

基于对象行为的情报关注模型研究^{*}

邹益民 张智雄 刘建华

摘 要 情报关注模型的构建是实施网络科技信息结构化监测的前提,在网络科技信息情报价值判断、科技动态监测中发挥着重要的作用。本文以能源科技领域监测动态快报为基础,对情报人员关注的网络科技信息的特征进行分析,发现对象及其行为在情报价值判断中具有重要的作用,对象及其行为具有机构特殊性,领域重要机构的数量具有收敛性。在此基础上构建基于对象行为特征的情报关注模型,该模型不但关注对象本身,而且关注对象的行为以及行为背后的事件。对情报关注模型应用效果进行的评测和分析的结果表明,该模型可以有效提高网络科技信息情报价值判断的准确性和精细化程度。图6。表1。参考文献22。

关键词 情报价值 关注模型 知识对象 对象行为 情报监测系统

分类号 G250

Research on Intelligence Attention Model Based on Object Behavior

Zou Yimin, Zhang Zhixiong & Liu Jianhua

ABSTRACT Intelligence attention model is a premise of implementing the structured monitoring of network scientific information, and plays an important role in identifying intelligence value of network scientific information and intelligence monitoring system. Analyzing the characteristic of scientific information based on monitoring the express of the energy technology domain, this paper has found that the object and its behavior have an important role in intelligence value judgment and have institutional particularity, and the number of important institutions in the domain has convergence. Using these characteristics, an intelligence attention model based on the object behavior has been proposed, which not only focuses on the objects themselves, but also is more concerned about the object behavior as well as the event behind the behavior to improve the degree of accuracy and refinement of intelligence value judgment. The application effect of this model was evaluated. 6 figs. 1 tab. 22 refs.

KEY WORDS Intelligence value. Intelligence attention model. Knowledge object. Object behavior. Intelligence monitoring system.

1 引言

在网络日益成为最重要的科学交流和传播渠道的今天,很多重要的科技战略、科研活动、科研成果都是通过网络实现信息的对外发布。例如:美国奥巴马政府 2011 年国家创新战略^[1]和 2012 年的

大数据计划^[2]、欧盟的创新积分榜^[3]、世界经合组织的主要科学和技术指标^[4]等与科学研究和科技战略决策相关的信息都可以直接通过网络获取。战略情报研究团队可以通过这些资源对科技发展重大趋势及战略倾向进行把握,及时发现领域内的重大科技问题和重大研究进展。但在良莠不齐的海量网络信息中,究竟有多少是战略情报人员关注

^{*} 本文系国家自然科学基金项目“基于语言网络的文本主题中心度计算方法研究”(编号:61075047)的研究成果之一。

通讯作者:邹益民,Email:zouyimin@mail.las.ac.cn

的信息,这些信息相对于战略情报人员的需求而言,其价值大小又如何确定,需要一个自动的网络科技信息情报价值判断方法体系来对其进行度量。其中如何针对战略情报人员的实际情报需求和领域先验知识构建情报关注模型是一个需要解决的关键问题。基于此,本文在对网络科技信息特征和监测快报分析研究的基础上,结合中国科学院战略情报团队的实际情报需求,提出了基于对象行为的情报关注度模型,并对其在网络科技情报价值判断方法体系中的应用效果进行了分析。

2 相关研究

通过网络信息支持科技动态监测和战略决策分析,日益成为情报机构的一项重要工作。国外相关的情报机构提出了开源情报(Open-source Intelligence, OSINT)^[5]的理念,致力于利用公共可以获得的信息资源实现情报分析。美国中央情报局(CIA)专门成立了DNI Open Source Center(OSC)^[6],主要针对以网络信息资源为主的开源情报信息进行收集、开发和利用。在国内,王飞跃^[7]也指出面向开源信息和大数据的科技态势解析与服务将为知识产生方式和科技决策支撑带来重大的变革。此外,Web智能(Web Intelligence)^[8]、舆情监测^[9]、话题追踪和探测(Topic Detection and Tracking)^[10]等相关的研究也为网络科技信息监测服务的建设,提供了重要的参考。

情报关注模型的构建是实现网络科技信息自动监测的前提,是情报领域先验知识、情报需求(兴趣)的结构化描述,是机器可读且可用于计算的知识组织。情报关注模型的构建在网络科技信息情报价值判断、科技动态监测中发挥着重要的作用。情报关注模型是用户关注模型的一种,因此对现有不同用户关注模型构建方式的研究将为本文的研究奠定基础。根据构建模型采用的数据和方法不同,目前相关领域用户关注模型的构建方式大致可归纳为以下四类:

