DOI:10.13530/j.cnki.jlis.190029

计算型情报分析的进展*

李广建　罗立群

摘　要　大数据理念和方法的不断应用,使情报分析在理论和实践上逐渐从定量化分析向计算化分析发展。本文将迄今为止的计算型情报分析的发展历程划分为三个阶段:经验计数分析阶段、精确计量分析阶段、计算型分析阶段,并分别从分析对象、代表成果、主要贡献、主要特征、方法与工具等维度对各阶段进行论述;对情报思维、计算思维、社会计算科学进行辨析,提出了计算型情报分析的主要研究目的和研究内容,并对计算型情报分析的三个主要内容——计算型情报分析模型、计算型情报分析算法、计算型情报分析框架和系统最新的研究进展和研究趋势进行系统的论述和分析。图 2。参考文献 71。

关键词　计算型情报分析　情报分析框架　情报分析模型　情报分析算法

分类号　G250.2

Progress in Computational Intelligence Analysis

LI Guangjian & LUO Liqun

ABSTRACT

In the 21st century, with the wide application of network technology and the rise of big data ideas and methods, intelligence analysis has gradually evolved from simple quantitative analysis to computational analysis in theory and practice. Computational intelligence analysis (CIA) is gradually taking shape. The development process of CIA can be divided into three stages: the empirical counting analysis stage, the accurate measuring analysis stage, and the computational analysis stage. This paper discusses these stages from 5 aspects: the analysis objects, the representative results, the main contributions, the main features, the methods and tools. By summarizing the concept "computation" in the area of computational thinking, social computation and social computing science, this paper puts forward the main contents of CIA which include CIA model, CIA framework and system, and CIA algorithm.

The CIA model is a kind of mathematical models, which is used to abstract and simulate the behavior from the specific task-oriented intelligence analysis subject, the specific intelligence object and the specific procedure of the intelligence transmission. With the introduction of big data analysis models, social network models, complex theoretical models and system dynamics models into intelligence analysis and their fruitful achievements, the construction and application of the CIA models have shown the following characteristics

* 本文系国家社会科学基金重大项目"大数据时代知识融合的体系架构、实现模式及实证研究"(编号:15ZDB129)的研究成果之一。(This paper is an outcome of the key project "The Architecture, Implementation Pattern and Empirical Research of Knowledge Fusion in the Big Data Era" (No. 15ZDB129) supported by National Social Science Foundation of China.)

通信作者:李广建,Email:ligj@pku.edu.cn,ORCID:0000-0002-2897-6246(Correspondence should be addressed to LI Guangjian, Email:ligj@pku.edu.cn,ORCID:0000-0002-2897-6246)

and trends in recent years: First, the CIA model is no longer limited to simple descriptive statistics modes, but tends to be a kind of more comprehensive or combined mathematical mode. Secondly, the construction of the CIA model is transformed from a static model to a complicated dynamic model, that is, variables changing over time and their relationships are introduced into the model. Thirdly, the construction and application of CIA models are based on massive data or big data. Fourthly, the model construction' s focuses shift from evaluating past achievements to predicting and interpreting future outcomes and scientific discoveries.

The CIA framework mainly refers to the technical framework for intelligence analysis, which consists of abstract components and messaging methods for the components' instances. The CIA framework can be considered as a constrained support structure designed to solve the open-ended problem in the process of intelligence analysis. A CIA system is an implementation of the CIA framework. In recent years, the development and research of CIA framework and system have shown the following characteristics and trends: First, the analytical framework and system are evolving from a single architecture to a distributed, cloud computing architecture. Such framework and system are more in line with the requirements of storage, computing, analysis of multi-source, heterogeneous, massive data. Secondly, the CIA framework and system have changed from a traditional relatively closed, dedicated mode to an open, military-civilian integration mode. Thirdly, the intelligence analysis processes in the framework or system can be customized and automated. Fourthly, the CIA framework and system are becoming more intelligent with the application of artificial intelligence technologies.

The CIA algorithm is a series of computation steps and instructions formed under specific intelligence analysis tasks or intelligence analysis scenarios by the guidance of intelligence thinking. The CIA algorithm is the prerequisites for automated intelligence analysis so that the intelligence analysis model can be calculated by machine and finally obtains the computational results. The CIA algorithm has recently shown some characteristics and trends: First, the explain ability of intelligence analysis algorithm is being initiated. Secondly, the open source intelligence-oriented perception algorithm is increasingly being valued. Thirdly, the algorithm for the fusion of human analyzed results and computer analyzed results has got more attention.

The CIA is an extension of traditional intelligence analysis and now it becomes one of the important directions of future development of information science. The CIA will gradually develop into a new discipline that integrates computational science, social science, cognitive science, complex science, biochemical science, engineering, and many other academic subjects. For researchers, opportunities and challenges will coexist. This requires us to meet the challenges of the future with more innovative, open and bold ideas. 2 figs. 71 refs.

KEY WORDS

Computational intelligence analysis. Intelligence analysis framework. Intelligence analysis model. Intelligence analysis algorithms.

0 引言

近年来,大数据理念和方法不断完善并得以广泛应用,对情报学特别是情报分析产生了深刻的影响,情报分析无论是在理论上,还是在实践上,都在强调定量化和自动化,并逐渐从简单的定量化分析发展到了计算化分析。国际上,以美

国为代表的发达国家对情报分析的计算化或计算型情报分析非常重视,理论上已经规划出了计算型情报分析研究体系,实践上对计算型情报分析投入大量的人力、物力、资金。在国内,学者们提出基于事实型数据的情报[1-2]、工程化情报[3]、计算型情报[4]、面向突发事件的快速响应情报[5]以及情报 3.0[6]等新的情报分析理念。这些研究成果的共同结论是,大数据时代下,情报分析应以数据为原始材料,经过一系列的计算与分析,深度解析数据中存在的内容及其关系,帮助分析人员解决情报问题,完成情报任务[7]。由此可见,国内对情报分析的计算化或计算型情报分析的研究已经开始起步,并且越来越重视。

1　计算型情报分析的发展历程

计算型情报分析本身有一个发展过程,经历从低级到高级的不同阶段,从重在将分析对象转化为可以度量的数值,到重在用数学公式来表示情报计算过程中的一般规律,再到重在方便计算代理进行计算的表示和模型构建。我们将迄今为止的计算型情报分析的发展历程划分为三个阶段,分别是经验计数分析阶段、精确计量分析阶段、计算型分析阶段,如图 1 所示。

发展阶段	经验计数分析阶段	精确计量分析阶段	计算型分析阶段
处理对象	文献量的关系	文献引文关系、网络的引用关系、文献内外部特征	大规模数据、多源数据
代表性成果	布拉福德定律、洛特卡定律、普赖斯定律	文献计量学、引文分析法、H指数、PageRank、文本挖掘	文献结构模型、主题模型、用户模型等
主要贡献	文献数量规律的描述,(作者写作规律,文献数量分布等)	发现重要的科学期刊、科学论文、学者	情报分析的建模
主要特征	数值化	公式化	计算化
方法与工具	观察 假设—验证	统计	数学建模

