

大数据外交：一场即将到来的外交革命？^{*}

董青岭

内容提要：在新技术变革的驱动下，大数据已然向外交决策、规划和实践领域渗透。在此背景下，世界一些国家的外交模式正在悄然发生变革，在经验驱动之外，数据驱动正日益成长成为一种全新的外交模式；从数字外交向数据外交的激进跃迁，展现了当今时代大数据喷涌而出、势不可挡的社会冲击力，同时也预示了传统的外交理论与外交指导原则可能需要根据技术的进步而加以修正。考虑到大数据的技术特征以处理海量数据、非结构数据和即时数据见长，并以数据可视化作为分析结果展现形式，文章主要考察了大数据兴起对当前外交领域的介入态势及其对外交决策模式的冲击，概述了当前对大数据应用于外交领域的主要争议与若干案例应用，并简要评析了大数据驱动外交模式创新的未来前景与关键障碍。

关键词：大数据 数字外交 冲突预测 危机感知

当前，大数据（Big Data）的兴起正在广泛改变着社会生活的各个领域，国际关系领域与外交活动也越来越难逃大数据的影响。在此背景下，围绕着“大数据能否介入外交领域以及介入后是否会引发一场外交革命”等话题，学界也展开了争论。^① 不少

^{*} 本文系对外经济贸易大学人才培养项目“网络空间威慑与数字时代的国家安全战略研究”（项目编号：14YQ09）、教育部人文社会科学研究青年基金项目“基于 WEB2.0 技术的西方国家对华新媒体外交态势及我国对策研究”（项目编号：12YJCGJW002）和中国博士后科学基金面上——等资助“基于决策者偏好的网络空间威慑学说与战略模型研究”（项目编号：2014M550032）的阶段性成果，感谢《欧洲研究》匿名评审专家的宝贵修改建议，文责自负。

^① 相关争论可参见 Conference on “From Big Data to Global Diplomacy: Today and Tomorrow”, Hold by Institute for the Study of Diplomacy, School of Foreign Service Georgetown University, April 10, 2014,

研究文献指出,大数据主要是指在一定时期内无法用通常的软件工具进行捕获、管理和处理的PB级以上的数据集合,这些数据通常具有体量巨大、类型繁多、价值密度低和增长速度快等特征,在数据来源上不仅包含传统数据库收集的结构化数据,同时还包含来自社交网络和物理传感器产生的非结构化数据,它需要新的处理模式加以处理方能拥有更强的洞察力和决策力。^①有鉴于此,很多热衷于接纳新技术变革的国际观察和政策分析人士认为,大数据以处理海量数据、非结构数据和即时数据见长,并以数据可视化作为分析结果的展现形式,可以在短时间内高速处理百万维度以上的数据并使数据以动态可视化形式展现,这就使得外交决策有可能确立在充沛数据分析和动态感知的基础之上,以前那些被忽视、被遗弃或不能进行技术分析的信息有可能会被重新发掘,并有机会进入外交决策过程、影响决策结果,由此导致传统外交思维、外交理论乃至外交战略被不同程度地修正、改造、完善或重塑,外交决策模式或将发生不同于以往的各种变化。^②就此而言,如果大数据对未来外交活动领域的冲击与挑战是不可避免的,那么大数据具体应如何应用于外交决策和外交战略的执行,它带来的诸多改变是否是革命性的?为探讨以上问题,本文着力讨论了大数据时代的可能数据应用场景与外交模式变革,并简要评析了大数据介入外交活动的未来前景与关键障碍。

一 从数字外交到数据外交:数据力量的崛起

简单而言,所谓“数字外交”(Digital Diplomacy)指的是外交活动中对因特网以及诸如手机等其他数字通信技术工具的重视和使用,通常又称“E-外交”,其目标在于借助数字平台将传统外交的垂直政治沟通方式(vertical communication)改变为垂直与平行(horizontal communication)并重,^③同时增加外交活动的人群接触面、接触频度以及多元参与性,并重视线上与线下活动的衔接。它肇始于2002年美国国务院“E-外交

① 概念总结可参见邹贺铨:“大数据时代的机遇与挑战”,《信息安全与通信保密》2013年第3期;蔡翠红:“国际关系中的大数据变革及其挑战”,《世界经济与政治》2014年第5期;任琳:“网络空间战略互动与决策逻辑”,《世界经济与政治》2014年第11期。

② 相关观点可参见陆钢:“大数据时代下的外交决策研究”,《社会科学》2014年第7期;Hmatom Fletcher, “E-nvoys: What Next For Diplomacy?”, October 22, 2014, <https://nakeddiplomat.wordpress.com/2014/10/22/e-nvoys-what-next-for-diplomacy/>; Witold Henisz, “Lessons from Moneyball: How Data Can Drive Corporate Diplomacy”, June 11, 2014, <http://nbs.net/lessons-from-moneyball-how-data-can-drive-corporate-diplomacy/>, 2015年3月13日访问。

③ 此处的“垂直政治沟通”指的是带有等级色彩地位或资源不平等者之间的沟通,即“政府对民众”(government to people)的沟通模式,而“平行政治沟通”则是指地位或资源平等者之间的政治沟通,即“政府对政府”(government to government)、“民众对民众”(people to people),作者注。