(1)基于关键词表的关注模型构建。在传统方式下关注模型的建模往往是利用用户提供的包

含一系列关键词的词表进行构建,关键词的选择取决于特定的应用需求,这种构建方式简单且易于操作,但过滤准确度不高。为了提高用户关注模型的效能,李文波等^[11]提出利用关键词组合信息来构建用户关注模型。但无论是基于关键词还是关键词组合的方式,都承受着多义词和同义词带来的问题。更重要的是,不是所有和关键词匹配度高的资源都是情报价值高的资源,例如:在科技政策与战略领域,不是所有包含“Obama”、“Department of Energy”或者它们组合的资源都是战略情报人员需要的资源。因此,如果用户对数据集敏感的话,基于关键词表的关注模型的构建并不能满足实际的需要。

(2)基于用户分类的关注模型构建。基于用户分类的关注模型构建首先将用户的兴趣分成若干个类别或者兴趣片段,而用户关注模型则可以由若干个类别或者兴趣片段构成,通过赋予不同的权重来表示关注模型对特定类别或兴趣片段的关注程度。例如:Liu等^[12]利用一组术语对一个类型进行描述,并用不同的权重来表示用户对该类型的关注程度;Loeb和Panagos^[13]把用户关注模型分解成若干个子模型,每个子模型代表一个特殊的背景(时间、地点)、情绪、任务、社会背景和一些其他影响用户兴趣和需求的参数。基于用户分类的关注模型构建的前提是对用户的关注进行预分类,这在很多场景下是不适用的,不具有通用性,也不能适应情报关注模型的构建。

(3)基于语篇特征抽取的关注模型构建。该方法是通过抽取用户标注语篇语料的特征,利用一组术语或模式对用户的关注进行描述。闵锦和黄莹菁^[14]从用户已标注的训练集中抽取特定领域的主题特征,采用向量空间模型(Vector Space Model, VSM)对关注主题的特征进行表示,以此来形成关注模型;Bing^[15]认为特征的选择应该融合词出现的频率、词的长度以及词在文本中的属性和语法特征,并结合通过统计学原理发现的训练集中的特征来建立关注模型;Algarni^[16]和Zhong等^[17]提出了基于模式的用户关注模型的构建方法,在训练集中获取用户关注的词语而非单个词的组合模式对用

户的关注进行建模。基于语篇特征抽取的关注模型的构建往往需要人工标注的训练集,不但非常耗费人力,有时候甚至是不现实的,更重要的是用户的关注会随着领域热点和关注点的变化而变化,这种关注模型的构建方法无法有效进行动态的调整。

(4) 基于本体的关注模型构建。该构建方式利用本体中各个概念间的关系对用户的先验知识和关注主题进行描述。例如:Daoud 等^[18] 根据同一用户一系列的检索词和开放式分类目录搜索系统本体 (Open Directory Project, ODP) 为每一个检索的会话 (Session) 创建一个加权图来描述用户的兴趣模型;Persona^[19] 也通过加权的 ODP 本体构建用户关注模型,对搜索引擎的结果进行过滤,并通过明确的相关反馈来更新用户关注模型;Vatani 等^[20] 利用 WordNet 对用户关注模型中的词进行处理,找出其同义词,在信息过滤中可以找出所有相关的资源。基于本体的关注模型将关键词扩展成本体中的概念以提高关注模型的性能,但它仍然未能解决基于关键词表的关注模型构建中的“假相关”(即包含特定概念或概念组合的资源并不一定是高价值的情报资源)问题。

3 网络科技信息的特征分析

通过上述分析可知,目前用户关注模型的构建方式并不能有效满足战略情报团队的实际需求。情报关注模型的构建需要对重要网络科技信息的特征进行深入分析,了解战略情报人员评价网络科技信息情报价值的流程和工作方式,还需要对战略情报人员的情报需求进行实际调研,将情报人员的评价标准和领域先验知识转化为机器可读的可计算的结构化模型。《动态监测快报》作为战略情报团队情报关注的集中体现,对其分析能够有效揭示战略情报人员关注的高价值网络科技信息的特征,为情报关注模型的构建提供有力的支持。本文以中国科学院先进能源科技战略团队为例,将其发布的2010年第1期到2012年第17期的《先进能源科技动态监测快报》作为研究对象,分析其内容特征和分布特征,并在此基础上实现情报关注模型的构建。

3.1 对象及其行为在情报价值判断中具有重要的作用

在对网络科技信息的研究中,笔者发现很多的科研对象和重要术语,例如:战略计划、科研人员、研究报告、战略声明、R&D 投入、重大项目和法案、研究主题等,一起内嵌在科学创新机构发布的网页、RSS 和富文档中。笔者将这些命名实体和术语称为知识对象或对象。通常这些知识对象携带了资源的关键信息。而对象行为指的是对象发出的一系列动作,例如:“Energy Department Announces Launch of Energy Innovation Hub”,这里对象行为为“Announce”,而行为的主体为对象“Energy Department”,而行为的客体为对象“Energy Innovation Hub”。

