

引用格式:江东,陈帅,付晶莹,等. GIS支持下的全球武装冲突数据体系建设评介[J].地球信息科学学报,2020,22(4):784-791. [Jiang D, Chen S, Fu J Y, et al. Review on the construction of global armed conflict data system supported by GIS[J]. Journal of Geo-information Science, 2020,22(4):784-791.] DOI:10.12082/dqxxkx.2020.200147

GIS支持下的全球武装冲突数据体系建设评介

江 东^{1,2}, 陈 帅^{1,2}, 付晶莹^{1,2}, 郝蒙蒙^{1,2*}

1. 中国科学院地理科学与资源研究所, 北京 100101; 2. 中国科学院大学资源与环境学院, 北京 100049

Review on the Construction of Global Armed Conflict Data System Supported by GIS

JIANG Dong^{1,2}, CHEN Shuai^{1,2}, FU Jingying^{1,2}, HAO Mengmeng^{1,2*}

1. Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China;
2. College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China

Abstract: Geopolitical environment security is an important part of national security. Accurate, refined and timely updated information of various types of geopolitical conflicts is the basis of global geopolitical risk assessment and early warning. Based on the armed conflict datasets, this paper comprehensively reviewed the development of the global geopolitical conflict data system, and analyzed the application progress of geo-spatial and temporal positioning, big data mining and other technologies. Under the support of Geographic Information System (GIS) technology, the understanding of the connotation of armed conflict and the general framework of the data system of armed conflict events has been deepened, a large number of high-quality datasets or data products have been produced. Geopolitical conflicts are gradually refined in the theme, the spatial positioning accuracy of conflict events developed from country to smaller scale geographical location (longitude and latitude). Supported by network information search and big data mining technology, the update frequency of armed conflict events increased from annually update to hourly update, the construction of global armed conflict data system has achieved initial results. The deep integration of geo-environmental system science and geographical science, the wide application of geographic big data and artificial intelligence technology will be the important driving forces for the development of this discipline.

Key words: geopolitical environment; armed conflict; big data; data system; scale

*Corresponding author: HAO Mengmeng, E-mail: haomm@igsnr.ac.cn

摘要: 地缘环境安全是国家安全的重要组成部分。准确、精细、及时更新的各类地缘冲突信息是全球地缘风险评估和预警的基础。本文全面回顾了以武装冲突为核心的全球地缘冲突数据体系建设的发展历程,剖析了地理时空定位、大数据挖掘等技术的应用进展。在地理信息系统(GIS)技术支持下,对武装冲突内涵和武装冲突事件数据体系总体框架的理解不断深入,生产出大批高质量的数据集合或数据产品;地缘冲突在主题上逐渐细化,冲突事件记载的空间定位精度从国家发展到更小尺

收稿日期:2020-03-27;修回日期:2020-04-19.

基金项目:中国科学院战略性先导科技专项(XDA23000000);中国科学院重点部署项目(KGFZD-135-17-009)。[**Foundation items:** Chinese Academy of Sciences Strategic Priority Research Program, No.XDA23000000; Chinese Academy of Sciences Key Deployment Project, No.KGFZD-135-17-009.]

作者简介:江 东(1972—),男,安徽寿县人,研究员,主要从事地缘环境系统模拟研究。E-mail: jiangd@igsnr.ac.cn

*通讯作者:郝蒙蒙(1990—),女,安徽萧县人,助理研究员,主要从事地缘环境系统模拟研究。E-mail: haomm@igsnr.ac.cn

度的地理位置(经纬度);在网络信息搜索与大数据挖掘技术支持下,武装冲突信息的更新频率从年度更新提高到每小时更新,全球武装冲突数据体系建设初见成效。地缘环境系统科学与地理科学的深度融合、地理大数据与人工智能技术的广泛应用将是推动本学科领域发展的重要动力。

