

政治科学预测方法研究

——以选举预测为例

王中原 唐世平

内容提要:预测是人类的基本认知活动。作为社会科学预测性研究的重要领域,选举预测是促进政治学理论和方法创新的动力。选举预测不仅限于民调,科学的选举预测可以划分为四类范式:意见聚合范式;模型范式;混合范式;大数据范式。每类范式包含若干种预测方法,不同预测方法在准确性、超前性、解释力、可重复性、中立性和预测成本等评判标准上各具优势,预测方法的选择是基于预测目标和应用场景的综合权衡。目前,选举预测领域呈现多元化、交叉互补、跨学科的发展趋势;其未来发展方向是“理论驱动”与“数据驱动”相结合、微观预测变量与宏观预测变量相结合、预测与解释相结合,以政治科学为基础,通过跨学科合作和前沿技术手段推进预测范式的创新,并在预测实践中不断更新升级。预测性研究与解释性研究可以构成“双螺旋”的共生互补关系,共同致力于增进人类对复杂政治活动的理解和研判,提升政治学研究的科学水平和政策应用价值。

关键词:选举预测 预测方法 政治学预测性研究 预测

一、引言

预测是人类的基本认知活动^①,是根据既有信息判断和指导下一步战略规划与行动决策的思维和过程。科学地预测未来是推进学术探索和不竭动力,也是撬动范式和方法革新的重要支点。作为社会科学预测性研究的重要领域,选举预测是带动政治学理论和方法创新的关键力量。运用科学方法对选举进行精准预测不仅可以指导竞选活动,满足公众对选举结果的好奇心,辅助制定相应的外交策略,同时还可以推进学科研究,提升政治学研究的科学品质和应用价值。

选举预测始于美国,随后发展到预测欧洲各国的选举,近年来对发展中国家的选举预测也方兴未艾^②。科学的选举预测不仅包括民调,还包括意见聚合、模型构造、大数据、仿真模拟等不

① Jakob Hohwy, *The Predictive Mind*, Oxford: Oxford University Press, 2013; Andy Clark, *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*, Oxford: Oxford University Press, 2015.

② M. S. Lewis-Beck and é. Bélanger, “Election Forecasting in Neglected Democracies: An Introduction”, *International Journal of Forecasting*, Vol. 28, No. 4, 2012; D. Walther, “Picking the Winner(s): Forecasting Elections in Multiparty Systems”, *Electoral Studies*, Vol. 40, 2015.

同范式。在西方,选举预测汇聚了一大批跨学科研究力量,形成了相当规模的选举预测市场,获得了大量来自社会和官方的资金支持。《国际预测研究期刊》(*International Journal of Forecasting*)、《选举研究》(*Electoral Studies*)等重要学术刊物相继发表一系列选举预测研究成果,2017年《科学》(*Science*)期刊专门刊发选举预测的相关前沿研究成果^①。相对而言,中国对西方国家选举研究特别是选举预测研究比较匮乏,公众和决策者对选举预测的认知尚停留在专家判断和民意调查阶段。这不仅不利于中国政治科学研究的推进,而且会让中国在相关外交领域陷入被动。

政治学研究是否应该参与预测活动、什么是科学的预测、科学的选举预测怎么做,围绕这几个问题,本文首先将从政治学预测性研究的视角出发,讨论科学预测的界定及其评判标准,据此对现有各类选举预测范式及其具体预测方法进行系统考察,分析各类方法的科学原理、预测效力和相对优劣势;其次,归纳选举预测研究的演进规律、困境局限、前沿领域以及突破方向;最后,本文将讨论政治学预测性研究的学科定位和发展前景,及其如何与传统解释性研究相互助益,共同推进中国政治学学科体系建设。

二、政治学的预测性研究

政治学研究乃至整个社会科学研究是否应该参与预测活动,一直是个充满争议的话题。其争议的内容涉及政治学研究的根本任务、政治事件的可预测性、政治预测的准确性、以及预测的应用途径等方面。

首先,社会科学的研究任务基本可分为解释(explanation)和预测(forecasting)两类。“解释”关注因果关系中的自变量(X)及其影响系数 β 和显著性,试图解释已经发生的事件或结果(Y)的影响因子和致因机制;“预测”则关注因果或相关关系中的输出(Y),旨在根据已经掌握的各类信息和关联结构去研判未来可能出现的结果。长期以来,政治学专注于“向后看”的描述分析、因果识别和假设检验等,聚焦于解释业已发生的政治现象,对“向前看”的科学预测研究缺乏热情和关注,甚至相当排斥。究其原因,一方面,研究者们认为政治学研究的根本任务在于解释而非预测,预测意味着“干预”,破坏了社会科学的“价值中立”;另一方面,预测性研究的科学方法薄弱、难度较大、风险极高。

其次,研究者们对政治事件和行为的可预测性充满质疑。一方面,人类社会是一个动态变化而非静态机械的复杂系统,有着多元交错的影响因素和作用机制,且受到不可观测要素和突发事件的影响,以致很多学者认定人类活动本质上是不可预测的;另一方面,以往预测效果较差,未能预测到诸多重大政治事件的发生,包括苏联解体、“911”恐怖袭击、阿拉伯之春、英国脱欧等“黑天鹅事件”,甚至有统计发现,专家预测活动的准确性跟喝醉的黑猩猩掷飞镖差不多^②。2016年,特朗普当选美国总统,一时间关于选举预测的悲观声音甚嚣尘上。例如,《纽约时报》评论道,“今夜数据死了”、“民调无法从统计上计算出不可预估的和高深莫测的人性细节”^③。此外,当前大量凭借个体感知的预测活动不仅因缺乏科学方法而频频出错,而且泛滥化和娱乐化^④,失去

① R. Kennedy, S. Wojcik and D. Lazer, “Improving Election Prediction Internationally”, *Science*, Vol. 355, 2017.

② Philip. E. Tetlock, *Expert Political Judgment: How Good Is It? How Can We Know?* Princeton, NJ: Princeton University Press, 2006, pp. 77 ~ 78.

③ J. Rutenberg, “A ‘Dewey Defeats Truman’ Lesson for the Digital Age”, *New York Times*, 9 November, 2016.