从实质上讲,网络科技信息的价值体现在其论述事件的情报受关注度上,而事件是由对象和对象的行为所构成的,事件的情报价值在很大程度上取决于对象行为的性质和结果。例如:“Barack Obama”是一个重要的对象,但并不是所有出现“Barack Obama”的网络科技信息都是高价值的资源,还要对语篇中 Obama 的行为和行为导致的事件进行识别和区分,如果报道的是 Obama 一次普通的出访活动,可能其战略情报价值就没有那么高;再者,如果是 Obama 发布的一则关于科技政策与战略领域的报告,那么能源科技领域对其关注的程度就会比较低。

战略情报人员对对象行为的关注在监测快报的统计结果中也得到了验证,比如:行为“Announce”、“Develop”、“Found”、“Toward”、“Invest”、“Create”、“Award”等在监测快报中就频繁出现。所以在构建情报关注模型时,不能只关注资源中所含对象的重要性,对象的行为特征以及行为背后的事件对资源情报价值的判断具有重要作用。以下是美国能源部发布的关于能源部长 Steven Chu 的两条新闻:

• Secretary Chu Announces \$14 Million for Six New Projects to Advance IGCC Technology

• Chu Visits Site of America's First New Nuclear Reactor in Three Decades

战略情报人员很容易区分第一条信息比第二条信息更具有情报价值,而如果通过关注对象或关注对象组合匹配的方式进行计算,可能的结果是第二条信息比第一条更重要。因为第二条信息中不但包含了重要的科研对象“Chu”以及“Nuclear Reactor”,同时,它还包含了两个具有情报价值指示性的形容词“First”和“New”。但如果考虑对象行为特征的话,“Announce”显然比“Visit”更具有情报价值。所以,在进行情报关注模型构建时,必须对对象的行为特征进行关注,提高情报价值判断的精准度。

3.2 对象及其行为具有机构特殊性

(1) 对象具有机构特殊性

对于一些对象,特别是主题概念类型的对象,可能是在整个目标领域都需要关注的,例如:在能源科技领域,包含“Solar Collector”、“Wind Generator”、“Nuclear Energy”、“Hydrogen Transportation”等与研究主题相关的资源,就容易受到战略情报人员的关注。另外,类似美国这种项目计划通常是由多个部门发起和组织的情况,其相关的科研主体往往也涉及多个机构。但在实际的环境中,更多的对象则具有机构的特殊性,这类对象可大致分为两类:一类是以机构的名称或者机构的领导人为主的科研主体,这类主体的行为往往代表了特定机构的最新动向,包含这些对象的资源在其所在机构发布则具有明显的权威性;另一类是由特定机构发起的科研活动,比如欧盟第七框架计划(Seventh Framework Programme, FP7)^[21]虽然在整个领域具有重要的影响力,但如果欧盟发布的网络科技信息中是有关 FP7 内容的,就显得尤为重要。此外,欧盟的创新积分榜^[3]、世界经合组织的主要科学和技术指标^[4]、国际空间站的航天任务^[22]等对象都具有鲜明的机构特殊性。

在构建情报关注模型时,如果不对这些具有鲜明机构特殊性的对象加以区分,则可能会导致那些虽然和这些对象有关(笔者称其为“假相关”),则无太多实质性内容的信息当作高情报价值的资源加以推荐,降低了战略情报人员利用信息的效率。

(2) 机构行为特征具有差异性

根据各个来源机构性质和职能的不同,可将战略情报人员关注的科研机构分为国家科技管理机构、科技咨询机构、国家研究委员会、新闻报道机构、科研资助机构及重要实验室等。对于不同类型的机构,其受情报人员关注的机构行为特征也不相同,这种差异性体现在情报人员对机构发布的不同类型网络科技信息的敏感度上。例如:情报人员更多地关注科研管理机构发布的战略规划、战略声明,关注科研资助机构发布的投资与预算等相关的信息,而对于重要实验室则更多地关注其发布的重大成果。

这种倾向性在对监测快报统计的结果中也得到了验证,笔者按照投资预算、战略声明、统计评价、重要成果、研究报告、战略规划、政策措施、项目与计划、机构调整 9 个类型对监测快报中的动态信息进行人工标注,并对各个机构受关注的资源类型进行了统计。美国能源部(Department of Energy, DOE)、劳伦斯伯克利国家实验室(Lawrence Berkeley National Laboratory, Berkeley Lab)、麻省理工学院能源计划(MIT Energy Initiative, MITEI)和国际能源署(International Energy Agency, IEA)在监测快报中的资源类型分布如图 1 所示。