工作任务组”(Taskforce on eDiplomacy)的创建^①;而数据外交(Data-driven Diplomacy)则是指包括数据挖掘、数据搜索、数据存储和数据算法等数据分析技术以及数据软件乃至数据本身在外交活动中的运用,其目标不在于构建政治沟通接触平台(platform for engagement),而是旨在创建用于辅助决策的情报整合与数据分析平台。数据外交的精髓是主张数据本身和数据分析技术在外交活动中应发挥辅助决策、优化外交执行的作用。这一概念的使用源于 2014 年 9 月美国公共外交咨询委员会(APCD)一份名为《数据驱动型公共外交》的评估报告^②。另外,也有学者将这一数据运用型外交活动称为“开源外交”(Open-Source Diplomacy)。^③

无论是“数据外交”还是“开源外交”,到目前为止都不是一个成熟和定型的概念。但以 2013 年肯尼斯·库克耶(Kenneth N. Cukier)和维克托·迈尔-舍恩伯格(Viktor Mayer-Schoenberger)发表在《外交事务》杂志上的《大数据崛起》一文为标志^④,大数据及其分析技术已然正式登陆外交决策与外交分析领域,并迅速表现出与数字外交协同并进、相互辅助的发展态势。可以说,继 2009 年“推特革命”(Twitter Revolution)之后,就在全世界的外交决策者和执行者们尚未从上一波数字外交“革命”的震撼性冲击中醒悟时,短短时间内全球外交活动的新一波创新驱动力又接踵而至。概括而言,从“数字外交”走向“数据外交”并不仅仅表现为两种技术变革对现代外交的简单介入,而更是表现为现代外交理念正在新技术驱动下从一种模式走向另一种模式,并最终走向多种模式的共存与相互补充。如果说以现代社交网络和智能移动工具为代表的通信技术进步,导致传统外交从单向宣传走向多向沟通、从垂直说教走向平等对话、从精英决策走向大众参与;那么,大数据及其分析技术的介入,则有可能导致传统外交从经验感知走向数据驱动、从粗放外交走向精准外交。

传统上,由于数据采集和数据分析手段的限制,很多时候外交决策和外交执行是建立在经验感知的基础之上,无论是数据体量、数据真实性还是数据生成方式和数据获取渠道都是极为有限的,小样本抽样调查、历史经验知觉感悟与学术研究因果逻辑推演是人们洞悉这个纷繁芜杂世界的主要决策基础,透过小样本调研与结构化数据分

① Antonio Deruda, *The Digital Diplomacy Handbook: How to Use Social Media to Engage with Global Audiences*, CreateSpace Independent Publishing Platform, 2015.

② United States Advisory Commission on Public Diplomacy, “Data-Driven Public Diplomacy: Progress Towards Measuring the Impact of Public Diplomacy and International Broadcasting Activities”, September 16, 2014, <http://www.state.gov/documents/organization/231945.pdf>, last accessed on 13 April 2015.

③ Jonathan Spalter, “Open-Source Diplomacy”, *Democracy: A Journal of Ideas*, No.23, 2012, pp.59-70.

④ Kenneth Neil Cukier and Viktor Mayer-Schoenberger, “The Rise of Big Data: How It’s Changing the Way We Think About the World”, *Foreign Affairs*, May/June 2013, pp.27-40.

析,获得的知识多是关于我们所生活之世界的线性因果结论,以至于很多时候世界的复杂性和不确定性被刻意忽略了。在有限获取数据和有限提取信息的条件下,外交决策与外交执行通常只能专注于问题的一个侧面而无法顾及全局,有时甚至无法洞悉和还原外交事件的真相。在此情形下,会经常性地出现理论与现实相脱节的决策偏差。

然而,在大数据时代,这一切行将改变。在新技术冲击下,国际生存环境的数字化和数据化态势日益凸显:一方面数据的发射、传播速度和传播频率呈指数倍数增加;另一方面数据的产生来源更趋多样化,两者共同构成了“数据洪流”。更为重要的是,数据革命不仅在工业化国家发生,在发展中国家也同样发生,而且趋势越来越明显。考虑到大数据对现代商业模式、知识创造方式和政治运行的基础支撑作用,各国政府相继推出“数据治国”战略,强调科学决策不仅要有基于历史经验的真知灼见,更需基于海量客观数据的态势感知与精准定位。在数字时代,能否快速准确地进行外交决策并合理配置外交资源已成为制胜关键。基于此,凭借大数据技术搭建基于数据驱动的外交策略理论和战略规划学说,目前已成为新一波国际竞争的潮流。

大数据具有从多样和多源数据中快速获取信息的能力,尤其是其技术特征擅长非结构化数据的处理、海量信息的动态可视化以及事件态势的模糊预测,这意味着数据驱动型外交与传统经验外交相比具有明显不同的特点:其一,数据驱动型外交力求掌握与研究对象有关的更多数据,着力刻画研究对象的整体特征而非局部细节;其二,数据驱动型外交力图容纳非结构化数据的存在,追求数据的混杂性而非精确性;其三,数据驱动型外交试图超越研究变量之间的因果解释,重在探究那些宏观上能引起变化的数据之间的关联关系。

作为一种新的战略资源和技术分析手段,大数据未必会带来国际关系和外交决策研究的根本性变革,但大数据辅助外交决策会明显提高决策效率和决策质量。大数据适合分析多元混杂数据,同时又不追求因果解释而是旨在探索看似不相关要素之间的相关关系,这在很大程度上可能比传统方法更适于捕捉复杂多变国际环境的不确定性。有鉴于此,在大数据时代,数据的决策价值会被日益重视和重新发掘,外交决策分析极有可能走向两种模式并存与相互验证:即基于历史经验的外交决策和基于数据驱动的外交决策两条路径。