图 1　计算型情报分析的发展历程

1.1　经验计数分析阶段

早期情报分析的基本处理单元是文献整体,通过对文献的研究间接地反映知识的状况,相应的,情报分析也是建立在对文献整体的认知之上的。因此,早期情报分析主要是借助于定量和量化方法对文献的外部特征进行计量[8],主要的方法是先观察,进而假设,再根据假设进行实验,如果实验的结果与假设不符合,则修正假设再实验。经验计量的典型代表是文献计量的三大定律:洛特卡定律、普赖斯定律和布拉福德定律。美国学者 A. J. 洛特卡在 20 世纪 20 年代经过对科学论文及其作者的分布关系进行长期的观察和计数,发现了科学工作者人数与其所著论文之间的数量关系,并用两组论文数据进行了验证[9]。在洛特卡定律的研究基础上,普赖斯进一步研究了科学家人数与科学文献数量之间的关系[10],提出了著名的普赖斯定律,其在《小科学,大科学》中论述:“在同一主题中,半数的论文为一群高生产能力作者所撰,这一作者集合的数量约等于全部作者总数的平方根。”洛特卡定律和普赖斯定律都是通过对文献的计数实现了对研究人员(文献作者)的量化研究,而布拉德福定律则是描述文献序性结构的定量经验定律[11],该定律指出,如果将科技期刊按其刊载某学科专业论文的数量多少,以递减顺序排列,可以把期刊分为专门面对这个学科的核心区、相关区和非相关区。各个区的文章数量相等,此时核心区、相关区、非相关区期刊

数量成 1∶n∶n^2的关系。维克利进一步将杂志分区的数目推广到大于三个的更普遍的情形，提出了布氏定律的维克利修正式[12]。从上述文献计量的三大定律发展过程来看，都是先经过观察和计数，获得一定范围内的经验估计和理论估计，再到被后人完善、修正并推广到更为一般或普遍的情况，成为真正意义上的文献计量定律。因而，这种经验计量分析的情报分析也可以看作是计数型情报分析，其主要特征包括以下几个方面。

(1)计数的对象是文献数量及作者数量(这两个量是文献最为显著的外部特征)。比如洛特卡定律描述了作者人数与其所著论文数量之间是一种倒平方的关系；普赖斯定律描述了不同文献作者之间发表论文数量的定量关系。

(2)分析结论大多源于研究人员在长期实践工作中的观察，是对经验和规律的总结，在提出假设的基础上，用单一的数学公式进行表示，所得出的公式结论在总体上是正确的，但并不是精确的计量结果或统计分布。

(3)使用的数学工具还比较简单和单一，主要是通过计数将计量的对象变为数值。数值化是计量研究的基础，定量化首先需要将计量对象进行数值表示，然后再通过计数对文献、作者、时间的维度进行定量分析，仅强调这些计量对象之间的等量关系，与其说是精确的度量，不如说是经验上的理论估计。

1.2 精确计量分析阶段

计数型情报分析开创了情报分析定量化的先河，为情报分析的精确计量化奠定了基础，也一直是情报学的重要研究内容，进入精确计量分析阶段，定量化的情报分析发展主要体现在三个方面。第一个方面是继续对已有的文献定律进行验证和完善，同时扩大并推广这些定律的应用领域，这部分研究为后来文献计量学的产生奠定了基础。第二个方面是综合运用数学、统计学、情报学的思想和方法，不断探索新的情报分析方法，这其中最为典型的是引文分析。20 世纪 50 年代，美国著名情报学家尤金·加菲尔德(Eugene Garfield)发明了引文索引，系统地提出了利用引文检索科技文献的新方法，开创了从引文角度来追踪科学发展动态的新领域，也开启了引文分析的大门。引文分析方法的中心就是测算被引数量，该方法最先被广泛用于评价科研成果的质量和影响力。根据加菲尔德的说法，“引文索引有助于历史学家衡量一篇文章的影响力——即它的‘影响因子’”[13]。此后，Google 的创始人 Larry Page[14] 发明了针对 Web 信息的引用(超链接)算法 PageRank，并通过对超链接集合中的元素用数字进行权重赋值，实现了“衡量集合范围内某一元素的相关重要性”的目的。目前，PageRank 的思想和方法除了在搜索引擎中发挥着巨大的作用以外，在社交网络分析、网络舆情分析、文献及科研机构影响力评价、网络用户兴趣分析等方面有着广泛的应用。以引文分析为代表的理论和方法，是精确计量分析阶段最为成熟的计算情报分析的理论和方法。第三个方面是 20 世纪 90 年代以来，伴随着计算机技术的广泛应用和深入发展，自然语言理解、信息抽取、自动标引、自动分类、自动文摘、关联规则等技术在情报学中的应用，使得情报计量和分析的对象从文献外部特征转向了文献的内部特征，计量的粒度也从文献的粗粒度演化到章节、信息片段、关键词等细粒度。建立在计算机自动化处理基础上的内容分析法，引起了情报学界的极大关注，成为研究热点，已经逐渐成为与上述两个方面相并列的核心内容[15]。精确计量分析阶段的主要特征包括以下几个方面。

(1)精确计量阶段的情报分析逐渐从基于分析人员业务背景及知识积累的业务驱动模式，向重视情报内容中反映出的数量关系从而得出相关结论的数据驱动模式转变。“用数据说话”几乎成了这一阶段情报分析成果的必备要求。但总的来说，除引文分析以外，其他的量化分析仍然处于辅助地位，对数据关系的认识和解释主要还是靠分析人员的主观判断。

(2)由于广泛地应用了自然语言处理技术和数据挖掘技术,这一阶段的情报分析已经将期刊论文以外的专利文献、政策文本、行业报告、社交网络数据、网页信息、网络日志、政府公开数据等都纳入了自己的分析范围,从这个阶段开始,情报分析开始从文献情报分析或科技情报分析走向了全领域情报分析,情报分析的价值正在凸显,其理论与方法已经渗透到多个不同学科领域。

(3)分析过程引入了一些更加复杂的统计和数学模型,特别是 20 世纪末 21 世纪初以来,定量化的内容分析需要一些复杂的统计模型来揭示内容之间的关系,例如,对文献内部词频的统计,k-means 方法可以实现对不同文献的自动聚类分析,支持向量机(SVM)可以实现对不同类别文献的自动分类分析。这一时期的计量过程更加复杂,计量的维度更多,从而能够得到更加丰富、更为深刻的情报分析结果。

