《国际政治研究》(双月刊)2020 年 第 3 期

新冠疫情与大数据:迈向人工智能时代的安全治理

董青岭

当前,全球仍在奋力抗击新冠病毒,世界正处于大变革、大调整和大分化的转折点上。以目前疫情蔓延态势来看,即便是有些国家实现了所谓的“群体免疫”,下半年除中、韩、朝和新西兰等少数几个国家外,绝大部分国家的社会开放与经济重启计划仍遥不可期。有医学专家甚至断言,新冠病毒或将与人类长期共存。而在此之前,由于中美之间的结构竞争和贸易摩擦,全球经济地理空间一直处于重新分化组合中,新冠疫情的暴发无疑加剧了中美之间的固有分歧。疫情之后,世界格局必将迎来新一轮深刻调整,新的世界秩序正在全球抗疫中悄然滋生。放眼未来,一个以开放、包容、协作和协商为特征的全球安全化进程或将戛然而止,代之以全球供应链和产业链的中断、国家间人员交往的限制性隔绝,以及各式各样种族主义和排外情绪的沉渣泛起。在新的政治形势和新的技术环境下,大数据被广泛应用于新冠疫情防控和社会稳定监测,以新冠疫情防疫为契机、以精准安全治理为目标、以大数据为核心驱动力,全球安全治理的人工智能时代正在悄然拉开帷幕。

一、全球人际网络与系统性安全风险

与传统安全威胁不同,新冠病毒的全球性蔓延和肆虐,预示了人类社会正

* 董青岭:对外经济贸易大学国家对外开放研究院研究员、国际关系学院教授。(邮编:100029)

在步入一个新的人际互联与社会组织形态，即全球性复杂巨系统的形成与出现。在这个史无前例的全球性复杂巨系统中，由于现代交通工具的使用和社交网络的普及，万物互联、人际相通，散落于全球各个角落的人、机构和组织越来越趋向于高频互动且休戚与共，全球已然形成一个统一的人际之网，各个国家和个人同呼吸、共命运。由于人际间的密集互动和社交圈子的重叠交错，不仅仅是信息和思想会沿着社交网络高速传播，就连病毒也会沿着网络节点四下蔓延，由此带来全球性系统风险。正是由于全球互联互通跨越了地理隔绝的临界门槛，近年来，“蝴蝶效应”和“黑天鹅事件”频频出现在全球政经领域、进而扰动全球系统平衡，典型案例如 2011 年的突尼斯“茉莉花革命”事件引发中东政权更迭连锁反应进而导致中东大变局、2016 年韩国梨花女子大学女生骑马事件牵出“闺密干政门”进而搅动韩国政局并带来东北亚变局。放眼未来，“安全”已不仅仅是传统理论所描述的地域性和结构性问题（如民族结构和经济利益分配格局），更多的安全问题可能起于人际间的微末互动（诸如病毒的传播）。换言之，在人类社会的复杂巨系统时代，“安全”一词已然泛化成“社会安全”和“人的安全”等范畴更为广泛的安全议题，很多风险看似局部且微不足道，然实则具有全局性的系统破坏效应，互联互通已使得全球每个角落里的微小事件都有可能演变成全球性安全灾难。尽管“逆全球化”进程愈演愈烈，但未来新的安全环境将更加凸显全球相互依赖，新的安全治理迫切需要更新安全思维、革新安全技术。

仅以新冠病毒为例，安全可能起于微末，影响却及于全球。基于大数据的安全治理首先假定：微观层面的互动会造就宏观层面的系统效应，通过捕捉微观主体之间的互动信息并观察其互动进程或可预测宏观层面的安全后果。正是基于这一系统性的风险认知，大数据新冠疫情防控主要着力于微观层面的末端安全治理，重在管控和切断人际网络中的病毒传播路径，目的是通过暂时中断微观层面的人际互动来防止宏观系统的整体性崩溃。显而易见，这是一种典型的系统性安全思维，而要完成这一史诗级的强制隔离与自我隔离，不仅需要全球层面休戚与共的共同命运认知，更需要不分种族、不分宗教、不分政见和不分地域国别的全球通力协作。然而，很不幸，在现实政治中，很多国家的安全思维依然停留在主权对抗、集团主义和本位主义至上的旧时代。实践表明，在人类社会演进到复杂巨系统时代，安全越来越是相互的，几乎没有国