鉴于数据技术在目前的发展速度和发展趋向,在可预见的未来,欧美国家一旦将几百万倍于今天数据规模的大体量数据纳入外交决策中,同时又基于大量诸如网络图表、搜索痕迹和视频声像等非结构化或半结构化数据进行混杂性计算,那么先前用于指导外交实践的理论总结是否依然成立?如果美欧国家在大数据支撑下能够洞察其

他国家所无法知晓的知识和信息情报,则其他国家所奉行的外交战略是否依然有效?美欧外交融入大数据元素后对国际关系格局和未来竞争又将产生怎样的影响?考量到大数据即将到来的种种政治应用场景,各国有必要重新思考先前所习以为常的外交理论与决策方法是否依然成立,并据此总结、构建基于数据驱动的外交指导原则和实践模式。

二 从技术变革到当前争议:大数据的适用性

毫无疑问,学界当下最时髦的热点话题莫过于大数据对社会科学渗透与冲击,对数据向来敏感而又难以获取的外交领域而言更是如此。大数据与传统数据不同,不仅体量大、产生速度快,而且数据噪音大、获取有价值信息难。但即便如此,在技术变革的驱动下,大数据对外交决策和外交执行领域的渗透和介入依然势不可挡,目前学界的研究和争议主要集中在以下三方面:

第一,大数据的适用性。作为一种新兴战略资产和新兴技术,大数据及其分析手段能否被应用于外交决策和外交执行领域?如果可以,大数据的应用能否带来颠覆性的知识发现和政策启示?针对这一问题,一种观点认为,作为一种全新的数字化生存方式,大数据正在改变我们的生活以及理解世界的方式。^①通过对推特(Twitter)、谷歌、脸谱(facebook)、微信、微博等新媒体平台信息的抓取和计算,政治研究者不仅可以跟踪大城市的抗议行动、发现恐怖主义行迹、明晰国家战略风险,还可对利益攸关人群进行精细划分、对政治态势进行整体感知、对危机进行预警和预测,从而有助于外交科学决策和精准政治营销。^②而另一种观点则认为,大数据不过是传统统计方法的更新,既非“科学”也非“革命”,并不一定会带来有关国际形势和外交战略的全新洞察,更不会诱导外交决策模式发生革命性变革。因为即使存在着大体量外交数据并可以自由抓取,各国出于安全考虑和保护隐私也会设置法律和技术障碍阻止数据的轻易扩

^① Kenneth Neil Cukier and Viktor Mayer-Schoenberger, “The Rise of Big Data: How It’s Changing the Way We Think About the World”, p.29.

^② 相关文献参见 Charles J. Dunlap, Jr., “The Hyper-Personalization of War: Cyber, Big Data, and the Changing Face of Conflict”, *Georgetown Journal of International Affairs*, 2014, pp.109–118; Francesco Mancini and Marie O’Reilly, “New Technology and the Prevention of Violence and Conflict”, *Stability: International Journal of Security & Development*, Vol. 2, No.3, 2013, pp.1–9; Stephen Ansolabehere and Eitan Hersh, “Validation: What Big Data Reveal About Survey Misreporting and the Real Electorate”, June 6, 2012, <http://polmeth.wustl.edu/media/Paper/hershpolemeth2012.pdf>, last accessed on 10 March 2015.

散,更加之海量数据存在的数据噪音太多导致数据开发价值极低。^①

第二,大数据的洞察力。目前,大数据在商业和金融领域应用较多也较为成功,但在外交领域应用尚不多见,成功的案例分析更是少之又少。概括欧美学界的相关研究,其核心作用有三:其一,大数据可以用来进行整合新媒体用户信息,明晰外交对象的行为特征、心理变化和情绪反应,辅助基于客观数据的外交资源配置和战略规划^②;其二,大数据近乎全样本的模糊预测可以用来感知冲突热点、预测态势演变,辅助危机预警、危机公关和有效开展预防性外交^③;其三,大数据的复杂算法与融合技术可有效处理国际关系中的大体量非结构、半结构化数据,同时又能以可视化(Visualization)方式展现数百万维度以上的数据关联,从而有助于国际关系学者了解研究对象的全貌,发现通过小样本分析和结构化数据推导不容易发现的事件关联和政治规律。^④

第三,大数据的风险性。在商业领域大数据的使用更多触及的是个人的隐私权,然而在国际关系和外交领域,数据流动跨越国界和数据挖掘常常涉及他国隐秘和国家安全,正如数据具有辅助决策的效能一样,数据使用和开发的政治风险性也显而易见。一些主张数据开放(open data)的学者认为,数据的价值在于开放、流动与不断被使用,如果通过数据分析可以挖掘更多决策信息、精准定位外交对象和客观预测竞争对手的战略意图,基于成规模大数据的外交战略将会极大地改善外交决策质量、提高资源配置效率并最终促进和扩大国家利益。^⑤然而,反对的观点也非常鲜明和尖锐,认为一旦大数据分析变成外交战略施展的常态,则一国的数据安全与个人隐私将难以保障,数据安全防范技术弱的国家更容易为技术强国所窥探、掌控和摆布,数据争夺将诱发

① David Lazer, Ryan Kennedy, Gary King and Alessandro Vespignani, "The Parable of Google Flu: Traps in Big Data Analysis", *Science*, Vol. 343, 2014, pp. 1203-1205.

② See Jonathan Spalter, "Open-Source Diplomacy", *Democracy: A Journal of Ideas*, No. 23, 2012, pp. 59-70; United States Advisory Commission on Public Diplomacy, "Data-Driven Public Diplomacy: Progress Towards Measuring the Impact of Public Diplomacy and International Broadcasting Activities", September 16, 2014, <http://www.state.gov/documents/organization/231945.pdf>, last accessed on 15 March 2015.