1.3　计算型分析阶段

进入 21 世纪,随着网络技术的广泛应用以及大数据理念和方法的兴起,计算型情报分析正在逐渐形成。计算型情报分析主要借助于数据建模的方式对情报问题和现象进行刻画和推理,运用计算的方式对大规模数据进行分析,从而获得有价值的结论。例如,Neil Johnson 等人[16]对恐怖主义和叛乱活动升级的模式进行建模,以此来估计致命性攻击的升级速度和时间。乔治城大学的 Kalev Leetaru[17]使用了数百万个开源指标的数据库借助于数学模型准确预测了本·拉登在巴基斯坦 Abbottabad 的地点。美国国防部高级研究计划署(DAPRA)花费数十亿美元资助 XDATA 项目,构建基于云计算和大数据的情报分析计算框架,用于从海量来源的异构数据中进行大规模的情报分析以获取有价值的情报,等等。计算型情报分析阶段的特征包括以下几个方面。

(1)情报分析的对象更加多元化和融合化。在精确计量阶段,情报分析的对象就已经开始多元化了,分析的数据不仅包括论文、专利等科技文献,还包括各种网络资源、统计数据等。但主要还是根据情报任务的需求,对各类型的数据分别做独立的分析。计算型情报分析阶段则更强调在数据多元化的基础上,进行数据的融合,综合利用多种数据相互补充、交叉认证,以获得对所研究事物的全面认识,提高情报分析的准确性和科学性。

(2)情报分析的模型化。计算型情报分析不再是简单的计数和计量,其目的不是简单的数据拟合,而在于逻辑推理。因此,计算型情报分析十分强调模型的作用,通过建立模型,化繁为简,把复杂的实际问题用数学模型进行抽象和表示,提高情报分析的洞察力。例如,通过对社交网络中流行病数据进行分析建立流行病预测模型,可以对某些区域的流行病流行发生的时间、范围、规模作出预测。另外,经过验证的模型还可以被其他研究依照不同的分析需求进行重组或再利用,从而提升情报分析的可重复性和科学性。

(3)情报分析过程的自动化。大数据时代,传统的以人工处理为主的情报分析方法,已经无法满足分析多样化海量数据的要求,必须要借助以计算机为代表的信息技术的运算能力来提高分析效率。在以前的发展阶段中,自动化计算处于辅助地位,人的分析处于主导地位,而在计算型分析阶段则正好相反,自动化计算上升到了主导地位,人的工作更多集中于与机器的协同,帮助机器理解人类的思维方式和意图,而不是直接干预分析的过程和对结果的解释,这样才能最大限度地保证情报分析结果的客观性和可靠性。

2　计算型情报分析的进展

从上述计算型情报分析的发展历程来看,情报分析正在从计量、定量走向计算,其“计算”的特征日渐突出。这里,我们首先梳理学界对“计算”的理解,进而总结计算型情报分析的内

容及其进展。

谈及计算的概念,就必须要提到“计算思维”(Computational Thinking)。“计算思维”这一概念最早由美国卡内基·梅隆大学计算机系主任周以真(Jeannette M. Wing)[18]教授提出,她认为,计算思维是一种分析思维,是借助计算这一基本概念来解决问题、设计系统以及理解人类行为的一种方法,具体地说,是对物理世界及其层次、关联关系的抽象(Abstraction)并用自动化(Automation)方式让计算机去处理、解析这些抽象,所谓的计算,就是对抽象的自动化(Computing is the automation of our abstractions)[19]。可见,计算概念包含了两个要素:抽象和自动化。Chenglie Hu[20]进一步明确指出,当前计算思维中的计算本质更像是追求“表示和模型”(Representations and Models),即通过模型构建或表示来说明一个信息被合适的计算代理(Computing Agents)执行的内部运行机制,其过程必须符合逻辑的、算法的、科学性的、数学的、分析的、面向工程的等原则。

涉及计算并且与计算型情报分析密切相关的另外两个概念是社会计算和计算社会科学。其中,社会计算中的计算概念是指“计算技术”(Computational Techniques),当前,社会计算从关注如何开发社会软件来提供他人使用[21],转变为关注通过多种计算技术辅助人们分析社会行为,同时也关注创建一个更强大的计算基础设施来支持人们的协同工作[22]。在计算社会科学领域中,Cioffi-Revilla[23]认为计算的概念是建立在信息处理视角下的“计算或可计算的方法”(Computation or Computational Approaches),即运用大量基于计算机的设备提供从信息抽取算法到计算机仿真模型的方法应用。我国学者也指出,计算社会科学的核心技术是数据挖掘[24]。

综上可以看出,当前环境下所说的计算,不是简单意义上的数量统计或工具应用,而是一个更为全面和概括性的概念,是学科发展、信息技术创新以及社会需求三者共同作用的结果。计算型情报分析也是如此,在本质上它是情报分析的基本规律在演化过程中,不断被渗透和应用了计算思维,并与社会计算和计算社会科学相交融而形成发展起来,因此,计算型情报分析的内容实际上就是计算概念在情报分析中的具体化,可以概括为情报分析模型、情报分析框架与系统、情报分析算法三个方面,如图2所示。以下从这三个方面讨论计算型情报学的当前研究进展。

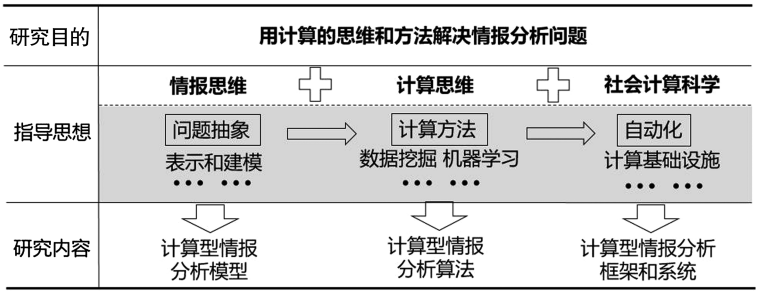

图2 计算型情报分析研究内容

2.1 计算型情报分析模型

模型方法一直以来都是情报分析研究和实践中的抽象与具体、理论与实践相互转化的桥梁[25],也是计算思维中“抽象”和“表示和模型”的具体体现。抽象的本质是将要解决的问题进行表示和数据建模,建模的过程就是理解问题的过程。计算型情报分析模型首先是一种数学模型[26],它用来对面向特定任务的情报分析主体的行为、特定的情报传递过程或特定的情报对象进行抽象和模拟。例如,构建科研人员研

