

·应急管理·

大数据时代的社交网络舆情主题图谱研究



□冯子桓 梁 循 牛思敏
[中国人民大学 北京 100872]

【摘要】 【目的/意义】 舆情主题图谱可以实现舆情事件节点之间相互关系的形式化描述, 构建舆情主题图谱对于有效分析和解决大数据时代海量、异构、复杂的舆情事件有重大意义。【设计/方法】 对大数据时代的社交网络舆情主题图谱的研究现状、主题图谱构建、舆情调控等进行较为详尽的介绍, 以社交网络分析相关技术方法为手段, 挖掘出主题图谱中更深层次的知识关联与更广范围的知识范畴, 从关键节点、关键路径、趋势模型等不同维度进行舆情主题图谱分析。【结论/发现】 对社交网络舆情主题图谱的有效管理和应用价值进行探究, 分析社交网络舆情主题图谱在经济、政治、生活等方面的实践应用的科学性、有效性和适用性等。

【关键词】 社交网络舆情; 主题图谱; 图谱构建; 舆情调控

[中图分类号] G206; C912.63

[文献标识码] A

[DOI] 10.14071/j.1008-8105(2021)-1008

Research on Social Network Public Opinion Theme Map in the Era of Big Data

FENG Zi-huan LIANG Xun NIU Si-min
(Renmin University of China Beijing 100872 China)

Abstract [Purpose/Significance] Public opinion theme map can realize the formal description of the interrelationship among public opinion events nodes, and the construction of public opinion theme map is of great significance to effectively analyze and solve massive, heterogeneous and complex public opinion events in the era of big data. [Design/Methodology] The current research status, the theme map construction, and opinion regulation of social network the theme map in the era of big data are introduced in detail, and on the basis of the construction of the theme map of public opinion, the technical methods related to social network analysis are used as a means to dig out deeper knowledge associations and broader knowledge categories in the topic mapping, and to analyze the theme map from different dimensions such as key nodes, key paths, and trend models. [Conclusions/Findings] Finally, it explores the effective management and application value of social network public opinion theme map, and analyzes the scientificity, effectiveness and applicability of the practical application of social network public opinion theme map in economy, politics and life.

Key words social network public opinion; theme map; map construction; public opinion control

[收稿日期] 2021-10-19

[基金项目] 国家自然科学基金项目(62072463); 国家社会科学基金项目(18ZDA309).

[作者简介] 冯子桓(1998-)男, 中国人民大学信息学院硕士研究生; 牛思敏(1999-)女, 中国人民大学信息学院博士研究生.

[通信作者] 梁循(1965-)男, 中国人民大学信息学院教授、博士生导师. Email: xliang@ruc.edu.cn.

引言

随着互联网的快速发展和数字时代的到来,人们传统地获取、分享、交流信息的方式正在发生根本性的改变,人们可以很方便地获取与共享网络上的资源,自由地通过网络进行交流与沟通。人们通过互联网的各种手段表达对社会事件的认识、态度和情感,这些手段包括微博、微信、短视频、新闻评论等形式。由于新技术、新应用不断出现,其传播途径和传播方式更加隐蔽和复杂,能否及时掌握网络舆论的最新动向,对于国家和社会稳定具有重要意义。

舆情主题图谱是一种以大规模采集的社交网络舆情数据为基础并对其进行语义抽取和标注,而构建的覆盖大规模舆情事件和实体的语义关联覆盖网络。它可以实现对舆情主题图谱数据的采集和主题图谱的构建,为社交网络舆情主题图谱的挖掘提供依据和支撑,对现实舆情事件相关的一切事物及其相互关系进行形式化的描述。

从舆情数据、舆情信息到舆情主题图谱的流程中,社交舆情大数据作为互联网大数据的一部分,具有其特殊性。社交网络舆情资源具有海量、分布、异构、多粒度等多种特性,为对其进行有效的数据挖掘,就需要首先对这些海量、分布、异构、多粒度的舆情资源进行知识抽取,为舆情主题图谱的构建提供数据基础。根据知识组织的相关性原理和有序性原理,社交网络舆情大数据的组织可以具体化为:社交网络舆情大数据主题图谱的构建与社交网络舆情大数据的分析两个主要方面。因此,如何构建社交网络舆情大数据主题图谱并对其进行有效分析便成为该领域需要解决的一个关键性问题。

本文以舆情和知识图谱为关键研究领域,针对舆情大数据多源异构性、多维关联性、多重可用性的问题,挖掘舆情信息、构建舆情主题知识图谱,并利用舆情主题图谱的应用价值,自下而上地从理论和实践多方面介绍大数据环境下社交网络舆情主题图谱的构建和分析调控。

一、大数据时代社交网络舆情主题图谱的国内外研究现状

(一) 舆情、社交网络舆情到知识图谱、社交网络舆情主题图谱

1. 社交网络舆情

舆情是指在一定的时间空间范围内,社会民众

对社会事件的产生与发展变化所持有的一种态度,是民众对社会中出现的各种现象和问题集中表达出来的意见和态度的总和^[1-3]。社交网络舆情是指公众在社交平台上对某种社会公共问题、现象或具体事件公开表达的具有倾向性和一定影响力的言论、态度的集合。

