

学科交叉与范式革新：迈向国际关系研究的智能分析时代*

董青岭 刘文龙

内容提要：在新文科交叉研究不断发展的背景下，国际关系研究迈向智能化分析的趋势初露端倪。放眼未来，随着社会生活数字化水平的提升，国际关系研究将越来越依赖于高密度、多模态数据的支撑和智能化分析手段的介入，国际关系研究的范式变革也越来越转向社会计算理论、人工智能技术与政治理论的融合。推动国际关系研究迈向智能化分析，首先要着力解决数据感知没有数据的问题，其次要解决智能计算算法和算力不可及的问题，全力推进国际关系研究转型为数据密集型科学研究范式的数据底座建设和算法底座建设，基于数据驱动和基于算法驱动重构国际关系研究的议题和议程。就此而言，未来国际关系研究将朝着以数据和算法为核心的学术创新演进；而在此过程中，只有解决大多数而不是极少数研究人员的数据获取和计算难题，国际关系研究才能真正步入智能化时代。

关键词：国际关系 范式转换 智能分析 数据底座 计算平台

引言

当前，伴随着社会生活数字化进程的推进和人工智能技术的崛起，国际关系学科正迎来新的范式转型机遇。放眼未来，国际关系研究的一个范式革新方向很可能是社会计算理论、人工智能技术与政治理论的融合，即国际关系研究的智能化。有关这一趋势，主要判断依据如下：第一，以内容生成为代表的“算法政治”^①的来临要求“以智

* 本文系国家社会科学基金重大项目“基于大数据的国家安全态势感知原理及冲突数据库建设”（项目编号：19ZDA131）的阶段性成果。感谢匿名审稿专家的宝贵修改建议，文责自负。

^① See Scott Timcke, *Algorithms and the End of Politics: How Technology Shapes 21st-Century American Life*, Bristol University Press, 2021, p.41.

能对抗智能”。传统国际关系研究主要着力于“真实数据”的剖析和“模拟数据”的推演,其中“真实数据”指的是由人工调查、人工统计和人工记录所形成的事实描述与关系揭示,如抽样民调和官方统计;而“模拟数据”则是指基于虚拟仿真和形式模型所产生的理想数据和假设数据,如模拟投票与谢林模型^①。但无论是真实数据还是模拟数据,在传统国际关系研究中,人是数据生产的主力军,其产出数据的类型相对单一,对数据处理技术的要求也不高。然而,自进入内容生成时代以来,以 ChatGPT 和 Deep-fakes 为代表的生成型数据已大量充斥网络,“机器”或称“算法智能体”正悄然崛起为新的数据生产者和信息传播者。与传统国际关系研究所要处理的数据类型不同,生成型数据主要由算法控制所产生,不仅产生速度快、体量也可无限大,而且并非肉眼清晰可辨,需要智能化手段的介入方能侦测和识别。

第二,非结构化数据处理能力的提升将助推国际关系研究走向“全息计算”^②。以前,囿于研究手段和技术工具的限制,传统国际关系研究的主要分析对象是结构化的统计数据(如截面数据和面板数据^③)和少量的非结构化文本数据(如官方文件和领袖言论)。诸如声音、图像和视频等数据类型基本被排除在大部分学者的分析视野之外,即便有分析,也主要集中在那些掌握先进技术并拥有高性能设备的分析机构(如互联网公司、军方、大型实验室和情报机构)手中。然而,由于先进的数据采集、存储和分析手段的介入,诸如定时爬虫、自动摘要和自动翻译技术以及向量数据库^④的使用,国际关系研究的分析对象已不再局限于传统的数据类型和数据体量。非结构化数据除传统的文本数据之外,声音、图像和视频等具有流数据性质的研究素材也逐步进入社会计算范畴,国际关系研究的可分析数据类型正在从单一模态向多模态和跨模态转变,未来国际关系研究有望踏入“全息计算”时代。

第三,新兴研究工具的应用将改变国际关系研究的数据处理模式和结果展示形

^① “谢林模型”由美国经济学家托马斯·谢林于1971年提出,主要描述同质性在空间隔离上的影响和作用,是一个用于分析城市人口分布这一复杂系统的经典的基于主体的模型(Agent Based Model, ABM)。它的提出标志着主体建模仿真(Agent Based Modeling and Simulation, ABMS)方法从理论走向应用,即通过利用计算机仿真技术分析系统在不同政策下的演变过程,是提炼理论解释现象、针对情景制定政策的有力工具。

^② “全息计算”,指的是在数字环境下,国际关系研究借助大数据和人工智能技术的支撑,可进行文本、视频、声音和图像等多模态和跨模态数据的综合分析,从而实现对所研究问题的多角度数据计算和态势评估。

^③ 截面数据是指在同一时间节点上或同一段时间内所收集的数据,描述多个观测对象在相同一段时间内或相同时间节点上的表现;面板数据是指对不同观测对象在不同时间段或时点上所收集的数据,描述多个观测对象随着时间变化而变化的情况。

^④ 向量数据库是一种专门用于存储和查询向量数据的数据库系统,与传统数据库相比,向量数据库使用向量化计算,能够高速地处理大规模的复杂数据;并可以处理高维数据,例如图像、音频和视频等,解决传统关系型数据库中的“痛点”问题;同时,向量数据库支持复杂的查询操作,也可以轻松地扩展到多个节点,以处理大规模的数据。