关键词: 地缘环境;武装冲突;地理大数据;数据体系;尺度

1 引言

武装冲突作为国际关系的一种极端表现形式,不仅会对国家安全和人民的生命财产造成严重危害,而且会改变国际关系的走势,对世界的地缘政治格局产生深远影响。在错综复杂、变幻莫测的国际形势下,清晰研判中国周边以及全球的地缘环境态势、全面掌握全球地缘政治格局的演变规律,对于保障我国的地缘环境安全以及“一带一路”倡议的顺利实施具有重要意义^[1]。在此背景下,地缘环境研究的重要性日益凸显,然而传统的地缘环境研究大多侧重于定性分析,与新时期国家地缘决策中对地缘环境信息快速获取、处理、评估、综合分析的需求尚有较大差距^[2-3]。究其原因,一方面是因为地缘环境系统本身的综合性和复杂性^[4],另一方面则是因为地缘环境基础数据积累薄弱。政治、社会、经济和生活中的突发事件,特别是武装冲突事件的动态更新和实时抽取,对于地缘环境的动态监测和模拟尤其重要。指标体系清晰、语义规范、精度可控的跨区域的全球武装冲突数据是地缘环境研究的基础^[5]。较为系统的武装冲突数据库直到20世纪下半叶才开始出现,早期的数据库大多依赖人工编码,数据采集和建库过程耗时耗力,并且存在数据粗糙、更新慢、主观性强等缺陷。20世纪90年代以来,随着计算机技术和地理信息技术的发展,机器自动编码技术开始应用于大规模武装冲突数据采集与建库,但数据质量参差不齐,真实性、可靠性有待提高。近年来,以网络时空大数据为信息源,利用地理空间分析技术和人工智能抽取社会事件信息的技术逐渐成熟,因其覆盖广、更新快的优势,将成为武装冲突事件动态更新与快速分析的主流技术,以提升数据准确性和及时性,满足地缘环境研究中迫切的数据需求。

本文回顾了以武装冲突为核心的全球地缘冲突数据体系建设的发展历程,重点从主题、空间和时间3个维度,全面梳理地缘冲突数据体系建设的关键技术与应用进展,同时对今后的主要发展方向进行了展望。

2 武装冲突主题的界定与分类

武装冲突数据获取的流程一般可以总结为3个步骤:① 确定编码规则和数据本体;② 确定数据来源、采集数据;③ 按照编码规则和数据本体定义对数据进整合处理,生成标准化的数据产品。对于武装冲突事件而言,本体首先给出武装冲突的基本定义,即一个事件被划分为武装冲突的条件,同时阐明武装冲突事件的分类规则和基本构成要素。

武装冲突数据库可以追溯到20世纪60—70年代,美国密歇根大学的政治学家J. David Singer发起了“战争相关指数”项目(Correlates of War Project, COW Project)^[6],致力于收集有关战争和国家间冲突的历史数据,旨在系统揭示战争和武装冲突的时空演变机制。该项目产生了第一个严格意义上的武装冲突数据库——COW数据库,有力推动了对战争和武装冲突起因的定量研究。COW数据库最主要的突破在于它制定了严格的编码规则和数据本体,明确了国家间战争(Inter-state War)的定义和数据整合原则,为后续的研发奠定了基本模式。随着人们对武装冲突认知的深入,武装冲突事件的外延不断拓展,其内涵也逐渐丰富,如冲突与和平数据库(Conflict and Peace Data Bank, COPDAB)共有16种事件类型^[7],而世界事件作用调查(World Event Interaction Survey, WEIS)的事件类型包含22个父类和63个子类^[8]。冷战后全球武装冲突呈现出新的特点和趋势,大规模的战争数量急剧减少,中小规模的武装冲突占据主导地位^[9],COW的定义将许多小规模武装冲突排除在外。后续的COW项目拓展了武装冲突事件的边界,从国家间战争(国家实体之间的战争, Inter-state War)扩展到国内战争(国家实体与境内非国家实体之间的战争, Intra-state War)^[10]、非国家战争(非国家实体之间的战争 Non-state war)和超国家战争(国家与境外非国家实体之间的战争 Extra-state War)^[11]。

1993年乌普萨拉冲突数据项目(Uppsala Conflict Data Program, UCDP)以“基于国家的武装冲突(State-based conflict)”为数据本体,将冲突定义为:

冲突双方至少有一方是一国政府,另一方与政府在控制政权或领土方面存在不相容的利益,并且双方使用了武力,在单一年份内与冲突相关的死亡人数达到25人^[12]。这种更低的死亡人数要求使UCDP收录了大量COW所不具备的小规模武装冲突事件。2002年Gleditsch等^[13]扩展了UCDP数据库的时间跨度,发布了UCDP/PRIO数据库,涵盖了1946—2001年全球多种规模的武装冲突。但是UCDP/PRIO依然要求冲突双方至少有一方为国家政府,使得一些武装组织之间的冲突或针对平民的暴力袭击事件被忽略。

Gochman和Maoz^[14]提出了国际军事争端数据(Militarized Interstate Disputes, MID),把冲突过程分解为3个层次:武力的威胁、武力的展示、武力的使用。而后在MID 3.0版本中又增加了非军事行为(No militarized action)和战争(War)2种类型^[15],每个MID事件包含了事件双方的国家、事件类型、事件的起始和终止日期以及敌对的程度等信息。MID主要的贡献在于它兼顾了其他类型的小规模冲突行为,更重要的是,它对于冲突过程的分解,便于研究人员更好地掌握地缘冲突从起源、发展到终止的全过程。为了进一步刻画冲突过程中的具体情况,后来的数据库则采取细节数据的方式,以一个具体的冲突事件为载体,并将参与冲突的双方、冲突的具体地点及其他冲突要素都列入其中,从而极大地丰富了冲突研究的内容,使研究者能够深入了解相关冲突事件发生时具体的时空背景,进行更微观层面的分析。例如,Raleigh等^[16]发布的武装冲突的位置和事件数据集(Armed Conflict Location & Event Data, ACLED)包含了8种类型的事件,包括战斗-领土未发生变化、战斗-叛军控制领土、战斗-政府恢复领土、战斗-建立总部、非暴力冲突事件、骚乱/抗议、针对平民的暴力、非暴力方式的领土变化。其中针对平民的暴力事件可以没有政府主体的参与,相比于COW和UCDP/PRIO覆盖的范围更加全面。