究行为的动力学模型,并借助该模型来分析和预测在科学交流网络中有潜在影响力和创新力的团队和个人。

情报分析有建模的优良传统并已积累了丰硕的成果,例如 H 指数、G 指数以及其变种,都是用来评价个体学者的已有学术成就和影响力的数学模型[27]。近年来,随着大数据及人工智能技术的发展,后现代数学方法和工具的应用,为情报分析注入了新活力。研究人员将大数据分析模型、社会网络模型、复杂理论模型、系统动力学模型等引入到了情报分析过程中,取得了非常丰富的成果。其中最为典型的代表是 2013 年以来发表在《科学》(*Science*)、《自然》(*Nature*)、《物理评论》(*Physical Review*)、《美国科学院院刊》(*PNAS*)等世界顶级杂志上的一系列有关创新成果识别、团队影响力、科学发展预测方面的论文,充分体现了计算型情报分析模型的最新发展动态。这些研究采用复杂网络、结构方程、随机过程、动力学等模型,或者是通过机器学习的方法,用大量数据训练出模型,再用历史数据或其他相关数据进行验证,从而使该模型能够描述、预测科学发展及科学家影响力的趋势;或者基于复杂网络理论,构建随机网络模型,并根据样本数据构建出与现实最为吻合的理想模型,来刻画研究科学创新的内在驱动机制和发展规律,据此解释、预测科学论文的影响力。

总的来看,近年来有关模型的构建和应用,呈现出如下特点和趋势。

第一,不再仅局限于单纯的描述性统计,而是趋向于综合或组合使用描述性统计、复杂网络模型、高级计量经济学方法、机器学习模型、数学分析和计算机模拟模型等,情报分析中的计算特征越来越突出。例如,Kyebambe 等人[28]综合运用专利引用分析方法和监督学习法通过自动标记数据来构建预测新兴技术的模型,该模型在一年内的预测精度达到 70%。Lee 等人[29]借助前馈多层神经网络模型来捕获特定时间段内能够识别新型技术的指标之间复杂非线性关系,从而识别出处于早期阶段的新兴技术,并运用动力学模型来预测一项技术随着时间的推移可能出现的变化。Nsoesie 等人[30]运用一种仿真优化的方法(SIMOP)预测流感流行曲线的时效,旨在结合使用模拟、分类、统计和优化技术来预测流行病曲线,并在流感爆发期间推断出潜在的模型参数,该方法预测了弗吉尼亚州、迈阿密、西雅图和周边大都市区的蒙哥马利县的社交网络上流行病演变情况,结果显示该模型能够在前七周就预测到高达 95%的流行病。

第二,分析模型的构建从静态模型向复杂的动态模型转化,也就是在模型中引入随时间变化的量及其关系。例如,Sinatra 等人[31]于 2016 年在《科学》杂志上发表了《个体科学家影响力演变的量化》,对科学家影响力(通过有影响力的出版物来衡量)的演变过程进行研究,加入了时间维度来量化科学家职业生涯中影响力和生产力的变化,发现这些影响力在科学家的出版物序列中是随机分布的。根据这个发现,他们通过随机影响力的规则构建了一个随机模型,将发表产量、个人能力和运气的影响作为参数,该模型能够揭示出科学研究成功的普遍模式,并能够准确地预测出科学家影响力的演变,包括 H 指数累计引用以及学术获奖等。Yian Yin[32]等人在对三个大型数据集(美国 NIH 基金、美国创业企业 VentureXpert 数据库、全球恐怖主义数据库)中有关成功和失败的数据分析的基础上,开发了用来表示成功和失败的动力学模型,该模型能够根据动力学信息所得到的早期信号将成功者和无法获得成功者区分和识别出来。这些例子都表明,当前包括文献分析在内的情报分析远远超出了以往计量分析过程中所做的简单统计分析,在以往揭示情报分析对象中相关属性之间的关系等结构特征的基础上,更强调构建能够揭示情报分析对象中相关属性之间随时间变化而变化的规律模型,从而能够得出更多、更有价值的分析结果。

第三,情报分析模型构建与应用以海量数

据或者是大数据为基础。例如,Brian Uzzi[33]等人在研究科学论文的影响力时,分析了 Web of Science 收录的 1950—2000 年间出版的 15 613 种期刊上的 1 790 万篇研究论文,所构造的引文随机网络模型中包含了 3.02 亿条边(引用)。通过将该引文网络中的引文频率与实际的引文频率作比较,进而计算出引用的 Z 分数(标准分数)来反映某个论文的新颖性。该项研究得出结论,创造了新知识又不脱离主流研究领域和范式的科研成果,具有较强的科技影响力。又例如,Dashun Wang 等人[34]在《长期的科学影响力的量化》一文中运用复杂网络中的动力学模型,开发了一种反映引文动态变化的论文引用机制模型(Mechanistic Model),用来发现长期的论文引用规律,这项研究使用的数据来自于美国物理学会(ASP)提供的 1893 至 2010 年的《物理评论》)语料,共 45 万篇论文及其引文,同时为了避免模型的学科依赖性,还使用《自然》《科学》等 12 种综合性及理、工、医学等领域的知名期刊在 1990、1995、2000 三年发表的全部论文以及 2011 以来对这些文章的引用文章。

第四,研究的重心从评价过去的成果向预测和解释未来的成果以及科学发现转变。例如,上述 Brian Uzzi 等人对科学论文影响力的研究,不仅发现了什么类型的研究更容易被学界接受,更为重要的是,他们所构建的论文引用随机网络模型,还可以用来预测具有创新性的研究成果及其未来的影响力,这是该研究的重要目标。另一个典型的例子是 2016 年由南加州大学维特比工程学院信息科学研究所举办的“混合预测竞赛”(Hybrid Forecasting Competition, HFC)项目[35-36],该项目的目标是对世界范围内的社会经济和地缘政治问题进行更准确的预测,例如疾病爆发、股市波动、选举和国家冲突等。该项目开发了一种融合机器自动预测和“众包”预测的模型,能够融合人类和机器系统的优势,通过人机交互,帮助分析人员对疾病爆发、股市波动、选举和国家冲突等进行有效的预测。

2.2 计算型情报分析算法

模型建立只是计算型情报分析的一个步骤,为了让计算机帮助求解,需要虚拟的符号来代替数学模型里的每个变量和运算规则,即把解题思路用程序语言完整地描述给机器。算法(Algorithm)就是解题方案准确而完整的描述,是一系列解决问题的清晰指令[37-38]。情报分析的算法是建立在通用的算法设计、实现思路和方法的基础之上,是在情报思维的指导下面向特定的情报分析任务或情报分析场景而形成的一系列计算步骤和指令,它是机器进行情报分析自动化的前提,它用机器可以理解的方式把情报分析的思路和过程进行描述,使得情报分析模型得以用机器加以计算并最终获得计算的结果。

近 10 年来,云计算技术、物联网技术、大数据技术、人工智能技术等领域均取得重大进展,极大地推动了情报分析算法的研究和实践,西方情报强国都积极地投入大量人力和资金研发先进的情报分析算法,诸如 Darpa Maven[39]、AI-Next[40]、Big-Mechanism[41]、World Modler[42]等项目。这一类的研究项目具有两方面的特点,第一,研究内容都是面向情报领域重大、基础的颠覆性技术和算法,这类算法在未来的情报活动中具有不对称优势;第二,研究的算法重点解决面向海量、异构、多源的数据挑战且提供的情报服务具有较高的实时性。除了专门研究算法的项目以外,研究人员还针对专门领域的情报分析任务开发了解决具体问题的算法,这些领域包括人道主义危机[43]、大规模暴力和骚乱[44]、大规模移民[45]、疾病爆发[46]、经济危机[47]、资源短缺和自然灾害[48],等等。