对社交网络舆情的有效掌控,是在新形势下应对社会新变化新发展的必然要求。但是,由于网络存在一定程度的虚拟性且社交网络的覆盖范围越来越大,舆情传播的途径和形式也越来越复杂,舆情管理面临着不小的压力和挑战^[4]。例如,在社交网络上,“造谣一张嘴,辟谣跑断腿”一直是辟谣者所面临的现实困境。要想和虚假信息做斗争,就必须对在线社会网络谣言进行实时检测,针对不同类别和应用场景的网络谣言,还需要从目标、对象和时间三个属性入手,同时对谣言源进行检测^[5],可见数据量之大,检测难度之高。

2. 社交网络舆情主题图谱

知识图谱是一个具有属性的实体通过关系链接而成的网状知识库,以符号形式来描述现实世界中的实体及其相互关系^[6]。其构建有两种方式:自顶向下构建和自底向上构建。其中最典型就是Google的Knowledge Vault^[7]。

构建知识图谱的第一步是知识抽取,主要分为以下几种方法。

(1) 实体和主题抽取。实体抽取的方法分为三种^[8]:基于规则和词典的方法、基于统计机器学习的方法、混合法。面向开放域的实体抽取则是从大量的互联网语料中采取一定的技术手段进行实体抽取^[9-11]。主题抽取也称话题检测,话题检测的主要任务是检测新话题,是一种无监督聚类方法,如果爬取到的话题与之前产生的某个话题类别相关,则将其融合;否则,产生一个新的话题类别^[12]。

(2) 概念和属性抽取。属性抽取的目标是从不同信息源中采集特定实体的属性信息。对于非结构化的公开数据,可以利用结构化数据作为训练集训练模型,再将训练出来的模型应用于非结构化的实体属性抽取^[13];还可以采用数据挖掘的方法从大量的数据中提取实体属性与属性值之间的关系模式^[14]。

(3) 情感关系抽取。对于情感关系抽取,学者们先后构建了基于规则的Twitter情感分析系统^[15]、基于层次结构多策略的情感分析框架^[16]、情感极性判别分析^[17]等。

实体关系抽取是图谱构建的重要环节。关系抽取的主要方法有:基于模式匹配的方法^[18-19]、基于

语义词典的方法^[20-22]、基于有监督的方法^[23-24]、基于种子或自举 (bootstrapping) 的方法、基于远程监督 (distant supervision) 的方法^[25-27]等。最后通过关联关系将离散的命名实体联系起来,形成网状的知识结构。

知识图谱作为人类对理解认识的一种结构化模拟,不仅可以很好地帮助计算机进行语义的理解,应用于知识库问答、文本生成、情感分析等多个方面,而且针对某一特定领域建立相关的知识图谱也极大地提高了用户的体验。构建具有领域特色的知识图谱,可以很好地应用于复杂分析和决策的场景,该技术在自然语言处理、推荐系统、预测分析、识别、分类等各种信息网络中具有深厚而有影响力的应用^[28]。在此,我们着重介绍针对社交网络舆情领域知识图谱的构建与分析。

(二) 网络舆情主题图谱国内外研究现状

1. 网络舆情分析方面

在Web of Science中,检索关键词“public opinion OR public sentiment”,选择范围为网络(web OR internet OR www OR network),时间跨度为1950年至今(2021-09-04),共检索到9965篇期刊论文和会议论文。