③ Pierre F. Landry and Mingming Shen, "Reaching Migrants in Survey Research: The Use of the Global Positioning System to Reduce Coverage Bias in China", *Political Analysis*, Vol. 13, No. 1, 2005, pp. 437-459; David W. Nickerson and Todd Rogers, "Political Campaigns and Big Data", *Journal of Economic Perspectives*, Vol. 28, No. 2, 2014, pp. 51-74.

④ See John Karlsrud, "Peacekeeping 4.0: Harnessing the Potential of Big Data, Social Media and Cyber Technology", in Jan-Frederik Kremer and Benedikt Müller eds., *Cyber Space and International Relations. Theory, Prospects and Challenges*, Berlin: Springer, 2013, pp. 141-160; Francesco Mancini and Marie O'Reilly, "New Technology and the Prevention of Violence and Conflict", *Stability: International Journal of Security & Development*, Vol. 2, No. 3, 2013, pp. 1-9.

⑤ See Jonathan Spalter, "Open-Source Diplomacy", *Democracy: A Journal of Ideas*, No. 23, 2012, pp. 59-70; Joel Gurin, "Open Governments, Open Data: A New Lever for Transparency, Citizen Engagement, and Economic Growth", *SAIS Review of International Affairs*, Vol. 34, No. 1, 2014, pp. 71-82.

更多“棱镜计划”。在此背景下,欧洲大陆学者多主张数据主权(data sovereignty)理论,并试图以此为支撑建立更为严密的数据收集法案和数据治理体系,^①而美英学者则更多支持数据自由流动说和数据开放论。^②

与美欧学界相比,我国学界有关大数据应用于外交领域的讨论近两年来也吸引了越来越多的年轻学者参与,相关文献和学术讲座主要集中在:(1)大数据国家战略布局。该议题侧重从国际关系未来竞争的视角探讨大数据对国家利益的重塑和国家现代治理能力的提升,主张将大数据提升至全球战略竞争层面进行顶层设计;^③(2)大数据国家安全风险。来自国际法、国家安全、公共行政和信息科学领域的诸多学者认为,大数据时代国家机密和公民个人隐私更容易为竞争对手所获取,社会公共安全和国家意识形态面临严峻挑战,主张学界应严肃探讨“数据主权”学说并呼吁建立国家数据安全标准和国际对话体制;^④(3)大数据外交决策分析。虽然这方面的文献尚不多见,但 2014 年发表在中国学术期刊上的几篇相关文章均认为,大数据的信息整合与挖掘功能有助于提升外交决策效率、改善外交决策质量。^⑤

综上所述,大数据渗入国际关系研究尤其是外交决策分析已是大势所趋,诸多学者的研究成果对大数据的技术特征、外交应用价值及政治风险进行了较为深入的前瞻性分析和探索性研究,但同时也在四个方面也存在明显不足:其一,在研究内容上,侧重对大数据技术特征和风险的笼统描述而缺乏大数据外交实战应用的个案分析;其二,在数据使用维度上,以推特、脸谱和微博、论坛等社交网络数据为主,缺乏多源数据支撑;其三,在数据开发程度上,以展示原始数据简单图表为主,缺乏算法较强的逻辑关联分析和深度数据挖掘;其四,在成果用途上,以舆情监测和政策建议为主,缺乏基于数据驱动的政治洞察和理论创见。

① Primavera De Filippi and Smari McCarthy, “Cloud Computing: Centralization and Data Sovereignty”, *European Journal of Law and Technology*, Vol.3, No2, 2012, pp.1-21.

② See UK Department for International Development, “Open Data Strategy”, April 2012- March 2014, <http://data.gov.uk/sites/default/files/DFID%20Open%20Data%20Strategy.pdf>, last accessed on 15 March 2015; The North-South Institute, “Open Data, Transparency and International Development”, November 2013, <http://www.nsi-ins.ca/wp-content/uploads/2013/11/2013-Open-Data-Summary-Report.pdf>, last accessed on 15 March 2015.

③ 邬贺铨:“大数据时代的机遇与挑战”;胡泳、郝亚洲:“数据治国与数据强国”,《新闻爱好者》2013 年第 7 期。

④ 蔡翠红:“国际关系中的大数据变革及其挑战”;沈国麟:“大数据时代的数据主权和国家数据战略”,《南京社会科学》2014 年第 6 期。

⑤ 任琳:“网络空间战略互动与决策逻辑”;陆钢:“大数据时代下的外交决策研究”。

三 从理论假说到案例实践:大数据语境下的外交创新

当前,大数据的介入正使许多国家的外交决策、规划和执行由“粗放外交”向“精细外交”、由“反应外交”向“预防外交”、由“普遍外交”向“定制外交”转变。如前所述,有三个应用趋向最为明显:第一,通过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,定制化精准推送政治营销广告和实施公共外交战略;第二,通过数据监控和云端处理,即时监测、锁定和跟进事态发展并自动生成事件报告和危机预警,动态掌控问题爆发点并提前推进基于数据预测的预防性外交战略;第三,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,优化外交决策、合理配置资源。

(一) 案例 1:大数据与公共外交:如何接触外国民众?