近年来,有关情报分析算法的研究和实践呈现出以下特点和趋势。

第一,情报分析算法开始注重可解释性。与大数据分析不同,情报分析除了注重相关性之外,还十分重视结论背后的原因[49]。近年来,情报分析领域越来越认识到,仅仅是发现相关性是不够的,大数据分析算法所分析出的数据

中的“关联”只是情报分析人员认知世界和识别事物的第一步,相关性背后的因果性探索仍有必要,数据的关联性预测并不能替代和取消情报分析中的解释功能,解释是情报分析任务中非常重要的要素,单纯的没有解释说明的分析结果本身是不可靠的。近年来,在医疗保健[50-51]、法律[52]、军事[53]等领域的情报分析十分重视所使用的情报分析算法的可解释性,并逐渐推广到其他领域[54-56]。例如,D Gunning[57]提出了一种“可解释的人工智能”(EXplainable Artificial Intelligence,XAI)研究计划,该计划针对现有机器学习算法存在的问题,如深度学习、SVM 算法、随机森林等算法虽然学习性能较好,但是可解释性较差,所得到的分析结论只能回答是什么,而不能回答为什么,旨在对这些机器学习方法进行改进,以使机器学习得到的模型具有可解释性,从而能够协助情报分析人员在情报分析的人机交互过程中理解分析的结果。该计划拟对自反射性深度学习(Reflexive and Rational Deep Learning)、叙事生成因果关系建模(Narrative Generation Causal Modeling)、三级解释的模式理论(3-Level Explanation Pattern Theory+)、验收测试自适应程序(Acceptance Testing Adaptive Programs)、交互式训练认知建模(Interactive Training Cognitive Modeling)、可解释的强化学习(Explainability RL)、基于展示和讲述解释的深度学习(Show and Tell Explanations Deep Learning)、基于论证和教育的深度学习(Argumentation and Pedagogy Deep Learning)、基于决策图的概率逻辑(Decision Diagrams Probabilistic Logic)、交互式可视化模拟学习(Interactive Visualization Mimic Learning)、基于贝叶斯教学的模式归纳(Bayesian Teaching Model Induction)等 11 类算法进行改进,使改进后的算法不仅具有较强的学习(预测)能力,而且还具备较强的解释能力,最终能够使用具有解释性的学习算法构建基于人类解释的模型。又例如,施乐帕洛阿尔托研究中心(PARC)研发了一种可解释的自主决策智能体 COGLE[58],其界面为用户提供了对 COGLE 推理的解释和见解。

第二,越来越重视面向开源情报的感知算法。随着计算机算力的不断提升和自然语言处理技术的突破,开源情报采集的方法已经从传统的爬虫向更加智能的情报感知系统转型,相应的实现算法不仅包括传统的 PageRank 算法、链接解析算法等,还广泛地使用情感分析、隐喻识别、实体识别、关系抽取的方法,并将这些方法与分析算法相融合,形成一种新的分析算法——感知算法,利用这种算法,情报分析系统或情报分析人员可以从开源、开放的网络环境中提前察觉到现实世界发生的有价值的“信号”,并尽可能敏锐和准确地预测出可能会发生的事件。Kejriwal 等人[59]提出了一种面向开源情报的感知算法,能够在缺乏语料资源(如孟加拉语、印度尼西亚语、旁遮普语、菲律宾宿雾族语和斯瓦希里语)的语言环境下从新闻和社交媒体中提取和处理群体意见和舆论,从而提前发现可能发生的人道主义灾难。从特定的 Web 信息源中获取有价值的犯罪线索,是打击人口贩运、武器走私等领域长期关注的研究问题[60],Szekely 等人[61]提出了一种从多源、异构、海量的 Web 页中提取关键实体信息并构建人口贩运知识图谱的算法,它利用语义技术融合从不同来源获得的数据,通过增量扩展提取了数十亿个三元组,形成了一个巨大的人口贩卖关系网的知识图谱,该项算法已经在多个执法机构和非政府组织中得到了应用,在帮助执法机构或其他组织查找贩运者并解救受害者方面发挥了作用。

第三,人机融合算法得到关注和重视。为充分发挥人的聪明才智,面向未来的情报分析和情报服务更加注重人机融合的分析过程,也就是说,在机器计算的同时,充分考虑人脑思考的结果,并将这种结果与机器计算的结果融合起来,形成最终的分析结论。只是这个融合的过程是由人机融合算法来自动完成的,而不是由人来主观完成的。例如,美国情报高级研究计划署(The Intelligence Advanced Research Pro-

jects Activity, IARPA) 开展了名为 FUSE (Foresight and Understanding from Scientific Exposition)的研究项目,设计一种能够融合人的预测与机器预测的算法,以识别当今未知但可能在三到五年内出现的颠覆性技术。FUSE 开发的自动化方法,利用已发表的科技文献和专利中的信息,根据既定的量化指标自动提名已知和新的技术术语,并对这些技术进行系统、连续和全面的评估,同时,利用众包的方式向特定专家群体提问,征询他们对新技术的预测意见,最终将机器预测与专家群体预测进行融合,以识别或预测出特定领域的新技术。

2.3 计算型情报分析框架和系统

计算型情报分析框架主要是指实现情报分析的技术框架,是为解决情报分析过程中开放性问题而设计的具有一定约束性的支撑结构,表现为一组抽象构件及构件实例间交互的方法[62]。计算型情报分析框架本身并不解决具体的情报分析问题,它只是为实现计算分析而给出一个用于插接、组合解决问题的相关组件的基础。一方面,它界定了实现计算分析的技术边界,进而将相关的软件组件约束在这个边界内,从而保持解决情报分析问题时手段的内聚性;另一方面,它还用来提供支撑解决情报分析问题的可选的配套软件组件或工具。利用计算型情报分析框架,可以根据具体情报分析问题来扩展、安插更多的组成部分,从而更迅速和方便地构建完整的面向情报任务和问题解决的情报分析系统。从某种意义上说,计算型情报分析框架是构建情报分析系统的元标准。

随着大数据的产生和广泛应用,数据的多源、异构、复杂、不确定性等特征给情报分析工作带来的挑战越来越大,情报分析技术正在从小数据集的分析向大数据的分析发展,情报分析技术正在从封闭的框架和软件生态向开放合作的框架和软件生态发展。在这种背景下,计算型情报分析框架成为计算型情报分析重要的研究和实践内容。这其中的典型代表是美国国防部高级研究计划署(Defense Advanced Research Projects Agency,DARPA)花费数十亿美元资助的 XDATA 项目[63],该项目旨在开发包括资源采集、清洗与转换、数据建模、数据分析、结果可视化、用户交互、信息查询等功能在内的面向各种领域和数据的情报分析的通用技术框架,目前该项目已经基本构建完成了一个技术成熟、体系完善的情报计算框架和生态体系,相关的情报部门可以在其基础框架之上快速地搭建个性化的面向具体领域情报工作的情报处理系统和平台。