从统计结果中不难看出情报人员更多地关注美国能源部的投资与预算,关注劳伦斯伯克利国家实验室和麻省理工学院能源计划的重要研究成果,而国际能源署的研究报告则受到更多的关注,这与它们机构实际的性质和职能是一致的。各个机构受战略情报人员关注的重点不同,而且差异较大,这就需要在情报关注模型的构建上充分考虑每个机构行为特征差异性,对其重要的行为给予特别的关注。

3.3 领域重要机构的数量具有收敛性

《先进能源科技动态监测快报》中动态信息来源 Top 20 的机构如图 2 所示,在和情报人员进行确认后发现这些机构和能源科技领域中重要的机构基本一致,这一结果也证明了以监测快报为分析对象能够较为准确和真实地对情报人员的关注进行反映。

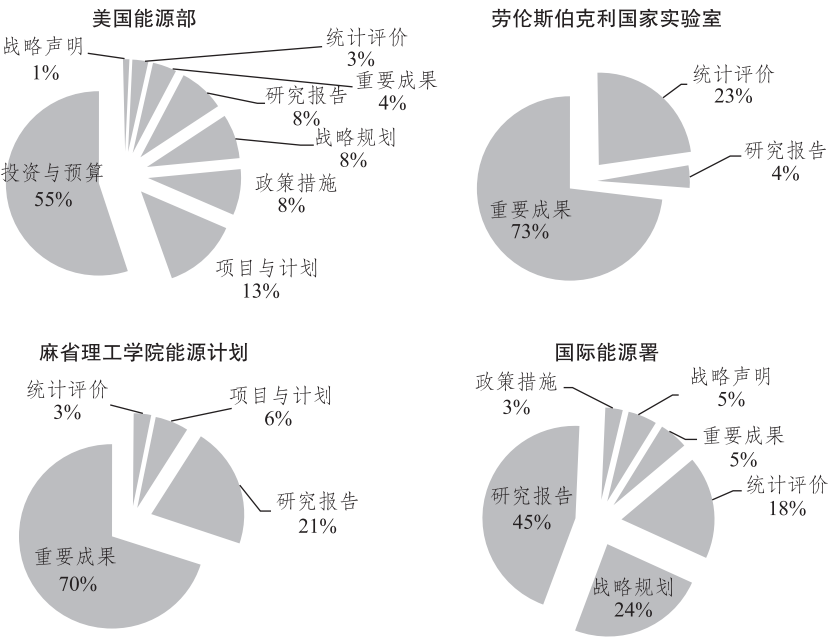


图1 各机构资源类型分布情况

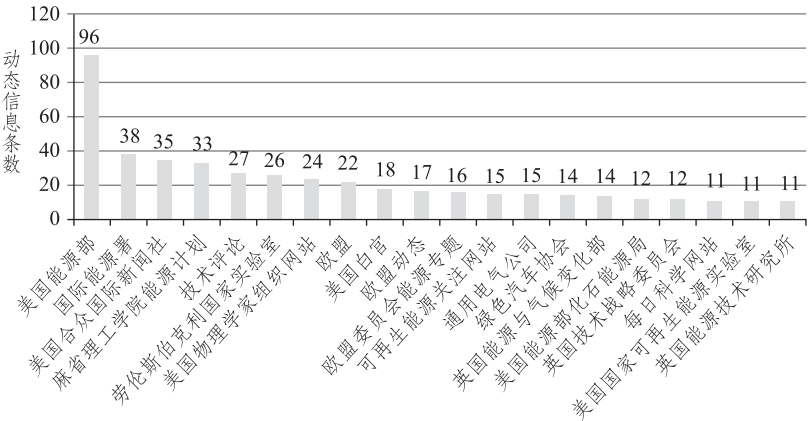


图2 《先进能源科技动态监测快报》动态信息来源 Top 20 机构

监测快报动态信息来源机构的分布情况如图3所示,从统计结果可以看出重要网络科技信息来源机构分布具有极大的不均衡性。48%的高价值网络科技信息来源于7%的机构,而在监测快报中仅出现一次的机构有164个,占来源机构的58%,来源于这些机构的动态信息只占17%。笔者对监测动态快报的跟踪也发现此类偶然出现的机构几

乎每期都会有,此类动态信息大多是情报人员间接获取的资源,所谓间接获取的资源是指情报人员通过阅读重要监测机构发布的资源,利用文中引用等线索信息获取到的资源。可见重要机构为情报人员提供了绝大多数的情报信息来源。另一方面,领域重要机构的数量具有收敛性,在能源科技领域的重点关注机构(出现次数大于10)共有21个,较关