冲突数据库在主题上逐渐细化,方便了研究者分析各类冲突事件发生的过程以及事件的前因后果,同时也提供了战争以外的小规模冲突事件分析的基础。例如,Partell^[17]利用MID数据库分析了冲突事件双方的反应和相互作用对于武装冲突事件升级的影响,并阐述了导致冲突升级的主要因素;Kreutz^[18]采用UCDP数据库对2种不同类型冲突(国

内战争和国家间战争)的发生和解决进行了定量研究,结果表明冲突并不仅仅以决定性的结果(如胜利或和平协定)而结束,在政府获胜或部署维和部队之后,再次发生冲突的可能性较低。Raleigh^[19]基于ACLED数据库分析了内战中对平民的暴力是如何发生的,表明在内战地区存在的多个暴力团体中小型反对派团体对平民实施了更程度的暴力。

2005年,“冲突与调解事件观察”(Conflict and Mediation Event Observation, CAMEO)的编码方案集上述研究成果之大成^[20],共定义了20大类和将近300小类的事件类型,并且建立了完整的地名和实体词典,基本能覆盖各种国际和国内政治事件所涉及的国家、次国家主体、族裔群体、地理区域和各种政府间组织和非政府组织。CAMEO已为学界和应用领域广泛认可,成为了是一种用于分析国家间政策相互影响的事件数据框架。

国内关于武装冲突数据库的研究起步于20世纪80年代。1981年吴长义^[21]通过翻译国外期刊整理了第二次世界大战之后全球发生的150余起主要的武装冲突;1998年贵琳^[22]通过摘编国外杂志建立了1997年世界范围内的地区冲突索引。这些研究普遍都是通过收集和整理国内外期刊杂志和报道来获取武装冲突数据,仅涉及那些规模或影响较大的武装冲突,数据体量较小,时间和空间覆盖范围有限,更新周期长且不可持续。2000年以来,由中国社会科学院世界经济与政治研究所出版的《全球政治与安全报告》^[23]系列年度报告对每年全球政治及安全形势的总体情况及变化进行回顾与分析,还介绍了中国周边安全环境、全球武装冲突及全球军事形势等内容。同样,《全球政治与安全报告》关注对象是全球重大的武装冲突,采用的分析方法以定性方法为主。2009年,阎学通等^[24]发表了《中美关系鉴览1950—2005》,并据此建立了“中国与大国关系数据库”,是国内最早出现的较为系统的冲突事件定量数据库。在事件的定义上,“中国与大国关系数据库”只关注与中国有关的双边关系事件,中国的事件记录明显多于外国的事件记录;数据主要源自《人民日报》和外交部网站,事件类型以与中国有关的经贸摩擦事件为主,对于军事、政治方面的事件记录较少^[25-27]。2017年周亦奇等^[28]综合利用人工编撰与机器学习的方法搜集资料,建立了“一带一路”政治安全风险数据库,包括武装冲突、暴力袭击和示威游行3种类型的冲突事件,主要关注的地

区为中南半岛、南亚、西亚和北非地区。第一阶段的数据建设主要收集2015年6月至2016年12月发生的事件。“一带一路”政治安全风险数据库采取计算机与人工相结合的方式,在人工确定信息来源(包括主要媒体和既有数据库)和信息边界后(即何为需要分析的冲突事件后),由计算机进行抓取,利用计算机的网络搜索能力,锁定其中含有冲突关键词(或触发词)的新闻及其段落,在此基础上直接进行信息提取,并在其中提取行为主体、地点和时间等关键信息,从而自动生成一个初级数据库。“一带一路”政治安全风险数据库采用的是人机互动的编码方式,虽然能够一定程度弥补人工编码效率低、机器编码准确性差的问题,但数据获取和筛选仍然依赖人的主观判断,数据库更新效率也受限于人的工作效率,因此还有很大的提升空间。