近年来,计算型情报分析框架与系统的发展与研究呈现出以下特点和趋势。

第一,由于大数据在计算和分析上带来的挑战,分析框架和系统正在从单一架构向分布式、云计算的构架发展,这类框架更加适合多源、异构、海量数据的存储、计算、分析等需求。Hadoop、MonoDB、Storm 等分布式存储/处理技术和框架的出现,为情报的分析和处理提供了方便、成熟的面向服务的调用接口(API),使得研究人员更加关注面向应用场景的框架设计与研究。例如,Sprague 框架是学界和业界最为广泛接受的决策支持系统(Decision Support Systems, DSS)框架[64]。为了能够适应于海量、异构数据存储和处理技术,Watson 等人[65]基于该框架提出了一种面向大数据情报决策的演变框架,能够将不同来源的数据通过大数据集群统一存储在数据湖(Data Lake)中。又例如,图计算在情报的分析与推理中具有重要的地位,许多实际的模型例如引文随机网络模型、社交网络模型等都涉及图计算。在某些情况下构建的网络图会有数十亿个顶点、数万亿个边缘,这对计算来说是一个非常巨大的挑战。Grzegorz 等人[66]提出了一种适合图任务的分布式计算框架——Pregel,能够在数千台通用型计算机上构建图分析集群并且能够支持高效、可扩展和容错的部署。Simmhan 等人[67]提出了一种叫 Goffish 的面向大规模图处理的计算框架,相对于 Pregel 它的性能提高了好几个数量级,该框架同时还设计

了分布式持久图存储,可以在商用服务器集群上进行大规模图分析。

第二,计算型情报分析框架的研发从传统的相对封闭、专用向开放、军民融合的方向转变。传统的情报分析框架通常都是由专门的机构开发,并面向特定的、专用的情报分析领域和机构,相对比较封闭,新技术的更新和应用较慢。大数据与人工智能时代,开源的情报分析社区集中了群体智慧和众包开发的优势,能够对情报分析框架进行较快的更新,极大地加速了新的数据分析与处理技术的研发。例如,XDATA 项目虽然由 DARPA 立项,但从一开始就不针对特定的情报或信息系统数据,十分强调框架的伸缩性和可配置性,不仅将业界许多知名的数据处理、可视化包、分布式计算等软件平台和工具库,像 Apache Spark、Mesos、DeepDive、BayesDB、Apache Nutch、NumPy 等都包含其中,并且将大部分研究成果以开源的方式在互联网上共享,因此,这个框架不仅可以应用于国防和军事情报分析,也可以在民用领域广泛应用并取得了一定的商业成果。比如 XDATA 项目的承包商之一 Uncharted 公司将其情报分析技术应用于美国的执法机构,用来提取和分析嫌疑人的行为特征[68]。

第三,情报分析流程定制与自动化是情报计算框架和系统发展的一个重要趋势。伴随着情报分析任务的复杂化、数据的多元化等特点,在真实的情报分析中,一般会涉及不同的数据映射以及多种处理算法的组合应用,这就要求情报分析计算框架和系统能够快速、敏捷地响应情报分析任务,通过人机交互或机器自动选择,实现分析过程的自动化。例如,Hauder 等人[69]提出了一种可以自动配置和定制分析流程的情报分析框架,该框架利用 Wings 工作流系统的独特功能来推理语义约束,能够在许多不同算法之间进行选择并尝试可能的组合算法,来帮助情报人员根据情报分析任务以及多元的数据特征,以可视化的交互方式对情报分析的流程进行定义,而不需要进行二次代码的编写。Gil 等人[70]提出了一种面向大规模分布式科学数据分析的自动生成计算工作流的语义框架。该框架能够自动选择应用程序组件和数据源,使用户能够指定分析流程中的组件细节和约束数量,包括分析流程的输入、中间或输出数据的约束、抽象工作流组件类以及可重用的分析流程模板。该框架最大的优点在于采用分布式架构,其中数据和组件目录与分析流程系统(工作流系统)相分离,从而保证了框架的可扩展性。

第四,随着人工智能技术的应用,计算型情报分析框架和系统正在朝着更加智能化的方向发展。特别是在数据建模更加常态化的今天,智能化的框架可以帮助情报分析人员根据数据集自动建立分析模型。传统的情报建模需要情报分析人员根据数据和情报任务的类型,借助自身的知识和对情报任务的理解进行人工建模,优点是可以构建出非常有效甚至是创新的模型,不足之处在于并非每一个分析人员都对所有可能的模型很熟悉,即使是熟悉的模型,也还需要花费大量的时间去调整参数和验证模型的可靠性。为解决这一问题,Richard Lippmann 等人[71]开发了一种数据驱动的模型发现框架,该框架能够借助强化学习技术根据输入的数据集进行自动数据清洗、自动搜索最优模型并进行参数的调整,从而帮助那些有领域专业知识但没有数据科学背景的情报分析人员快速构建真实、复杂过程的情报分析模型。这将使情报专家能够在不需要数据科学家的情况下创建数学模型,并通过自动化提高情报分析的效率。

3　结语

计算型情报分析作为一门新型的研究领域在近年来得到了快速的发展,尽管它的发展历史很短,还很年轻,但是越来越受到情报界的关注和重视。计算型情报分析是对传统情报分析的扩展,是未来情报学发展的重要方向之一。我们认为计算型情报分析将会逐渐发展成为一门融合计算科学、社会科学、认知科学、复杂科

学、生化科学、工程等多学科的新型学科，对于研究人员来说将是机遇与挑战并存，这要求我们用更加创新、开放、大胆的思想去迎接未来的挑战。

参考文献

[1] 贺德方.基于事实型数据的科技情报研究工作思考[J].情报学报，2009(5)：764-770.(He Defang. Reflections on scientific and technological information research based on factual data [J]. Journal of the China Society for Scientific and Technical Information, 2009(5): 764-770.)

[2] 贺德方.事实型数据：科技情报研究工作的基石[J].情报学报，2010(5)：771-776.(He Defang. Factual data: the cornerstone of scientific and technological information research [J]. Journal of the China Society for Scientific and Technical Information, 2010(5): 771-776.)

[3] 贺德方.工程化思维下的科技情报研究范式——情报工程学探析[J].情报学报，2014，33(12)：1-13.(He Defang. Research paradigm of scientific and technological information under engineering thinking: analysis of information engineering [J]. Journal of the China Society for Scientific and Technical Information, 2014, 33 (12): 1-13.)