注机构(出现次数介于4到9之间)的共有34个。网络科技信息情报价值判断是一个对信息非常敏感的领域,而且对象及其行为具有鲜明的机构特殊性,使得有必要针对具体机构实施精细化的分析,而重要机构数量上的收敛性为这种针对领域重要机构的精细化分析提供了可能。

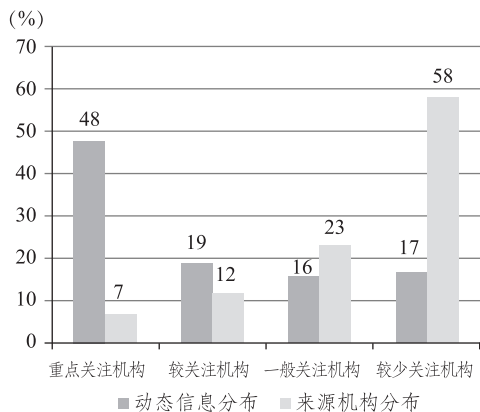


图3 情报动态信息和来源机构分布情况

4 构建基于对象行为的情报关注模型

对象及其行为能够有效揭示信息中所蕴含的事件,在网络科技信息情报价值判断中具有重要作用,并且对象及其行为具有机构的特殊性,使得有必要针对具体机构进行精细化分析,而领域重要机构在数量上的收敛性又为这种精细化分析提供了可能。所以在构建情报关注模型时,首先需要分析对象及其行为的特征,对对象行为模式进行结构化表示;然后,在对象行为模式构建的基础上,结合对象及其行为所具有的机构特殊性以及重要机构数量的收敛性,构建基于对象及其行为的情报关注模型,将战略情报人员的情报需求和领域先验知识转化为机器可读可计算的结构化模型。

4.1 对象行为模式

对象行为模式是对对象及其行为的一种结构化描述,根据行为的施动者和受动者不同又将对象行为模式分为主体对象行为模式和客体对象行为

模式。所谓主体对象行为模式是指以行为主体对象为中心构造的模式,中心对象为行为的施动者。而与其相对的是客体对象行为模式,和主体对象行为模式不同的是,其中心对象为行为的受动者,所有的行为都指向该对象。在主体(或客体)对象行为模式中允许对象行为或对象为空,此时对对象行为模式的关注即转化为对特定对象或者行为集的关注。

另外,在对2011年9月1日至2012年8月31日美国能源部新闻栏目发布的所有205条资源,同该段时间内被《先进能源科技动态监测快报》采纳的资源进行对比分析中,笔者发现对象DOE的“Announce”、“Award”、“Advance”、“Spur”、“Develop”等行为往往受到战略情报人员更多的关注;有些行为,例如:“Finalize”、“Cite”、“Statement”、“Visit”、“Tour”、“Talk”等则很少受到关注;而有些对象行为则比较中性,例如:“Build”、“Reach”、“Speed”、“Guide”等。因此,笔者又将对象的行为分为积极、中性和消极三种。积极行为对网络科技信息的情报价值具有很好的指示作用,消极行为也为情报价值的判断提供了一定的线索,将在一定程度上提高情报价值判断的准确度。对象DOE的主体对象行为模式如图4所示。

对象DOE的这些行为特征也是以投资和预算为主的科研管理机构的主要行为特征,而在情报人员关注的发布重大成果的研究机构中,如果其行为为“Develop”、“Found”、“Could”、“Toward”、“Create”、“Spark”、“Identify”、“Reveal”、“Boost”、“Make”、“Open”等时,则其将受到更多的情报关注。

此外,行为还具有强度特性。在现实世界中,行为的剧烈程度将直接影响其造成的结果,同样在情报关注模型中对象行为也有相应的强度。如果在事件的主体对象、行为或客体对象中包含了“Significantly”、“Revolutionary”、“Innovative”、“Breakthrough”、“Groundbreaking”等修饰词时,也将在很大程度上加强行为的强度,使得该网络科技信息更具情报关注价值。