3 武装冲突地理空间分布信息获取与表达

经过几十年的探索,武装冲突数据库在主题和分类方面不断细化,并逐渐形成了统一的共识(如CAMEO),但在武装冲突事件的空间地理位置信息方面相对薄弱,一般只记录了战争双方的国家,并没有战争发生的具体地点信息(如经纬度、城市、乡镇等),这种粗粒度的信息对于更加详尽的地缘关系研究作用有限。近年来的研究表明,武装冲突受自然(地形地貌、气象、生态、自然灾害等)、社会(人口、宗教等)、经济(经济水平、交通条件等)等多种要素的综合驱动^[29-31]。精细的空间位置信息不仅可以分析事件在地理空间上的集聚特征和地域分异规律,同时自然要素与社会经济要素及事件信息在统一的空间表达体系下,为武装冲突的自然、人文多种影响因素的融合与集成分析提供了可能。

在应用需求和科学研究需求的联合驱动下,在武装冲突数据体系建设过程中,地理空间的概念得到强化,地理信息系统在位置信息获取、管理、分析方面的优势逐渐得到体现。美国马里兰大学创建的全球恐怖袭击事件数据库(Global Terrorist Data, GTD)记录了1970—2018年全球恐怖袭击的数据^[32],包括:事件发生的时间、地点、行为主体、受袭击对象、采用的袭击手段以及造成的伤亡人数等,其中的地点为具体的经纬度信息。武装冲突的位置和事件数据集ACLED也提供了冲突事件发生时的具

体地点、所在国家、省市等^[6],明确的地理位置信息对于研究武装冲突的时空演变具有重要意义。

类似于ACLED, Melander和Sundberg也发布了一个新的地理编码数据集,被称为UCDP空间定位事件数据集(Georeferenced Event Dataset, UCDP-GED)^[33],它的编码方式和ACLED十分类似,对冲突事件的地理位置描述可以精确到村庄级别。UCDP-GED包含3种类型的冲突事件:“武装冲突(Armed conflict)”,“非国家冲突(Non-state conflict)”(双方都不是国家)和“单方面暴力(One-sided violence)”(有组织的团体袭击手无寸铁的平民)。UCDP-GED的基本分析单元是有组织的致命事件:在某一特定地点和时期内,某一组织对另一组织或平民使用武力,直接造成至少1人死亡的事件。只要满足这个条件,UCDP-GED就会记录该事件,这比UCDP之前所采用的至少造成25人死亡的要求更低。UCDP-GED最初的版本记录了1989—2010年非洲发生的22 000起冲突事件,之后的版本开始覆盖全球并且每年更新。

早期的地缘冲突事件是基于国家尺度的,其较粗的空间分辨率难以和其他空间要素直接匹配,限制了冲突事件的定量研究。随着空间分析技术的引入以及对冲突事件更加精细尺度的记录,一方面使得分析地缘冲突事件的发展动态和时空变化特征成为可能,如Hao等^[30]利用GTD数据库分析了中南半岛恐怖袭击事件的时空变化规律,得到了恐怖袭击发生的高风险区域;Weidmann和Ward^[34]通过考虑冲突事件的时空变化,表明暴力很可能随着时间的推移而重现并在空间上扩散,利用这些信息可以极大地提高冲突预测的能力;另一方面,冲突事件空间分辨率的提高,也使得冲突事件能够和其他空间连续变量相互兼容,与其发生的自然和社会经济背景联系起来进行分析,从而研究地缘冲突发生的内在机理。例如,Onda^[35]将冲突事件与降雨联系起来,证明了干旱与内战爆发之间的因果关系;Cederman的研究表明随着群体与首都距离的增加以及地形不平整,可能导致发生冲突的风险增加^[36];Hao等^[29]采用一系列经济指标分析了印度经济状况与武装冲突之间的联系,模拟了武装冲突的状态和发展趋势。

2015年开始,中国科学院部署建设了“‘一带一路’地缘环境大数据平台”,采用标准的冲突事件CAMEO分类体系,以全球、“一带一路”、我国周边

重点区域3个层次,建立了一套完整的“一带一路”地缘环境数据产品。除了详细记录了各类武装冲突,同时收集整合了地缘冲突的主要影响因素数据,包括自然地理、资源环境、气候气象、社会经济、人文发展等专题,为地缘冲突的分析与决策研究提供了基础数据支撑。

4 武装冲突时间属性的智能抽取

早期的武装冲突数据库都是采用人工从报刊杂志等媒体上采集数据,时效性差,数据的更新频率一般以年为单位。WEIS和COPDAB最早将政治事件中以自然语言表示的事件发起者、目标者和行为动作转化为相应的角色分类和事件代码,以便用于数据整合和实证分析,这种模式非常适合于在每天发生的新闻事件中提取关键信息,为后来的机器编码奠定了基础。基于WEIS的事件分类体系,Philip A. Schrodtt主导的堪萨斯事件数据系统(Kansas Event Data System, KEDS)项目首次尝试使用计算机将文本解析为事件数据库^[37],这种解析方法主要依赖于模式匹配,通过在文章中匹配特定的关键词来识别事件类型和事件主体。2000年,KEDS项目开发了增强替代指引文本分析(Textual Analysis By Augmented Replacement Instructions, TABARI)软件^[38],作为KEDS的继承者,TABARI成为占主导地位的国际事件机器编码系统,后来的自动编码软件很多都是以TABARI为原型开发的。TABARI具有很快的编码速度,可以在6 min内完成对2600万条文本的编码。