从学科分布来看,网络舆情问题吸引了计算机科学、行为科学、社会学、心理学、工程、新闻与媒体等领域专家学者的密切关注。

从研究者的分布情况来看,如图1所示,相关研究者主要集中在中国、美国、英国、西班牙、加拿大、德国等国家。

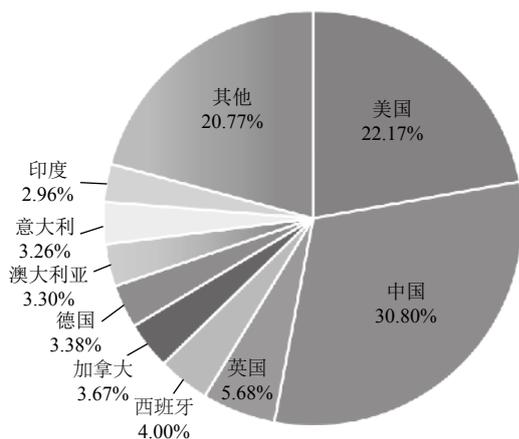


图1 关于网络舆情研究的国家分布状况

从时间分布来看,这几年网络舆情主题越来越受到国际学者的关注,如图2所示,尽管2021年发文量暂时还未统计完,2011~2020年一直呈上升的态势。

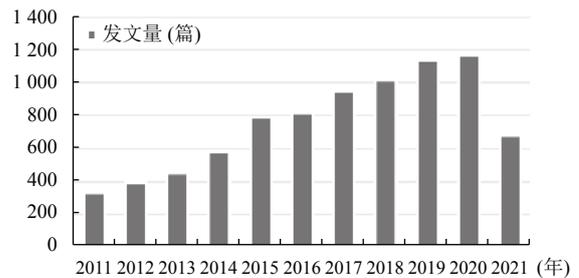


图2 关于网络舆情研究的时间分布状况

国内学者如兰月新、陈福集、黄微、李国祥等对网络舆情进行了积极探索。研究主题主要集中在网络舆情、网络舆论、网络舆情应对的研究。新媒体与移动网络的流行使舆情在豆瓣、微博、微信上传播更加迅速,舆情往往会在多个社交平台上联动发酵。高校大学生作为活跃的网民群体引起了学者对高校网络舆情的关注。“新媒体”“情感分析”“电子政务”等热词开始进入网络舆情的研究领域,推动该主题的研究往纵深方向发展。

2. 知识图谱方面

在Web of Science中,检索主题为“Knowledge Graph”,时间跨度为1950年至今(2021-09-04),共检索到24251篇期刊论文。

从学科分布来看,知识图谱吸引了计算机科学、数学、工程、商业、经济学、医学等领域专家学者的密切关注。从研究者的分布情况来看,如图3所示,相关研究者主要集中在中国、美国、德国、法国、西班牙、英国、加拿大等国家。

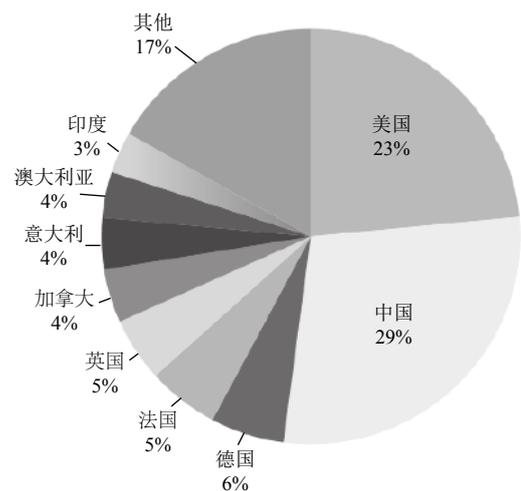


图3 关于知识图谱研究的国家分布状况

从时间分布来看,这几年知识图谱主题越来越受到国际学者的关注,如图4所示,2011~2019年一直呈上升的态势,尽管2020年发文较少,但期刊发文数仍超过2500篇。

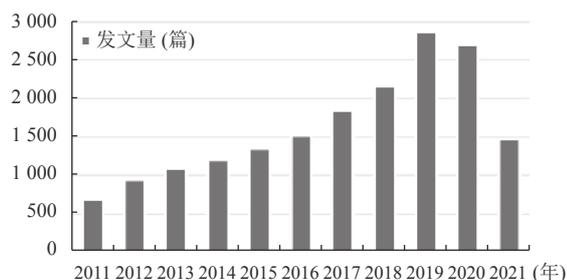


图 4 关于知识图谱研究的时间分布状况

从中国知网中的知识图谱主题发文章量来看,从2014年开始,该主题成为国内研究热点领域之一,发文章持续攀升。目前,知识图谱除了在图书情报学领域得到广泛和深入应用外,正快速地向其他学科或领域(如计算机软件、教育理论、自动化管理、企业经济、体育等)扩散,并且取得了非常丰富的科研成果,且研究成果呈现递增趋势。总体而言,知识图谱当前应用主要集中在公开数据集的研究以及知识库的构建,同时对各学科的研究热点与前沿、研究主题及研究基础等进行可视化分析。而利用知识图谱原理和技术对网络舆情信息进行可视化关联分析的研究比较少。

3. 基于社交网络分析相关技术舆情主题图谱挖掘方面

从学术研究方面来看,现有的人工智能技术大多以数据驱动为主,目前仍停留在计算智能和感知智能。通过一系列复杂算法和大规模的数据计算去得到规律进行学习与决策,很容易造成数据依赖,导致模型的效果不佳^[29],并且这种方式的可解释性差,不可控性凸显。同时,越来越多的学者意识到人类对于问题的解决思路,并不像现有的基于统计的机器学习算法、深度学习技术那样需要经过大规模的数据运算,而是简单的通过概念之间的关系进行推理就可以得到答案。这使得研究者更加清晰地认识到人工智能不应该局限于计算智能与感知智能,要想实现智能、鲁棒、可解释、可推理多方完善的人工智能系统,将现有的技术与基于知识规则的认知智能相结合才是更好的选择。例如,在进行表征学习时,混合知识图谱嵌入和卷积运算或神经网络技术的优势,可以用于研究和提取社交网络结构中有意义的特征,以帮助节点分类,社区检测和链接预测问题^[30],并且使用这种融合技术进行嵌入相较于传统的节点嵌入性能更好,但计算成本相对而言更高^[31];对于企业而言,结合知识图谱与文本挖掘技术,将其业务数据和互联网舆情数据进行对比、关联分析,可发掘研判与应对规律,利于舆论

