms for crowdsourcing[C]// 2015 IEEE Conference on Computer Communications (INFOCOM). IEEE, 2015: 2830–2838.
- [20] Alam S L, Campbell J. Temporal motivations of volunteers to participate in cultural crowdsourcing work[J]. Information Systems Research, 2017, 28(4): 744–759.
- [21] Koh T K. Adopting seekers' solution exemplars in crowdsourcing ideation contests: Antecedents and consequences[J]. Information Systems Research, 2019, 30(2): 486–506.
- [22] Zhang S, Singh P V, Ghose A. A structural analysis of the role of superstars in crowdsourcing contests[J]. Information Systems Research, 2019, 30(1): 15–33.
- [23] Zheng Y, Li G, Li Y, et al. Truth inference in crowdsourcing: Is the problem solved?[J]. Proceedings of the VLDB Endowment, 2017, 10(5): 541–552.
- [24] Daniel F, Kucherbaev P, Cappiello C, et al. Quality control in crowdsourcing: A survey of quality attributes, assessment techniques, and assurance actions[J]. ACM Computing Surveys (CSUR), 2018, 51(1): 7.
- [25] Oh O, Agrawal M, Rao H R. Community intelligence and social media services: A rumor theoretic analysis of tweets during social crises[J]. MIS Quarterly, 2013, 37(2): 407–426.
- [26] Yuan K, Liu G, Wu J. Whose posts to read: Finding social sensors for effective information acquisition[J]. Information Processing & Management, 2019, 56(4): 1204–1219.
- [27] Xie W, Zhu F, Xiao J, et al. Social network monitoring for bursty cascade detection[J/OL]. ACM Trans Knowl Discov Data, 2018, 12(4). <https://doi.org/10.1145/3178048>.
- [28] An J, Weber I. Whom should we sense in “social sensing” analyzing which users work best for social media now-casting[J]. EPJ Data Science, 2015, 4(1): 22.
- [29] 刘云浩. 群智感知计算 [J]. 中国计算机学会通讯, 2012, 8(10): 38–41.
- [30] Tong Y, Zhou Z, Zeng Y, et al. Spatial crowdsourcing: A survey[J]. The VLDB Journal, 2019, 29: 217–250.
- [31] Yang D, Xue G, Fang X, et al. Incentive mechanisms for crowdsensing: Crowdsourcing with smartphones[J]. IEEE/ACM Transactions on Networking (TON), 2016, 24(3): 1732–1744.
- [32] Yao S, Hu S, Zhao Y, et al. Deepsense: A unified deep learning framework for time-series mobile sensing data processing[C]// Proceedings of the 26th International Conference on World Wide Web, 2017: 351–360.
- [33] Montori F, Jayaraman P P, Yavari A, et al. The curse of sensing: Survey of techniques and challenges to cope with sparse and dense data in mobile crowd sensing for internet of things[J]. Pervasive and Mobile Computing, 2018, 49: 111–125.
- [34] Guo B, Chen C, Zhang D, et al. Mobile crowd sensing and computing: When participatory sensing meets participatory social media[J]. IEEE Communications Magazine, 2016, 54(2): 131–137.
- [35] Radu V, Lane N D, Bhattacharya S, et al. Towards multimodal deep learning for activity recognition on mobile devices[C]// Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct. ACM, 2016: 185–188.
- [36] Zhang Y, Chen Q, Zhong S. Privacy-preserving data aggregation in mobile phone sensing[J]. IEEE Transactions on Information Forensics and Security, 2016, 11(5): 980–992.
- [37] Xu G, Li H, Liu S, et al. Efficient and privacy-preserving truth discovery in mobile crowd sensing systems[J]. IEEE Transactions on Vehicular Technology, 2019, 68(4): 3854–3865.
- [38] Samarati P, Sweeney L. Protecting privacy when disclosing information: k -anonymity and its enforcement through generalization and suppression[R]. SRI Computer Science Laboratory, 1998.
- [39] Machanavajjhala A, Kifer D, Gehrke J, et al. L-diversity: Privacy beyond k -anonymity[J]. ACM Transactions on Knowledge Discovery from Data, 2007, 1(1). <https://doi.org/10.1145/1217299.1217302>.
- [40] Wang M, Jiang Z, Zhang Y, et al. T-closeness slicing: A new privacy-preserving approach for transactional data publishing[J]. INFORMS Journal on Computing, 2018, 30(3): 438–453.
- [41] Li X B, Qin J. Anonymizing and sharing medical text records[J]. Information Systems Research, 2017, 28(2): 332–352.
- [42] Li X B, Sarkar S. Class-restricted clustering and microperturbation for data privacy[J]. Management Science, 2013, 59(4): 796–812.
- [43] Lin C, Song Z, Song H, et al. Differential privacy preserving in big data analytics for connected health[J/OL]. Journal of Medical Systems, 2016, 40. doi: 10.1007/s10916-016-0446-0.
- [44] Zhang Z, Qin Z, Zhu L, et al. Cost-friendly differential privacy for smart meters: Exploiting the dual roles of the noise[J/OL]. IEEE Transactions on Smart Grid, 2017, 8(2): 619–626. doi: 10.1109/TSG.2016.2585963.

- [45] Yang D, Qu B, Cudré-Mauroux P. Privacy-preserving social media data publishing for personalized ranking-based recommendation[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2018, 31(3): 507–520.
- [46] Gong N Z, Liu B. Attribute inference attacks in online social networks[J]. *ACM Transactions on Privacy and Security (TOPS)*, 2018, 21(1): 1–30.
- [47] Heimbach I, Hinz O. The impact of sharing mechanism design on content sharing in online social networks[J]. *Information Systems Research*, 2018, 29(3): 592–611.
- [48] Xu H, Teo H H, Tan B C, et al. The role of push-pull technology in privacy calculus: The case of location-based services[J]. *Journal of Management Information Systems*, 2009, 26(3): 135–174.
- [49] Buckman J R, Bockstedt J C, Hashim M J. Relative privacy valuations under varying disclosure characteristics[J]. *Information Systems Research*, 2019, 30(2): 375–388.
- [50] Kummer M, Schulte P. When private information settles the bill: Money and privacy in google market for smartphone applications[J]. *Management Science*, 2019, 65(8): 3470–3794.
- [51] Schneider M J, Jagpal S, Gupta S, et al. A flexible method for protecting marketing data: An application to point-of-sale data[J]. *Marketing Science*, 2018, 37(1): 153–171.
- [52] Parker G G, Van Alstyne M W, Choudary S P. Platform revolution: How networked markets are transforming the economy? And how to make them work for you[M]. WW Norton & Company, 2016.
- [53] Parmar R, Mackenzie I, Cohn D, et al. The new patterns of innovation[J]. *Harvard Business Review*, 2014, 92(1): 2.
- [54] Borgman C L. The conundrum of sharing research data[J]. *Journal of the American Society for Information Science and Technology*, 2012, 63(6): 1059–1078.
- [55] Koutroumpis P, Leiponen A, Thomas L D. The (unfulfilled) potential of data marketplaces[R]. ETLA Working Papers, 2017.
- [56] Naor M, Pinkas B, Pinkas B. Efficient oblivious transfer protocols[C]// *Proceedings of the 12th Annual ACM-SIAM Symposium on Discrete Algorithms*. Society for Industrial and Applied Mathematics, 2001: 448–457.
- [57] Gentry C, Boneh D. A fully homomorphic encryption scheme: Volume 20[M]. Stanford University Stanford, 2009.
- [58] Clifton C, Kantarcioglu M, Vaidya J, et al. Tools for privacy preserving distributed data mining[J]. *ACM Sigkdd Explorations Newsletter*, 2002, 4(2): 28–34.