随着信息技术的发展,信息的载体由传统的纸质媒介转变成为互联网,人们获取数据、存储数据和分析数据的能力都经历了爆炸式的增长,大数据时代已经来临。基于大数据的事件抽取是从非结构化文本中识别事件类型并抽取事件各元素的任务,需要确定特定类型事件发生的时间、地点、参与者等信息。事件抽取任务可分解为4个子任务:触发词识别、事件分类、论元识别和角色分类任务。抽取方法可以分为基于模式匹配的事件抽取和基于机器学习的事件抽取2类,模式匹配在特定领域的性能表现较好,但是移植性较差,早期的系统多基于模式匹配方法。使用机器学习方法的事件抽取只需要训练语料,不需要太多的专家知识,移植性较好。机器学习方法已经成为了目前事件抽取

的主流,基于TABARI和智能抽取技术,产生了综合危机早期预警系统(Integrated Crisis Early Warning System, ICEWS)和全球事件、语调与语言数据库(Global Database of Events, Language and Tone, GDELT)。这两大武装冲突数据库无论是在规模体量还是在覆盖范围上都是空前的,并且可以实现连续的、近乎实时的更新,已成为新一代武装冲突数据库的翘楚。

ICEWS是由美国国防部高级研究计划局(Defense Advanced Research Projects Agency, DARPA)资助开发的一种危机早期预警系统^[39],该项目在2008年开始启动,旨在帮助美国政策分析人士了解过去和现在,以及预测未来美国可能必须应对的各种国际危机,包括国际和国内危机、种族和宗教暴力以及叛乱和骚乱。它从数字化的新闻媒体、社交媒体和其他来源获取大量的文本语料,然后通过计算机编码自动生成事件数据库。ICEWS使用了一个叫BBN ACCENT的编码器,记录了从1995年至今全球发生的重要的冲突事件,提供了关于事件类型、参与者和时间空间等信息。从2018年10月开始,ICEWS实现了每日更新。

GDELT由美国乔治城大学教授Kalev Leetaru于2012年创建并发布,是目前全球规模最大、内容最全面、分类粒度最细并且免费开源的冲突与调解事件数据库,包含了1979年到现在从全球新闻媒体报道中抽取得到的各类冲突与调解事件^[40]。GDELT数据库实时监测世界上65种语言的电视、广播、报纸、网络媒体甚至学术论文中的新闻事件,对其进行分析提取,提炼出新闻事件相关的人物、组织、时间、地点和事件类型等关键信息。该数据库包含了1979年至今的几亿条事件记录,每15 min便会进行一次更新。GDELT数据库的主表称为Event表,提供了事件的时间、参与者、事件类型、地理位置4个方面的属性信息,事件类型被分解成不同的层级,原始的事件参与者被分解到各变量对应的字段中进行存储,为研究人员与数据进行交互提供了方便。同时还包含一系列用于地理位置查询的字段,提供动作的发出者与接受者所在位置和动作发生地点的具体地理信息。从2015年2月起,GDELT数据库新增了2个数据表:Mentions数据表与Global Knowledge Graph (GKG)表。Mentions表通过全球媒体追踪每条报道的来源,记录Event表中提及的所有事件,便于梳理全球媒体系统中的各

类事件,使其之间错综复杂的网状关系清晰地呈现出来。Event表中提及的每一个事件都对应着一条Mentions表记录。GKG数据表应用了一系列十分复杂的自然语言处理算法,将每个独立的主体、地理位置、数据、主题、消息来源和全球的事件联结为一个巨大的网络,这个网络每天都捕捉最新的全球动态,包括此时此刻正在发生什么事件、什么人或组织参与其中以及人们对这些事件看法如何。

随着大数据和人工智能技术在冲突事件数据集中的广泛应用,产生了许多类似GDEL T和ICEWS这种大规模实时更新的冲突数据集,这些数据集通过对多源新闻媒体网址进行实时监测和分析,产生的冲突数据集在规模上是空前的,但在准确性上与传统的人工编码数据集还有一定的差距,这也符合大数据体量大(Volume)、来源多样(Variety)、增长速度快(Velocity)、价值密度低(Value)的“4V”特点。对冲突事件记录的时间分辨率的提高和数据量的指数增长,结合大数据挖掘和机器学习技术,使得这些数据在宏观上依然能够很好地反映冲突事件的发展趋势,在一定程度上弥补了其在准确性上的不足,这也为冲突事件的实时分析和预测提供了可能性,产生了一系列的研究成果,如利用GDEL T数据库来预测社会动荡^[41-42]、抗议^[43]、暴力事件^[44]和政治危机^[45],以及预测股市的变化^[46-48],这些研究可以为相关内政外交政策的制定以及社会治理提供及时高效的决策支持信息。