引导、舆情应对的高效开展^[32]。

4. 国内外知名企业及学者相关研究方面

我们熟知的Facebook、腾讯、微博等,无论是企业本身在知识图谱的构建与应用,还是众多学者在此领域进行的相关研究,都在进行不断地尝试与突破。

早在2013年,Facebook就推出图谱搜索(Graph Search),Graph Search不同于Google的基于关键字匹配的搜索方式,Graph Search更像语义识别搜索,将与个人直接相关的事物以搜索结果的形式呈现,而非像Google搜索结果那样仅展示相关网站链接。但Graph Search还处于比较早期的开发阶段,还有许多问题亟待解决。

腾讯作为国内最大的关系社区,在构建多领域特色知识图谱方面则显得更有建树。腾讯知识图谱(Tencent Knowledge Graph, TKG)是一个集成图数据库、图计算引擎和图可视化分析于一体的一站式平台。TKG拥有丰富的应用场景,涉及金融、物联网、安全等方方面面。例如在泛安全领域,我们可以处理结构化、半结构化、非结构化数据识别数据中隐藏模式和关系^[33-34],将从大数据中深度挖掘关联关系,转化为关系图谱数据,进行司法风险与舆情风险发现^[35],有力支撑公安机关展开情报研判分析、犯罪团伙跟踪以及重大事情预警等,从而更好地提升警民良性互动与政府公信力^[36]。

新浪微博相继推出图谱搜索体验服务。与此同时,为解决微博搜索中存在的一些问题,使得结果中呈现更多知识,不少研究人员也进行了多样化的探索。例如,一些学者研究微博社区知识图谱构建方法并针对其构建效果、演化特征、应用效果进行了多元分析^[37];通过构建微博舆情主题知识图谱进行主题挖掘,不仅可以有效识别出主题,还可以准确定位每个主题下的意见领袖^[38-39];在文献^[40]中,作者也进一步构建微博主题识别与演变知识图谱,进而发现微博主题演变路径,揭示了微博舆情传播中的主题演变情况以及发生演变的动态变化过程,可以更好地对微博舆情传播进行实时监控;通过构建面向社交网络用户意见的知识图谱,用于理解不同的用户意见,并以知识图谱的形式识别复杂的关系,使得社会科学研究人员可以更轻松地获取各种视角的意见,并对数据进行进一步分析^[41]。

二、大数据驱动的社交网络舆情主题图谱研究定位

舆情主题图谱的概念来源于知识图谱,针对社

交网络舆情信息的大数据化、动态化、多维度等特征,从社交网络舆情大数据获取的新方法和新工具研究出发,对社会安全、突发事件预警等某一特定领域、特定需求进行实践,就是本文探索构建大数据驱动的社交网络舆情主题图谱的出发点所在。

舆情主题图谱是舆情信息工作的有效方法体系,探索构建大数据驱动的社交网络舆情主题图谱的构建理论方法和技术,以此为基础研究基于社交网络舆情主题图谱的舆情分析和推理方法,进而从舆情点、舆情链路、舆情主题、全局层面等多个维度讨论社交网络舆情主题图谱的调控策略,实现社交网络舆情的动态收集与监控工作,编制行业舆情调研报告,为各级政府部门提供决策参考。

三、大数据驱动的社交网络舆情主题图谱知识表示

社会网络中舆情传播来源多、传播速度快、传播范围广,使得网络中舆情的监控和管理更加困难。大量积累的、复杂的数据也给查询、处理等带来了挑战,并且随着时间的推移、数据的更新,图中蕴含的节点和边的数量可能越来越庞大。任何一种复杂网络图系统都具有一定的特殊性,这给研究增加了难度。大规模图结构数据中的节点数量已经达到百万甚至数亿的级别,并且呈现多样化、异质化和动态性等特征^[42]。这些特征丰富了表示学习的内容。

知识表示学习的代表模型主要包括距离模型、双线性模型、神经张量神经网络模型、矩阵分解模型、翻译模型等^[43-47]。知识表示的方法主要是将实体与关系嵌入到低维向量空间中,如TransE模型。研究人员在TransE模型的基础上进行拓展或改进,从不同角度尝试解决复杂关系建模问题^[48-50],采用多种数学方法来表示实体及实体之间的关系,如使用高斯分布的均值表示实体或关系在语义空间中的中心位置,使用协方差表示实体或关系的不确定度^[51]。