④ 张千帆:《我们为何如此热衷预测?》,英国金融时报(FT)中文网,2020年1月1日。

了公众信任。

上述质疑和批评提醒政治学者必须对预测性研究保持审慎和谦卑,但这绝不是放弃预测性研究的理由。首先,研判和指导实践是社会科学研究的重要目标,即使描述性研究和因果推断中也带有预测和启发未来的意涵,正如汤普森(Thompson)和德尔(Derr)所言,“好的解释可以预测”,进而验证解释的可靠性^①。解释和预测并非是孤立的,更不是对立的,相反可以相互助益、互为补充^②。其次,我们应该将错误运用预测方法导致的问题与预测本身的问题区分开来,预测失准说明目前预测方法尚不够科学,需要进一步探索和改进,但不能将“婴儿和洗澡水”一同丢弃。

在开展预测性研究之前,需要明确什么可以预测、什么难以预测。海因德曼(Hyndman)和雅典娜梭普洛斯(Athanasopoulos)指出,“可预测性”取决于三个限定条件:第一,我们对产生结果的影响因素理解多少?第二,目前拥有多少数据?第三,预测活动是否会改变预测对象的运行轨迹^③?在此基础上,本文增设“预测手段和方法是否科学多元”作为第四个限定条件。由此将政治事件分为三类:高度可预测、高度不可预测和审慎可预测(参见表1)。高度可预测的政治事件是我们对其变量关系和影响机制理解充分,数据的体量、质量和可获得度都很高,预测活动不会影响预测对象的运行轨迹,预测的手段和方法科学多元的政治事件。高度不可预测的政治事件则相反,“黑天鹅事件”之所以极难预测是因为其缺少历史数据,研究者对其致因机制知之甚少,同时缺乏科学的预测方法。诸多政治事件介于二者之间,属于“审慎可预测”的范畴,即我们对该类事件有一定的研究积累,并掌握一定量的可靠数据,预测活动之于对象运行轨迹的影响基本可控,同时预测方法处在不断丰富和发展当中。这也是政治学预测性研究应当致力攻克的领域,选举正是属于“审慎可预测”的政治事件范畴。

表 1 政治事件可预测性的等级和限定条件

	理解程度 (变量关系和影响机制)	数据质量 (信噪比/丰富性/可及度)	预测是否改变 对象轨迹	预测手段和方法 是否科学、多元
高度可预测	充分	高	否	是
审慎可预测	部分	中	可控	发展中
高度不可预测	甚少	低	是	否

总之,科学的预测性研究是政治学乃至整个社会科学发展相对滞后的学术领域。目前,无论公众还是学者对该领域尚存在诸多质疑和误解,打消这些疑虑的关键在于发展出更加精准可靠的预测方法。近年来,大数据、人工智能、仿真模拟等方法的兴起为社会科学预测性研究提供了新的可能和契机^④。伴随着数据的指数级累积、计算机算法的突破、跨学科研究的兴盛,诸多“向前看”的定量预测方法应运而生,并被运用于政治学、经济学、社会学、国际关系等领域,涵盖选

① N. S. Thompson and P. Derr, “Contra Epstein, Good Explanations Predict”, *Journal of Artificial Societies and Social Simulation*, Vol. 12, No. 1, 2009.

② J. M. Hofman, A. Sharma, and D. J. Watts, “Prediction and Explanation in Social Systems”, *Science*, Vol. 355, No. 6324, 2017.

③ Rob J Hyndman and George Athanasopoulos, *Forecasting: Principles and Practice*, Melbourne: OTexts Publisher, 2018, p. 4.

④ D. Lazer. , et al. , “Computational Social Science”, *Science*, Vol. 323, No. 5915, 2009; 孟天广:《政治科学视角下的大数据方法与因果推论》,《政治学研究》,2018 年第 3 期。

举、族群冲突、革命叛乱、政治风险等政治学议题的研究^①。科学预测可以成为政治科学探索和发展的前沿领域。既运用政治学理论、知识和方法指导预测活动,又通过预测驱动、丰富和优化政治学研究,将预测性研究与解释性研究相结合,有望提升政治学的科学水平和应用前景。

三、选举预测的科学方法

选举是西式民主政治的核心事件,熊彼特(Schumpeter)认为“竞争性选举”是民主体制最关键的构成要件^②。一国的选举结果不仅将对其国内政治和治理产生重大影响,甚至会波及国际局势。对西式选举的科学预测不仅能够推动政治科学研究,而且可以服务于外交和经贸决策。选举预测是社会科学预测性研究中历史较为悠久、方法相对成熟的领域。然而近年来,发生于英国脱欧公投、美国总统大选、马来西亚国会选举中的一系列黑天鹅事件让我们重新审视选举预测。同时,选举误判导致相关国家(包括中国)政策应对和外交事务陷入被动局面,使得更新选举预测方法更具必要性和紧迫性。

科学的选举预测怎么做、如何构建选举预测的综合评判体系、当前选举预测方法有哪些、各自的优势和缺陷怎样,选举预测的发展困境、前沿领域和突破方向为何,这些问题都呼唤政治学研究的学理探索。“预测已死”严重夸大了选举预测遇到的困难和瓶颈^③,预测困难或失准并不是我们放弃探索预测的理由,而是我们推动科学预测方法创新的契机。

选举预测是指在选举结果发布之前预判结果的研究活动^④,其具体任务可分为预测总体得票率、预测席次或选举人团、预测选举胜负、以及预测政府组成方式等方面。通常精度要求越高,预测链条越长,预测的难度就越大,预测失误的风险也越高。选举预测可以追溯到19世纪下半叶的美国,早期因为技术手段和数据资源的缺乏,预测活动主要依靠非科学的个人判断。20世纪中叶以来,随着研究方法和技术路线的革新,各种科学的量化预测方法相继出现。然而,大量非科学的预测手段仍长期存在,其中包括通过政客、观察家或学者的个体判断进行预测;通过历史上与大选结果相近的同趋势小选区来预测;通过“征兆”预测,例如“倘若复活节在四月,民主党会赢”,“姓氏最长的候选人当选”,等等^⑤。这些方法虽能在某届选举中预测较准,但其缺乏科学依据,要么混淆了相关性与因果性,要么受制于个体层面的认知偏差,不具备解释性、透明性和可重复性。选举预测不是拍脑袋和碰运气,必须依赖科学的研究路径和预测机理。因此,本文将聚焦科学的量化预测方法^⑥,并将在构建选举预测评价体系的基础上考察现有各类量化预测范式的原理和优劣。