家和个人可以独善其身,安全治理需要更宏观的系统视角和更细致入微的工作落脚点,而这恰恰是大数据思维和大数据技术所擅长的。

二、新冠疫情防控与大数据安全治理

作为一种技术性解决方案,大数据安全治理的对象是面向微观主体及他们之间的高频互动,在算法逻辑上重在发掘相关性、寻找相似性。其中,相关性指的是事物之间的关联关系,但并不一定是严格的因果关系;相似性指的是具有相近属性越多的事物往往具有趋同的特征表现或行为趋向。在当前新冠疫情防控中,大数据因其处理数据体量大、抽取信息速度快和预测结果稳定性强,而主要应用于如下场景:

其一,追踪和甄别密切接触者。它的基本逻辑是相关性分析,它假定接触者是易感染的高危人群,通过相关性分析和社交网络距离度量,大数据可以发现和甄别密接者,进而通过隔离和救治,可有效防止因一人感染而危及整个人际网络。传统上,追踪密接者主要是采用纸质填表和入户访谈来完成的,花费时间多、耗费精力大,采集到的信息往往却并不都是精确可用,面对当下汹涌而来的数据洪流更是难以应对。现在借助大数据自动采集手段(诸如搜索引擎主题词统计、健康二维码扫描、手机地理位置定位)可以很好地变被动数据挖掘为主动数据采集,这不仅可以有效规避密接者的刻意隐瞒,而且还可以全覆盖搜索密接者的人际网络,更可实现数据的即时采集和即时分析,从而为抗疫政策的制定和执行赢得宝贵时间。总体上,这一过程通常可分为大数据接触者识别、开列接触者名单和隔离接触者三个步骤。

其二,辅助诊断与药物研发。传统上,一种临床诊断方案的成熟或一种药品疫苗的研发,都需要投入大量的人力、物力和时间。在没有疫情暴发的平常时期,这些工作往往具有充足的人力资源和充分的时间保障。然而,一旦疫情大规模暴发并汹涌而来,传统人工手段就变得力不从心。就此而言,当前新冠肺炎诊断主要依赖的是CT影像和核酸检测,然而,人工阅片量大、耗时长,医生很容易因过度疲劳而导致误诊、漏诊。大数据机器学习通过自动构建模式识别,可以有效缩短看诊时间并提高看诊效率。相关新闻报道如阿里云创立了新冠肺炎CT辅助诊断系统方案,平均最长20秒钟处理一个病例,读片速度

是医生的 50 倍，每天可分析约 1.3 万例样本。此外，阿里云还提供了新冠病毒全基因组检测方案，以解决全国范围内核酸检测能力不足、PCR 方法假阴性率高，以及病毒可能发生变异等关键问题。全基因组分析可以通过序列分析和序列拼接，分析与病毒序列的同源性，定制化地给出最终诊断报告；可以构建进化树，智能分析病毒传播或演化的时间图谱，智能分析患者感染事件；还可以预测病毒蛋白二级结构和三维结构。①

其三，通过数据汇集支撑决策。大数据通过数据集和算法模型的构建，可以前瞻性地预测疫情走势与高危人群的地理分布，进而可以为政府决策提供数据支撑，以帮助政府确定需要医疗资源的地区和人群、优化防疫流程和资源配置，进而为重新开放社会和重启经济复兴计划提供指导性建议。譬如，疫情之初，我国各省相继发布的各种健康手机应用，询问人们是否出现发烧、咳嗽、呼吸急促或嗅觉丧失等症状，而这些症状或与新冠病毒相关，从而可以帮助预测未来几天内医院可能登记的感染病例人数，并提供病毒扩散区域的早期痕迹。再如，2020 年 4 月 16 日，由艾伦人工智能研究所领衔的国际合作网络联合发布一个有关新冠病毒的开放研究数据集（CORD-19），②该数据集是一个统一的免费资源，包含超过 4.4 万篇学术文章、超过 2.9 万篇有关新冠病毒（COVID-19）和冠状病毒家族的病毒的资料全文，全世界的机器学习社区都可以使用它，来推进新冠病毒的研究。