四、大数据驱动的社交网络舆情主题图谱更新推理

(一) 图谱的知识更新

网络舆情中存在的信息和知识量随时间的变化呈指数式增长,因此舆情主题知识图谱的内容也需要不断更新,与时俱进,其构建过程是一个不断迭代更新的过程。

从逻辑上看,知识库的更新主要包括模式层的更新与数据层的更新。模式层的更新包括但不限于增加、删除或修改本体中元素的概念,同时还需更新元素的概念属性以及概念之间上下级关系等。其中,概念属性的更新操作将直接影响到所有与实体有直接或间接联系的概念和实体^[52]。模式层的增量更新方式多数情况下是在人工干预的情况下完成的。数据层的更新指的是实体元素的更新,比如修改实体的基本信息和属性值。由于数据层的更新一般影响面较小,因此通常以自动的方式完成。

(二) 图谱的知识推理

知识推理是知识图谱构建与发展的重要手段和关键环节。知识推理是指从知识库中已有的多对实体关系数据出发,经过计算机复杂推理,从旧知识中发现新的知识,发现并建立已有实体间的新联系,形成新的三元组,从而拓展和丰富知识网络。通过知识推理,不断扩大知识库的覆盖范围。

知识库推理可以分为基于符号逻辑的推理和基于图的推理。基于符号的推理一般是基于经典逻辑或者经典逻辑的变体。基于符号的推理可以利用一定的规则从一个已有的知识图谱中推理出新的实体间关系,还可以对知识图谱进行逻辑的冲突检测。基于图的推理方法主要基于神经网络模型或Path Ranking算法。例如,可以根据两个实体间的连通路径来判断两个实体是否属于某个关系。

(三) 图谱质量评估

质量评估也是知识库构建技术的重要组成部分。公开数据杂乱无规则,而且受当前阶段中文处理技术水平限制,有时从互联网公开数据中抽取到的知识元素有可能存在一些问题,如数据错误,获取到的知识元素准确率并不高,经过知识推理得到的知识同样也是没有质量保证的。因此在将其加入知识库之前,需要有一个质量评估的过程。通过质量评估,设定一定的规则量化图谱中存在的知识可信度,舍弃置信度较低的知识,来保障知识库的质量以提高知识图谱的推理能力^[53]。随着社交网络复杂性的提高和实体数的增多,数据间的冲突日益增多,对图谱质量进行评估,对于全局知识图谱的构建起着重要的作用。

五、大数据驱动的社交网络舆情主题图谱构建

图谱构建是大数据驱动的社会网络舆情主题图谱构建和分析的起点,通过对社会网络舆情大数据

的语义揭示和语义关联,对社交网络舆情大数据进行知识组织,形成语义关联的覆盖网络,为图谱挖掘提供高质量的数据资源,为图谱分析提供研究框架。以大规模采集的社交网络舆情数据为基础,对舆情数据进行语义抽取和标注,构建一个覆盖大规模舆情事件和实体的语义关联覆盖网络,实现对舆情主题图谱数据的采集和主题图谱的构建,为社交网络舆情主题图谱的挖掘提供依据和支撑。

(一) 数据采集与语义内容识别

借助行为追踪法、增量获得法、不同平台匹配法等方法进行数据的采集,为社交网络舆情主题图谱的构建提供更为丰富、有效、及时的数据支撑。数据预处理后,应用自然语言处理、文本挖掘、命名实体识别等方法和技术,对基于频繁词组挖掘的命名实体的识别技术和概念描述方法进行研究,以达到有效识别社交网络舆情大数据的实体名称技术性、实体关系、实体类别等语义内容。

其中,命名实体识别和分类是社交网络舆情主题图谱构建过程中的关键性技术。英文的命名实体识别相对容易,实体边界也比较容易确定,在英文文本中专有名词的第一个字符一般为大写。中文的专有名词识别就存在很大的困难,中文中的专有名词没有特殊标识。利用深度学习方法,尽可能地提取专有名词的特点,同时利用上下文的知识,充分利用这些技术和资源解决中文实体边界识别问题。

(二) 语义关联与语义标注

借助形式化描述、实体关联、属性映射、类映射和机器学习等方法和技术,构建社交网络舆情大数据的语义数据,对社交网络舆情事件中的实体、属性等对象之间的顺序关系、因果关系、时间关系、相似关系、参考关系等多类型逻辑关系进行标注。标注完成后,基于相似度计算、规则推理等方法和技术,根据语义标注的结果,对社交网络舆情大数据的语义关联进行研究,从而生成一个社交网络舆情大数据关联网络。

六、大数据驱动的社交网络舆情主题图谱分析

图谱分析是大数据驱动的社交网络舆情主题图谱构建和调控策略的基础环节,主要通过社交网络舆情主题图谱的分析和推理方法的研究,从关键节点、关键路径、网络结构影响力分析和趋势预测等多个维度对社交网络舆情主题图谱进行全面深入的挖掘,达到将数据转化为知识的目的。依据语义

关联覆盖网络,在对社交网络舆情主题图谱关键节点、关键路径、网络结构影响力分析和趋势预测等多个维度进行挖掘的基础上,探索大规模舆情事件和实体的综合分析和推理方法,达到面向大规模舆情事件和实体的、针对用户需求的社交网络舆情主题图谱知识发现的目的。