与传统公共外交不同,在大数据时代有两个重要趋势正显著改变着全球公共外交的图景:其一,目前几乎所有行业和社会生活领域都在经历一场所未有的“数字化革命”,并由此产生了巨量、实时的新型数据,其中流数据、地理空间数据和传感器数据等非标准化数据占据了绝大部分比重。这些数据并不能完美地适用于传统的结构化关系型数据库,也非我们习惯已久的 SPSS、SAS 以及 R 语言等计算分析软件所能够处理;其二,当前世界各国超级计算中心(Supercomputer Center)的建立以及诸如 Hadoop、Spark 和 E-Charts 等先进数据分析软件的开源普及,使得各国际组织和各国外交决策部门以及为数众多的学术团队能够以前所未有的复杂度、速度和准确度从海量数据中获得知识洞察力和政策启示。在此情形下,对外交官和政策观察者来说,研究大数据并不断扩大公共外交的界限以精细化方式接触他国民众,在当前无疑是绝佳的历史时机。

正是震撼于大数据惊人的信息采集与数据挖掘能力,美国白宫和国务院成功地将其运用于奥巴马总统对巴西的访问。2011 年 3 月 19 日,美国总统奥巴马出访巴西首都巴西利亚,此行目标在于寻求建立一个以巴西为盟友的美洲地区新联盟。为了获取巴西民众的政治支持,白宫在奥巴马的出访行程中设置了与当地民众接触的活动。但是行前奥巴马团队对巴西各界民众如何看待美国、如何看待奥巴马家庭及其本人,以及巴西民众对美国拉美政策调整会做出何种反应,都心里没底。在此情形下,美国国务院下属的“E-外交”办公室指示其数字技术团队挖掘了近两年来巴西民众在 Twitter、Orkut 以及 Blogger 上的上千万条社交数据,通过关联运算和人群搜索发现,在涉美

政治传播中不同的新闻平台、不同的自媒体评论人关注度差别迥异。在此情形下,如何制定一项清晰的公共外交战略对巴西民众展开卓有成效的外交公关活动,便成为白宫和国务院的首要之责。大数据分析至少为奥巴马外交团队提供了以下信息:(1)巴西民众讨论涉美话题最活跃的社交媒体或平台依次为 Twitter、Newsbusters、Bloggers 和 Orkut;(2)最有可能传播奥巴马新闻的关键社交达人是 LeiscaRJ、Celosathayde、Maria_fro、JairoRoberto 和 Dominiofeminino;(3)巴西民众对奥巴马之行最感兴趣的话题可能依次为现场参观、国内政治、泛美关系、迈克尔与奥巴马的关系、访问贫民区和联合国问题。^①由此,早在 3 月 18 日奥巴马到达巴西的前一天,白宫和国务院就已知晓了巴西民众对奥巴马之行的哪些话题最感兴趣、哪些人最有可能传播这些消息以及会把这些消息和相关评论通过什么平台和渠道传递给谁。也就是说,在奥巴马未启程之前,他的外交团队就已经为他精准推荐了谈论话题并圈定了关键传播源、传播渠道乃至传播模式,近乎未卜先知地做到了对巴西民众的精准接触与高效政治营销。到目前为止,这一案例堪称大数据介入公共外交的一次经典实战运用。

(二) 案例 2: 大数据与冲突预防: 如何进行危机预警并开展预防性外交?

所谓冲突预防,指的是国际组织或国家行为体用以阻止他国或地区冲突爆发、降低冲突烈度和控制冲突蔓延的一系列和平活动。在实践中,冲突预防通常包括早期预警、危机管控、冲突解决、冲突后的维持和平、塑造和平以及援助重建等活动。2013 年,美国国际和平研究所(United States Institute of Peace)的一份报告指出,作为一个实践领域,大数据冲突预防预示了一种前所未有的应用前景。大数据本身就是一场预防革命,它不仅与冲突预防密切相关,而且最适合应用于冲突预防。与传统预防外交的数据贫乏和预案不周不同,大数据的即时信息抓取能力、远距离监控能力以及深度挖掘能力,可以使冲突全貌乃至冲突的每一个细节被观察者一览无余。^②全球冲突预防专家大卫·基尔卡伦(David Kilcullen)和亚历克斯·考特尼(Alexa Courtney)共同撰文指出:“掌控大数据并使之动态可视化有助于我们在最复杂、最危险的环境中精确认知冲突的类型及其特征,并最终有助于我们采取最富成效的人道主义和发展援助

^① Ali Fisher and David Montez, “Evaluating Online Public Diplomacy using Digital Media Research Methods: A Case Study of #ObamainBrazil”, InterMedia White Paper, 2011, http://www.audiencescapes.org/sites/default/files/InterMedia_ObamainBrazil%20and%20New%20Media%20Research_Fisher%20and%20Montez.pdf, last accessed on 15 March 2015.

^② Sheldon Himelfarb and Megan Chabalowski, “Media, Conflict Prevention and Peacebuilding: Mapping the Edges, USIPeace Briefing”, United States Institute of Peace, October 2008, <http://www.usip.org/publications/media-conflict-prevention-and-peacebuilding-mapping-edges>, last accessed on 15 March 2015.