(一) 选举预测的科学评判标准

评判选举预测成功与否需要一套系统、客观的科学标准,准确性(accuracy)无疑是最重要的

① 参考《科学》期刊上关于“预测”的系列文章:2017年2月3日, <http://science.sciencemag.org/content/355/6324>, 2019年3月11日; 庞珣:《定量预测的风险来源与处理方法——以“高烈度政治动荡”预测研究项目的再分析为例》,《国际政治科学》,2017年第3期; 董青岭:《大数据安全态势感知与冲突预测》,《中国社会科学》,2018年第6期。

② 约瑟夫·熊彼特:《资本主义、社会主义与民主》,第415页,商务印书馆,1999年版。

③ R. Kennedy, S. Wojcik and D. Lazer, “Improving Election Prediction Internationally”, *Science*, Vol. 355, 2017.

④ 在预测研究中,出于训练模型的需要,也会对历史结果进行回溯性预测(after-the-fact forecasting)。

⑤ Michael S. Lewis-Beck and Tom Rice, *Forecasting Elections*, Washington, DC: Congressional Quarterly Press, 1992, pp. 3~6.

⑥ 此处并非否定定性预测,质性选举预测也可以通过流程设计和分析推断来提高预测精度。参见 S. M. Herzog and R. Hertwig, “The Wisdom of Many in One Mind: Improving Individual Judgments with Dialectical Dootstrapping”, *Psychological Science*, Vol. 20, 2009.

衡量指标。除了“准确性”之外,政治学预测性研究还须兼顾预测的超前性(lead time)、解释力、透明性、可重复性、中立性等指标,并根据具体预测目标和研究任务在不同标准之间做好综合权衡。

1. 如何认识准确性?准确是最关键的指标,也是实现预测功能的基本要求。然而,社会科学定义的准确性与媒体和大众理解的准确性略有不同。首先,对于不同预测任务,准确性的标准存在差异,例如预测选举胜负与预测得票比率对准确性的精度要求不同,媒体和大众更关心谁胜谁负,学术研究则更加追求预测精度。其次,预测的成败和统计学意义上的准确是不同概念,科学的选举预测给出的是某种可能的结果、有关这个结果的概率以及预测的误差估计。第一,预测结果是一种概率而非固定不变的数值;第二,任何预测都存在偏误,统计学上已经发展出包括均方根误差(RMSE)、平均绝对误差(MAE)、平均绝对百分误差(MAPE)在内的预测误差度量方法^①。选举预测的特殊性在于,误判了胜负的预测结果,很可能落在统计误差允许的范围之内,即统计学上准确^②。最后,没有绝对准确只有相对准确,所有预测方法都存在局限性,无法达到百分百精准,只能追求相较于此前预测或者其他预测方法更加准确。

2. 选举预测的综合评价指标。第一,平衡准确性与超前性。准确性是选举预测的必要但非充分条件;除了准确性之外,选举预测还须兼具超前性。通常来说,距离投票日越近,预测者获得的信息越完全,预测准确性越高。然而,预测的功能在于指导决策,这就要求足够的超前性,以预留充裕的时间窗口研制对策^③。随着选举日临近,预测的政策应用价值递减,选举当日的民调(如“出口民调”)除了提供媒体报道的噱头外,已不具政策价值。有学者将选举预测方法分为“长视野”(long view)和“短视野”(short view)两类,并认为依赖模型的长视野方法侧重“超前性”,以民调为代表的短视野方法追求“准确性”^④。因此,选举预测需要在准确性和超前性之间做出某种权衡,在保证准确性的前提下最大限度地提升超前性。第二,平衡预测力和解释力。选举预测作为一项学术研究,除了追求预测力之外,还必须兼顾解释力,即通过预测促进假设检验、机制解析和学理推断,提升人们对选举制度和选举行为的洞察力。量化预测的模型设定、变量选取、数据分析不仅要满足精度追求,同时其预测过程和结果需具备学理上的可解释性。一个高度复杂的预测模型可能具有超强预测力,但如若无法从学理上对其变量、参数和模型构造进行解释,就会产生“预测黑箱”,无法获知中间机制和理论意涵^⑤,也无法服务于选后分析。例如,机器学习方法可以通过特征工程(feature engineering)构造出各种特征变量以提升预测力,但其模型过度复杂,往往超出社会科学的解释范畴。因此,选举预测需要兼顾预测力和解释力,保持预测的学理价值。第三,平衡计算和判断。量化的选举预测将数据(data)变成信息(information)进而变成知识(knowledge),计算机算法可以让预测活动摆脱专家个体判断的偏误,提高预测的科学

① R. J. Hyndman and A. B. Koehler, “Another Look at Measures of Forecast Accuracy”, *International Journal of Forecast*, Vol. 22, 2006.

② 这种情况在“领先者当选”(First Past the Post)的选举制度之下尤其明显,通常严肃的媒体报道中会说“得票太接近以至于尚无法预判胜负”(too close to call)。

③ 林策认为较理想的超前时间为3~4个月。D. Linzer, “Dynamic Bayesian Forecasting of Presidential Elections in the States”, *Journal of the American Statistical Association*, Vol. 108, 2013.

④ M. S. Lewis-Beck and C. Tien, “Election Forecasting: The Long View”, in D. Pervin, ed., *Oxford Handbooks Online*, Oxford: Oxford University Press, 2016.

⑤ J. M. Hofman, A. Sharma, and D. J. Watts, “Prediction and Explanation in Social Systems”, *Science*, Vol. 355, No. 6324, 2017.