(一) 关键节点分析

政企对网络舆情的及时发现与快速收集所需舆情信息并做出“适时、适实、顺势”的响应,离不开对舆情事件传播过程中的关键节点的分析。准确找到事件的核心并针对事件进行有效引导,以确保舆情传播朝着健康的社会舆论方向发展是关键所在。

对于最常用的搜索引擎Google和Baidu来说,当用户在搜索引擎中查找某个具体人物或事件信息时,由于网络上的信息庞杂,还存在着大量的虚假和垃圾信息,使得有价值的信息被湮没,有时查询到的结果并不理想。为了得到想要的信息,只能采取人工方式,逐个浏览搜索到的网页,费时又费力。

将知识图谱引入人物画像分析,能够自动地从大量的无规则互联网数据中抽取和整理出人物的全局信息,将零散的、碎片化的数据汇集起来形成人物画像,并以结构化的形式存储和显示,极大地提高了研究人员分析人物全貌信息的效率,为精准营销、制定数据驱动产品提供了便利。Fader等以新闻文本为数据集,介绍了一种人物画像挖掘技术,从人物社会关系、参与事件、人物热度及情感分析三个维度进行研究,其研究成果可应用于热门人物探测、人物搜索和特定目标追踪等领域^[54]。

针对关键节点进行基于知识图谱的人物画像挖掘,便于对高舆情风险人群进行动态跟踪。例如,从情感分析角度来看,通过聚类和分类得出个体情感倾向和群体情感倾向,便于政企掌握情感异动,以便及时采取相关措施,疏导负面舆情。

(二) 关键路径分析

关键路径分析主要指的是网络路径分析与网络社区分析。社会网络分析的核心是社会网络中的节点,这些节点可以代表社会中的任何单位,例如个人、组织、群体等。节点之间的连接被称为节点之间的关系,节点以及它们的关系就构成了社会网络。在社交网络社区中,个人观点会受到多个邻居同时影响,但系统中所有节点的意见最终会趋向一致^[55]。目前的研究中,既有探索如何让算法更高效地对数据进行处理,提高算法的执行效率,也有考虑社交网络节点和网络连接中的社会属性^[56-58]。

常见的关键路径分析场景有舆情事件传播路径分析、敏感信息溯源分析等。

(三) 网络结构影响力分析

社交网络影响力可以通过用户之间的社交活动体现出来, 表现为用户的行为和思想等受他人影响发生改变以及用户行为和思想对他人的影响情况。在社交网络分析中, 引入社会属性权重来表示节点的社会属性差异, 给节点和边赋予相应的属性权重, 来表示信息在社交网络中传播的特点和规律, 可以提高关键节点挖掘算法执行效率^[59]。

网站影响力的评价方法可以分为定性和定量两种。定性的评价方法主要有问卷调查法和效益评价。定量的评价方法主要有链接分析和流量分析两种。一个站点被链接的次数越多, 证明该站点影响越大。网站的访问量越大, 则网站影响力越大, 该网站所蕴含的信息价值越高。

(四) 趋势模型分析: 舆情预测

借助舆情主题图谱是进行舆情预测的一种有效表现方式。对于网络舆情信息进行有效获取, 正确分析和统计、预测, 对于政府应对舆论危机将十分有利。网络舆情的发展一般存在蛰伏期、爆发期、持续期、消亡期四个阶段。对于网络舆情的有效预测, 协助政企准确判定舆情走向, 进而制定相应的应对管理办法。以企业舆情管理为例, 传统的舆情管理方式已不再适用, 企业舆情也需要现代化的技术形式进行管理。在新型的企业舆情管理中, 企业舆情部门的决策必须紧紧结合社会网络的结构、内容和用户三个核心点, 完善企业舆情管理制度, 及时应对突发事件。

有效的网络舆情预测有利于舆情危机蛰伏期的预防预警、有利于舆情危机爆发期的应急反应和干预、有利于舆情危机持续期的信息分析与研判、有利于舆情危机消亡期的信息再跟踪与反馈。

现有的网络舆情预测分析, 主要是在时间序列模型分析的基础上, 采用数学方法建立预测模型, 是一种定量分析的方法。网络舆情预测相关数学模型主要有基于统计学的预测模型和基于机器学习的预测模型两种。图5是网络舆情预测分析遵循的主要路线。

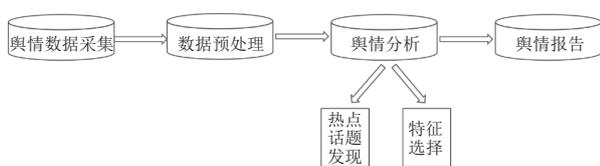


图5 网络舆情预测分析路线图

七、大数据驱动的社交网络舆情调控与应用前景

舆情调控是舆情主题图谱构建和调控策略的桥梁。基于社交网络舆情主题图谱, 动态提取社交网络舆情特征, 并构建社交网络舆情预警模型, 然后针对不同的舆情主体、传播路径和舆情主题, 提出自适应的舆情点调控、舆情链路调控和舆情主题调控策略, 并在三者的基础上, 提出舆情全局精准调控策略。利用爬虫、信息检索以及专家知识库等技术, 构建舆情特征——调控策略知识库, 然后以此为基础, 针对社交网络舆情的不同特征, 提出社交网络舆情自适应的点调控策略、链路调控策略和主题调控策略, 最后在三者的基础上提出全局精准调控策略, 达到对社交网络舆情主题图谱从点、线、面, 全局精准调控的目的。