- [59] McMahan B, Moore E, Ramage D, et al. Communication-efficient learning of deep networks from decentralized data[C]// *Artificial Intelligence and Statistics*. 2017: 1273–1282.
- [60] Zhu L, Liu Z, Han S. Deep leakage from gradients[C]// *33rd Conference on Neural Information Processing Systems*, 2019.
- [61] Breiman L. Bagging predictors[J]. *Machine Learning*, 1996, 24(2): 123–140. doi: 10.1023/A:1018054314350.
- [62] Weingessel A, Dimitriadou E, Hornik K. An ensemble method for clustering[C]// *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*, 2003.
- [63] Karypis G, Aggarwal R, Kumar V, et al. Multilevel hypergraph partitioning: Applications in VLSI domain[J]. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 1999, 7(1): 69–79.
- [64] Zimek A, Campello R J, Sander J. Ensembles for unsupervised outlier detection: Challenges and research questions a position paper[J]. *ACM SIGKDD Explorations Newsletter*, 2014, 15(1): 11–22.
- [65] Freund Y, Schapire R E. A decision-theoretic generalization of on-line learning and an application to boosting[J]. *Journal of Computer and System Sciences*, 1997, 55(1): 119–139. doi: 10.1006/jcss.1997.1504.
- [66] Ho T K. Random decision forests[C]// *Proceedings of the Third International Conference on Document Analysis and Recognition (Volume 1)*. Washington, DC, USA: IEEE Computer Society, 1995: 278–282.
- [67] Dietterich T G, Bakiri G. Solving multiclass learning problems via error-correcting output codes[J]. *Journal of Artificial Intelligence Research*, 1995, 2(1): 263–286.
- [68] Brown G, Wyatt J L, Tiño P. Managing diversity in regression ensembles[J]. *Journal of Machine Learning Research*, 2005, 6: 1621–1650.
- [69] Mendes-Moreira J A, Soares C, Jorge A M, et al. Ensemble approaches for regression: A survey[J]. *ACM Computing Surveys*, 2012, 45(1): 10:1–10:40. doi: 10.1145/2379776.2379786.
- [70] Wolpert D H. Stacked generalization[J]. *Neural Networks*, 1992, 5(2): 241–259.
- [71] Chen T, Guestrin C. Xgboost: A scalable tree boosting system[C]// *KDD'16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: ACM, 2016: 785–794. doi: 10.1145/2939672.2939785.
- [72] Kotschieder P, Fiterau M, Criminisi A, et al. Deep neural decision forests[C]// *IJCAI'16: Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*. AAAI Press, 2016: 4190–4194.

- [73] Topchy A, Jain A K, Punch W. Combining multiple weak clusterings[C]// Proceedings of the Third IEEE International Conference on Data Mining. Washington, DC, USA: IEEE Computer Society, 2003: 331–338.
- [74] Strehl A, Ghosh J. Cluster ensembles — A knowledge reuse framework for combining multiple partitions[J]. *Journal of Machine Learning Research*, 2002, 3(Dec): 583–617.
- [75] Fred A L, Jain A K. Combining multiple clusterings using evidence accumulation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, 27(6): 835–850.
- [76] Wu J, Liu H, Xiong H, et al. K-means-based consensus clustering: A unified view[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2014, 27(1): 155–169.
- [77] Liu H, Wu J, Liu T, et al. Spectral ensemble clustering via weighted k-means: Theoretical and practical evidence[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2017, 29(5): 1129–1143.
- [78] Buckman J, Roy A, Raffel C, et al. Thermometer encoding: One-hot way to resist adversarial examples[C]// Proceedings of the 6th International Conference on Learning Representations (ICLR 2018), 2018.
- [79] Blei D M, Ng A Y, Jordan M I. Latent dirichlet allocation[J]. *Journal of Machine Learning Research*, 2003, 3(Jan): 993–1022.
- [80] Wang C, Blei D M. Collaborative topic modeling for recommending scientific articles[C]// KDD'11: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY, USA: ACM, 2011: 448–456.
- [81] Lin H, Zhu H, Zuo Y, et al. Collaborative company profiling: Insights from an employee's perspective[C]// Proceedings of the 31st AAAI Conference on Artificial Intelligence, 2017: 1417–1423.
- [82] Melnyk I, Banerjee A, Matthews B, et al. Semi-markov switching vector autoregressive model-based anomaly detection in aviation systems[C]// KDD'16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY, USA: ACM, 2016: 1065–1074. <http://doi.org/10.1145/2939672.2939789>.
- [83] Lin H, Liu G, Wu J, et al. Fraud detection in dynamic interaction network[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2019. doi: 10.1109/TKDE.2019.2912817.
- [84] Li X, She J. Collaborative variational autoencoder for recommender systems[C]// KDD'17: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY, USA: ACM, 2017: 305–314. doi: 10.1145/3097983.3098077.
- [85] Miura Y, Taniguchi M, Taniguchi T, et al. Unifying text, metadata, and user network representations with a neural network for geolocation prediction[C]// Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Vancouver, Canada: Association for Computational Linguistics, 2017: 1260–1272.
- [86] Hu B, Shi C, Zhao W X, et al. Leveraging meta-path based context for top-N recommendation with a neural co-attention model[C]// KDD'18: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York, NY, USA: ACM, 2018: 1531–1540. doi: 10.1145/3219819.3219965.
- [87] Wang J, Wu J, Wang Z, et al. Understanding urban dynamics via context-aware tensor factorization with neighboring regularization[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2019. doi: 10.1109/TKDE.2019.2915231.
- [88] Wang H, Zhang F, Wang J, et al. Ripplenet: Propagating user preferences on the knowledge graph for recommender systems[C]// CIKM'18: Proceedings of the 27th ACM International Conference on Information and Knowledge Management. New York, NY, USA: ACM, 2018: 417–426. doi: 10.1145/3269206.3271739.
- [89] Wang J, Lin Y, Wu J, et al. Coupling implicit and explicit knowledge for customer volume prediction[C]// Thirty-First AAAI Conference on Artificial Intelligence. 2017.
- [90] Wu M, Goodman N. Multimodal generative models for scalable weakly-supervised learning[C]// NIPS'18: Proceedings of the 32nd International Conference on Neural Information Processing Systems. USA: Curran Associates Inc., 2018: 5580–5590.
- [91] Kendall A, Gal Y, Cipolla R. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7482–7491.
- [92] Pan S J, Yang Q. A survey on transfer learning[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2009, 22(10): 1345–1359.
- [93] Xu Y, Pan S J, Xiong H, et al. A unified framework for metric transfer learning[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2017, 29(6): 1158–1171.
- [94] Long M, Zhu H, Wang J, et al. Deep transfer learning with joint adaptation networks[C]// Proceedings of the 34th International Conference on Machine Learning-Volume 70. JMLR.org, 2017: 2208–2217.
- [95] Chang H, Han J, Zhong C, et al. Unsupervised transfer learning via multi-scale convolutional sparse coding for biomedical applications[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 40(5): 1182–1194.

- [96] Luo Z, Zou Y, Hoffman J, et al. Label efficient learning of transferable representations across domains and tasks[C]// Advances in Neural Information Processing Systems, 2017: 165–177.
- [97] Wang L, Geng X, Ma X, et al. Cross-city transfer learning for deep spatio-temporal prediction[C]// Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI-19), 2019: 1893–1899. <https://doi.org/10.24963/ijcai.2019/262>.
- [98] Kaelbling L P, Littman M L, Moore A W. Reinforcement learning: A survey[J]. *Journal of Artificial Intelligence Research*, 1996, 4: 237–285.