式。首先,在数据获取领域,各种智能传感器和内容生成工具的嵌入式应用正在取代传统的数据采集范式,各种仿真模拟数据和细分领域垂直数据库的建立正在奠定国际关系学迈向数据密集型科学研究的基础;其次,在数据分析领域,以 Python 和 R 为代表的面向对象的编程语言正在取代以 C 语言为代表的面向过程的编程语言,这无疑增加了算法模型的复用性和可调用性,面向对象编程意味着国际关系研究即将迈入算法应用加速知识生产的时代;再次,在结果展示层面,各种即插即用可视化分析软件与数据仪表盘的出现,正在改变传统研究的数据透视方式、数据理解方式和数据展示形式,数据与图形的交互正在成为国际关系研究的新风潮。借助智能化分析手段,基于数据驱动的知识发现和政治洞察正在成为当下国际关系研究的一个重要范式转型趋向,新一代国际关系研究将会严重依赖计算社会科学、政治理论和其他学科的交叉融合,智能分析技术的介入将为纷繁芜杂社会问题的解决提供智慧化解决方案,基于智能技术的即时研判无疑会更加适应瞬息万变国际环境的态势感知。

总体而言,一个新的历史时代正在开启、一种新的研究范式正在加速浮现,国际关系研究的未来很可能是转向以数据和算法为核心的知识创新。但在智能化国际关系研究落地前,有两个关键问题亟待解决:其一,数据的即时获取与稳定供应,如果智能化国际关系研究的核心特征是数据密集型科学研究,那么需要解决基于数据的国际关系研究没有数据可计算的问题;其二,数据的高效计算与模型复用,即需要通过智能软件和计算模块的嵌入,解决数字环境下国际关系研究超大体量数值计算和多模态数据处理问题。目前,计算社会科学与国际关系学的交叉研究刚刚起步,未来国际关系研究的智能化图景虽然已隐约可见,但未来之路仍然面临重要挑战。

一 范式转换:数字化进程与国际关系研究的可计算性

当前,数字化进程已被认为是社会科学革命的崭新驱动力量。早在 2007 年,图灵奖得主吉姆·格雷就指出:“科学正在进入一个崭新的阶段”。“由于信息技术的影响,关于科学的一切几乎都在变化中。实验、理论和计算科学范式都正在被数据泛滥和正在出现的数据密集型科学范式——第四范式所影响。”^①这是一场继实验归纳、模型推演和仿真模拟之后的“科学方法的革命”。放眼未来,基于数据思考、设计和实施

^① [英]托尼·海伊、[美]斯图尔特·唐斯雷、[美]克里斯汀·托尔:《第四范式:数据密集型科学发现》,潘教峰、张晓林等译,科学出版社 2012 年版,第 xxiii 页。

科学研究已是新范式潮流所趋。事实上,伴随着社会生活各领域数字基础设施的完善,数字化进程已然对当前国际关系研究的核心议题和主导范式构成巨大冲击,新出现的国际关系研究范式转换趋向突出表现为三个方面。

第一,新数据、新技术与新的知识增长点。概而言之,传统国际关系研究所触及的研究素材和数据体量有限,以经验归纳和案例实证为主要分析路径,其中经验归纳重在从历史案例和政治互动中总结规律和启示,而案例实证则侧重于从逻辑推演的视角提出命题假说并给予统计说明和案例佐证。总体上,传统研究范式高度吻合吉姆·格雷所说的经验证据(“第一范式”)、理论推演(“第二范式”)和仿真计算(“第三范式”)。^①以上三种范式类型皆适用于数据稀疏时代的政治观察,核心特征是有限数据观察与低维数值计算,基本可满足研究人员在有限信息环境下对于主要矛盾的捕捉和决策时间上的非即时反应需求。

然而,伴随着信息超载时代的来临,传统国际关系研究范式显然已无法实现信息丰裕环境下的快速决策与有效决策。在此背景下,新的信息处理技术、新的政治分析工具和新的研究路径的引入,正在推动当下国际关系研究向以数据和算法为驱动的范式演变。纵览文献,社会计算与国际关系的融合之作俯拾皆是,择其要者,列举如下:(1)基于数据挖掘的公共外交战略。2011年,美国总统奥巴马出访巴西,其外交团队通过挖掘社交网络数据实现了公共外交全过程的可预测、可监控与可评估。^②(2)基于数据感知的国家安全战略。借助实时数据监测和各种算法模型的嵌入式分析,无论是突发事件、社会舆情、网络攻击抑或是军事冲突等安全议题,都越来越可做到提前预警与预防。^③此外,数字技术的滥用和误用所引发的数字安全问题,如推荐算法、社交机器人和深度伪造等,也越来越成为全球数字治理与数字外交的核心议题。^④概言之,在数字化浪潮的涌动下,未来国际关系研究将瞄向新的议题,使用新的工具开启新

① [英]托尼·海伊、[美]斯图尔特·唐斯雷、[美]克里斯汀·托尔:《第四范式:数据密集型科学发现》。

② 董青岭:《大数据外交:一场即将到来的外交革命?》,载《欧洲研究》,2015年第2期,第137页。

③ 高望来:《北约人工智能反恐新常态及其困局》,载《欧洲研究》,2021年第2期,第135-154页;刘乐:《社会网络与“伊斯兰国”的战略动员》,载《外交评论(外交学院学报)》,2016年第2期,第82-109页;谢许潭:《国际反恐新战场:应对“伊斯兰国”媒体宣传的挑战》,载《外交评论(外交学院学报)》,2016年第1期,第82-103页。

④ 高奇琦、张莹文:《主体弥散化与主体责任的终结:ChatGPT对全球安全实践的影响》,载《国际安全研究》,2023年第3期,第3-27页;蔡翠红、张若扬:《“技术主权”和“数字主权”话语下的欧盟数字化转型战略》,载《国际政治研究》,2022年第1期,第9-36页;刘国柱:《深度伪造与国家安全:基于总体国家安全观的视角》,载《国际安全研究》,2022年第3期,第3-31页;张涛、余丽:《算法在国际政治中的“非中性”作用》,载《国际论坛》,2022年第5期,第41-57页;张煌、杜雁芸:《全球自主武器系统的发展及其对战略稳定性的影响》,载《国际论坛》,2021年第3期,第79-96页;董青岭:《新战争伦理:规范和约束致命性自主武器系统》,载《国际观察》,2018年第4期,第51-66页。