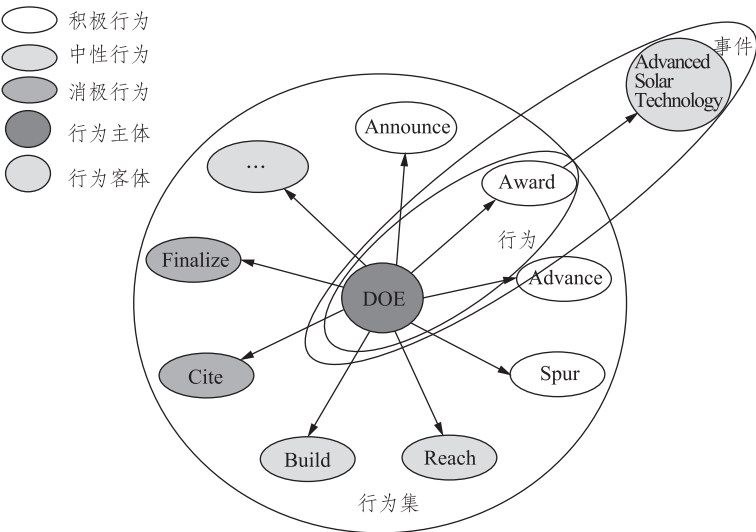


图 4 主体对象行为模式

4.2 情报关注模型

在对象行为模式构建的基础上,结合对象行为所具有的机构特殊性以及重要机构数量的收敛性,笔者提出了基于对象及其行为的情报关注模型(见图5)。该模型主要由三部分组成:行为强度词典

典、领域对象行为模式集和机构对象行为模式集。其中行为强度词典是建立在对大量语料统计和分析的基础上,并经过战略情报专家的确认,例如:“Groundbreaking”、“Revolutionary”和“Breakthrough”等行为强度词。领域对象行为模式集为全

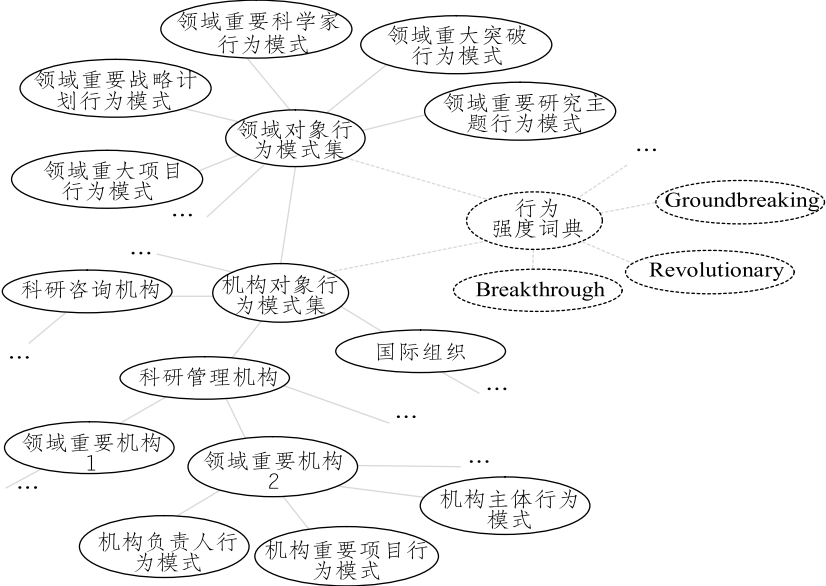


图 5 基于对象行为的情报关注模型

领域关注的对象行为模式,例如:在能源领域的科技和科技政策等领域,“Barack Obama”和“OSTP”等对象行为模式集就会受到较多的关注。为了突出机构的特殊性,在情报关注模型中设置了机构对象行为模式集,例如:在美国能源部发布的网络资源中包含“DOE”、“Steven Chu”和“Daniel Pone-man”等对象的行为模式时则更具情报价值。领域对象行为模式集与机构对象行为模式集和具体的领域以及情报人员关注的方向密切相关。

情报关注模型为战略情报人员发现高价值情报资源提供了一个结构化框架,在具体模型的构建上,情报人员可利用自身的领域先验知识,对其关注的对象及其行为进行建模,并可根据领域热点的变化调整对象及其行为的关注度。此法与利用情报人员提供的训练集构建关注模型的方法相比更具灵活性,而且准确度和精细化程度更高。基于对象行为的情报关注模型,不但关注对象本身,而且更多地关注对象的行为以及行为背后的事件,并将情报关注的目标定位到具体的机构,摆脱了单一利用情报人员提供的对象进行情报价值判断带来的精度较低的问题。

5 应用和效果分析

基于对象行为的情报关注模型对战略情报人

员的情报关注和领域先验知识进行了结构化描述,而网络科技信息情报价值的计算,还需要对网络资源中的重要情报线索进行识别。基于此,笔者提出了基于对象网络的重要对象及其行为识别算法,该算法将非结构化的网络信息资源,转化为可计算的知识单元,例如:句子“White House Announces National HIV/AIDS Strategy”将被转化为 < White House, Announces, National HIV/AIDS Strategy > 三元组的形式,在此基础上构建对象网络,利用对象在网络中的分布规律实现网络资源中重要对象及其行为的识别(限于篇幅,此不赘述),并将其与情报关注模型中的行为模式进行匹配。匹配的结果同网络科技信息来源的权威性、载体类型(PDF、DOC等)、正文在页面所占的比例、正文长度等外部指标一起用于判断网络科技信息的情报价值。