与舆情图谱结合, 在信息推送和场景应用等方面建立相应的功能模块, 为个人用户、企业和政府提供强大的舆情分析服务。以大数据思维检测和分析海量的社交网络上的海量舆情数据, 通过构建社交网络舆情主题图谱, 为各级政府或相关机构提供民生热点、热点事件、热点话题等重点关注舆情的发现与分析服务, 以帮助政府或相关机构了解特定舆情事件的传播路径、传播态势并把握其发展趋势。依据不同场景, 可分为经济网络监控、面向舆情的政府知识管理、突发事件预警。

(一) 经济网络监控

经济网络舆情监测的基本类别包括用户消费类、投资类、创业类、生态环保类。用户消费类主要包括线上消费数据分析和市场价格分析, 从网民反映的消费问题、生活消费等数据监测反映当地物价走势, 判断当前经济形势和预测未来经济发展; 投资类分析主要包括金融投资分析和房地产投资分析, 从企业投资和投资金融服务等方面反映民众对当前投资环境的判断和评价, 监测民众最关注的住房、投资舆情; 创业类分析主要是分析民众对创新创业的热情程度, 以此反映整个社会创新创业的情况; 生态环保监测社交网络上的生态环保热点问题, 从网民反映的生态环保问题、对环保关注程度等方面监测生态环保情况。

(二) 面向舆情的政府知识管理

政府舆情知识管理可以分为三个部分: 事件知识管理、决策类知识管理和响应类知识管理。事件知识管理主要针对网络舆情本体, 探究事件发生的

性质, 勾画人物和相关机构关系脉络。大致相当于给政府阐述“何人何地何时间发生何事”。在静态描述完毕后, 对舆情进行跟踪, 持续跟进事件发展动态和变化, 让政府可以获取事件的演化机理和相关影响。决策类知识管理主要通过信息的收集、加工和分类, 帮助政府做出决策和判断。响应类知识管理则包括政府针对事态变化需要进行调度的机构, 做出行动的部门和获取的资源。构建面向舆情的政府知识管理的目的在于引导和控制舆情变化, 降低不良信息对社会造成的损害, 防止别有用心之人对社会的正常秩序产生干扰。

(三) 突发事件预警

社会突发事件可能会造成严重社会危害, 需要实时对重点舆情话题进行检测, 及时采取应急处置措施予以应对。舆情主题图谱中的数据来源非常广阔, 因此能从中发掘出各种不同的事件和关系。针对突发事件的危害性, 有必要专门将这一部分提取出来进行相应的分析和策划。根据舆情主题敏感度、发展趋势和危害程度等级, 提前向有关部门发出实时预警。

八、总结与展望

本文首先对大数据时代的社交网络舆情主题图谱的研究现状进行了较为全面的总结, 并在研究基础上对社交网络舆情主题图谱构建、分析与调控等开展了详细的介绍。从关键节点、关键路径、趋势模型等不同维度进行社交网络舆情主题图谱的多元分析。挖掘更深层次的知识关联与更广范围的知识范畴。对社交网络舆情态势感知与“适时、适实、适势”的响应与引导具有一定的理论意义和实用价值。

参考文献

- [1] 游丹丹, 陈福集. 我国网络舆情预测研究综述[J]. 情报科学, 2016, 34(12): 156-160.
- [2] 李纲, 陈璟浩. 突发公共事件网络舆情研究综述[J]. 图书情报知识, 2014(2): 111-119.
- [3] 陈华, 梁循, 阮进. 网络与舆情关联分析系统的设计实现[C]. 苏州: 第三届全国信息检索与内容安全学术会议, 2007: 54-58.
- [4] 张柳. 社交网络舆情用户主题图谱构建及舆情引导策略研究[D]. 长春: 吉林大学, 2021.
- [5] 陈燕方, 李志宇, 梁循, 等. 在线社会网络谣言检测综述[J]. 计算机学报, 2018, 41(7): 1648-1677.
- [6] 刘峤, 李杨, 段宏, 等. 知识图谱构建技术综述[J]. 计

算机研究与发展, 2016, 53(3): 582-600.