性和精准度。然而,纯粹技术驱动的预测也存在较大风险,甚至带来灾难性后果^①。首先,量化预测在数据收集和编码、变量选取和测量、模型建构和优化等阶段都少不了专家知识和政治学理论的指导;其次,计算机无法精准处理数据中的“噪音”(noise)^②,需要人工监督。当然,科学预测必须谨慎把握“学者自由度”(researcher degrees of freedom)。一方面,判断的标准和过程必须有理可依、有据可循,保证预测过程的透明性和预测结果的可重复性;另一方面,判断标准和计算过程必须公开,接受学术共同体的监督。总之,科学的预测应该综合计算和判断,将“知识驱动”与“数据驱动”相结合。

(二) 量化选举预测范式的类型和演进

选举预测研究聚焦西式民主国家。源于美国的各种选举预测方法被逐渐应用到对英国、法国、德国等国家以及欧盟议会的选举预测当中。近年来,选举预测方法在不断丰富的同时,也延展到对“边缘国家”“新兴民主国家”和“竞争性威权国家”的选举预测当中^③。不同国家的政治体制、选举制度、经济社会环境不同,对选举预测带来诸多挑战,同时推动着预测方法推陈出新。

经过八十多年发展,选举预测形成了一系列科学量化的预测方法,并随着计算社会科学的兴起呈现出新的创新态势。选举预测方法不仅限于民调,本文将选举预测的科学方法划分为四大类:意见聚合范式(Aggregators);模型范式(Models);混合范式(Synthesizers);大数据范式(Big Data)。每一类范式下又可细分出不同的预测方法,每种方法都各具比较优势,也面临不同的问题和挑战。

1. 意见聚合范式。意见聚合范式通过测算调查对象对候选人/政党的支持意见或基于意见的支持行为来预测候选人的获选概率,包括民意调查(polls)、政治博彩市场(prediction market)、群体智慧(wisdom of crowds)、以及专家调查(expert survey)等方法。虽然这些方法的数据来源和技术路线不尽相同,但其共性在于根据抽样或特定分布来聚合不同个体意见(或基于意见的客观行为),由此克服个体判断的偏差,获得更为可靠的预测结果。

首先,最为人熟知的当属选民意调查。第一种是“简单民调”,即基于某种选民分布结构的抽样民意调查,通过邮寄、电话、网络、入户等方式了解样本选民的投票意向,由此推测整体的投票趋向^④。然而,简单民调面临抽样调查本身的各种问题,包括拒访率过高、样本量过小、样本结构不科学、测量方式不合理、成本太高,等等^⑤。此外,不同民调机构往往具有各自的政治倾向性,即所谓“机构效应”(polling house effects)^⑥。这些问题都可能导致民调结果失准,也是近年来民调预测表现欠佳的原因。第二种是“综合民调”,为了克服单一民调的弊端,研究者们尝试将不同民调数据进行加权综合,以期让不同民调的“机构效应”相互抵消,减少误差和噪音^⑦。该方法

① B. Jasny and R. Stone, “Prediction and Its Limits”, *Science*, Vol. 355, Issue 6324, 2017.

② Nate Silver, *The Signal and the Noise: Why So Many Predictions Fail-but Some Don't*, London: Penguin Press, 2012, p. 17.

③ 参见 *International Journal of Forecasting* 上的专刊“Election Forecasting in Neglected Democracies”, 2012年10月~12月。

④ 最早的选举民调可追溯到20世纪30年代,以盖勒普(Gallup)的民调方法为代表,但选举民调的广泛运用要到有电话广泛普及的20世纪60年代。

⑤ R. S. Kenett, D. Pfeiffermann, and D. M. Steinberg, “Election Polls—A Survey, A Critique, and Proposals”, *Annual Review of Statistics and Its Application*, Vol. 5, 2018.

⑥ A. Gelman and G. King, “Why Are American Presidential Election Campaign Polls So Variable when Votes Are So Predictable?” *British Journal of Political Science*, Vol. 23, 1993.

⑦ M. Blumenthal, “Polls, Forecasts, and Aggregators”, *PS: Political Science & Politics*, Vol. 47, No. 2, 2014.

近年来越发流行^①,最具代表性的当属奈特·希尔沃(Nate Silver)在“538网站”(FiveThirtyEight)做出的综合民调预测,其根据历史准确率、样本量、时效性等来加权多个民调数据,曾在2012年美国总统大选中成功预测了所有50个州的胜负结果,类似方法也被运用到英国、法国和荷兰的议会选举预测当中。第三种是“民调映射”,即测算历史上选前某个关键时点的民调支持率与最终得票率的相关关系得到映射函数,进而推算当届选举的结果,例如,根据美国劳动节当日的民调来预测选举结果^②。

其次,除民调外,学界还发展出诸多意见聚合的预测方法:其一,根据政治博彩市场的投注风向预测选举,其预测机理是选举投注是民众根据选情预判做出的投资决策,相当于让其回答“您认为谁会赢”?因此具有风向标意义。1884~1940年的统计数据显示博彩市场的预测准确率达到73%,其准确率甚至超过部分民调^③。当代最知名的政治预测市场是带有学术研究性质的“爱荷华选举对赌市场”(Iowa Electronic Markets)。其二,“群体智慧法”,又称“公民预测法”(citizens forecasts),即通过询问选民“您认为谁会赢”而非“您会投给谁”来预测选举结果,这种“期望调查”相当于把个体的预测汇集起来,集合群体的智慧,研究显示群体预判在意见多元、决策独立以及可综合的条件下比个体预判更加精准^④。其三,专家调查法,即通过调查选举专家群体来测算结果。虽然单个专家或政客对选举的预判常存在偏差,但借助一些科学方法(例如德尔菲法、群体提案评估法)对专家群体进行意见调查通常具有一定的预测力^⑤。

通过科学规则聚合个体意见或基于意见的行为来预测选举结果是当前最为流行、最受关注的预测范式。该范式一定程度上克服了个体判断的偏误,具有较强预测力,并且以其简单易懂、实时迅速、动态更新等优势长期占据预测市场和媒体报道的中心,仅民调一项在美国的市场规模就达180亿美元^⑥。然而,这类预测也具有明显局限:第一,非理论驱动,意见聚合是某个时点上群体判断的截面图,除了直观反映大家的意向和预期之外,缺乏对变量关系和影响机制的解释,对从学理角度理解选举贡献较少;第二,受制于技术手段局限,数据本身的代表性、测量的信度和效度、样本量等问题都将影响预测绩效;第三,无法预测投票率,意见聚合是通过样本支持率推测总体得票率,然而支持意向不等于选票,近年来投票率下降成为明显趋势,并成为决定选举的关键因素。上述问题都可能导致严重的预测偏差,也是近年来选举民调等意见聚合方法饱受质疑的原因。

2. 模型范式。模型范式依据影响投票决定的某些关键影响因子来构建模型以预测选举结果,其中典型的模型方法包括指数模型方法(index model)、结构性因素模型方法(structuralist model)和选举周期模型方法(election circle model)等,虽然模型构造不尽相同,但都是依托某种

① N. Jackson, “The Rise of Poll Aggregation and Election Forecasting”, in L. Rae Atkeson and R. M. Alvarez, eds., *The Oxford Handbook of Polling and Survey Methods*, Oxford: Oxford University Press, 2016.