5 结论与展望

在地理信息系统(GIS)技术支持下,对武装冲突内涵和武装冲突事件数据体系总体框架的理解不断深入,生产出一大批高质量的数据集合或数据产品;地缘冲突在主题上逐渐细化,冲突事件记载的空间定位精度从国家发展到更小尺度的地理位置(经纬度);在网络信息搜索与大数据挖掘技术支持下,武装冲突信息的更新频率从年度更新提高到每小时更新,全球武装冲突数据体系建设初见成效。

具有时空属性的地理大数据,包括社交网络、移动设备、新闻媒体及各类网页信息^[49-50],以其持续更新、实时获取等优良特性给地缘环境研究带来了新的机遇,也是构建武装冲突数据库最理想的数据来源。然而,这些数据大多都是以非结构化的形式

展现的,数量巨大、结构复杂、类型众多,无法直接被计算机所理解,如何从中快速挖掘有用信息也带来新的挑战^[50]。虽然ICEWS数据库和GDEL T数据库已经被大量学术研究和政策建议所广泛采用和认可,但是其信度和效度却一直饱受质疑^[51-52]。Wang等^[52]的研究表明,GDEL T数据库中仅有21%的数据反映了真实的抗议事件,这很大程度限制了基于机器自动编码的武装冲突数据库的进一步应用。为此,机器自动编码的武装冲突数据库的准确性亟待提高。近年来,自然语言处理(Natural Language Processing, NLP)、机器学习(Machine Learning, ML)与深度学习(Deep Learning, DL)技术发展迅速,成功应用在许多领域。自然语言处理目前已经被广泛应用于机器翻译、情感分析、推荐系统及信息抽取等方面,这也为从海量的互联网非结构化文本中快速抽取出武装冲突事件的要素并将其以结构化的形式存储提供了可能。随着各种机器学习算法和空间信息的引入,信息抽取可以利用跨句、跨文档的特征信息,相比于传统的模式匹配和句子级的信息抽取方法,能有效减少误差,进一步提高信息抽取的精度。同时,GIS中地理编码技术的运用可以有效提高地缘冲突事件抽取的效率,实现冲突事件的自动、实时定位;地理大数据挖掘和空间分析技术的运用为地缘冲突事件的时空演变规律和内在机理挖掘提供了强大的分析工具。地缘环境系统科学与地理科学的深度融合、地理大数据与人工智能技术的广泛应用将是推动本学科领域发展的重要动力。

参考文献(References):

- [1] 陆大道,杜德斌.关于加强地缘政治地缘经济研究的思考[J].地理学报,2013,68(6):723-727. [Lu D D, Du D B. Some thoughts on the strengthening of geopolitical and geo-economic studies[J]. Acta Geographica Sinica, 2013, 68(6):723-727.]
- [2] 胡志丁,陆大道,杜德斌,等.未来十年中国地缘政治学重点研究方向[J].地理研究,2017,36(2):205-214. [Hu Z D, Lu D D, Du D B. Key research direction of China's geopolitics in the next decade[J]. Geographical Research, 2017,36(2):205-214.]
- [3] 王礼茂,牟初夫,陆大道.地缘政治演变驱动力变化与地缘政治学研究新趋势[J].地理研究,2016,35(1):3-13. [Wang L M, Mou C F, Lu D D. Changes in driving forces of geopolitical evolution and the new trends in geopolitics studies[J]. Geographical Research, 2016,35(1):3-13.]