的知识发现。

表 1 国际关系研究:传统范式 VS 第四范式

	传统范式研究	“第四范式”研究
研究范式	实验观察、理论推导、仿真模拟	数据密集型科学研究
交叉模式	理论的移植与融合(如经济学、社会学、心理学和民族学理论被引入国际关系分析)	场景的方法应用(如社会网络分析、情感分析和知识图谱等算法被用于武装冲突预测和反恐怖等场景)
研究素材	以小体量、稀疏型结构化数据为主,数据普遍具有静态、非即时性	以大体量、高密度多模态数据为主,结构化与非结构化数据混合,数据普遍具有动态流式特征,需要做出即时反应
数据收集	以人工整理和关系型数据库为主,数据模态单一、体量不大,田野调查、抽样统计和面对面访谈是常见的数据采集形式	以网络抓取、传感器收集和向量数据库为主,数据形式复杂、清洗难度高,定时爬虫、自动摘要和自动标注是常见的数据整理形式
数据管理	以本地化和集中式存储为主	以云存储和分布式存储为主
研究工具	面向过程编程(如 C 语言)	面向对象编程(如 R 语言、Python 等)
研究方法	传统计量与统计分析	神经网络与深度学习
算力需求	单机电脑、有限量服务器	算力集群(超算中心或分布式计算)

注:表由作者自制。

第二,新思维、新视角与新的研究旨趣。由于数字化进程便利了全球交往、增强了世界政治的联动性,国际关系现象的复杂性与非线性特征正在被重新审视。当然,一并考量的还有目前盛行的一些主流研究方法和研究路径,首当其冲的便是传统定量分析。国际关系领域中的传统定量分析主要以变量控制和因果回归为主导特征,其在某种意义上预设国际关系世界如同自然界一样具有可控实验特征,重在从条件假设出发透过变量控制寻找社会因果关系,有限变量间的线性比较是其核心的解题思路。然而,现实中的国际关系世界具有复杂系统特征,牵一发而动全身,各种变量在开放体系下不断演进且相互扰动,导致以线性思维为主要解题思路的传统研究路径越来越难以有效应对政治进程的复杂性。在此背景下,社会计算的崛起与引入有可能会给当下的

国际关系研究带来不一样的问题探索路径。譬如,2017年《科学》杂志曾专门组织了一期题为“预测及其局限”的特刊,撰稿者集中讨论了社会计算分析各模型在冲突预测、社会危机管理和资源调配中的应用优势及其局限。^①通过借鉴社会计算的理论、工具与模型,当下的国际关系研究能更好、更迅速地捕捉世界政治的复杂性,从而实现自我修正、技术革新与理论进步。

第三,新交叉、新场景与新的解题思路。在数智时代,国际关系研究的一个重要发展趋向是越来越朝向工程性应用学科演进。首先,伴随着全球互联互通一体化进程的纵深推进,越来越多的国际问题需要跨学科知识的整合才能加以应对,跨学科交叉研究已是国际关系学的典型特征。譬如,全球气候问题的解决,不仅需要气象学家、天文学家、农林学家和地理学家的参与,更需要统计学家、经济学家和政治学者的协同智慧。其次,面向新的技术环境和新的知识融入,以前无法涉足的研究问题正在被新的研究手段和新的研究路径所触碰。例如社会舆情的实时监测、政治人物面部微表情的识别以及文本、声音和视频等跨模态数据的整合分析,都已成为当下国际关系研究的前沿话题。再次,为适应信息超载时代的快速决策需求,集合专家智慧与智能化分析手段的计算平台建设,也越来越成为智能化国际关系研究的标准基础设施配置。基于数据中心或超级计算平台,国际关系学者对于纷繁复杂国际问题的反应将越来越迅捷,也将越来越具有俯瞰全局、深入细节的能力。就此而言,国际关系研究的数智时代已然来临,借助跨学科知识的交融和多元方法的交叉,国际关系研究走向智能化与半智能化分析已初见轮廓。例如,有学者通过调取“全球事件、语言与语调数据库(GDELT)”中的约400万条数据,综合运用向量自回归模型、脉冲响应函数等多元方法,系统性地探讨了影响中美互动的多重因素。^②此外,还有学者借助地理信息系统、夜间灯光和事件数据,运用时空数据模型对非洲的政治暴力冲突、中国对非援助绩效

^① Barbara R. Jasny and Richard Stone, “Prediction and Its Limits,” *Science*, Vol.355, 2017, pp.468-469, <https://www.science.org/doi/10.1126/science.355.6324.468>; Lars-Erik Cederman and Nils B. Weidmann, “Predicting Armed Conflict: Time to Adjust Our Expectations?” *Science*, Vol.355, 2017, pp.474-476, <https://www.science.org/doi/10.1126/science.aal4483>; Philip E. Tetlock, et al., “Bringing Probability Judgments into Policy Debates via Forecasting Tournaments,” *Science*, Vol.355, 2017, pp.481-483, <https://www.science.org/doi/10.1126/science.aal3147>; Susan Athey, “Beyond Prediction: Using Big Data for Policy Problems,” *Science*, Vol.355, 2017, pp.483-485.