② J. E. Campbell, “Forecasting the 2012 American National Elections”, *PS: Political Science & Politics*, Vol. 45, 2012.

③ P. W. Rhode and K. S. Strumpf, “Historical Presidential Betting Markets”, *The Journal of Economic Perspectives*, Vol. 18, No. 2, 2004; R. S. Erikson and C. Wlezien, “Markets vs. Polls as Election Predictors: An Historical Assessment”, *Electoral Studies*, Vol. 31, 2012.

④ A. E. Murr, “‘Wisdom of crowds’? A Decentralised Election Forecasting Model That Uses Citizens’ Local Expectations”, *Electoral Studies*, Vol. 30, 2011.

⑤ S. Armstrong, et al., “Forecasting Elections Using Expert Surveys: An Application to U. S. Presidential Elections”, *Paper Presented at the Annual Meeting of the American Political Science Association*, 2007.

⑥ J. Bohannon, “The Pulse of The People: Can Internet Data Outdo Costly and Unreliable Polls in Predicting Election Outcomes?” *Science*, Vol. 355, No. 6324, 2017.

变量关系和影响机制来预测选举。

第一,指数模型方法是通过收集候选人履历数据或能力评估数据来构建其潜力指数,进而预测其当选概率。该方法的预测机理是候选人资质和能力是影响投票的关键因素,其理论支撑是“选举个人化”^①。一种方法是根据候选人履历来建构其资质指数(bio-index),例如,阿姆斯特朗(Armstrong)和格雷费(Graefe)通过收集总统候选人的59个简历指标合成候选人的资质指数、建立指数预测模型,发现该模型可以成功预测1896~2008年29次美国总统选举中的27次,准确率甚至超过民调和博彩市场^②;另一种方法是基于对候选人在关键议题上的能力评估来构建候选人能力指数模型,据此预测选举结果^③,其议题通常包括经济、就业、移民、反恐、福利等。该方法预设选民会围绕选举当年的重大事项展开投票(issue voting),历史回测表明,该方法与政治博彩市场的预测力不相上下^④。指数模型方法很好地“变量化”了候选人资质和能力指标,在候选人因素日趋凸显的新媒体时代更具预测力,并对候选人提名具有指导价值。

第二,结构性因素模型方法是根据影响选举的关键结构性因素来构建回归预测模型。通常被纳入模型的预测因子包括经济变量(例如GDP增长、失业率)和选民对在任者的政绩评估,等等。该类模型融合了政治学中的“经济投票”理论、“在任者优势”理论(incumbent advantages)、“回溯性投票”理论,其预测机理是选民会根据当下的宏观经济状况以及在任者的过往执政表现来投票。例如,一个经典的预测模型表达式为 $\text{Vote} = f(\text{Incumbent Popularity}, \text{Economic Growth})$,即依据某个时点上的在任者受欢迎程度和经济增长状况两个宏观指标来预测选举^⑤。多年来,此类模型不断更新升级,发展出适用于不同国家的次模型,并被证明在总统制和在议会制下都具有一定的预测力^⑥。

第三,选举周期模型方法是根据历届选举的波动规律来预测选举,其预设同一个国家的同一种选举在时间上具有某种稳定的周期性规律,当届选举与此前选举之间不是孤立的,而是存在某种延续性。此方法对应的是“政治重组”的周期性理论,该理论认为选民的政治倾向和对政党的偏好会随着一定的时间周期移转或回归,进而带来政治的重组和轮替,这种选举“钟摆效应”在以美国和英国为代表的成熟西式民主国家表现较为明显^⑦。基于选举的时间序列特征,学者们构建出选举预测的自回归模型。例如,诺波特(Norpoth)根据美国总统大选的周期变化规律构建出二阶自回归预测模型($\text{VOTE}_t = 49.2 + 0.525 \text{VOTE}_{t-1} - 0.474 \text{VOTE}_{t-2}$),发现当届选举的结

① Alan Renwick and Jean-Benoit Pilet, *Faces on the Ballot: The Personalization of Electoral Systems in Europe*, Oxford: Oxford University Press, 2016.

② J. S. Armstrong and A. Graefe, “Predicting Elections from Biographical Information About Candidates: A Test of the Index Method”, *Journal of Business Research*, Vol. 64, No. 7, 2011.

③ A. Graefe and J. S. Armstrong, “Forecasting Elections from Voters ‘Perceptions of Candidates’ Ability to Handle Issues”, *Journal of Behavioral Decision Making*, Vol. 26, No. 3, 2013.

④ A. Graefe, “Issue and Leader Voting in U. S. Presidential Elections”, *Electoral Studies*, Vol. 32, No. 4, 2013.

⑤ M. S. Lewis-Beck, et al., “General Election Forecasts in the United Kingdom: A Political Economy Model”, *Electoral Studies*, Vol. 23, No. 2, 2004; H. Norpoth and T. Gschwend, “The Chancellor Model: Forecasting German Elections”, *International Journal of Forecasting*, Vol. 26, No. 1, 2010.

⑥ M. S. Lewis-Beck and C. Tien, “Election Forecasting for Turbulent Times”, *PS: Political Science & Politics*, Vol. 45, No. 4, 2012; P. Hummel and D. Rothschild, “Fundamental Models for Forecasting Elections at The State Level”, *Electoral Studies*, Vol. 35, 2014.