- [4] 葛全胜, 江东, 陆锋, 等. 地缘环境系统模拟研究探讨[J]. 地理学报, 2017, 72(3): 371-81. [Ge Q S, Jiang D, Lu F. Views on the study of geopolitical environment system simulation[J]. Acta Geographica Sinica, 2017, 72(3): 371-381.]
- [5] 江东, 王倩, 丁方宇. 大数据时代的地缘环境研究[J]. 科技导报, 2018, 36(3): 41-48. [Jiang D, Wang Q, Ding F Y. Research on the geo-environment in the era of big data[J]. Science & Technology Review, 2018, 36(3): 41-48.]
- [6] Singer J D. The "Correlates of War" project: Interim report and rationale[J]. World Politics, 1972, 24(2): 243-270.
- [7] Azar E E. The conflict and peace data bank (COPDAB) project[J]. Journal of Conflict Resolution, 1980, 24(1): 143-152.
- [8] McClelland C A. Warnings in the international event flow: EFI and ROZ as threat indicators[J]. International Interactions, 1979, 5(2-3): 135-202.
- [9] 唐永胜, 刘东哲, 陈晓东. 冷战后全球武装冲突的特点及演变[J]. 现代国际关系, 2008(8): 1-9. [Tang Y S, Liu D Z, Chen X D. Post-cold war international armed conflict: Features and evolution[J]. Contemporary International Relations, 2008(8): 1-9.]
- [10] Melvin S, Singer D. Resort to arms: International and civil wars: 1816-1980[M]. New York: Sage, 1980.
- [11] Sarkees M R, Wayman F W. Resort to War: 1816-2007 Correlates of War[M]. Washington DC: CQ Press, 2010.
- [12] Wallensteen, Peter, Axell, et al. Armed conflict at the end of the Cold War, 1989-1992[J]. Journal of Peace Research, 1993, 30(3): 331-346.
- [13] Gleditsch, Petter N, Wallensteen, et al. Armed conflict 1946-2001: A new dataset[J]. Journal of Peace Research, 2002, 39(5): 615-637.
- [14] Gochman C S, Maoz Z. Militarized interstate disputes, 1816-1976: procedures, patterns, and insights[J]. Journal of Conflict Resolution, 1984, 28(4): 585-616.
- [15] Ghosn F, Palmer G, Bremer S A. The MID3 data set, 1993-2001: procedures, coding rules, and description[J]. Conflict Management Peace Science, 2014, 21(2): 133-154.
- [16] Raleigh C, Linke A, Hegre H, et al. Introducing ACLED: An armed conflict location and event dataset[J]. Journal of Peace Research, 2010, 47(5): 651-660.
- [17] Partell P J. Escalation at the outset: An analysis of targets' responses in militarized interstate disputes[J]. International Interactions, 1997, 23(1): 1-35.
- [18] Kreutz J. How and when armed conflicts end: Introducing the UCDP conflict re-termination dataset[J]. Journal of Peace Research, 2010, 47(2): 243-250.
- [19] Raleigh C. Violence against civilians: A disaggregated analysis[J]. International Interactions, 2012, 38(4): 462-481.
- [20] Schrodt P A, Gerner D J, Hermreck, et al. The CAMEO (conflict and mediation event observations) actor coding framework[M]. Washington DC: American Political Science Association, 2005.
- [21] 吴长义. 第二次大战后的150多次武装冲突[J]. 现代外国哲学社会科学文摘, 1981(6): 60-61. [Wu C Y. More than 150 armed conflicts after the Second World War[J]. Digest of Foreign Social Sciences, 1981(6): 60-61.]
- [22] 贵琳. 一九九七年世界地区冲突索引[J]. 国际研究参考, 1998(1): 22-24. [Gui L. 1997 world regional conflict index[J]. International Research Reference, 1998(1): 22-24.]
- [23] 张宇燕, 李东燕, 邹治波. 全球政治与安全报告(2020)[M]. 北京: 社会科学文献出版社, 2020. [Annual report on international politics and security(2020) [M]. Beijing: Social Sciences Academic Press, 2020.]
- [24] 阎学通. 中外关系鉴览: 1950-2005: 中国与大国关系定量衡量[M]. 北京: 高等教育出版社, 2010. [Yan X T. Sino Foreign Relations: 1950-2005: quantitative measurement of China's relations with major countries[M]. Beijing: Higher Education Press, 2010.]
- [25] 董青岭. 从事件赋值走向关系赋值: 双边关系的定量衡量——评《中外关系鉴览 1950-2005——中国与大国关系定量衡量》[J]. 外交评论(外交学院学报), 2011, 28(2): 154-155. [Dong Q L. From event assignment to relationship assignment: quantitative measurement of bilateral relationship: A review of the review of Sino foreign relations 1950-2005: quantitative measurement of China's relations with major countries[J]. Foreign Affairs Review, 2011, 28(2): 154-155.]
- [26] 郭锐, 王箫轲. 国际关系定量研究与数据库建设——评《中外关系鉴览 1950-2005——中国与大国关系定量衡量》[J]. 世界经济与政治, 2011(7): 128-139, 59-60. [Guo R, Wang X K. Quantitative research on international relations and database construction: A review of the review of Sino foreign relations 1950-2005: Quantitative measurement of China's relations with major countries[J]. World Economics and Politics, 2011(7): 128-139, 59-60.]
- [27] 周建仁. 国家双边关系的定量衡量: 突破与局限——评《中外关系鉴览 1950-2005: 中国与大国关系定量衡量》[J]. 欧洲研究, 2011, 29(1): 143-152, 61. [Zhou J R. Quantitative measurement of bilateral relations: breakthrough and limitation: A review of the review of Sino foreign relations 1950-2005: Quantitative measurement of China's relations with major countries[J]. Chinese Journal of European Studies, 2011, 29(1): 143-152, 61.]