三、常态防疫与基于数据驱动的安全战略

放眼环球，鉴于欧美各国当前的防疫政策，新冠疫情在短期内恐难结束，在不少国家甚至已有二次暴发的风险。考虑到全球社会业已形成一个互联互通的复杂巨系统，单一国家的抗疫胜利并不意味着疫情的真正结束，全球抗击新冠肺炎即将步入常态化防疫阶段。2020 年 4 月 8 日，中共中央政治局召开的常务委员会会议指出，“常态化”防疫意味着必须加快建立与疫情防控相适应的常态化经济社会运行秩序，要让各个行业、各个人群慢慢找回正常的运转节奏。然而，伴随着经贸摩擦的不断升级和新冠疫情的持续蔓延，中美关系进

① 倪思洁：《疫情之下，大数据和人工智能做了什么》，《中国科学报》2020 年 4 月 19 日。

② *COVID-19 Open Research Dataset*，https://www.semanticscholar.org/cord19，2020-04-20.

一步趋于恶化、全球经济结构进入新一轮深度调整,而中国周边一些国家也呈现出政局不稳定迹象,构建大数据安全风险监测平台、创新基于数据驱动的安全战略,将有助于我国及时掌握内外环境的瞬息万变,进而制定科学的安全应对之策、从容应对各种内外挑战。

2020年下半年,全球新冠疫情如果持续蔓延,我国社会发展与国家安全将面临一系列挑战:其一,中国外部发展环境空前严峻。新冠疫情的长期持续将使全球经济陷入深度衰退,贸易保护主义必将愈演愈烈。自新冠疫情暴发以来,全球国际贸易额急剧下滑,世界各国工业生产明显放缓,部分企业甚至出现生产经营困难。能否确保经济平稳复苏,将是疫情之后世界各国面临的共同难题;其二,中国外部安全环境日趋恶化。疫情期间,朝美关系、伊核问题和印巴冲突等国际热点问题有升温趋势,中东、中亚和外高加索地区动荡加剧,恐怖主义、跨国犯罪和种族排外等问题日益突出,国际安全形势严峻。中国周边一些国家处于经济社会转型期,社会矛盾因新冠疫情空前加剧,政局渐趋动荡,与中国的经济和政治摩擦将日益增多;其三,中国面临国际舆论压力日渐增大。当前,西方国家防疫渐趋政治化,部分西方学者和政客宣扬"强国必霸""疫情责任"等论调,质疑中国的防疫举措与和平发展。同时,意识形态领域的斗争更趋深刻复杂,国际上一些势力频繁利用民主、人权和疫情问题"抹黑"甚至"妖魔化"中国。如何消除外部世界对中国的敌视和意识形态偏见,创造一个客观友善的国际舆论环境,将是疫情之后中国面临的长期挑战。

鉴于上述安全环境的深刻变革,基于数据驱动的安全战略创新或将成为疫情之后各国安全战略竞争的新常态。仅就国际关系领域而言,竞争焦点或将集中在如下方面:

其一,面向实时数据自动采集的新型数据库建设,旨在利用大数据技术重构安全研究的底层数据基础。与传统数据库如战争相关因素数据库(COW)、乌普萨拉武装冲突数据库(UCDP)和全球恐怖主义数据库(GTD)不同,新一代数据库建设将着力应对当下汹涌而来的数据洪流,这些新型数据不仅数据体量巨大、数据产生速度快,而且数据维度和数据颗粒度也远超以前时代所能想象。在此情景下,以自动摘要和自动编码技术为核心的新一代数据库建设正在取代传统人工摘录和人工编码数据库,在这方面,目前业已成型并被广泛使用的数据库如全球事件、语言与语调数据库(Global Database of Events,