- [99] Sutton R S, Barto A G. Reinforcement learning: An introduction[M]. MIT press, 2018.
- [100] Szepesvári C. Algorithms for reinforcement learning[J]. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 2010, 4(1): 1–103.
- [101] Silver D, Schrittwieser J, Simonyan K, et al. Mastering the game of go without human knowledge[J]. *Nature*, 2017, 550(7676): 354–359.
- [102] Kober J, Bagnell J A, Peters J. Reinforcement learning in robotics: A survey[J]. *The International Journal of Robotics Research*, 2013, 32(11): 1238–1274.
- [103] Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning[J]. *Nature*, 2015, 518(7540): 529–533.
- [104] Wiering M. Multi-agent reinforcement learning for traffic light control[C]// Machine Learning: Proceedings of the Seventeenth International Conference (ICML'2000), 2000: 1151–1158.
- [105] Wei H, Zheng G, Yao H, et al. Intellilight: A reinforcement learning approach for intelligent traffic light control[C]// Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. ACM, 2018: 2496–2505.
- [106] Wei H, Chen C, Zheng G, et al. Presslight: Learning max pressure control to coordinate traffic signals in arterial network[C]// Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. ACM, 2019: 1290–1298.
- [107] Nevmyvaka Y, Feng Y, Kearns M. Reinforcement learning for optimized trade execution[C]// Proceedings of the 23rd International Conference on Machine Learning. ACM, 2006: 673–680.
- [108] Moody J, Wu L, Liao Y, et al. Performance functions and reinforcement learning for trading systems and portfolios[J]. *Journal of Forecasting*, 1998, 17(5–6): 441–470.
- [109] Wang J, Zhang Y, Tang K, et al. Alphastock: A buying-winners-and-selling-losers investment strategy using interpretable deep reinforcement attention networks[C]// The 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2019: 1900–1908.
- [110] Kara A, Dogan I. Reinforcement learning approaches for specifying ordering policies of perishable inventory systems[J]. *Expert Systems with Applications*, 2018, 91: 150–158.
- [111] Tesauro G, Jong N K, Das R, et al. A hybrid reinforcement learning approach to autonomic resource allocation[C]// 2006 IEEE International Conference on Autonomic Computing. IEEE, 2006: 65–73.
- [112] Maxwell M S, Restrepo M, Henderson S G, et al. Approximate dynamic programming for ambulance redeployment[J]. *INFORMS Journal on Computing*, 2010, 22(2): 266–281.
- [113] 刘冠男, 曲金铭, 李小琳, 等. 基于深度强化学习的救护车动态重定位调度研究 [J]. *管理科学学报*, 2020, 23(2): 38–52.
Liu G N, Qu J M, Li X L, et al. Dynamic ambulance redeployment based on deep reinforcement learning[J]. *Journal of Management Sciences in China*, 2020, 23(2): 38–52.
- [114] Tang X, Qin Z T, Zhang F, et al. A deep value-network based approach for multi-driver order dispatching[C]// Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. ACM, 2019: 1780–1790.
- [115] Xu Z, Li Z, Guan Q, et al. Large-scale order dispatch in on-demand ride-hailing platforms: A learning and planning approach[C]// Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. ACM, 2018: 905–913.
- [116] Botvinick M, Ritter S, Wang J X, et al. Reinforcement learning, fast and slow[J]. *Trends in Cognitive Sciences*, 2019, 23(5): 408–422.
- [117] 张军阳, 王慧丽, 郭阳, 等. 深度学习相关研究综述 [J]. *计算机应用研究*, 2018, 35(7): 1–12.
Zhang J Y, Wang H L, Guo Y, et al. Review of deep learning[J]. *Application Research of Computers*, 2018, 35(7): 1–12.
- [118] Liu W, Wang Z, Liu X, et al. A survey of deep neural network architectures and their applications[J]. *Neurocomputing*, 2017, 234: 11–26.
- [119] Dargan S, Kumar M, Ayyagari M R, et al. A survey of deep learning and its applications: A new paradigm to machine learning[J]. *Archives of Computational Methods in Engineering*, 2019. <https://doi.org/10.1007/s11831-019-09344-w>.
- [120] Lecun Y, Bengio Y, Hinton G. Deep learning[J]. *Nature*, 2015, 521(7553): 436–444.

- [121] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770–778.
- [122] Pouyanfar S, Sadiq S, Yan Y, et al. A survey on deep learning: Algorithms, techniques, and applications[J]. ACM Computing Surveys (CSUR), 2019, 51(5): 92.
- [123] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735–1780.
- [124] Cho K, Van Merriënboer B, Gulcehre C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation[C]// Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Doha, Qatar: Association for Computational Linguistics, 2014: 1724–1734. <https://www.aclweb.org/anthology/D14-1179>. doi: 10.3115/v1/D14-1179.
- [125] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate[J]. arXiv preprint arXiv: 1409.0473, 2014.
- [126] Wu Z, Pan S, Chen F, et al. A comprehensive survey on graph neural networks[J]. arXiv: Learning, 2019.
- [127] Zhou J, Cui G, Zhang Z, et al. Graph neural networks: A review of methods and applications[J]. arXiv preprint arXiv: 1812.08434, 2018.
- [128] Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks[C]// International Conference on Learning Representations, 2017.
- [129] Veličković P, Cucurull G, Casanova A, et al. Graph attention networks[J]. arXiv preprint arXiv: 1710.10903, 2017.
- [130] Bengio Y, Ducharme R, Vincent P, et al. A neural probabilistic language model[J]. Journal of Machine Learning Research, 2003, 3(Feb): 1137–1155.
- [131] Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality[C]// Advances in Neural Information Processing Systems, 2013: 3111–3119.
- [132] Arora S, Liang Y, Ma T. A simple but tough-to-beat baseline for sentence embeddings[C]// 5th International Conference on Learning Representations, 2017.
- [133] Le Q, Mikolov T. Distributed representations of sentences and documents[C]// ICML'14: Proceedings of the 31st International Conference on International Conference on Machine Learning-Volume 32. JMLR.org, 2014: 1188–1196.
- [134] Kalchbrenner N, Grefenstette E, Blunsom P. A convolutional neural network for modelling sentences[C]// Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, 2014: 655–665.
- [135] Lai S, Xu L, Liu K, et al. Recurrent convolutional neural networks for text classification[C]// AAAI'15: Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence. AAAI Press, 2015: 2267–2273.
- [136] Perozzi B, Alrfou R, Skiena S. Deepwalk: Online learning of social representations[C]// Acm Sigkdd International Conference on Knowledge Discovery & Data Mining, 2014.
- [137] Aditya Grover J L. node2vec: Scalable feature learning for networks[C]// Acm Sigkdd International Conference on Knowledge Discovery & Data Mining, 2016.
- [138] Tang J, Qu M, Wang M Z, et al. Line: Large-scale information network embedding[C]// International Conference on World Wide Web, 2015.
- [139] Zuo Y, Liu G, Lin H, et al. Embedding temporal network via neighborhood formation[C]// KDD'18: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018: 2857–2866.
- [140] Wang D, Peng C, Zhu W. Structural deep network embedding[C]// Acm Sigkdd International Conference on Knowledge Discovery & Data Mining, 2016.
- [141] Bordes A, Usunier N, Garcia-Durán A, et al. Translating embeddings for modeling multi-relational data[C]// NIPS'13: Proceedings of the 26th International Conference on Neural Information Processing Systems-Volume 2, 2013: 2787–2795.
- [142] Wang Z, Zhang J, Feng J, et al. Knowledge graph embedding by translating on hyperplanes[C]// AAAI'14: Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, 2014: 1112–1119.