为了使最终的情报价值计算结果更加直观,笔者将归一化后的情报价值划分为一星到五星五个区间,并用星级的大小来表示情报价值的大小。目前,网络科技信息情报价值自动判断作为网络科技自动监测系统的重要功能,已在中国科学院“科技战略与政策”、“空天科技”、“资源与环境”、“能源科技”和“信息科技”五个战略情报团队得到应用,科技战略与政策领域中美国白宫科学与技术政策办公室的监测结果如图6所示。

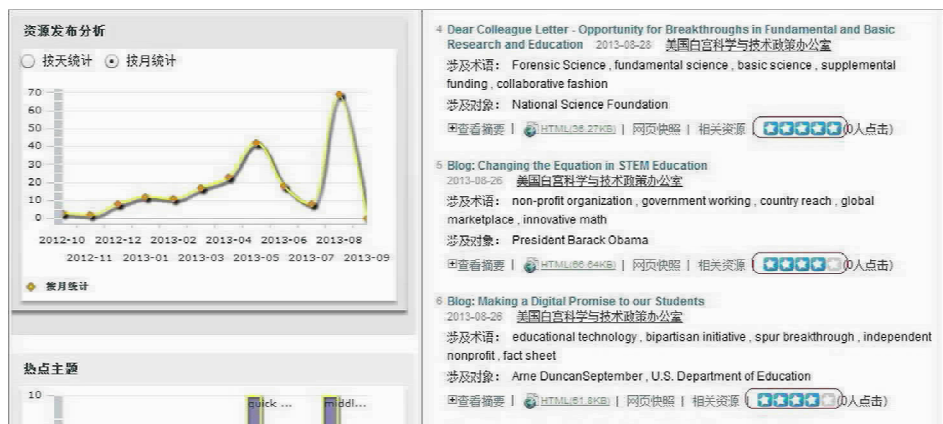


图6 美国白宫科学与技术政策办公室的监测界面

动态监测快报中出现的资源是经过情报人员确认的高价值情报资源,笔者将其作为基准数据,对网络科技信息情报价值判断方法的效果进行评价。以2012年9月份发布的《资源环境科学动态监测快报》、《地球科学动态监测快报》、《先进能源科技动态监测快报》、《信息科技动态监测快报》为例,情报价值评价方法对这些快报中所采纳资源评

价的星级分布情况如表1所示。如果把算法的评价结果中的四星和五星认为可接受,三星认为尚可接受,而二星和一星认为是不能接受。那么算法评价结果的可接受度为59.42%,尚可接受度为86.96%,不可接受度为13.04%,评价结果为二星级以上的资源到达了100%。

表1 网络科技信息情报价值判断结果

价值评级 监测快报类型	五星(%)	四星(%)	三星(%)	二星(%)	一星(%)
资源环境科学动态监测快报	25.00	41.67	25.00	8.33	0.00
地球科学动态监测快报	10.00	45.00	30.00	15.00	0.00
先进能源科技动态监测快报	14.81	44.44	29.63	11.11	0.00
信息科技动态监测快报	20.00	40.00	20.00	20.00	0.00

注:网络科技自动监测系统,目前只对英文的资源进行处理,非英语和访问受限等资源并未包含在监测系统当中,表中所示为监测系统中包含的监测快报资源的评价结果。

6 总结

在良莠不齐的海量网络资源中发现高情报价值的信息,需要一个自动的网络科技信息情报价值判断方法体系,其中针对战略情报人员的实际需求和领域先验知识构建情报关注模型是一个需要解决的关键问题。本文以《先进能源科技领域监测快报》为基础对情报人员关注的网络科技信息的内容特征和分布特征进行分析,发现对象及其行为在情

报价值判断中具有重要的作用、对象及其行为具有机构特殊性以及领域重要机构的数量具有收敛性等特征。基于此提出了基于对象行为特征的情报关注模型,该模型不但关注对象本身,而且更多地关注对象的行为以及行为背后的事件,提高了网络科技信息情报价值判断的准确度和精细化程度。从应用效果来看,能够有效帮助情报人员发现高价值的网络资源。在下一步工作中,将对不同性质的行为特征在情报价值判断中的作用进行更加深入的分析,进一步提高情报价值判断的准确度。

参考文献

[1] A strategy for American innovation; Securing our economic growth and prosperity[EB/OL]. [2013-01-09]. <http://www.whitehouse.gov/sites/default/files/uploads/InnovationStrategy.pdf>.

[2] Obama administration unveils “big data” initiative; Announces \$200 million in new R&D investments[EB/OL]. [2013-01-09]. http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release_final_2.pdf.