Language and Tone，简称GDELT），[①]是一个面向全球、免费开放的滚动型即时新闻事件数据库，由美国乔治城大学教授卡里夫·利塔鲁（Kalev Leetaru）于2013年创建，它不仅对新闻事件中的人物、组织、事件、语气等事件要素进行标签化提取，同时，还通过自动编码技术自动标注新闻事件的地理位置信息（即经纬度坐标），并且每15分钟实时更新一次。

其二，面向特定安全问题解决的算法模型研发，目的在于将安全决策理论与计算机智能分析相结合改善决策质量。譬如，通过协同过滤算法（collaborative filtering）筛选新冠疑似患者、通过邻近算法（K-NearestNeighbor）进行新冠人群特征聚类分析、通过排序算法（PageRank）进行新冠传播网络链接分析，以及通过随机森林算法（Random Forest）进行高低危人群分类预测等。概括来说，基于算法（Algorithm）的大数据安全应用重在规避数据噪音、挖掘数据关联，进而建立特征模式识别和进行分类预测。目前，大数据算法在安全研究中的应用主要集中在以下三个场景：第一，精准定位。通过抓取数据痕迹和聚类分析，精准圈定事件地域、事件人群及人群属性特征，定制化推送安全信息和实施精准安全管控；第二，早期预警。通过数据监控和云计算，即时监测、锁定、跟进事态进展并自动生成事件报告和危机预警，动态掌控问题爆点，提前推进基于预测的预防性战略执行；第三，关联预测。通过多源数据收集和数据组合算法，在各种结构化和非结构化数据资源中发掘事件关联关系和节点因素，优化决策、合理配置资源。

综上所述，在新冠疫情结束之后，全球安全议程即将迈入基于数据驱动的安全战略创新，与传统因果性分析相比，大数据安全态势感知更加侧重挖掘关联关系和相似关系，旨在实现安全议题的实时监测与即时预测，这在一定程度上支持了复杂科学的“人际相互扰动论”和“系统演化论”，以及有关人际互动的“信息交换论”。尽管世界是不确定和复杂的，但现实安全仍然是可以被感知、被预测的，借助大数据分析技术，安全监测平台将自动监测并智能化分析内外安全情势的变化，进而创新风险评估与安全态势感知方法。透过大数据的海量信息提取能力和高速运算能力，安全决策者们可以即时感知风险来源、明晰战略方向，进而实施实现精准化战略制定和精细化策略推进。

当然，凡事皆有两面，大数据在进行精准安全防控的同时，也正带来一系

① *The GDELT Project*, https://www.gdeltproject.org/, 2020-06-10.

列新的安全隐患,大数据安全应用本身正在成为安全化进程中的一个重大安全问题:其一,大数据精准防控潜伏个人隐私泄露风险。以新冠疫情防控为例,个人信息正在被政府和政府授权的机构以前所未有的规模采集、整理和储存,这些信息不仅包括个人的身份信息如年龄、性别、居住地、家庭、联系方式和医疗史,甚至还包括个人的行为轨迹、地理定位、社交网络和消费习惯,一旦这些信息被泄露,将会造成无可挽回的的生命、财产损失,严重的甚至会引发社会恐慌与政府信誉滑坡;其二,大数据安全防控蕴含信息讹诈风险。由于信息特别是个人信息被史无前例的深度挖掘和计算,一些掌控信息或被授权处理信息的人和机构可能会获取数据赋权优势,即利用掌握的他人信息进行权力压迫和利益勒索。譬如,在未来战争中,敌方对手会通过数据手段跟踪作战士兵的在线行为习惯、家庭成员的日程安排乃至子女状况,进而制造“个性化精准威胁”,从而胁迫士兵无法专注作战任务。

总体而言,数据科学技术的进步无疑会极大便利我们的生活,但同时一旦数据和算法被滥用也会反伤我们自身:从智慧医疗到无人驾驶汽车、从智能家居到 3D 打印、从数字政府到无人作战系统,所有这些都有可能被黑客或别有用心的组织势力所攻击,进而带来灾难性后果和系统整体性动荡。在大数据时代,我们的社会比以往任何时刻都更加脆弱、更加不安全,但基于数据驱动的安全战略已然成为新时代的安全治理取向。