- [143] Lin Y, Liu Z, Sun M, et al. Learning entity and relation embeddings for knowledge graph completion[C]// AAAI'15: Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, 2015: 2181–2187.
- [144] He S, Liu K, Ji G, et al. Learning to represent knowledge graphs with Gaussian embedding[C]// CIKM'15: Proceedings of the 24th ACM International Conference on Information and Knowledge Management, 2015: 623–632.
- [145] Li X, Zhao K, Cong G, et al. Deep representation learning for trajectory similarity computation[C]// 2018 IEEE 34th International Conference on Data Engineering (ICDE). IEEE, 2018: 617–628.
- [146] Pan S, Ding T. Social media-based user embedding: A literature review[C]// Proceedings of the 28th International Joint Conference on Artificial Intelligence. AAAI Press, 2019: 6318–6324.

- [147] Siddharth N, Paige B, Van De Meent J W, et al. Learning disentangled representations with semi-supervised deep generative models[C]// Advances in Neural Information Processing Systems, 2017: 5925–5935.
- [148] Hu Z, Yang Z, Liang X, et al. Toward controlled generation of text[C]// ICML'17: Proceedings of the 34th International Conference on Machine Learning-Volume 70, 2017: 1587–1596.
- [149] Vlastelica M, Paulus A, Musil V, et al. Differentiation of blackbox combinatorial solvers[C]// International Conference on Learning Representations, 2020.
- [150] Lowry S, Macpherson G. A blot on the profession[J]. *British Medical Journal*, 1988, 296(6623): 657–658.
- [151] Caliskan A, Bryson J J, Narayanan A. Semantics derived automatically from language corpora contain human-like biases[J]. *Science*, 2017, 356(6334): 183–186.
- [152] Kroll J A, Barocas S, Felten E W, et al. Accountable algorithms[J]. *University of Pennsylvania Law Review*, 2016, 165(3): 633–706.
- [153] Danks D, London A J. Regulating autonomous systems: Beyond standards[J]. *IEEE Intelligent Systems*, 2017, 32(1): 88–91.
- [154] Nguyen A, Yosinski J, Clune J. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 427–436.
- [155] Guidotti R, Monreale A, Ruggieri S, et al. A survey of methods for explaining black box models[J]. *ACM Computing Surveys (CSUR)*, 2019, 51(5): 93.
- [156] Ribeiro M T, Singh S, Guestrin C. Why should i trust you? Explaining the predictions of any classifier[C]// Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2016: 1135–1144.
- [157] Ribeiro M T, Singh S, Guestrin C. Anchors: High-precision model-agnostic explanations[C]// Thirty-Second AAAI Conference on Artificial Intelligence, 2018.
- [158] Xu K, Ba J, Kiros R, et al. Show, attend and tell: Neural image caption generation with visual attention[C]// International Conference on Machine Learning, 2015: 2048–2057.
- [159] Montavon G, Lapuschkin S, Binder A, et al. Explaining nonlinear classification decisions with deep taylor decomposition[J]. *Pattern Recognition*, 2017, 65: 211–222.
- [160] Landecker W, Thomure M D, Bettencourt L M, et al. Interpreting individual classifications of hierarchical networks[C]// 2013 IEEE Symposium on Computational Intelligence and Data Mining (CIDM). IEEE, 2013: 32–38.
- [161] Zhou B, Khosla A, Lapedriza A, et al. Learning deep features for discriminative localization[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 2921–2929.
- [162] Selvaraju R R, Cogswell M, Das A, et al. Grad-cam: Visual explanations from deep networks via gradient-based localization[C]// Proceedings of the IEEE International Conference on Computer Vision, 2017: 618–626.
- [163] Sabour S, Frosst N, Hinton G E. Dynamic routing between capsules[C]// Advances in Neural Information Processing Systems, 2017: 3856–3866.
- [164] Wang J, Feng K, Wu J. SVM-based deep stacking networks[C]// 33rd AAAI Conference on Artificial Intelligence (AAAI'19), 2019.
- [165] Craven M, Shavlik J W. Extracting tree-structured representations of trained networks[C]// Advances in Neural Information Processing Systems, 1996: 24–30.
- [166] Krishnan R, Sivakumar G, Bhattacharya P. Extracting decision trees from trained neural networks[J]. *Pattern Recognition*, 1999, 32(12): 1999–2009.
- [167] Chipman H, George E, McCulloch R. Making sense of a forest of trees[C]// Proceedings of the 30th Symposium on the Interface. Fairfax Station, VA: Interface Foundation of North America, 1998: 84–92.
- [168] Schetinin V, Fieldsend J E, Partridge D, et al. Confident interpretation of bayesian decision tree ensembles for clinical applications[J]. *IEEE Transactions on Information Technology in Biomedicine*, 2007, 11(3): 312–319.
- [169] Lou Y, Caruana R, Gehrke J. Intelligible models for classification and regression[C]// Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2012: 150–158.
- [170] Lou Y, Caruana R, Gehrke J, et al. Accurate intelligible models with pairwise interactions[C]// Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2013: 623–631.
- [171] Henelius A, Puolamäki K, Boström H, et al. A peek into the black box: exploring classifiers by randomization[J]. *Data Mining and Knowledge Discovery*, 2014, 28(5–6): 1503–1529.
- [172] Sonnenburg S, Zien A, Philips P, et al. Poims: Positional oligomer importance matrices-understanding support vector machine-based signal detectors[J]. *Bioinformatics*, 2008, 24(13): i6–i14.

- [173] Wang J, Wang Z, Li J, et al. Multilevel wavelet decomposition network for interpretable time series analysis[C]// Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. ACM, 2018: 2437–2446.
- [174] Horel E, Giesecke K. Towards explainable AI: Significance tests for neural networks[J]. arXiv preprint arXiv: 1902.06021, 2019.
- [175] Yang C, Rangarajan A, Ranka S. Global model interpretation via recursive partitioning[C]// 2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS). IEEE, 2018: 1563–1570.
- [176] 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.
- [177] Johansson F, Shalit U, Sontag D. Learning representations for counterfactual inference[C]// International Conference on Machine Learning, 2016: 3020–3029.
- [178] Deerwester S, Dumais S T, Furnas G W, et al. Indexing by latent semantic analysis[J]. Journal of the American Society for Information Science, 1990, 41(6): 391–407.
- [179] Hofmann T. Probabilistic latent semantic analysis[C]// UAI'99: Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1999: 289–296.
- [180] Tang J, Zhang M, Mei Q. One theme in all views: Modeling consensus topics in multiple contexts[C]// KDD'13: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY, USA: ACM, 2013: 5–13.
- [181] Zuo Y, Wu J, Zhang H, et al. Complementary aspect-based opinion mining[J]. IEEE Transactions on Knowledge and Data Engineering, 2018, 30(2): 249–262.
- [182] Quan X, Kit C, Ge Y, et al. Short and sparse text topic modeling via self-aggregation[C]// IJCAI'15: Proceedings of the 24th International Conference on Artificial Intelligence. AAAI Press, 2015: 2270–2276.
- [183] Zuo Y, Wu J, Zhang H, et al. Topic modeling of short texts: A pseudo-document view[C]// KDD'16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY, USA: ACM, 2016: 2105–2114.
- [184] Yoshii K, Goto M, Komatani K, et al. An efficient hybrid music recommender system using an incrementally trainable probabilistic generative model[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2008, 16(2): 435–447.
- [185] Yin H, Zhou X, Shao Y, et al. Joint modeling of user check-in behaviors for point-of-interest recommendation[C]// Proceedings of the 24th ACM International Conference on Information and Knowledge Management. ACM, 2015: 1631–1640.