^② 庞珣、刘子夜:《基于海量事件数据的中美关系分析——对等反应、政策惯性及第三方因素》,载《世界经济与政治》,2019年第5期,第53-79页;曹玮:《中美战略竞争下的战略第三方选择:基于海量事件数据的中美印、中美俄、中美欧VAR模型分析》,载《当代亚太》,2021年第5期,第39-74页。

等问题进行了不同于以往研究路径的探索性分析。^①

综上所述,伴随着数据科学技术的进步和数字基础设施的完善,当下国际关系研究正在迎来新旧范式的转换。旧范式主要以实验归纳、理论推演和有限数据的仿真模拟为主要解题思路,重在粗线条的捕获和理解研究对象的核心特征;而新范式则侧重“数据密集型科学研究”和“智能化分析手段的介入”,重在解决数字时代的信息超载与敏捷反应问题。

二 准入门槛:国际关系研究迈向智能分析的核心障碍

每一个时代的科学研究,有每一个时代的科学使命和亟待解决的关键问题。数智时代的国际关系研究,整体趋势是迈向“第四范式”,尽管这一概念已被提出很久,但各个学科包括国际关系学并未真正进入这一范式。简单来说,国际关系研究的“第四范式”主要体现为“数据密集型科学研究”和“智能化分析手段的介入”两个关键特征,即以“高密度数据”和“多模态数据”为分析对象,旨在通过多源、多模态、大体量数据的挖掘、能够建立起基于数据驱动的政治洞察和态势感知;在研究策略上以“深度学习”为工具选择、以“计算平台”为算力支撑,目的在于通过算法模型的部署与应用,实现由算法驱动的智慧决策和敏捷反应。就此而言,国际关系研究要想真正迈向与计算社会科学的融合,“数据的稳定供应”和“高性能计算”是两个亟待解决的关键问题。

第一,数据的有效获取和稳定供应难题。高密度数据和即时性数据是智能化国际关系研究的数据基石,但目前却因软硬件问题而难以实现有效供应。原因在于:其一,数据采集程序的编写和维护难度较高。由于网站结构和网页设计千差万别,任何一个数据采集程序的编写都不能一次性解决各种网页数据的抓取任务,并且有些网页在设计之初还设定有蜘蛛协议和反爬机制,这就导致网络数据采集程序的编写只能在协议规定的框架下面向特定网页逐个建设,然后通过爬虫调度器和管理器来运维爬虫队列,继而成为一项极为复杂且规模化的工作任务。另外,由于有些网站的底层设计会不定时更新,这也导致某一网页数据采集程序的设计和应用不能一劳永逸,需要专业

^① 陈冲:《机会、贪婪、怨恨与国内冲突的再思考——基于时空模型对非洲政治暴力的分析》,载《世界经济与政治》,2018年第8期,第94-127页;黄振乾:《中国援助分配的政治经济学——对21世纪中国援非项目的空间考察》,载《世界经济与政治》,2021年第9期,第102-127页;黄振乾:《中国援助与受援国绩效合法性——基于地理信息数据的实证考察》,载《世界经济与政治》,2022年第3期,第30-58页。

队伍维护并进行实时更新。其二,数据采集的软硬件环境搭建成本高、投入大。除了高性能的传感设备,要想高效获取和存储数据,研究团队通常还需要有便于使用的处理器和存储设备。也许对于传统国际关系分析而言,一台单机电脑就足以满足各种研究的分析需求;而对于智能化国际关系研究而言,高性能处理器、大容量存储器以及高效数据传输设备等都是必不可缺的数字设施。要想让这些设备组合在一起并保持良好运行,软硬件环境的搭建不仅建设难度大,同时建设费用也极为高昂,这在一定程度上成为多数国际关系学者迈向智能化分析的障碍。

第二,算法的研发部署与算力支撑问题。一方面,在目前的学术训练体系下,既懂算法又懂政治分析的交叉人才奇缺。经历过数理统计、计算机科学和政治理论的交叉训练的研究人员本来就凤毛麟角,能够熟练使用编程手段进行大体量数据分析和复杂建模的人员更是少之又少。在此情形下,国际关系研究要想迈向智能化分析,第一阶段应是着力于解决大多数研究人员的“技术准入门槛问题”,即如何通过一项零代码解决方案将大部分国际关系学者普遍带入智能化分析场景。目前来看,模型预训练并使之界面化集成封装,可能是一项立竿见影的务实选择。在这一方案中,算法模型由专业技术人员来开发和训练,零基础编程知识的国际关系研究人员主要是通过界面选择而不是代码编程来实现算法调用。据此,智能化国际关系研究便可在一定程度上解决绝大部分研究人员不会编程、编程难的问题。当然,这一方案也有缺陷,它虽然可以解决无编程能力的学者进入智能化国际关系分析的“门槛准入”问题,但对于一些具有复杂性和创新性的研究任务仅通过调用通用模型是不够的,复杂任务的解决仍然需要高级编程知识和政治理论的交叉融合。

另一方面,算力平台建设也是国际关系研究迈向智能化分析不得不考虑的另一基石。这主要体现为:其一,智能化国际关系研究所要处理的数据类型不仅多样化,而且体量巨大,既涵括诸如文本、声音、图像和视频等非结构化数据,同时也包含诸如面板和截面等结构化数据,需采用更加复杂的算法模型组合进行综合运算,其所需算力已远超传统计算机软件的处理能力;其二,算力平台建设主要体现为软、硬件环境的一体化,不仅开发投入和运行维护成本高,其资源消耗和人员开支也非常大,非单个国际关系学者或学术团队所能支撑。就此而言,智能化的国际关系研究,其算力基础设施最好是分布式、轻量级、轻资产部署,重在突出将成熟的社会计算理论与技术引入国际问题解决,而不是像自然科学研究那样重在筹建重量级、重资产和集中式部署的超大计

算设备。就此而言,面向社会问题解决的国际关系研究算力基础设施建设应与面向科学发现的新工科算力基础设施建设有所不同。

三 解题思路:“数据资源的可得”与“计算资源的可及”