⑦ S. Merrill, B. Grofman and T. Brunell, “Cycles in American National Electoral Politics, 1854 – 2006: Statistical Evidence and an Explanatory Model”, *American Political Science Review*, Vol. 102, No. 1, 2008; M. Lebo and H. Norpoth, “The PM and the Pendulum: Dynamic Forecasting of British Elections”, *British Journal of Political Science*, Vol. 37, 2007.

果可以根据上届和上上届选举的结果来预测^①,该模型成功预测了2016年美国总统大选结果。

综上,模型范式具有两大优势:第一,预测模型纳入了变量关系,依托选举理论指导,具备较强可解释性;第二,模型范式可以提前较长时间(甚至半年)得到预测结果,具有更强政策应用价值。模型范式的不足在于:第一,只注重少量宏观预测因子,忽略微观层面的选民因素,而选举研究表明投票活动除了受外在因素影响之外,选民自身的人口学特征、认同结构和社会化过程等都对其影响甚大;第二,模型范式生成的是稳态的预测结果,无法捕捉竞选活动以及突发事件引发的波动;第三,不同模型在理论视角、变量选取和测量方法上差异较大,难以检验不同模型之间的竞争性理论假设^②。总之,模型范式虽然优势明显,但也面临诸多挑战。

3. 混合范式。混合范式通过一定规则综合不同预测手段来优化预测效力,以期克服单一预测方法的局限^③。其基本预设是集合多个方法的预测结果比运用单个方法更加精准,具体可细分为一阶混合方法和二阶混合方法。

一阶混合方法通常是将“回归预测模型”与“选举民调”相结合^④。例如,经典的预测表达式为:

$$\begin{aligned} \text{Vote} = & \text{Structural Weight} * [\beta_1 \text{Incumbent Approval} + \beta_2 \text{Economic Growth}] \\ & + \text{Polling Weight} * [\text{Polls}_{t-x}] + \varepsilon \end{aligned}$$

即先依据选民对在任政府的认可度和经济增长状况两个结构性变量来构建预测模型,用以预判候选人的基准得票率,再通过实时民调捕捉选情波动,最后将两个结果通过一定的加权规则集合加总,以期得到更加精准的预测结果。该方法综合了“模型范式”和“意见聚合范式”的各自优势,同时弥补了民调和模型的部分弱点。研究表明,该方法在英国、德国、爱尔兰等国家的选举预测中比单个静态模型或单个民调预测都准确^⑤,虽然民调和回归模型在不同国家对预测精准度的贡献率各有不同。

二阶混合方法则更为复杂,其试图对运用各种不同方法得到的预测结果进行整体加权集合以提高预测精准度^⑥。该方法通常分两步展开:第一步是综合同一类方法内部的不同预测数据(例如不同机构的民调结果)得到此类方法的总体预测结果,以此削弱单一预测主体可能存在的预测偏差;第二步再凭借一定的加权规则来集合各类不同方法的总体预测结果,得到一个大集总的预测,由此降低单一预测范式可能存在的偏差。当前,最具影响力的当属“PollyVote”方法,其集合了来自民调、博彩市场、专家调查、群体智慧、候选人指数、经济模型等不同方法的预测数据,通过上述两步集合方法得到一个集总预测结果,回溯表明该预测方法比其他单

① H. Norpoth, “The Electoral Cycle”, *PS: Political Science and Politics*, Vol. 47, No. 2, 2014.

② B. E. Lauderdale and D. Linzer, “Under-performing, Over-performing, or Just performing? The Limitations of Fundamentals-based Presidential Election Forecasting”, *International Journal of Forecasting*, Vol. 31, No. 3, 2015.

③ J. S. Armstrong, “Combining Forecasts”, in J. S. Armstrong ed., *Principles of Forecasting: A Handbook for Researchers and Practitioners*, Norwell, MA: Kluwer Academic Publishers, 2001.

④ M. S. Lewis-Beck and R. Dassonnevilleb, “Comparative Election Forecasting: Further Insights from Synthetic Models”, *Electoral Studies*, Vol. 39, 2015.

⑤ M. S. Lewis-Beck and R. Dassonnevilleb, “Forecasting Elections in Europe: Synthetic Models”, *Research and Politics*, Vol. 1, 2015.

⑥ D. Rothschild, “Combining Forecasts for Elections: Accurate, Relevant, and Timely”, *International Journal of Forecasting*, Vol. 31, No. 3, 2015; A. Graefe, et al., “Combining Forecasts: An Application to Elections”, *International Journal of Forecasting*, Vol. 30, No. 1, 2014.

个方法都更加接近真实结果^①。

混合范式的优点在于：第一，将基于结构性因素的稳态预测与来自意见聚合的动态预测相结合，在提前数月得到模型结果的同时可根据选情动态调整；第二，综合不同类型的信息源，克服单一方法存在的局限。混合范式也面临诸多问题：第一，存在系统性偏差的风险，特别是当被集合的多个预测源都朝同一个方向产生偏误时，预测偏差将被放大（例如2016年美国总统大选）；第二，对不同预测方法的权重估计需要更加科学、透明的标准，权重设定失误可能产生新的偏误；第三，预测源选取可能存在研究者偏见，综合并不总是最好的选择^②；第四，综合的预测源越多，其预测结果越难从学理角度进行解释。如果预测错误，也较难回溯验证。

4. 大数据范式。大数据范式通过收集社交媒体上网民关于选举的语言和行为数据，并借助机器学习算法预测选举结果。近年来，随着移动终端和社交网络的普及，人们在虚拟空间里积累了海量的行为和交往大数据，运用计算机技术挖掘社交媒体数据以预测电影票房、产品销量、信贷风险等成为趋势，该方法也被运用于选举预测^③，并呈现出预测潜力。