- [186] Jacobs B J, Donkers B, Fok D. Model-based purchase predictions for large assortments[J]. Marketing Science, 2016, 35(3): 389–404.
- [187] Jiang M, Cui P, Liu R, et al. Social contextual recommendation[C]// CIKM'12: Proceedings of the 21st ACM International Conference on Information and Knowledge Management. New York, NY, USA: Association for Computing Machinery, 2012: 45–54.
- [188] Zhu H, Chen E, Xiong H, et al. Mining mobile user preferences for personalized context-aware recommendation[J]. ACM Transactions on Intelligent Systems and Technology (TIST), 2015, 5(4): 58.
- [189] Liu B, Kong D, Cen L, et al. Personalized mobile app recommendation: Reconciling app functionality and user privacy preference[C]// Proceedings of the Eighth ACM International Conference on Web Search and Data Mining. ACM, 2015: 315–324.
- [190] Kingma D P, Welling M. Auto-encoding variational bayes[C]// 2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14–16, 2014, Conference Track Proceedings, 2014.
- [191] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]// Advances in Neural Information Processing Systems, 2014: 2672–2680.
- [192] Li C, Zhu J, Shi T, et al. Max-margin deep generative models[C]// Advances in Neural Information Processing Systems, 2015: 1837–1845.
- [193] Ho J, Ermon S. Generative adversarial imitation learning[C]// Advances in Neural Information Processing Systems, 2016: 4565–4573.
- [194] Wang H, Wang J, Wang J, et al. Graphgan: Graph representation learning with generative adversarial nets[C]// Thirty-Second AAAI Conference on Artificial Intelligence, 2018.
- [195] Feuerriegel S, Gordon J. Long-term stock index forecasting based on text mining of regulatory disclosures[J]. Decision Support Systems, 2018, 112: 88–97.

- [196] 部慧, 解峥, 李佳鸿, 等. 基于股评的投资者情绪对股票市场的影响 [J]. 管理科学学报, 2018, 21(4): 86–101.
Bu H, Xie Z, Li J H, et al. Investor sentiment extracted from internet stock message boards and its effect on Chinese stock market[J]. Journal of Management Sciences in China, 2018, 21(4): 86–101.
- [197] Jung Y, Suh Y. Mining the voice of employees: A text mining approach to identifying and analyzing job satisfaction factors from online employee reviews[J]. Decision Support Systems, 2019, 123: 113074.
- [198] Wu Z, Cao J, Wang Y, et al. HPSD: A hybrid pu-learning-based spammer detection model for product reviews[J]. IEEE Transactions on Cybernetics, 2020, 50(4): 1595–1606.
- [199] Ibrahim N F, Wang X. A text analytics approach for online retailing service improvement: Evidence from Twitter[J]. Decision Support Systems, 2019, 121: 37–50.
- [200] Brown P F, Desouza P V, Mercer R L, et al. Class-based n-gram models of natural language[J]. Computational Linguistics, 1992, 18(4): 467–479.
- [201] Severyn A, Moschitti A. Twitter sentiment analysis with deep convolutional neural networks[C]// Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 2015: 959–962.
- [202] Chiu J P, Nichols E. Named entity recognition with bidirectional LSTM-CNNs[J]. Transactions of the Association for Computational Linguistics, 2016, 4: 357–370.
- [203] Alsumait L, Barabara D, Domeniconi C. On-line LDA: Adaptive topic models for mining text streams with applications to topic detection and tracking[C]// 2008 Eighth IEEE International Conference on Data Mining. IEEE, 2008: 3–12.
- [204] Shahnaz F, Berry M W, Pauca V P, et al. Document clustering using nonnegative matrix factorization[J]. Information Processing & Management, 2006, 42(2): 373–386.
- [205] Kingma D P, Welling M. Auto-encoding variational Bayes[J]. arXiv preprint arXiv:1312.6114, 2013.
- [206] Dos Santos C, Gatti M. Deep convolutional neural networks for sentiment analysis of short texts[C]// Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers. 2014: 69–78.
- [207] He R, Lee W S, Ng H T, et al. An interactive multi-task learning network for end-to-end aspect-based sentiment analysis[C]// Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy: Association for Computational Linguistics, 2019: 504–515.
- [208] Trisedya B D, Weikum G, Qi J, et al. Neural relation extraction for knowledge base enrichment[C]// Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy: Association for Computational Linguistics, 2019: 229–240.
- [209] Moon S, Shah P, Kumar A, et al. OpenDialKG: Explainable conversational reasoning with attention-based walks over knowledge graphs[C]// Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy: Association for Computational Linguistics, 2019: 845–854.
- [210] Zhang X, Yang Y, Yuan S, et al. Syntax-infused variational autoencoder for text generation[C]// Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy: Association for Computational Linguistics, 2019: 2069–2078.
- [211] Xu K, Lai Y, Feng Y, et al. Enhancing key-value memory neural networks for knowledge based question answering[C]// Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), 2019: 2937–2947.
- [212] Zhuang Y, Wang H. Token-level dynamic self-attention network for multi-passage reading comprehension[C]// Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy: Association for Computational Linguistics, 2019: 2252–2262.
- [213] Peters M E, Neumann M, Iyyer M, et al. Deep contextualized word representations[C]// Proceedings of NAACL-HLT, 2018: 2227–2237.
- [214] Radford A, Narasimhan K, Salimans T, et al. Improving language understanding by generative pre-training[J]. <http://www.nlp.ir.org/wordpress/wp-content/uploads/2019/06/Improving-language-understanding-by-generative-pre-training.pdf>, 2018.
- [215] Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[C]// Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), 2019: 4171–4186.
- [216] Yang Z, Dai Z, Yang Y, et al. Xlnet: Generalized autoregressive pretraining for language understanding[C]// Advances in Neural Information Processing Systems, 2019: 5754–5764.
- [217] Lan Z, Chen M, Goodman S, et al. Albert: A lite bert for self-supervised learning of language representations[C]// International Conference on Learning Representations, 2019.

- [218] Pazzani M J, Billsus D. Content-based recommendation systems[M]// *The Adaptive Web*. Springer, 2007: 325–341.
- [219] Linden G, Smith B, York J. Amazon.com recommendations: Item-to-item collaborative filtering[J]. *IEEE Internet Computing*, 2003, 7(1): 76–80.
- [220] Koren Y, Bell R, Volinsky C. Matrix factorization techniques for recommender systems[J]. *Computer*, 2009(8): 30–37.
- [221] Mnih A, Salakhutdinov R R. Probabilistic matrix factorization[C]// *Advances in Neural Information Processing Systems*, 2008: 1257–1264.
- [222] Vargas S, Castells P. Rank and relevance in novelty and diversity metrics for recommender systems[C]// *Proceedings of the Fifth ACM Conference on Recommender Systems*, 2011: 109–116.
- [223] Oestreicher-Singer G, Sundararajan A. Recommendation networks and the long tail of electronic commerce[J]. *MIS Quarterly*, 2012, 36(1): 65–83.
- [224] He X, Liao L, Zhang H, et al. Neural collaborative filtering[C]// *Proceedings of the 26th International Conference on World Wide Web*, 2017: 173–182.
- [225] Cheng H T, Koc L, Harmsen J, et al. Wide & deep learning for recommender systems[C]// *Proceedings of the 1st Workshop on Deep Learning for Recommender Systems*. ACM, 2016: 7–10.
- [226] Wang R, Fu B, Fu G, et al. Deep & cross network for ad click predictions[C]// *Proceedings of the ADKDD'17*. ACM, 2017: 12.
- [227] Zhou G, Zhu X, Song C, et al. Deep interest network for click-through rate prediction[C]// *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 2018: 1059–1068.