放眼未来,国际关系的发生场域将越来越深植于数智环境,国际关系研究迈向智能计算越来越具有可实现性,但也需满足如下前提条件:一要有可供计算分析的数据;二要有高效处理数据的计算平台。就此而言,面向新的技术环境和新一轮的学科交叉,谁拥有丰裕的高质量数据,谁就可以基于数据领衔知识生产,进而主导研究议题和研究议程;谁率先建立起高精尖的算法平台,谁就可以基于算法支撑实现智能计算,从而在决策上先人一步、掌握制胜先机。然而,考虑到人员、知识、技术与资金分布的不均匀性,以智能计算作为国际关系研究新的知识赛道也必将加大在各个学者和学术团队之间业已存在的“数字鸿沟”。不过,“不进者则退”,要想在激烈的学术竞争中领先,智能化国际关系研究需要着力于强有力的数字基础设施建设,其中“数据资源的可得”与“计算资源的可及”是问题解决的关键。

第一,数据的可得性问题。与传统国际关系研究不同,智能化国际关系研究旨在通过智能化分析手段的嵌入式应用辅助科学决策、推进资源优化配置和提升政策执行效果,亦即外交政策制定与执行过程的智慧化,其典型特征突出表现为面向特定政治场景的一整套算法模型的研发与应用,譬如针对海外舆情的持续性跟踪和监测(典型案例如斯诺登事件披露的美国“棱镜”计划)、针对政治选举的建模分析与预测(典型案例如“538网站”与“剑桥分析”)以及针对国际民心争夺的舆论战与信息战(如以社交机器人和深度伪造为核心应用的在线虚假宣传与政治“污名化”)等。因此,推进国际关系研究朝向智能化发展所需的数据供应通常不是单次性和间断性的,而是需要尽可能连续、持续且稳定。就此而言,智能化国际关系研究的旨趣更多追求的是社会计算理论在政治场景中的实战应用,其终端产品通常体现为具有工程应用性质的系统建设或政治分析平台。若需维持这个系统或平台的运行,即一套算法的政治应用,则要不断地为该系统和平台投喂数据,此时需筹建一个数据中心或调用某几个垂直数据库方能实现研究任务。但截至目前,在国际关系研究领域内,无论这种大型数据中心的建设还是面向专门问题领域的垂直数据库研发,都深受资金投入和人员运维成本的掣肘,已有的数据中心建设模式主要分为三类。

一是账号购买型。这类数据中心本质上是一种数据库集成平台,其本身并不生产数据内容,而只是将涉及专业领域的相关数据资源购入平台,进行账号汇集并设置使用权限,仅供本单位成员或加盟成员有条件、小范围使用。例如,各个大学或研究机构的图书馆数据中心将购买的数据资源放置在平台上,单位成员可通过IP地址或账号登录的方式获取使用权限。这种数据供应通常可归类为科研单位基础公共设施建设的一部分,科研单位承担了数据资源的购入成本,研究人员本身并无使用数据的经济支出。此类数据供应稳定与否严重依赖于建设单位的经济状况,主动权掌握在数据供应商手中,如果数据供应商不满足于数据交易价格,就有可能抬高数据价格或中断数据供应。

二是协商共建型。这种数据中心建设主要发生在两个或多个单位之间的业务合作或战略协作,目的在于整合合作单位之间的数据优势与研究优势,针对某一具体问题共建共享特定数据资源、协同开展科学研究。此种类型的合作常见于政府与企业、政府与高校以及企业与高校之间的产学研合作。典型案例如复旦大学消费市场大数据实验室,该实验室由商务部、上海市商务委员会、上海市统计局共同授牌,旨在通过个人支付终端的大数据汇总和多源数据、多支付方式(银行卡、微信、支付宝)的计算融合,实现对线上线下、境外境内全维度消费的监测。^①此种形式的政府与高校共建数据,主要是由政府提供数据、高校负责分析并提供相应对策建议。上述类型的共商共建型数据库建设,一方面能够帮助政府充分利用“外脑”、实现研究资源的外部引入;另一方面能够帮助高校获取市场上难以获得的数据资源,依托一线数据进行更加精准的政策研究和更具创新性的科学研究。但是,为保证数据的安全性和保密性,这类数据共享一般仅供内部机构或合作授权单位之间使用,数据在有限可控范围内分享,其他研究人员想要获取相应数据资源则有一定的难度。

三是自主研发型。此类数据中心建设的目标旨在通过自主可控的系统研发彻底解决数据“卡脖子”问题,完全掌握数据采集和数据使用的主导权。建设模式主要分为人工整理与机器采集两种。其中,人工数据采集是目前数据中心建设中成本最高、效率最低的数据供应模式,通常的做法是雇佣人员以数据切块分包的形式摘录形成垂直数据库。该模式的优点是技术门槛低、只要投入足够的人员并进行简单的业务培训

^① 复旦发展研究院:《消费市场大数据实验室(商务部)》, <https://fddi.fudan.edu.cn/xfscdsjsyswswbww/list.htm>。

即可执行,缺点是数据质量不统一、数据采集规模不大、数据时效性低。有鉴于此,考虑到数字环境下的信息过载和数据采集时效,有越来越多的研究团队寻求使用机器手段取代人工数据采集,主流方式是借助网络爬虫和自动语义分析进行信息自动采集、自动摘录、自动编码、自动标注和自动存储,从而在节省大量人力、物力支出的同时又增强了数据供应的稳定性和及时性。此类数据库建设较为典型的案例如“全球事件、语言与语调数据库”(GDELT)。^① 该数据库基于定时爬虫和自动翻译技术能够对全球各国超过100种语言的各种事件和主题信息进行实时监控和捕捉,其中包括新闻报道、危机事件、社交媒体言论等内容,并识别这些内容中的人员、地点、组织、主体、信息来源、新闻情感值和行为特征等信息,将其按照特定的编码规则进行整理分类。^②