- [28] 周亦奇,封帅.安全风险分析的方法创新与实践——以“一带一路”政治安全风险数据库建设为例[J].国际展望,2017,9(5):147-166,73-74. [Zhou Y Q, Feng S. Innovation and practice of one method of safety risk one case study of "one belt, one road" political security risk database[J]. Global Review, 2017,9(5):147-166,73-74.]
- [29] Hao M, Fu J, Jiang D, et al. Simulating the linkages between economy and armed conflict in India with a long short-term memory algorithm[J]. Risk Analysis, 2020.
- [30] Hao M, Jiang D, Ding F, et al. Simulating spatio-temporal patterns of terrorism incidents on the Indochina Peninsula with GIS and the random forest Method[J]. ISPRS International Journal of Geo-information, 2019,8(3):133.
- [31] 冯长强.基于GIS的地缘环境分析方法与建模研究[D].郑州:解放军信息工程大学,2017. [Feng C Q. Research on geo-environment analysis method and modeling based on GIS[D]. Zhengzhou: PLA Information Engineering University, 2017.]
- [32] Lafree G, Dugan L. Introducing the global terrorism database[J]. Terrorism and Political Violence, 2007,19(2):181-204.
- [33] Sundberg R, Melander E. Introducing the UCDP georeferenced event dataset[J]. Journal of Peace Research, 2013, 50(4):523-532.
- [34] Weidmann N B, Ward M D. Predicting conflict in space and time[J]. Journal of Conflict Resolution, 2010,54(6): 883-901.
- [35] Onda C. The effect of the spatial resolution of conflict data on the analysis of drought as a determinant of civil war onset: Africa, 1980-2001[J]. Consilience, 2008(1):52-64.
- [36] Cederman L-E, Buhaug H, Rød J K. Ethno-nationalist dyads and civil war: A GIS-based analysis[J]. Journal of Conflict Resolution, 2009,53(4):496-525.
- [37] Schrodtt P A. Parallel event sequences in international crises[J]. Political Behavior, 1990,12(2):97-123.
- [38] Gerner D J, Schrodtt P A. A new event data framework for the analysis of foreign policy interactions[M]. New Orleans: International Studies Association, 2000.
- [39] O'brien S P. Crisis early warning and decision support: contemporary approaches and thoughts on future research [J]. International Studies Review, 2010,12(1):87-104.
- [40] Leetaru K, Schrodtt P A. GDELT: Global data on events, location and tone, 1979-2012[J]. In ISA Annual Convention, 2013,2(4):1-49.
- [41] Galla D, Burke J. Predicting social unrest using GDELT [C]. Proceedings of the International Conference on Machine Learning and Data Mining in Pattern Recognition, Springer, 2018.
- [42] Qiao F, Li P, Zhang X, et al. Predicting social unrest events with hidden Markov models using GDELT[J]. Discrete Dynamics in Nature and Society, 2017(2017):1-13.
- [43] Qiao F, Li P, Deng J, et al. Graph-based method for detecting occupy protest events using gdeltdataset[C]. Proceedings of the 2015 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery, IEEE, 2015.
- [44] Yonamine J E. Predicting future levels of violence in Afghanistan districts using GDELT[D]. State College: The Pennsylvania State University, 2013.
- [45] Keneshloo Y, Cadena J, Korkmaz G, et al. Detecting and forecasting domestic political crises: A graph-based approach[C]. Proceedings of the 2014 ACM conference on Web science, 2014.
- [46] Fallahi F. Machine learning on big data for stock market prediction[D]. Carbondale: Southern Illinois University, 2017.
- [47] Jakel T. Using sentiment data from the global database for events, language and tone (GDELT) to predict short-term stock price developments[D]. Enschede: University of Twente, 2019.
- [48] Alamro R, Mccarren A, Al-Rasheed A. Predicting Saudi stock market index by incorporating GDELT using multivariate time series modelling[C]. Proceedings of the International Conference on Computing, 2019, Springer.
- [49] 王树良,丁刚毅,钟鸣.大数据下的空间数据挖掘思考[J].中国电子科学研究院学报,2013(1):8-17. [Wang S L, Ding G Y, Zhong M. On spatial data mining under big data[J]. Journal of China Academy of Electronics and Information Technology, 2013(1):8-17.]
- [50] 杨宗亮,张玉茜,李建飞.一种基于地理空间大数据的网络舆情监测软件架构[J].测绘通报,2017(3):96-100. [Yang Z L, Zhang Y Q, Li J F. An architecture of public sentiment monitoring based on geospatial big data[J]. Bulletin of Surveying and Mapping, 2017(3):96-100.]
- [51] Ward M D, Beger A, Cutler J, et al. Comparing GDELT and ICEWS event data[J]. Analysis, 2013,21(1),267-297.
- [52] Wang W, Kennedy R, Lazer D, et al. Growing pains for global monitoring of societal events[J]. Science, 2016,353 (6307):1502.