- [228] Lu S, Xiao L, Ding M. A video-based automated recommender (var) system for garments[J]. *Marketing Science*, 2016, 35(3): 484–510.
- [229] Ansari A, Li Y, Zhang J Z. Probabilistic topic model for hybrid recommender systems: A stochastic variational bayesian approach[J]. *Marketing Science*, 2018, 37(6): 987–1008.
- [230] Zhang F, Yuan N J, Lian D, et al. Collaborative knowledge base embedding for recommender systems[C]// *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2016: 353–362.
- [231] Liu G, Fu Y, Chen G, et al. Modeling buying motives for personalized product bundle recommendation[J]. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 2017, 11(3): 1–26.
- [232] Zhang L, Liu G, Wu J. Beyond similarity: Relation embedding with dual attentions for item-based recommendation[J]. *arXiv preprint arXiv: 1911.04099*, 2019.
- [233] He J, Fang X, Liu H, et al. Mobile app recommendation: An involvement-enhanced approach[J]. *MIS Quarterly*, 2019, 43(3): 827–850.
- [234] Gao H, Tang J, Hu X, et al. Content-aware point of interest recommendation on location-based social networks[C]// *Twenty-ninth AAAI conference on Artificial Intelligence*, 2015.
- [235] Xie M, Yin H, Wang H, et al. Learning graph-based poi embedding for location-based recommendation[C]// *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, 2016: 15–24.
- [236] Kawaguchi K, Uetake K, Watanabe Y. Effectiveness of product recommendations under time and crowd pressures[J]. *Marketing Science*, 2019, 38(2): 253–273.
- [237] Dzyabura D, Hauser J R. Recommending products when consumers learn their preference weights[J]. *Marketing Science*, 2019, 38(3): 417–441.
- [238] Oestreicher-Singer G, Sundararajan A. The visible hand? Demand effects of recommendation networks in electronic markets[J]. *Management Science*, 2012, 58(11): 1963–1981.
- [239] Lin Z, Goh K Y, Heng C S. The demand effects of product recommendation networks: An empirical analysis of network diversity and stability[J]. *MIS Quarterly*, 2017, 41(2): 397–426.
- [240] Liang C, Shi Z, Raghu T. The spillover of spotlight: Platform recommendation in the mobile app market[J]. *Information Systems Research*, 2019, 30(4): 1296–1318.
- [241] Adomavicius G, Bockstedt J C, Curley S P, et al. Effects of online recommendations on consumers' willingness to pay[J]. *Information Systems Research*, 2018, 29(1): 84–102.
- [242] Prawesh S, Padmanabhan B. The “most popular news” recommender: Count amplification and manipulation resistance[J]. *Information Systems Research*, 2014, 25(3): 569–589.
- [243] Song Y, Sahoo N, Ofek E. When and how to diversify — A multicategory utility model for personalized content recommendation[J]. *Management Science*, 2019, 65(8): 3737–3757.
- [244] Adomavicius G, Kwon Y. Optimization-based approaches for maximizing aggregate recommendation diversity[J]. *INFORMS Journal on Computing*, 2014, 26(2): 351–369.

- [245] Jiang H, Qi X, Sun H. Choice-based recommender systems: A unified approach to achieving relevancy and diversity[J]. *Operations Research*, 2014, 62(5): 973–993.
- [246] Beutel A, Chen J, Doshi T, et al. Fairness in recommendation ranking through pairwise comparisons[C]// *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019: 2212–2220.
- [247] Wang W, Xu J, Wang M. Effects of recommendation neutrality and sponsorship disclosure on trust vs. distrust in online recommendation agents: Moderating role of explanations for organic recommendations[J]. *Management Science*, 2018, 64(11): 5198–5219.
- [248] Gu S, Kelly B, Xiu D. Empirical asset pricing via machine learning[J]. *Review of Financial Studies*, 2020, 33(5): 2223–2273.
- [249] Grudnitski G, Osburn L. Forecasting S&P and gold futures prices: An application of neural networks[J]. *Journal of Futures Markets*, 1993, 13(6): 631–643.
- [250] Tay F E, Cao L. Application of support vector machines in financial time series forecasting[J]. *Omega*, 2001, 29(4): 309–317.
- [251] Huang W, Nakamori Y, Wang S Y. Forecasting stock market movement direction with support vector machine[J]. *Computers & Operations Research*, 2005, 32(10): 2513–2522.
- [252] Li J, Bu H, Wu J. Sentiment-aware stock market prediction: A deep learning method[C]// *2017 International Conference on Service Systems and Service Management*. IEEE, 2017: 1–6.
- [253] Tsaih R, Hsu Y, Lai C C. Forecasting S&P 500 stock index futures with a hybrid AI system[J]. *Decision Support Systems*, 1998, 23(2): 161–174.
- [254] Pai P F, Lin C S. A hybrid arima and support vector machines model in stock price forecasting[J]. *Omega*, 2005, 33(6): 497–505.
- [255] Yu L, Wang S, Lai K K. Forecasting crude oil price with an emd-based neural network ensemble learning paradigm[J]. *Energy Economics*, 2008, 30(5): 2623–2635.
- [256] Lu C J, Lee T S, Chiu C C. Financial time series forecasting using independent component analysis and support vector regression[J]. *Decision Support Systems*, 2009, 47(2): 115–125.
- [257] Wang S, Yu L, Kin K L. Crude oil price forecasting with TEI@I methodology[J]. *Journal of Systems Science & Complexity*, 2005, 18(2): 145–166.
- [258] 闫妍, 徐伟, 邵慧, 等. 基于 TEI@I 方法论的房价预测方法 [J]. *系统工程理论与实践*, 2007, 27(7): 1–9.
Yan Y, Xu W, Bu H, et al. Method for housing price forecasting based on TEI@I methodology[J]. *Systems Engineering — Theory & Practice*, 2007, 27(7): 1–9.
- [259] Franses P H. Merging models and experts[J]. *International Journal of Forecasting*, 2008, 24(1): 31–33.
- [260] Olson D, Mossman C. Neural network forecasts of canadian stock returns using accounting ratios[J]. *International Journal of Forecasting*, 2003, 19(3): 453–465.
- [261] Schumaker R P, Chen H. Textual analysis of stock market prediction using breaking financial news: The azfin text system[J]. *ACM Transactions on Information Systems*, 2009, 27(2): 12.
- [262] Ding X, Zhang Y, Liu T, et al. Deep learning for event-driven stock prediction[C]// *Proceedings of the 24th International Joint Conference on Artificial Intelligence*. AAAI Press, 2015: 2327–2333.
- [263] Armstrong J S. Principles of forecasting: A handbook for researchers and practitioners: Volume 30[M]. Springer Science & Business Media, 2001.
- [264] Chen C, Li Y, Bu H, et al. Forecasting on trading: A parameter adaptive framework based on q-learning[C]// *2018 International Conference on Service Systems and Service Management*. IEEE, 2018: 1–6.
- [265] Agarwal S, Chen V Y S, Zhang W. The information value of credit rating action reports: A textual analysis[J]. *Management Science*, 2016, 62(8): 2218–2240.
- [266] Mai F, Tian S, Lee C, et al. Deep learning models for bankruptcy prediction using textual disclosures[J]. *European Journal of Operational Research*, 2019, 274(2): 743–758.
- [267] Lin M, Prabhala N R, Viswanathan S. Judging borrowers by the company they keep: Friendship networks and information asymmetry in online peer-to-peer lending[J]. *Management Science*, 2013, 59(1): 17–35.
- [268] Wei Y, Yildirim P, Van Den Bulte C, et al. Credit scoring with social network data[J]. *Marketing Science*, 2015, 35(2): 234–258.