[3] Innovation union scoreboard[EB/OL]. [2013-01-09]. http://ec.europa.eu/enterprise/policies/innovation/facts-figures-analysis/innovation-scoreboard/index_en.htm.

[4] Main science and technology indicators (MSTI): 2012/1 edition[EB/OL]. [2013-01-09]. <http://www.oecd.org/science/innovationinsciencetechnologyandindustry/mainscienceandtechnologyindicatorsmsti2012edition.htm>.

- [5] Stalder F, Hirsh J. Open source intelligence[J]. First Monday, 2002(7)[2013-01-09]. <http://firstmonday.org/ojs/index.php/fm/article/view/961/882>.
- [6] Intelligence: Open source intelligence [EB/OL]. [2013-01-09]. <https://www.cia.gov/news-information/featured-story-archive/2010-featured-story-archive/open-source-intelligence.html>.
- [7] 王飞跃. 知识产生方式和科技决策支撑的重大变革——面向大数据和开源信息的科技态势解析与决策服务[J]. 中国科学院院刊, 2012(27): 527-537. (Wang Feiyue. Decision service and academic analytics for development of S&T based on open source intelligence and big data[J]. Bulletin of the Chinese Academy of Sciences, 2012(27): 527-537.)
- [8] Zhong N. Toward web intelligence[M]//Advances in web intelligence. Heidelberg, Germany: Springer Verlag, 2003: 5.
- [9] 谈国新, 万一. 突发公共事件网络舆情监测指标体系研究[J]. 华中师范大学学报: 人文社会科学版, 2010(3): 66-70. (Tan Guoxin, Wan yi. Research on emergencies public network opinion monitoring indicator system[J]. Journal of Huazhong Normal University (Humanities and Social Sciences), 2010(3): 66-70.)
- [10] Allan J. Introduction to topic detection and tracking[M]//Topic detection and tracking. Heidelberg, Germany: Springer Verlag, 2002: 12.
- [11] 李文波, 孙乐, 诺明花, 等. 基于核方法的敏感信息过滤的研究[J]. 通信学报, 2008(29): 57-62. (Li Wenbo, Sun Le, Nuo Minghua, et al. Sensitive information filtering based on kernel method[J]. Journal on Communications, 2008(29): 57-62.)
- [12] Liu F, Yu C, Meng W. Personalized web search for improving retrieval effectiveness[J]. IEEE Transactions on Knowledge and Data Engineering, 2004(16): 28-40.
- [13] Loeb S, Panagos E. Information filtering and personalization: Context, serendipity and group profile effects[C]// Consumer Communications and Networking Conference (CCNC), Las Vegas, USA, 2011: 393-398.
- [14] 闵锦, 黄萱菁. 基于主题和态度分类的文本过滤系统[J]. 计算机工程, 2007(33): 163-164. (Min Jin, Huang Xuanjing. Text filtering system based on topic and sentiment classification[J]. Computer Engineering, 2007(33): 163-164.)
- [15] Bing R. Information filtering algorithm based on feature vector [C]// International Conference on Intelligence Science and Information Engineering (ISIE), Wuhan, China, 2011: 468-471.
- [16] Algarni A. Relevance feature discovery for text analysis[D]. Brisbane: Queensland University of Technology, 2011.
- [17] Zhong N, Li Y, Wu S T. Effective pattern discovery for text mining[J]. IEEE Transactions on Knowledge and Data Engineering, 2012(24): 30-44.
- [18] Daoud M, Lechani L T, Boughanem M. Towards a graph-based user profile modeling for a session-based personalized search[J]. Knowledge and Information Systems, 2009(21): 365-398.
- [19] Tanudjaja F, Mui L. Persona: A contextualized and personalized web search[C]//Annual Hawaii International Conference on System Sciences (HICSS-35), Hawaii, USA, 2002: 1232-1240.
- [20] Vatani N, Shiri M E. A personalized information filtering method[J]. International Journal of Computer Science and Security (IJCSS), 2012(6): 1-8.
- [21] Seventh framework programme[EB/OL]. [2013-01-09]. http://cordis.europa.eu/fp7/home_en.html.
- [22] NASA missions[EB/OL]. [2013-01-09]. <http://www.nasa.gov/missions/index.html>.

邹益民 浙江师范大学经济与管理学院。通讯地址: 浙江省金华市迎宾大道 688 号。邮编: 321004。

张智雄 中国科学院国家科学图书馆研究馆员。通讯地址: 北京市北四环西路 33 号。邮编: 100190。

刘建华 中国科学院国家科学图书馆馆员, 博士研究生。通讯地址同上。

(收稿日期: 2013-01-27; 修回日期: 2013-04-09)