总体而言,在有限资金的约束下,国际关系研究智能化要实现数据稳定供应仍然需要探索可行路径,至于要建设大型数据中心还是建设个性化的垂直数据库,是选择账号购买、协商共建还是自主研发,各个国际关系学者和研究团队可因时因地做出自己的选择。除此以外,生成式人工智能的出现也提供了新的数据来源,诸如虚假信息的批量伪造与ChatGPT的自动内容生成,业已成为扰动政治运行的重要数据变量。

第二,计算的可及性问题。随着智慧社会的建设与发展,在国际关系研究中融入智能计算进而探索新的知识发现,已成为当下国际关系研究的重要前沿趋势。但就目前的学术训练体制特别是方法训练而言,将大部分国际关系学者培养成既懂编程知识同时又精于政治分析的交叉研究人才,成本过高且现实可操作性有限。一种简洁有效的思路是:如果短期内大面积培养国际关系学者的数据科学素养技术门槛过高,那么可以将国际关系研究一些常用算法进行界面化封装并集成为统一的计算平台,这样无需技术培训、仅需提供使用指南就可实现算法调用。换言之,在不考虑每一个国关学者都必然是编程高手的前提下,国际关系研究有没有可能通过一些计算工具的便捷化应用开启智能化分析进程呢?就此而言,着力便捷可用的数字基础设施建设是推进国际关系研究实现计算可及性的可探索路径。

其一,建设算法工具箱、打造“算法菜谱”,提升算法模型的可调用性和复用性。一个解题方案是促进算法模型在国际关系研究中的应用低代码化,主导思路是通过面向对象的编程思维来解决代码模块的可调用问题,即面向同一对象任务组织功能模块

^① The GDELT Project, “Querying, Analyzing and Downloading,” <https://www.gdeltproject.org/data.html>.

^② The GDELT Project, “The GDELT Story: Learn about How GDELT Came to Be, Who Made It, and How You Can Use It,” <https://www.gdeltproject.org/about.html>.

以免重复编程,主要用于服务那些已具备编程知识的研究人员更加便捷地调用他人算法,目的在于解决国际关系研究中的算法复刻问题。另一种解题方案是建设具有工具箱性质的零代码化开源计算平台,核心任务是模型预训练和代码模块的界面化封装,以零代码和“傻瓜式”的功能菜单选择来服务那些编程知识零基础的国际关系研究人员,以实现其借助算法模型来推进自身研究的技术愿望。低代码服务于少数业务熟练的国际关系数据精英,而零代码则是要着力解决大部分国际关系研究人员进入智能化分析的“门槛准入”问题。

截止到目前,真正由国际关系学者在本领域理论指导下,建立起来的面向某一问题场景的算法模型极少,绝大部分国际关系学者所应用的算法模型主要是改造、复刻和移植其他领域的通用模型,例如 Github、Kaggle、TensorFlow 和 PyTorch 等^①平台汇集了各个学科领域的算法分享,尽管这些平台的建设并不是专门针对国际关系研究开发的,但平台上的很多算法模型却具有通用性和可移植性,其中不乏与国际关系研究相关的代码参考,内容涉及国际关系研究各个子领域,如恐怖主义侦测、社会风险感知、政治极化分析和双边关系建模评估等。此类平台建设的核心思想是汇集各种算法以供分享、学习和讨论,进而通过提供低代码化的应用部署方案以促进模型在各个应用场景中的可调用。此外,还有平台如 CSDN、稀土掘金、开源中国等还设有相应的问答机制,鼓励不同程序研发者相互交流经验、解答算法模型在应用过程中所遇到的各种困难。显而易见,这些平台的建设对于推进智能化国际关系研究极具参考意义。

其二,强化算力集群建设、打造“算力底座”,增强算力的可及性与可满足性。在智能化国际关系研究中,如果数据中心建设旨在解决研究智能计算的数据供应问题,那么算法工具箱就是要解决素材加工的可选择菜单问题,而算力集群建设是要为工具菜单的执行即算法提供可实现仪器和设备。就此而言,算力是实现一切算法的基础和保障,但考虑到算力基础设施的建设难度与维护成本,目前学界主要有两大算力解决方案。

方案一是构建超级计算中心,核心思想是通过高速网络将数千甚至数万个计算节点连接在一起形成大规模计算集群,这些计算节点配备了高性能的处理器、大容量的内存和高速存储系统,从而依靠连接在一起的强大计算资源实现大规模数据处理和并

^① Github, “Let’s Build from Here: The Complete Developer Platform to Build, Scale, and Deliver Secure Software,” <https://github.com/about>.

行计算。一旦建成,此种超算中心通常可用于模拟天气系统、推演社会系统演化、计算弹道轨迹等超大体量数值运算。当前世界各国都在争相建设超级计算中心,如美国国家超级计算应用中心(NCSA)、德国高性能计算中心(HLRs)、英国国家超级计算中心(NSCC)、中国国家超级计算无锡中心等。^①一言以蔽之,超级计算中心的建设不仅是计算机算力的比拼,更是国家间综合科技实力的竞争,在国际关系研究中的潜在意义不言而喻。但鉴于其昂贵的建设成本和高难度的系统研发,目前对于一些有预算约束的重大国际关系研究任务而言,宜租用算力但不宜自建。

方案二是采用分布式计算系统,核心思想是将一个超级计算任务切成多个小块分配给多个计算机或处理器进行处理,最后把这些计算结果综合起来得到最终结果。作为一种计算方法,分布式计算是相对集中式计算而言的。每个计算机或处理器都可以独立工作,相互之间通过网络进行通信,不仅可以节约整体计算时间,同时还因为增加了计算节点拓展了计算能力,具有成本低、灵活部署和安全计算之特性。在没有大规模集中式计算资源的情况下,它非常适合部署在任务重、资金少且须做出快速反应的国际关系研究领域。目前,流行分布式计算的系统主要有Hadoop、Spark和Flink,其中Hadoop系统适合处理离线的静态数据,Spark适合处理离线的流式数据,而Flink则适合处理在线的实时数据。相较于集中式计算解决方案,考虑到国际关系研究的数据多样性、技术支撑能力以及成本预算约束,分布式计算的性能特征可能更适用于智能化国际关系研究的计算能力建设。此外,基于分布式计算构思,云算力租用也是不错的落地选择。