行动。”^①

当然也有人不好大数据在预防性外交领域的应用前景,认为大数据应用不过是一些理想主义者的美好想象,不但无助于冲突预防反而会常常适得其反。因为基于客观数据的监控往往无法洞悉事件的本质也难以触及人们的真实情感,这并不关乎数据的体量大小,海量的数据分析要么因数据虚假和数据噪音过大而难以反映事件的真实状况,要么被权力当局或利益集团滥用而成为镇压异己的工具。大数据冲突预防导致的过度干预行为最终会使得本不至于失控的事态不断蔓延、不断升级。^② 就此而言,大数据应用在很多情况下不是预防冲突的灭火器而是火上浇油的助燃器。

尽管存在争议,但大数据运用于冲突预防并辅助预防性外交的案例日渐增多,在现实国际政治中有迹可寻。其一,联合国“全球脉动”(Global Pulse)计划。该计划正式启动于2009年,联合国秘书长潘基文为帮助各国政策制定者实时了解冲突蔓延、危机进程并及时调整政策、提升社会保障能力而推出。该计划强调推进三个战略:(1)研究、创新方法和技术,分析实时信息数据,定位前期隐患;(2)集中整合开放源码技术包,分析实时数据并共享预测结论;(3)建立脉动实验室全球集成网络,推动脉动项目在各个国家层面的试点。^③ 总体而言,该计划的目标在于使用数字化的预警信号来预先指导全球和各国援助计划,预防出现贫困倒退、疾病蔓延、宗教纷争和部族仇杀的局面。目前,该计划除提供实时数据分析报告外,还提供免费的软件工具,供专家挖掘实时数据、分享知识和经验以做出基于实证的决策。据悉,冲突频仍的乌干达和印度尼西亚是首批参与这一计划的国家。

其二,哈佛人道主义危机定位与早期预警计划(HHI Program on Crisis Mapping and Early Warning)。^④ 作为哈佛人道主义倡议(Harvard Humanitarian Initiative)的一项科研计划,该计划发起于2007年,旨在将信息通信技术应用于冲突预防和灾难救援以丰富冲突预防知识和提高灾难救援效果。该计划的实施主要是通过数据挖掘、定位、人群搜索等数据分析手段,辨识不同冲突场景下的紧急人道主义救援模式与最优决策程序。该项目启动以来已吸引了大量的外交人员、技术团体以及国际组织和非

^① David Kilcullen and Alexa Courtney, "Big Data, Small Wars, Local Insights: Designing for Development with Conflict-affected Communities", <http://voices.mckinseyonsociety.com/big-data-small-wars-local-insights-designing-for-development-with-conflict-affected-communities/>, last accessed on 13 December 2014.

^② Emmanuel Letouzé, "Can Big Data From Cellphones Help Prevent Conflict?", April 10, 2013, <http://www.tuicool.com/articles/IfQ7vy>, last accessed on 15 March 2015.

^③ Global Pulse, "Big Data for Development: Challenges & Opportunities", May 2012, <http://www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGglobalPulseJune2012.pdf>, last accessed on 15 March 2015.

^④ Jennifer Leaning and Patrick Meier, "The Untapped Potential of Information Communication Technology for Conflict Early Warning and Crisis Mapping", Working Paper Series, *Harvard Humanitarian Initiative* (HHI), 2009.

政府机构参与,项目的诸多数据应用分享(data sharing)活动有效地促进了不同背景的人道主义救援团体之间的经验交流与知识共享,同时该项目向所有感兴趣的机构和个人开放研究数据源,以促进人类社会对各类冲突动因的深层理解,并探索和改善人道主义救援知识。

其三,Ushahidi 目击者数据平台。^① 这是迄今所有冲突预防大数据平台中最具草根性质、同时也是最具开源特征的冲突预警平台。该平台由肯尼亚博客作者奥瑞·奥科罗(Ory Okolloh)所开发,最初是通过邮件和电话自动收集群众的暴力事件目击报道,并在电子地图上实时显现地理位置。之后,该平台在一位高效的草根领导下,利用开源技术发展成为一个目击者可以通过社交媒体上传暴力、死亡和冲突信息的大数据集散中心,虽然没有官方授权、没有正式的指令机制甚至没有复杂的信息传送协议,但是其危机预警能力却速度惊人甚至超过媒体和当地政府。

就技术特征而言,Ushahidi 目击者平台擅长收集那些可用但分散的信息,即任何人都能设立自己的服务账号并通过短信来收集、定位信息,然后将个人目击所得的碎片信息过滤、聚类、按照时间序列编织成一幅整体性画面。自 2007 年诞生于肯尼亚选举暴力事件以来,Ushahidi 目击者平台目前已在全球 150 多个国家的冲突预防中得到应用,其成功预警案例包括:追踪刚果民主共和国内的各类暴力事件,监督印度和墨西哥的投票地点以及预防投票者作弊,报道东非国家的人道主义救援药品供应量以及海地和智利的震后搜救等。^② 简言之,Ushahidi 目击者平台在民主政策的制定、危机管理和维护公共卫生等领域都发挥了日益重要的预警作用,这是一部人人都可以参与和分享的开源数据平台,它的应用和普及正使得诸多国家的预防性外交活动成为可能。

(三) 案例 3:大数据与国家安全:基于数据挖掘的情报竞争

有关大数据在舆情监测和情报竞争领域的应用,国内外诸多文献已深入探讨,现实案例俯首可拾,其中最著名的莫过于“棱镜计划(PRISM)”和“X-DATA”项目。

“棱镜计划”正式名号为“US-984XN”,由美国国家安全局于 2007 年小布什政府时期开始执行,其目标是进入包括微软、思科、雅虎、谷歌、苹果等九家公司在内的美国国际网路公司的中心服务器挖掘数据、收集情报,监听内容涉及电子邮件、视频和语音交谈、影片、照片、谈话内容、文件传输、登录通知以及社交网络细节。具体工作原理可简单概括为:(1)进入数据公司服务器,在互联网的骨干网路由器上安装监听设备,监听来往流量,包括邮件、聊天记录、文件传输、社交网络资料等所有明文传输的内容;

^① Ushahidi - Official Site, <http://www.ushahidi.com>, last accessed on 15 March 2015.

^② Ushahidi, <http://www.realtechsupport.org/UB/MRHH/papers/CollectiveIntelligence/Ushahidi.pdf>, last accessed on 15 March 2015.