- [269] Oskarsdottir M, Bravo C, Sarraute C, et al. The value of big data for credit scoring: Enhancing financial inclusion using mobile phone data and social network analytics[J]. *Applied Soft Computing*, 2019, 74: 26–39.
- [270] Berg T, Burg V, Gombovic A, et al. On the rise of fintechs: Credit scoring using digital footprints[J]. *Review of Financial Studies*, 2020, 33(7): 2845–2897.
- [271] Baesens B, Setiono R, Mues C, et al. Using neural network rule extraction and decision tables for credit-risk evaluation[J]. *Management Science*, 2003, 49(3): 312–329.

- [272] Martens D, Baesens B, Van Gestel T, et al. Comprehensible credit scoring models using rule extraction from support vector machines[J]. *European Journal of Operational Research*, 2007, 183(3): 1466–1476.
- [273] He J, Zhang Y, Shi Y, et al. Domain-driven classification based on multiple criteria and multiple constraint-level programming for intelligent credit scoring[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2010, 22(6): 826–838.
- [274] Lazer D, Pentland A, Adamic L, et al. Computational social science[J]. *Science*, 2009, 323(5915): 721–723.
- [275] Mann A. Core concept: Computational social science[J]. *Proceedings of the National Academy of Sciences*, 2016, 113(3): 468–470.
- [276] Watts D J. Computational social science: Exciting progress and future directions[J]. *The Bridge on Frontiers of Engineering*, 2013, 43(4): 5–10.
- [277] May R M, Lloyd A L. Infection dynamics on scale-free networks[J]. *Physical Review E*, 2001, 64(6): 066112.
- [278] Newman M E. The structure and function of complex networks[J]. *SIAM Review*, 2003, 45(2): 167–256.
- [279] Li M, Wang X, Gao K, et al. A survey on information diffusion in online social networks: Models and methods[J]. *Information*, 2017, 8(4): 118.
- [280] Golder S A, Macy M W. Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures[J]. *Science*, 2011, 333(6051): 1878–1881.
- [281] Zhao J, Dong L, Wu J, et al. Moodlens: An emoticon-based sentiment analysis system for Chinese tweets[C]// *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2012: 1528–1531.
- [282] Schouten K, Frasinca F. Survey on aspect-level sentiment analysis[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2015, 28(3): 813–830.
- [283] Kosinski M, Stillwell D, Graepel T. Private traits and attributes are predictable from digital records of human behavior[J]. *Proceedings of the National Academy of Sciences*, 2013, 110(15): 5802–5805.
- [284] Mathioudakis M, Koudas N. Twittermonitor: Trend detection over the twitter stream[C]// *Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data*. ACM, 2010: 1155–1158.
- [285] Xie W, Zhu F, Jiang J, et al. Topicsketch: Real-time bursty topic detection from twitter[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2016, 28(8): 2216–2229.
- [286] Wu K, Yang S, Zhu K Q. False rumors detection on sina weibo by propagation structures[C]// *2015 IEEE 31st International Conference on Data Engineering*. IEEE, 2015: 651–662.
- [287] Vosoughi S. Automatic detection and verification of rumors on twitter[D]. *Massachusetts Institute of Technology*, 2015.
- [288] Vosoughi S, Roy D, Aral S. The spread of true and false news online[J]. *Science*, 2018, 359(6380): 1146–1151.
- [289] Zhang C, Liu L, Lei D, et al. Triovevent: Embedding-based online local event detection in geo-tagged tweet streams[C]// *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2017: 595–604.
- [290] Sakaki T, Okazaki M, Matsuo Y. Tweet analysis for real-time event detection and earthquake reporting system development[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2012, 25(4): 919–931.
- [291] Ginsberg J, Mohebbi M H, Patel R S, et al. Detecting influenza epidemics using search engine query data[J]. *Nature*, 2009, 457(7232): 1012–1014.
- [292] Sharpe J D, Hopkins R S, Cook R L, et al. Evaluating google, twitter, and wikipedia as tools for influenza surveillance using Bayesian change point analysis: A comparative analysis[J]. *JMIR Public Health and Surveillance*, 2016, 2(2): e161.
- [293] Wang J, Wang X, Wu J. Inferring metapopulation propagation network for intra-city epidemic control and prevention[C]// *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 2018: 830–838.
- [294] Eagle N, Macy M, Claxton R. Network diversity and economic development[J]. *Science*, 2010, 328(5981): 1029–1031.
- [295] Blumenstock E J. Fighting poverty with data[J]. *Science*, 2016, 353(6301): 753–754.
- [296] Toole J L, Lin Y R, Muehlegger E, et al. Tracking employment shocks using mobile phone data[J]. *Journal of The Royal Society Interface*, 2015, 12(107): 20150185.
- [297] Hu X, Zhao J, Li H, et al. Online footprints of workforce migration and economic implications[J]. *Physica A: Statistical Mechanics and its Applications*, 2019, 528: 121497.
- [298] Blumenstock J, Cadamuro G, On R. Predicting poverty and wealth from mobile phone metadata[J]. *Science*, 2015, 350(6264): 1073–1076.
- [299] Luo S, Morone F, Sarraute C, et al. Inferring personal economic status from social network location[J]. *Nature Communications*, 2017, 8: 15227.

- [300] Berkman L F, Syme S L. Social networks, host resistance, and mortality: A nine-year follow-up study of alameda county residents[J]. *American Journal of Epidemiology*, 1979, 109(2): 186–204.
- [301] Hobbs W R, Burke M, Christakis N A, et al. Online social integration is associated with reduced mortality risk[J]. *Proceedings of the National Academy of Sciences*, 2016, 113(46): 12980–12984.
- [302] Issenberg S. How president Obama’s campaign used big data to rally individual voters[J]. *Technology Review*, 2013, 116(1): 38–49.
- [303] Nickerson D W, Rogers T. Political campaigns and big data[J]. *Journal of Economic Perspectives*, 2014, 28(2): 51–74.
- [304] Xie Z, Liu G, Wu J, et al. Big data would not lie: Prediction of the 2016 Taiwan election via online heterogeneous information[J]. *EPJ Data Science*, 2018, 7(1): 32.
- [305] Burnap P, Gibson R, Sloan L, et al. 140 characters to victory? Using twitter to predict the UK 2015 general election[J]. *Electoral Studies*, 2016, 41: 230–233.
- [306] Grassegger H, Mikael K. The data that turned the world upside down[EB/OL]. https://www.vice.com/en_us/article/mg9vvn/how-our-likes-helped-trump-win.
- [307] Leetaru K. Culturomics 2.0: Forecasting large-scale human behavior using global news media tone in time and space[J]. *First Monday*, 2011, 16(9). <https://doi.org/10.5210/fm.v16i9.3663>
- [308] Chadefaux T. Early warning signals for war in the news[J]. *Journal of Peace Research*, 2014, 51(1): 5–18.
- [309] Wang J, Wu N, Zhao W X, et al. Empowering A* search algorithms with neural networks for personalized route recommendation[C]// *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 2019: 539–547.
- [310] Wu N, Wang J, Zhao W X, et al. Learning to effectively estimate the travel time for fastest route recommendation[C]// *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. ACM, 2019: 1923–1932.
- [311] Wang D, Zhang J, Cao W, et al. When will you arrive? Estimating travel time based on deep neural networks[C]// *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.
- [312] Feng J, Li Y, Zhang C, et al. Deepmove: Predicting human mobility with attentional recurrent networks[C]// *Proceedings of the 2018 World Wide Web Conference*, 2018: 1459–1468.
- [313] Li Y, Shahabi C. A brief overview of machine learning methods for short-term traffic forecasting and future directions[J]. *SIGSPATIAL Special*, 2018, 10(1): 3–9.