早期的大数据预测比较简单直接，通过收集社交网络（如Twitter、Facebook）上针对某个政党或候选人的评论量、转发量、点赞数、粉丝数，以及热搜、维基百科浏览量等数据，来测量民众对政党/候选人的关注强度，进而推测其选举支持率^④。该方法被证明在某些选举中具有一定预测力，但整体预测效果欠佳且不稳定。原因在于，上述网络行为数据无法准确测量真实支持率，其中夹杂了大量“噪音”，包括机器自动推送（bots）、网络水军、黑公关等。此外，网络关注热度无法直接转换为选票，这受到投票资格和投票意愿的影响，且有正向和负向关注之分。为了克服这些缺陷，研究者们开始运用“自然语言处理”技术和无监督或半监督的机器学习方法对社交数据进行情感分析（sentiment analysis），试图探索网络意见表达（例如网民评论）的语意与情感取向，并综合不同网络信息源来推算选民对政党或候选人的支持率，其改进后的预测效力有时堪比民意调查^⑤。

大数据范式的优势在于：第一，实时快捷，能够每天甚至每小时及时更新，真正做到实时播报（nowcast），即刻掌握选情动态；第二，信息源丰富，成本较低，同时可供分析和解释；第三，借助计算机处理技术，可以挖掘选民心理与投票行为的关系，捕捉线上动员的选举后果。大数据预测也面临一些挑战：第一，网络数据通常缺乏代表性，社交网络上活跃的群体往往只是选民中的一部分（偏年轻、都市和教育程度高的群体），同时存在迫于政治正确较少发声的“害羞选民”（shy voters），代表性偏差会影响预测效力；第二，网络语言属于自然语言，自然语言处理技术目前尚不成熟，例如政治暗语和场景化语言的语义甄别、多语种语言采集和

① A. Graefe, et al., “Combined Forecasts of the 2012 Election: the Pollyvote”, *Foresight: The International Journal of Applied Forecasting*, Vol. 28, 2013.

② M. Hibon and T. Evgeniou, “To Combine or Not to Combine: Selecting among Forecasts and Their Combinations”, *International Journal of Forecasting*, Vol. 21, 2005.

③ Andrea Ceron, Luigi Curini, and Stefano M. Iacus, *Politics and Big Data: Nowcasting and Forecasting Elections with Social Media*, New York, NY: Routledge, 2017.

④ Joseph DiGrazia, et al., “More Tweets, More Votes: Social Media as a Quantitative Indicator of Political Behavior”, *PLOS ONE*, Vol. 8, No. 11, 2013.

⑤ J. Bohannon, “The Pulse of The People: Can Internet Data Outdo Costly and Unreliable Polls in Predicting Election Outcomes?” *Science*, Vol. 355, No. 6324, 2017; M. Huberty, “Can We Vote with Our Tweet? On the Perennial Difficulty of Election Forecasting with Social Media”, *International Journal of Forecasting*, Vol. 31, No. 3, 2015.

分析难度较大，数据“信噪比”低会影响预测质量^①；第三，社交媒体的普及率在不同国家存在差异，在较多发展中国家无法获取足量的网络数据用于预测。此外，自从“剑桥分析”（Cambridge Analytica）利用脸书数据影响美国总统大选^②，社交媒体平台对数据获取设置了更高门槛，这些都影响了大数据预测的发展。但是，随着算法技术的突飞猛进和数据的指数级激增，加之计算社会科学的发展，该范式前景广阔。

四、选举预测范式、方法的评估比较和发展趋势

选举预测是政治学预测性研究的前沿领域，经过八十余年的探索和积累，已经发展出丰富多元的科学预测方法，并处在不断创新升级当中。本文归纳出四类选举预测范式：意见聚合范式；模型范式；混合范式；大数据范式。每一类范式下又可细分出不同类型的预测方法，每种预测方法的技术路线、预测机理、数据基础和应用场景各有差异，同时在选举预测综合评价体系的各个维度上各显优劣（见表 2）。

表 2 不同选举预测方法优势比较

范式	具体方法	准确性	超前性	解释性	可重复性	中立性	成本优势
意见聚合范式	民意调查	****	***	***	***	**	**
	政治博彩市场	****	**	*	***	***	****
	群体智慧	****	***	***	***	***	**
	专家调查	**	***	*	***	***	****
模型范式	指数模型	***	*****	*****	*****	*****	*****
	结构因素模型	**	*****	*****	*****	***	*****
	选举周期模型	**	*****	***	*****	***	*****
混合范式	一阶混合	*****	**	**	*****	*****	**
	二阶混合	*****	**	*	*****	*****	**
大数据范式	关注度数据	**	***	*	***	**	****
	机器学习情感分析	***	***	***	*****	**	***

注释：* 号数目越多代表优势越显著

具体来说,混合范式在准确性上要高于意见聚合范式、模型范式和大数据范式,因为其综合了模型范式的稳态预测和意见聚合范式的动态预测,在可重复性和中立性方面也表现较好,但由于集合方式复杂而欠缺可追溯性和解释力。模型范式在超前性、解释性、可重复性、中立性和成本方面都占据优势,然而由于预测时间超前不得不牺牲一部分准确性。意见聚合范式实时更新,具备较高的准确性,但在超前性和解释力上较弱,同时调查类预测成本高昂。大数据范式结合前沿算法,具有较强的成本优势和追踪预测能力,虽然目前在机制解释、因果识别和准确性上略有不足,但有望随着技术突破和理论引导得到改进提升。总之,预测范式和方法的选择是基于不同

① J. Grimmer and B. M. Stewart, “Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts”, *Political Analysis*, Vol. 21, No. 3, 2013.

② 房宁、丰俊功：《2016 年美国总统选举中的技术革命与选民行为控制》，《比较政治学研究》，2017 年第 2 期。

预测目标和应用场景的综合权衡。

整体而言,现有选举预测范式存在一些共同缺陷,部分方法(例如民调)甚至遭遇发展瓶颈。首先,多数预测方法缺少选举研究理论的系统支撑,未能将“理论驱动”与“数据驱动”相结合,虽然模型范式纳入了一些有意义的预测变量,但其模型过度简约、理论依据单薄,难以检验竞争性理论假设;其次,现有预测范式缺少对选民个体层面数据的有效运用,未能纳入对选举产生关键影响的选民因素(例如性别、年龄、族群、教育、收入、宗教、职业),过度关注有限的结构性因素或候选人指标等因素,选民作为选举决策中最重要的能动主体(agent)却是缺失的;最后,未能充分将预测和解释相结合,政治学预测性研究应该致力于贯通解释和预测,在预测结果的同时探索变量之间的因果关系和影响机制,检验和优化选举理论,增进人们对选举活动的认知。