综上所述,无论是集中式超算中心还是分布式计算系统,当前的国际关系研究如果想真正地迈向智能化分析,计算能力建设不可或缺。但至于选择哪种方案,取决于研究者个人和研究者团队的计算需求和资源支撑能力。大型任务或长期研究可以考虑构建自主可控的计算能力,短期任务或零散计算可以优先选择租用算力。

四 未来议程:智能化国际关系研究的可能图景

展望未来,人工智能技术的应用将会遍布各个领域,国际关系研究走向智能分析已初露端倪。但一旦融入智能分析元素,无论是目前的国际关系理论,还是当下盛行

^① 李顺顺、颜颖:《高质量发展调研行|国家超级计算无锡中心:超算既要“高大上”,又要接地气》,新华网,2023年5月17日,https://www.xhby.net/index/202305/t20230517_7941840.shtml。

的国际关系研究方法,都将为之重塑。随着数字基础设施的建设与完善,特别是“数据中心”建设与“算力平台”的打造,国际关系研究的智能化水平将会得到稳步提高,届时国际关系研究的社会工程应用学科属性将更加凸显,理论与算法的融合将促进国际关系研究更接地气、更以解决问题为导向。就此而言,智能化国际关系研究的终极目标是使用先进的计算手段为现实国际问题的解决找寻智慧化解决方案,它不同于计算机领域中的纯算法研发与纯性能改善,也不同于传统国际关系研究中的简单计量统计与数学建模,其本质是国际关系研究的软硬件一体化,寻求的是通过平台建设和系统研发促进某一特定问题的持续性解决,其交叉融合的落脚点在于将政治理论融于一套技术上可执行的智慧解决方案。总体而言,在智能化浪潮的驱动下,未来国际关系研究将会产生新的知识增长点并形成新的理论高地。

首先,数字基础设施建设将成为国际关系研究知识竞争的新赛道。工欲善其事,必先利其器。为了抢占下一波国际关系研究知识创新的先发优势,未来各个研究团队和专家学者可能会不惜重金投入数字基础设施建设。在迈向智能化分析的道路上,无论是掌握数据还是掌握算法,都将意味着拥有知识增长的核心资源,进而以数据和算法为依托加速人才培养、提升研究质量、扩大社会服务范围。目前来看,前期将围绕两个问题展开最为关键的竞争。

一是重构数据底座、稳定数据供应,强化阵地建设。构建新型数据采集系统、建立自主可控的数据采集能力,是国际关系研究迈向智能化分析的数据基石。面向未来的数字环境,新一代数据采集系统构建应着力应对数据洪流与信息超载,在功能设计上应能透过数据的涌动感知世界风云的瞬时变化,并基于数据关系的挖掘提供敏捷且精准的政治洞察。就此而言,新一代数据系统建设应至少满足以下任务的解决:(1)收集并存储结构化和非结构化等多种数据形式,面对同一任务能够调度多模态和跨模态数据,进而利用多模数据的统一计算完成任务解题;(2)实现数据即时滚动更新并自动摘要和自动标注,筑牢国际关系研究的数据底座,推进国际关系研究朝向时空数据的细颗粒度分析演进;(3)面向跨语种阅读和多语言计算,需嵌入翻译系统或翻译模块,最大程度解决国际关系研究人员的跨语言阅读问题,消除语言障碍、提高阅读速度。

二是打造计算平台、普及算法应用,强化计算能力。在国际关系研究中构建开源计算平台,主要目的是为了通过算法模型的封装提升模型的复用性并给予数据计算充

足的算力支持,以解决零基础编程人员如何步入智能化分析的“门槛准入”问题。相比于在国际关系研究人员中花费大量精力和财力专门培养编程技能,开源计算平台的建设是一个可以填平“数据鸿沟”的低成本解决方案,其核心的解题思路是雇用专业人员进行模型预训练和界面化封装,然后按一定接口标准将其集成部署到一个开源框架上,从而构成一个“算法超市”或“算法工具箱”,以便各方“零门槛”调用和“菜单式”使用。就此而言,平台的建设目标不是为了创新算法模型,而是为了在国际关系研究和计算科学之间搭建学科交叉的桥梁,将计算科学领域中业已成熟的技术集成化、智能化和可复用化,并将其应用到国际关系研究中以推进实际问题的技术解决。一言以蔽之,放眼未来的技术发展和数字化环境,着力构建面向本领域专题任务解决的专业性开源计算平台具有便利算法应用、解决“算力卡脖子”之功能。

其次,新兴研究议题之争将开辟国际关系研究新的知识增长空间。未来,智能技术的应用将深嵌学科创新和社会发展的方方面面,基于新技术而形成的新型研究方法也将越来越多渗透蔓延至国际关系研究各个问题领域。伴随着一系列新型数据系统的建成并投入使用,国际关系研究的数据基础将变得更加坚实且丰厚,同时诸多开源计算平台建设也将赋能国际关系学者,使其拥有前所未有的社会感知与计算能力。在此背景下,智能化国际关系研究将涌入更多知识竞争者,基于新的数据类型和新的研究手段产生一系列新的研究话题。目前来看,以下话题不仅具有智能分析特征,而且影响巨大。

一是基于内容生成技术的国际关系虚假信息研究。伴随着生成式人工智能的大发展,深度伪造和大语言模型正在为国际关系研究提供新的研究素材。譬如在近期的乌克兰危机中,俄乌双方都将信息深度伪造作为认知作战和民心争夺的工具。根据媒体报道,2022年3月16日,乌克兰总统泽连斯基的一段伪造视频出现在各大社交网站上,内容是乌克兰总统“呼吁士兵们放下武器、放弃对俄战斗”,泽连斯基随后发布自拍视频澄清:“乌克兰不会向俄罗斯投降”。^①2023年6月5日,多家俄罗斯电台遭到黑客攻击并播放了伪造的俄罗斯总统普京讲话,声称“乌军已入侵俄罗斯”,随后克

^① James Pearson and Natalia Zinets, “Deepfake Footage Purports to Show Ukrainian President Capitulating,” Reuters, March 17, 2022, <https://www.reuters.com/world/europe/deepfake-footage-purports-show-ukrainian-president-capitulating-2022-03-16/>.