- [314] Wang J, Mao Y, Li J, et al. Predictability of road traffic and congestion in urban areas[J]. *PLoS One*, 2015, 10(4): e0121825.
- [315] Fu R, Zhang Z, Li L. Using LSTM and GRU neural network methods for traffic flow prediction[C]// *2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC)*. IEEE, 2016: 324–328.
- [316] Wang J, Gu Q, Wu J, et al. Traffic speed prediction and congestion source exploration: A deep learning method[C]// *2016 IEEE 16th International Conference on Data Mining (ICDM)*. IEEE, 2016: 499–508.
- [317] Geng X, Li Y, Wang L, et al. Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting[C]// *2019 AAAI Conference on Artificial Intelligence (AAAI 19)*, 2019.
- [318] Yao H, Wu F, Ke J, et al. Deep multi-view spatial-temporal network for taxi demand prediction[C]// *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.
- [319] Yao H, Tang X, Wei H, et al. Revisiting spatial-temporal similarity: A deep learning framework for traffic prediction[C]// *AAAI Conference on Artificial Intelligence*, 2019.
- [320] Ye J, Sun L, Du B, et al. Co-prediction of multiple transportation demands based on deep spatio-temporal neural network[C]// *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 2019: 305–313.
- [321] Guo S, Chen C, Wang J, et al. Rod-revenue: Seeking strategies analysis and revenue prediction in ride-on-demand service using multi-source urban data[J]. *IEEE Transactions on Mobile Computing*, 2019.
- [322] Guo S, Chen C, Wang J, et al. Fine-grained dynamic price prediction in ride-on-demand services: Models and evaluations[J]. *Mobile Networks and Applications*, 2019: 1–16.
- [323] Ji S, Zheng Y, Wang W, et al. Real-time ambulance redeployment: A data-driven approach[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2019.
- [324] Kool W, Van Hoof H, Welling M. Attention, learn to solve routing problems![C/OL]// *International Conference on Learning Representations*, 2019. <https://openreview.net/forum?id=ByxBFsRqYm>.
- [325] 郑宇. 城市计算概述 [J]. *武汉大学学报 (信息科学版)*, 2015, 40(1): 1–13.
Zheng Y. Introduction to urban computing[J]. *Geomatics and Information Science of Wuhan University*, 2015, 40(1): 1–13.
- [326] Wang J, He X, Wang Z, et al. CD-CNN: A partially supervised cross-domain deep learning model for urban resident recognition[C]// *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

- [327] Wang J, Wu N, Lu X, et al. Deep trajectory recovery with fine-grained calibration using Kalman filter[J]. IEEE Transactions on Knowledge and Data Engineering, 2019. doi: 10.1109/TKDE.2019.2940950.
- [328] Du B, Liu C, Zhou W, et al. Catch me if you can: Detecting pickpocket suspects from large-scale transit records[C]// Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2016: 87–96.
- [329] Kim S, Lee J G. Utilizing in-store sensors for revisit prediction[C]// 2018 IEEE International Conference on Data Mining. IEEE, 2018: 217–226.
- [330] Lin H, Liu G, Li F, et al. Where to go? Predicting next location in IOT environment[J]. Frontiers of Computer Science, 2019. doi: 10.1007/s11704-019-9118-9.
- [331] Liu Y, Sui Z, Kang C, et al. Uncovering patterns of inter-urban trip and spatial interaction from social media check-in data[J]. PloS One, 2014, 9(1): e86026.
- [332] Liu S, Liu Y, Ni L M, et al. Towards mobility-based clustering[C]// Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2010: 919–928.
- [333] Yuan J, Zheng Y, Xie X. Discovering regions of different functions in a city using human mobility and pois[C]// Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2012: 186–194.
- [334] Pang L X, Chawla S, Liu W, et al. On detection of emerging anomalous traffic patterns using GPS data[J]. Data & Knowledge Engineering, 2013, 87: 357–373.
- [335] Fan Z, Song X, Shibasaki R. Cityspectrum: A non-negative tensor factorization approach[C]// Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing. ACM, 2014: 213–223.
- [336] Wang J, Gao F, Cui P, et al. Discovering urban spatio-temporal structure from time-evolving traffic networks[C]// Asia-Pacific Web Conference. Springer, 2014: 93–104.
- [337] Wang Y, Zheng Y, Xue Y. Travel time estimation of a path using sparse trajectories[C]// Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2014: 25–34.
- [338] Zheng Y, Liu F, Hsieh H P. U-air: When urban air quality inference meets big data[C]// Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2013: 1436–1444.
- [339] Wang J, Chen C, Wu J, et al. No longer sleeping with a bomb: A duet system for protecting urban safety from dangerous goods[C]// Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2017: 1673–1681.
- [340] Lian D, Ge Y, Zhang F, et al. Scalable content-aware collaborative filtering for location recommendation[J]. IEEE Transactions on Knowledge and Data Engineering, 2018, 30(6): 1122–1135.
- [341] Wu R, Luo G, Shao J, et al. Location prediction on trajectory data: A review[J]. Big Data Mining and Analytics, 2018, 1(2): 108–127.
- [342] 方文凤, 周朝荣, 孙三山. 移动群智感知中任务分配的研究 [J]. 计算机应用研究, 2018, 35(11): 3206–3212.
Fang W F, Zhou C R, Sun S S. Research on task assignment for mobile crowd sensing[J]. Application Research of Computers, 2018, 35(11): 3206–3212.
- [343] Xiao M, Wu J, Huang L, et al. Multi-task assignment for crowdsensing in mobile social networks[C]// 2015 IEEE Conference on Computer Communications (INFOCOM). IEEE, 2015: 2227–2235.
- [344] Xiao M, Wu J, Huang L, et al. Online task assignment for crowdsensing in predictable mobile social networks[J]. IEEE Transactions on Mobile Computing, 2016, 16(8): 2306–2320.
- [345] Yang F, Lu J L, Zhu Y, et al. Heterogeneous task allocation in participatory sensing[C]// 2015 IEEE Global Communications Conference (GLOBECOM). IEEE, 2015: 1–6.
- [346] Wang Z, Huang D, Wu H, et al. Qos-constrained sensing task assignment for mobile crowd sensing[C]// 2014 IEEE Global Communications Conference. IEEE, 2014: 311–316.
- [347] Nazari M, Oroojlooy A, Snyder L, et al. Reinforcement learning for solving the vehicle routing problem[C]// Advances in Neural Information Processing Systems, 2018: 9839–9849.
- [348] 吴垚, 曾菊儒, 彭辉, 等. 群智感知激励机制研究综述 [J]. 软件学报, 2016, 27(8): 2025–2047.
Wu Y, Zeng J R, Peng H, et al. Survey on incentive mechanisms for crowd sensing[J]. Journal of Software, 2016, 27(8): 2025–2047.
- [349] Reddy S, Estrin D, Hansen M, et al. Examining micro-payments for participatory sensing data collections[C]// Proceedings of the 12th ACM International Conference on Ubiquitous Computing, 2010: 33–36.
- [350] Wen Y, Shi J, Zhang Q, et al. Quality-driven auction-based incentive mechanism for mobile crowd sensing[J]. IEEE Transactions on Vehicular Technology, 2014, 64(9): 4203–4214.
- [351] Kawajiri R, Shimosaka M, Kashima H. Steered crowdsensing: Incentive design towards quality-oriented place-centric crowdsensing[C]// Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, 2014: 691–701.
- [352] Zhou J, Cao Z, Dong X, et al. Security and privacy in cloud-assisted wireless wearable communications: Challenges, solutions, and future directions[J]. IEEE Wireless Communications, 2015, 22(2): 136–144.