[4] 李广建，江信昱.论计算型情报分析[J].中国图书馆学报，2018，44(2)：4-16.(Li Guangjian, Jiang Xinyu. On computational intelligence analysis [J]. Journal of Library Science in China, 2018, 44 (2): 4-16.)

[5] 苏新宁，朱晓峰.面向突发事件应急决策的快速响应情报体系构建[J].情报学报，2014，33(12)：53-77.(Su Xinning, Zhu Xiaofeng. Construction of quick response intelligence system for emergency decision making [J]. Journal of the China Society for Scientific and Technical Information, 2014, 33 (12): 53-77.)

[6] 吴晨生，张惠娜，刘如，等.追本溯源：情报3.0时代对情报定义的思考[J].情报学报，2017，36(1)：1-4.(Wu Chensheng, Zhang Huina, Liu Ru, et al. Tracing back to the source: reflections on the definition of intelligence in the age of intelligence 3.0 [J]. Journal of the China Society for Scientific and Technical Information, 2017, 36 (1): 1-4.)

[7] 李广建，江信昱.情报分析计算化：背景、作用及关键问题[J].图书情报工作，2017，61(16)：24-30.(Li Guangjian, Jiang Xinyu. Computationalization of information analysis: background, function and key issues [J]. Library and Information Service, 2017, 61 (16): 24-30.)

[8] 马费成. 情报学的进展与深化[J]. 情报学报，1996，15(5)：338-344.(Ma Feicheng. Progress and deepening of information science [J]. Journal of China Society for Scientific and Technical Information, 1996, 15 (5): 338-344.)

[9] Lotka A J. The frequency distribution of scientific productivity [J]. Journal of the Washington Academy of Sciences, 1926, 16(12): 317, 324.

[10] 庞景安.科学计量研究方法论[M].北京：科学技术文献出版社，1999.(Pang Jingan. Scientometrics research methodology [M]. Beijing: Science and Technology Literature Publishing House, 1999.)

[11] Vickery B C. Bradford's law of scattering[J]. Journal of Documentation, 1948, 4(3): 198-203.

[12] 顾明远.教育大辞典[M].上海：上海教育出版社，1998.(Gu Mingyuan. Dictionary of education[M]. Shanghai: Shanghai Education Press, 1998.)

[13] Garfield E. Citation indexes for science: a new dimension in documentation through association of ideas[J]. International Journal of Epidemiology, 2006, 35(5): 1123-1127.

[14] Brin S, Page L. The anatomy of a large-scale hypertextual Web search engine[J]. Computer Networks and ISDN Systems, 1998, 30(1-7): 107-117.

[15] 刘红煦，曲建升. 情报学定量研究现状与趋势分析[J]. 情报理论与实践，2015，38(12)：10-14.(Liu

Hongwei, Qu Jiansheng. Analysis of the status quo and trend of quantitative research in information science[J]. Information Studies: Theory & Application, 2015, 38(12):10-14.)

[16] Johnson N, Carran S, Botner J, et al. Pattern in escalations in insurgent and terrorist activity[J]. Science, 2011, 333(6038): 81-84.

[17] Leetaru K. Culturomics 2.0: forecasting large-scale human behavior using global news media tone in time and space[J]. First Monday, 2011, 16(9).

[18] Wing J M. Computational thinking[J]. Communications of the ACM, 2006, 22(9): 45-49.

[19] Wing J M. Computational thinking and thinking about computing[J]. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 2008, 366(1881): 3717-3725.

[20] Hu C L. Computational thinking: what it might mean and what we might do about it[C]// ACM.Germany, 2011.

[21] Schuler D. Social computing[J]. Communications of the ACM, 1994, 37(1): 28-29.

[22] Wang F Y, Carley K M, Zeng D, et al. Social computing: from social informatics to social intelligence[J]. IEEE Intelligent Systems, 2007, 22(2): 79-83.

[23] Cioffi-Revilla C. Introduction to computational social science: principles and applications[M]. Germany: Springer, 2014.

[24] 孟小峰, 李勇, 祝建华. 社会计算:大数据时代的机遇与挑战[J]. 计算机研究与发展, 2013, 50(12): 2483-2491.(Meng Xiaofeng, Li Yong, Zhu Jianhua. Social computing: opportunities and challenges in the age of big data[J]. Journal of Computer Research and Development, 2013, 50(12):2483-2491.)

[25] 徐敏,李广建.情报分析模型综述[J].情报理论与实践,2018,41(2):14-21.(Xu Min,Li Guangjian.A survey of intelligence analysis models[J]. Information Studies: Theory & Application, 2018,41(2):14-21.)

[26] Roderick M.Mathematical and computational modeling: with applications in natural and social sciences, engineering, and the arts[M]. Wiley,2015.

[27] 叶鹰. h 指数和类 h 指数的机理分析与实证研究导引[J]. 大学图书馆学报, 2007, 25(5):2-5.(Ye Ying. Mechanism analysis and empirical research guidance of h-index and h-like index[J]. Journal of Academic Libraries, 2007, 25(5): 2-5.)

[28] Kyebambe M N, Cheng G, Huang Y, et al. Forecasting emerging technologies: a supervised learning approach through patent analysis[J]. Technological Forecasting and Social Change, 2017, 125: 236-244.

[29] Lee C, Kwon O, Kim M, et al. Early identification of emerging technologies: a machine learning approach using multiple patent indicators[J]. Technological Forecasting and Social Change, 2018, 127: 291-303.

[30] Nsoesie E O, Beckman R J, Shashaani S, et al. A simulation optimization approach to epidemic forecasting[J]. PloS One, 2013, 8(6): e67164.

[31] Sinatra R, Wang D, Deville P, et al. Quantifying the evolution of individual scientific impact[J]. Science, 2016, 354(6312): aaf5239.

[32] Yin Y, Wang Y, Evans J A, et al. Quantifying dynamics of failure across science, startups, and security[J]. arXiv preprint arXiv:1903.07562, 2019.

[33] Uzzi B, Mukherjee S, Stringer M, et al. Atypical combinations and scientific impact[J]. Science, 2013, 342(6157): 468-472.

[34] Wang D, Song C, Barabási A L. Quantifying long-term scientific impact[J]. Science, 2013, 342(6154): 127-132.

[35] Tetlock P E, Mellers B A, Scoblic J P. Bringing probability judgments into policy debates via forecasting tournaments[J]. Science, 2017, 355(6324): 481-483.

[36] Barnes A. Making intelligence analysis more intelligent: using numeric probabilities[J]. Intelligence and National

Security, 2016, 31(3): 327-344.

[37] Cormen T H, Leiserson C E, Rivest R L. 算法导论[M]. 第2版.潘金贵,等,译.北京:机械工业出版社,2006. (Cormen T H, Leiserson C E, Rivest R L. Introduction to algorithms [M].2nd Edition. Pan Jingui, et al, trans. Beijing: Mechanical Industry Press, 2006.)