里姆林宫发言人佩斯科夫发布声明指出:“所有这些信息都是假的”。^①就此而言,传统研究主要以真实案例和模拟数据为基础,侧重于从真实数据中归纳,从模拟数据中演绎;然而,深度伪造和大语言模型的诞生,直接改变了国际关系研究的素材基础。伴随着机器生成数据的批量使用,现实世界的运行越来越为虚假数据所扰动。在此背景下,国际关系研究不仅要着力理解人与人之间的关系,更要关注人机互动乃至机器与机器之间的互动。就此而言,内容生成不仅是国际关系新的研究对象,同时也是社会计算的新方法。借助深度伪造和大语言模型,机器可以自动生成各种真假难辨的研究素材,也可仿真构建虚拟的镜像世界,从而使得社会实验和社会计算朝向虚拟现实演进。可以预见,内容生成技术的引入将有助于研究人员创新理解国际问题的思路,为政策制定和战略决策提供更加精细且符合逻辑的智力支持。

二是基于深度学习模型的国际问题预测研究。在高密度数据和充裕算力支持下,未来国际关系研究将能够进行庞大的机器学习模型训练,在一定程度上有望实现对未来事件和趋势的精准预测,这将有助于研究人员和政策制定者提出更有针对性、更具前瞻性的对策建议。借助神经网络和机器学习,国际关系研究可以深入拓展如下研究:(1)事态监测和预测,利用机器学习算法对某一特定领域的多源数据进行深度学习,预测未来的政治发展、国家间关系、军事冲突和恐怖主义等问题;(2)经济建模与评估,使用神经网络模拟开放环境下的专家决策,预测未来经济走向、制定投资战略和提出政策指导意见等;(3)社会情绪感知与情感计算,借助自动化语义理解和自然语言处理技术构建机器学习模型,计算新闻文本和社交媒体数据中的情感倾向,捕捉社情民意、预测民心走向。简言之,面向一个数据密集型的科研环境,机器学习手段的介入已是必然趋势。

三是基于面部微表情识别的政治微心理分析。传统国际关系研究因为技术水平的限制,主要以分析统计数字和文本阅读为主,声音、图像和视频等非结构化数据一度被排斥在国际关系研究的素材之外。然而,伴随着社交网络的盛行与数字政治的来临,人们越来越擅长使用声音、图像和视频作为政治表达形式,这些非结构化数据的批量涌现无疑丰富和拓展了国际关系研究的素材基础,同时也导致了以非结构化数据为研究对象的新研究路径的出现。其中,以图像和视频为分析对象的研究议程发展最为

^① Sam Fellman, “Russian TV Airs Apparent Deepfake Video of Putin Ordering Martial Law Amid Reports Ukraine Is on the Attack,” *Business Insider Nederland*, June 6, 2023, <https://www.businessinsider.com/russia-tv-air-apparent-deepfake-video-of-putin-ordering-martial-law-2023-6>.

强劲。在此背景下,有关政治人物的面部微表情识别也随之成为可能,并成为当下国际关系研究备受关注的的一个前沿探索领域。借助 GPU 图形处理器和基于卷积神经网络的机器学习,一个政治人物可能非常善于掩饰自己的内心,但针对其面部微表情的识别技术已经进展到可度量、可计算与可视心理曲线阶段。就此而言,在国际关系领域基于微表情识别展开的微心理分析不仅可用于精准公共外交和认知作战,同时还可被用于国际谈判和政治选举中的情绪侦测。这意味着国际关系研究通过智能化手段的赋能将开启新的政治分析路径,并产生新的政治见解。

除此以外,伴随着高密度数据的可获得和开源计算平台的建设日渐成为可能,未来国际关系研究利用智能计算来加强安全态势感知也将成为重要的研究议题。当然,凡事有其用,必有其反。智能计算在政治场域的应用,无疑也会引发一系列数据安全、算法安全和数字伦理问题,这也是未来需要持续关注 and 讨论的重要话题。

结 语

综上所述,智能化国际关系研究的本义是政治理论与智能分析技术的融合,集中体现为基于数据驱动的政治分析和基于算法的政治行动。坚实的数据底座和高性能的计算平台,是智能化国际关系研究得以展开的两个重要基石。在某种意义上,数据感知和智能计算是智能化国际关系研究不可或缺又相辅相成的两个侧面。但考虑到目前学界的学术训练体制、技术发展水平和资源支持现状,国际关系研究迈向智能化分析可采取阶段化策略,因时因地制宜、先易后难、以点带面。总体上,未来国际关系研究将在智能技术加持下快步迈向工程性应用,形成更加贴近现实、更加具有技术含量的知识增长模式。但在此过程中,只有解决那些零基础编程知识的多数人的“算法准入”问题,使更多研究人员拥有数据可得性和计算可及性,国际关系学科才能真正进入智能化分析时代。

(作者简介:董青岭,对外经济贸易大学国家安全计算实验室教授;刘文龙,对外经济贸易大学国际关系学院博士研究生。责任编辑:张海洋)