“科学预测基于对现有信息的分析,对未来尚未发生的事件进行判断,是风险较大的研究工作”^①。选举预测对精度的要求远超过其他形式的预测,因此风险更大。任何预测方法都可能犯错,但科学预测对错误有着不同的理解和因应方式。首先,方法运用不当导致的问题与方法本身的问题应该区分开来,误差和错误应该区分开来;其次,需要建立一套评价预测结果的客观标准,科学预测除了精确度之外,还需兼顾超前性、解释力、可重复性等;再次,失败的预测可以揭示我们知识和方法的短板,推动收集更优质的数据、优化变量选取和模型建构、反思和改进选举研究,科学的预测是一种可以被检验、被追溯、被改进的预测;最后,量化选举预测也面临不可观测因素、过度拟合(overfitting)以及样本外(out-of-example)等问题,研究者需要保持谨慎。

选举预测呈现出以下发展趋势:第一,数据来源、技术手段和预测机理不断开拓创新,老方法不断朝精细化方向改进完成自我升级,新方法旨在弥补老方法的短板和缺陷,不同预测方法相互竞争赶超,形成“百花齐放”的局面;第二,不同范式之间以及同一范式中的不同预测方法之间逐渐整合,各自发挥比较优势又相互助益;第三,选举预测越来越成为一个跨学科的研究领域,在这个领域,政治学、经济学、统计学、社会学、心理学、计算机科学甚至脑科学等不同学科交相融合,需要我们吸纳各个学科的理论和方法以及前沿的技术手段来不断推进预测研究。

选举预测的若干前沿议题值得学界共同探索:第一,“预测选举预测”,即预测不同的选举预测方法在什么时候、在哪些条件下更加准确,这属于更高阶的预测活动(high-order prediction)^②;第二,提升运用不完全信息和低信噪比数据做预测的能力,预测活动无法等到拥有高质量的数据才去做,必须探索如何高效地收集清洗数据、识别有效信息、减少噪音并将信息转化为知识和预测力;第三,加大对发展中国家选举以及各国地方选举的研究和预测。新兴民主化国家的选举往往对区域性政治动荡、投资贸易摩擦、族群宗教冲突等产生重大影响,此外,地方选举的波动性和情境性远高于国家选举,现有的各类预测方法盛行于西方成熟民主国家并聚焦国家层面,如何拓展到新型选举当中、需要做哪些调试和修正,值得学界深入探究。

五、结论

“选举预测是一项高风险的工作,但却是一项值得探索的科学事业”^③。选举预测是社会科

① 庞珣:《定量预测的风险来源与处理方法——以“高烈度政治动荡”预测研究项目的再分析为例》,《国际政治科学》,2017年第3期。

② V. S. Subrahmanian and S. Kumar, “Predicting Human Behavior: The Next Frontiers”, *Science*, Vol. 355, No. 6324, 2017.

③ J. Scott Armstrong ed., *Principles of Forecasting: A Handbook for Researchers and Practitioners*, Norwell, MA: Kluwer Academic Publishers, 2001.

学预测性研究的关键议题,是政治学理论和方法创新的前沿领域,预测结果不仅可以满足人们预知未来的好奇心、推进政治科学研究,同时可以指导科学决策。目前科学量化的选举预测方法可划归为四大类:意见聚合范式;模型范式;混合范式;大数据范式。这些预测范式及其具体的预测方法克服了非科学预测的随意性、不可解释性、不可重复性等问题,为我们预测选举提供了丰富的工具。选举预测接下来的突破方向是将“理论驱动”与“数据驱动”相结合、微观的个体预测变量与宏观的结构性预测变量相结合、预测与解释相结合,通过跨学科合作和借助前沿技术手段推进预测范式创新,并在预测实践中不断更新升级,提升预测过程的透明度和预测结果的解释力。在这些方面,结合选举理论和计算机技术的微观仿真模拟预测方法将是有益的尝试^①。

加强对选举预测以及整个政治学预测性研究的重视并非旨在替代原有的解释性研究。相反,我们认为预测和解释是“双螺旋”式的共生成长关系:解释性研究可以为预测性研究提供理论和知识基础,开启预测的议题和领域,指导预测的数据收集、变量选取和模型建构,以及解释最终预测结果;反过来,预测性研究可以检验和优化理论,发现新的研究方向,提升政治学研究的政策应用价值,等等^②。两者相互助益,共同服务于增进人类对复杂政治现象和政治行为的理解。

科学的预测性研究是政治学乃至整个社会发展相对滞后的领域,中国政治学者在相关领域涉足更少。相比西方国家(特别是美国)对预测性研究的高度重视和资金扶持,中国在相关学术领域投入严重不足。科学的预测性研究呼唤中国政治学者的参与和尝试,同时离不开学界和政府的宽容、重视和投入。作为社会科学发展的前沿领域,科学预测可以成为中国政治学探索的“富矿”。大数据、人工智能、仿真模拟等方法的兴起为预测性研究提供了新契机。选举预测的研究经验表明,解释性研究和预测性研究可以相得益彰。中国政治学学科可以在不断完善解释性研究的同时,探索预测性研究,在诸如政治风险预测研究、海外冲突预测研究、公共服务需求预测研究、公共危机预测研究等领域做出努力,运用政治学理论、知识和方法指导预测活动,通过科学预测不断丰富、完善和推进政治学研究,提升政治学的科学水平和应用价值。

作者:王中原,复旦大学社会科学高等研究院(上海市,200433)

唐世平,复旦大学国际关系与公共事务学院(上海市,200433)

(责任编辑:林立公)

① 该方法已经运用到对中国台湾地区选举和美国国会选举的预测当中,并取得了良好效果,参见复旦大学复杂决策分析中心唐世平教授团队的研究成果:<http://www.ceda.fudan.edu.cn/index.php?c=article&a=type&tid=3>,2020年1月15日。

② J. M. Hofman, A. Sharma, and D. J. Watts, “Prediction and Explanation in Social Systems”, *Science*, Vol. 355, No. 6324, 2017.