(2)在谷歌、雅虎、微软等搜索引擎上植入键盘追踪程序,记录并过滤用户地址、搜索关键词,分析用户日常行为、发现并监视恐怖行为。^①

“X-DATA”项目是美国国防部高级计划研究局(DAPRA)的大数据研发项目,主要内容为大数据在军事情报和国防领域项目的实战运用。项目启动于2012年,计划每年投入资金2500万美金,持续四年,项目旨在“保持美国的技术领先地位、防止被潜在对手意想不到的超越”。目前,“X-DATA”研发主要涉及四个技术领域(Technical Area, TA):(1)可扩展的分析和数据处理技术(TA1);(2)可视化的用户界面技术(TA2);(3)软件集成研究(TA3);(4)评价(TA4)。简而言之,DARPA在“X-DATA”项目上的主要目标是通过开发大容量数据分析所需的可扩展算法,以处理分布式数据存储库中的不规则数据;通过开发高效的人机互动设备和可视用户界面技术,以在多样化任务中更好、更快地执行操作。^②

此外,大数据在情报反恐领域的应用也较为引人注目。美国早已将声纹信息管理系统应用于军事、情报和国家安全等重要部门。通过综合利用有关恐怖分子的各种信息,包括通话、交通、购物、交友、电子邮件、聊天记录和视频等,对恐怖行为进行事前预警和事后分析排查,越来越成为美国的反恐手段之一。譬如,美国政府的国家情报网格(NATGRID)计划主要是利用大数据分析技术来研究分析情报部门和执法机构等21个不同信息来源的大数据,以帮助跟踪嫌疑人,防止恐怖袭击。NATGRID设立于孟买恐怖袭击之后,其主要职责是就针对美国的恐怖袭击进行数据挖掘、实时分析,以清晰勾勒出犯罪嫌疑人的行踪和面貌。该项目的一个重要组成部分就是开发某种技术程序以跟踪一个人的数字指纹,从而使执法机关更容易追踪犯罪嫌疑人,使政府部门可以更深入分析事件。

四 大数据应用于外交:未来前景与关键障碍

有关大数据应用于外交决策可以明显提高决策质量和优化外交资源配置,有两个案例甚为典型:其一,外交百科(Diplopedia),这是一部大数据采集和知识分享案例。

^① PRISM Collection Manager, “PRISM/US-984XN Overview or The SIGAD Used Most in NSA Reporting Overview”, April 2013, https://www.eff.org/files/2013/11/21/20131022-mondeprism_april_2013.pdf, last accessed on 15 March 2015.

^② See Patrick Meier, “DARPA’s Crisis Early Warning and Decision Support System”, *Conflict Early Warning and Early Response*, March 20, 2010, <https://earlywarning.wordpress.com/2010/03/20/early-warning-decision-support>, last accessed on 15 March 2015; 庄林、沈彬:“美国国防部大数据项目研发与应用”,《国防科技》2013年第3期,第59-60页。

外交百科隶属美国国务院下属的“E-外交办公室”,是一个运行于美国政府内网、不对公众开放的在线外交事务“百科全书”,其功能和创立宗旨是鼓励外交人员或政府专家、学者以自身经验和所见所闻提供情报信息、解决视角、分享知识和外交经验,通过经验共享的方式来促进问题解决、改善外交决策和外交执行状况。截止到 2012 年 2 月,外交百科已有超过 5000 名作者发表 16300 余篇专业外交情报文档,堪称美国国务院内部的开源数据平台和信息自动收割机。^①

其二,外交骇客(DiploHack),这同样也是一个外交知识与业务协作平台,主要用于荷兰大使馆与瑞典大使馆之间合作应对一些共同关心的全球性挑战。外交骇客是专业外交人员、社会企业家、技术专家、新闻记者、学术界、非政府组织和商业机构之间聚集知识和经验的平台,其创立目标如下:(1)创新外交协作模式和提供革新性的外交手段;(2)通过协同创新使外交官和外交业务本身熟悉、适应市民社会以及技术产业发展;(3)通过协同创新使外交官熟悉社会企业、社交媒体和大数据,同时鼓励外交官深入理解技术产业背后的原初文化并改进外交实践;(4)积极探索社交媒体和大数据等新技术应用于外交活动的社会增值效应;(5)探寻社交网络与大数据在人际互动、国际关系特别是在公共外交中扮演的角色和作用。^②

综上所述,在外交领域尤其是在外交决策、外交规划和外交执行环节,大数据应用的年代已经来临,世界各国正在竞相进行基于数据驱动的外交理论创新与实践模式布局。在不久的将来,数据驱动外交或将成为业界常态。在国际关系领域,探寻大数据对传统外交理论的影响已成为学术时尚,譬如加拿大哥伦比亚大学正在启动一个名为 IRDTP(The International Relations and Digital Technology Project)的大数据国际关系研究项目。^③作为一个跨学科、跨领域的合作研究计划,该项目旨在以革新的思维和理念探究数字技术对国际关系运行及其理论假说的重塑。虽然具体的研究计划尚不明晰,但却标志着大数据及其分析技术已渗入国际关系研究和外交决策学理分析领域。

可以预见,随着数据累积速度的加快和低成本数据分析技术的普及,将会有更多国家的外交洞察建立在数据驱动的基础之上。但这并不等于说,基于数据的分析一定或必然优于基于经验的推理。概括而言,大数据辅助外交决策与外交执行之间至少存在着以下障碍和困境:

第一,数据障碍。首先,大数据研究的第一要求就是数据体量要大、数据来源要

^① About: Diplopedia, <http://www.state.gov/documents/organization/116059.pdf>, last accessed on 15 March 2015.