[38] Rogers H, Rogers H. Theory of recursive functions and effective computability[M]. New York: McGraw-Hill, 1967.

[39] Darpa.Darpa maven project[EB/OL]. [2019-06-01].http://maven.gr/tag/darpa/, 2019-05-19/2019-05-19.

[40] Darpa.Darpa AI-Next[EB/OL]. [2019-06-01].https://www.darpa.mil/news-events/2018-09-07, 2019-05-19/2019-05-19.

[41] Darpa. Darpa Big Mechanism[EB/OL]. [2019-06-01]. https://www.darpa.mil/program/big-mechanism, 2019-05-19/2019-05-19.

[42] Darpa. Darpa World Modelers[EB/OL]. [2019-06-01]. https://www.darpa.mil/program/world-modelers, 2019-05-19/2019-05-19.

[43] Kejriwal M, Gilley D, Szekely P, et al. THOR: text-enabled analytics for humanitarian operations[C]//Companion of the The Web Conference 2018 on The Web Conference 2018. International World Wide Web Conferences Steering Committee, 2018: 147-150.

[44] Johnson N, Carran S, Botner J, et al. Pattern in escalations in insurgent and terrorist activity[J]. Science, 2011, 333(6038): 81-84.

[45] Moat H S, Preis T, Olivola C Y, et al. Using big data to predict collective behavior in the real world 1[J]. Behavioral and Brain Sciences, 2014, 37(1): 92-93.

[46] Rekatsinas T, Ghosh S, Mekaru S R, et al. SourceSeer: forecasting rare disease outbreaks using multiple data sources[C]//Proceedings of the 2015 SIAM International Conference on Data Mining. Society for Industrial and Applied Mathematics, 2015: 379-387.

[47] Preis T, Kenett D Y, Stanley H E, et al. Quantifying the behavior of stock correlations under market stress[J]. Scientific Reports, 2012, 2: 752.

[48] Preis T, Moat H S, Bishop S R, et al. Quantifying the digital traces of Hurricane Sandy on Flickr[J]. Scientific Reports, 2013, 3: 3141.

[49] 李广建,化柏林.大数据分析与情报分析关系辨析[J].中国图书馆学报, 2014, 40(5):14-22.(Li Guangjian, Hua Bolin. Discrimination of big data analysis and intelligence analysis[J]. Journal of Library Science in China, 2014, 40(5): 14-22.)

[50] Katuwal G J, Chen R. Machine learning model interpretability for precision medicine[J]. arXiv preprint arXiv: 1610.09045, 2016.

[51] Che Z, Purushotham S, Khemani R, et al. Interpretable deep models for ICU outcome prediction[C]//AMIA Annual Symposium Proceedings. American Medical Informatics Association, 2016, 2016: 371.

[52] Tan S, Caruana R, Hooker G, et al. Detecting bias in black-box models using transparent model distillation[J]. arXiv preprint arXiv:1710.06169, 2017.

[53] Knight W. The US military wants its autonomous machines to explain themselves [EB/OL]. [2019-06-01]. https://www.technologyreview.com/s/603795/the-us-military-wants-its-autonomous-machines-to-explain-themselves,2019-05-19/2019-05-19.

[54] Equifaxlaunches neuroDecision © technology[EB/OL]. [2019-06-01].https://investor.equifax.com/news-and-events/news/2018/03-26-2018-143044126.

[55] Henelius A, Puolamäki K, Ukkonen A. Interpreting classifiers through attribute interactions in datasets[J]. arXiv preprint arXiv:1707.07576, 2017.

[56] Unfairness by algorithm: distilling the harms of automated decision-making, Jun. 2017[EB/OL].[2019-06-01].https://fpf.org/wp-content/uploads/2017/12/FPF-Automated-Decision-Making-Harms-and-Mitigation-Charts.pdf.

[57] Gunning D. Explainable artificial intelligence (xai)[J]. Defense Advanced Research Projects Agency (DARPA), nd Web, 2017.

[58] PARC. Explainable AI: an overview of PARC's COGLE project with DARPA[EB/OL]. [2019-06-01]. https://www.parc.com/blog/explainable-ai-an-overview-of-parcs-cogle-project-with-darpa/, 2019-05-19/2019-05-19.

[59] Kejriwal M, Peng J, Zhang H, et al. Structured event entity resolution in humanitarian domains[C]//International Semantic Web Conference. Springer, Cham, 2018: 233-249.

[60] Kejriwal M, Szekely P, Knoblock C. Investigative knowledge discovery for combating Illicit activities[J]. IEEE Intelligent Systems, 2018 (1): 53-63.

[61] Szekely P, Knoblock C A, Slepicka J, et al. Building and using a knowledge graph to combat human trafficking [C]//International Semantic Web Conference. Springer, Cham, 2015: 205-221.

[62] 陈晓红. 信息系统教程[M]. 北京: 清华大学出版社, 2003.(Chen Xiaohong. Information system tutorial [M]. Beijing: Tsinghua University Press, 2003.)

[63] XDATA. Defense advanced research projects agency[EB/OL]. [2019-06-01]. https://www.darpa.mil/program/xdata, 2019-05-10/ 2019-05-10.

[64] Decision support systems: putting theory into practice[M]. Englewood Cliffs, NJ.: Prentice-Hall, 1986.

[65] Watson H J. Revisiting Ralph Sprague's framework for developing decision support systems[J]. CAIS, 2018, 42: 13.

[66] Malewicz G, Austern M H, Bik A J C, et al. Pregel: a system for large-scale graph processing[C]//Proceedings of the 2010 ACM SIGMOD International Conference on Management of data. ACM, 2010: 135-146.

[67] Simmhan Y, Kumbhare A, Wickramaarachchi C, et al. Goffish: a sub-graph centric framework for large-scale graph analytics[C]//European Conference on Parallel Processing. Springer, Cham, 2014: 451-462.

[68] GeoTime-Pattern of Life in Time and Space.Uncharted software [EB/OL].[2019-06-01].https://uncharted.software/products/law-enforcement/.

[69] Hauder M, Gil Y, Liu Y. A framework for efficient data analytics through automatic configuration and customization of scientific workflows[C]//E-Science (e-Science), 2011 IEEE 7th International Conference on. IEEE, 2011: 379-386.

[70] Gil Y, Gonzalez-Calero P A, Kim J, et al. A semantic framework for automatic generation of computational workflows using distributed data and component catalogues[J]. Journal of Experimental & Theoretical Artificial Intelligence, 2011, 23(4): 389-467.

[71] Lippmann R, Campbell W, Campbell J. An overview of the darpa data driven discovery of models (D3M) program[C] //29th Conference on Neural Information Processing Systems (NIPS 2016). Barcelona, Spain, 2016.

李广建　北京大学信息管理系教授,博士生导师。北京 100871。

罗立群　北京大学信息管理系博士后,讲师。北京 100871。

(收稿日期:2019-06-11)