^② DiploHack, <http://www.diplohack.org>, last accessed on 15 March 2015.

^③ IRDTP, <http://www.irdtp.org>, last accessed on 15 March 2015.

广、数据类型要多样化,然而目前无论在市场还是高校、研究所均不存在专业的国际关系与外交数据收集平台,大数据研究的前端基础“数据”严重缺失。在此情形下“巧妇难为无米之炊”,即使再为高明的数据算法和再为精致的研究议程也难以进行大数据运算,因而也就难以洞察目前常规数据分析之外的知识创见。其次,数据收集和整理是一个高难度且又枯燥冗长的过程,不仅需要大数据挖掘和应用算法方面的专业人才参与,更需大量的人力、物力乃至资金投入,在目前数据科学技术高成本、非普及的状态下,有限的人文科研经费难以支持真正意义上的外交研究计划,但基于国家力量的外交决策和外交执行则另当别论。再次,外交领域的大数据挖掘和数据算法人才严重缺失,放眼全球外交界,各国争相拟定数据人才培养计划并投入巨额资金以期抢占大数据时代的外交竞争制高点,在这方面美国的核心国际关系院校如哈佛大学贝尔福中心、乔治城大学爱德蒙·沃尔什外交学院都已开设相关课程并授予相应学位^①,而我国大学和科研院所里的大数据专业外交人才培养却还未正式起步。

第二,技术障碍。大数据分析是一门融合技术,通常基于数据驱动的决策分析和外交执行过程,需要两方面人才的知识结合与分工合作:(1)需要职业外交官拟定定向数据需求;(2)需要专业数据科技人员设计数据采集和挖掘方案,勾勒算法模型和进行数据运算;(3)最后由专业外交官和专家解读数据分析结果和进行知识洞察,规划外交战略、改善外交实践。总体而言,大数据介入外交决策和外交实践既需要很高的业务知识门槛(由专业外交官设计研究问题和数据应用场景),同时又需要极高的数据分析水平(由数据科学专业人才精准采集数据和设计算法),可以说,大数据时代的数据驱动型外交是建立在跨学科、跨领域协同合作、协同创新的基础上。就此而言,大数据外交的有效执行仅凭职业外交知识或数据科学知识是不够的,两个领域的人员和知识需要协作和沟通。因而,基于数据驱动的外交决策和实践如何跨越知识屏障,将成为高质量外交的关键决定因素。

第三,制度障碍。外交不同于商业领域和社会生活领域,其数据收集标准将会遇到更为严格的制度约束和法律障碍。首先,这一学科的数据收集涉及国家安全、国防机密和他国公民隐私等问题,虽然当前很多国家都提倡数据开放(open data)和政府开放(open government),然而即使数据是开源和开放的,一国外交机构对他国基础人口、基本经济和精确国防开支数据的收集会不会演变成两国间的情报间谍活动,在没有任何国际性条约约束的前提下确实令人忧虑。进而,这又引发了一个重要的数据安全问

^① 大数据相关培训课程参见 Big Data Training Courses, <http://techjobs.sulekha.com/big-data-training>; Big Data & Global Affairs, <http://courses.georgetown.edu/? CourseID=MSFS-546>; Big Data Analytics, <http://www.extension.harvard.edu/courses/big-data-analytics>, last accessed on 15 March 2015。

题:即使是出于善意竞争的需要,一国对其外交目标对象国进行数据收集和分析,哪些数据可以收集、哪些不可以收集?哪些数据国家间可以自由交换、哪些不可以?其次,任何精细化的数据收集最终都将触及个人隐私问题。对于外交决策和外交实践来说,数据收集的颗粒度越小,知识发现和政策洞察的可能性就越大。换言之,对一个国家人口数据的采集到省的颗粒度显然不如到市县的颗粒度洞察力强;对一个恐怖分子地理位置知悉相差数十里和数米的价值显然是不一样的。然而,数据采集通过什么方式、采集到什么程度又用于何种目的才算侵犯个人隐私,才能为各国隐私保护法所允许?这将是一个严肃的国内法律问题,同时也是一个跨越主权疆界到另外一国主权管辖区域内采集数据是否合法的国际法问题,更是涉及科学技术是否可以无边界、无约束使用的道德伦理问题。

五 结语

概而言之,在新技术应用的驱动下,外交决策、规划与战略执行的大数据时代正悄然逼近。正如其他许多社会生活领域无法回避大数据的侵袭一样,大数据介入国际关系和外交领域也将是大势所趋。在这方面,世界各国政府以不同方式表达了对大数据潜能的关注和重视,不仅纷纷拟定数据开发和数据开放战略,同时各国也通过各种技术手段和制度屏障加强了对各自数据资产的保护。在此背景下,大数据跃然进入外交领域,势必带来一系列外交模式的创新,同时也将产生许多诸如数据安全和隐私保护等以前所无法想象的新型国际摩擦。与传统的小样本分析和经验决策相比,大数据时代的外交决策和战略执行将具有明显的多源异构数据分析和数据驱动特征,在容纳了大体量非结构化和半结构化数据之后,先前的国际关系假设和外交指导原则,可能受到来自外交实践领域一定程度的实证冲击和挑战。总体而言,大数据应用于外交领域的未来前景广阔且政策启示意义重大,但也面临数据、技术和制度等多层面的障碍,需要加以克服。

(作者简介:董青岭,对外经济贸易大学国际关系学院副教授;责任编辑:张海洋)