- [7] DONG X, GABRILOVICH E, HEITZ G, et al. Knowledge vault: a web-scale approach to probabilistic knowledge fusion[C]. New York: The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'14), 2014: 601-610.
- [8] 孙镇, 王惠临. 命名实体识别研究进展综述[J]. 现代图书情报技术, 2010(6): 42-47.
- [9] LIN Y, LIU Z, SUN M, et al. Learning entity and relation embeddings for knowledge graph completion[C]. Austin: The Twenty-Ninth AAAI Conference on Artificial Intelligence, 2015: 2181-2187.
- [10] 杨博, 蔡东风, 杨华. 开放式信息抽取研究进展[J]. 中文信息学报, 2014, 28(4): 1-11+36.
- [11] BANKO M, CAFARELLA M J, SODERLAND S, et al. Open information extraction from the web[EB/OL]. (2016-03-15). <https://www.docin.com/p-1489469183.html>.
- [12] PAPKA R, CROFT W B, BARTO A G, et al. On-line new event detection, clustering, and tracking[EB/OL]. [2021-06-08]. <http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=0EB104652B9EEA5E4DFDC776A5BF7318?doi=10.1.1.50.3192&rep=rep1&type=pdf>.
- [13] WU F, WELD S. Autonomously semantifying wikipedia[C]. Lisbon: The sixteenth ACM Conference on Information and Knowledge Management, 2017: 41-50.
- [14] 王宇, 谭松波, 廖祥文, 等. 基于扩展领域模型的有名属性抽取[J]. 计算机研究与发展, 2010, 47(9): 1567-1573.
- [15] JOSHI A, BALAMURALI R, BHATTACHARYYA P. C-feel-it: a sentiment analyzer for micro-blogs[C]. Portland: The ACL-HLT, 2011: 127-132.
- [16] 谢丽星, 周明, 孙茂松. 基于层次结构的多策略中文微博情感分析和特征抽取[J]. 中文信息学报, 2012, 26(1): 73-83.
- [17] MUKHERJEE S, BHATTACHARYYA P. Feature specific sentiment analysis for product reviews[C]. Berlin: International Conference on Intelligent Text Processing and Computational Linguistics, 2012: 475-487.
- [18] MYLOPOULOS J, BUITELAAR P, OLEJNIK D, et al. OntoLT: A Protégé Plug-In for Ontology Extraction from Text[J/OL]. (2003-01-01). https://www.researchgate.net/publication/277285900_OntoLT_A_Protege_Plug-In_for_Ontology_Extraction_from_Text.
- [19] HEARST M. Automatic acquisition of hyponyms from large text corpora[C]. Nantes: The 14th Conference on Computational Linguistics, 1992: 539-545.
- [20] 刘群, 李素建. 基于《知网》的词汇语义相似度计算[C]. 中国台北: 第三届汉语词汇语义学研讨会, 2002: 59-76.
- [21] NAVIGLI R, VELARDI P. Learning domain ontologies from document warehouses and dedicated web sites[J]. Computational Linguistics, 2004, 30(2): 151-179.

- [22] KUREMATSU M, IWADE T, NAKAYA N, et al. DODDLE II: a domain ontology development environment using a MRD and text corpus[J]. *IEICE Transactions on Information and Systems*, 2004, 87(4): 908-916.
- [23] WU F, WELD S. Automatically refining the wikipedia infobox ontology[C]. *Beijing: The 17th International Conference on World Wide Web*, 2008: 635-644.
- [24] BUNESCU C, MOONEY J. Subsequence kernels for relation extraction[C]. *Vancouver: Advances in Neural Information Processing Systems*, 2005: 171-178.
- [25] ZHANG X, ZHANG J, ZENG J, et al. Towards accurate distant supervision for relational facts extraction[C]. *Sofia: The 51st Annual Meeting of the Association for Computational Linguistics*, 2013: 810-815.
- [26] WEI X, RAPHAEL H, LE Z, et al. Filling knowledge base gaps for distant supervision of relation extraction[C]. *Sofia: The 51st Annual Meeting of the Association for Computational Linguistics*, 2013: 13-17.
- [27] TAKAMATSU S. Reducing wrong labels in distant supervision for relation extraction[C]. *Jeju Island: The 50th Annual Meeting of the Association for Computational Linguistics*, 2012: 721-729.
- [28] 蒋祖存, 徐双敏, 牟珈琪. 基于CiteSpace的国内公共卫生事件研究现状与趋势[J]. *黄冈师范学院学报*, 2021, 41(5): 114-121.
- [29] 冯静, 齐艳平. 知识图谱在智慧城市治理过程中的应用[J]. *中国国情国力*, 2021(7): 50-53.
- [30] MOLOKWU B, KOBTI Z. Social network analysis using RLVEC: representation learning via knowledge-graph embeddings and convolutional neural-network[C]. *Yokohama: The Twenty-Ninth International Joint Conference on Artificial Intelligence*, 2020: 5198-5199.
- [31] PARNAMI A, DESHPANDE M, MISHRA K, et al. Transformation of node to knowledge graph embeddings for faster link prediction in social networks[DB/OL]. (2021-11-17). <https://arxiv.org/abs/2111.09308>.
- [32] 庄莉, 宋立华. 企业业务数据及外部舆情信息的融合分析[J]. *福建电脑*, 2021, 37(8): 52-55.
- [33] ELEZAJ O, YAYILGAN Y, KALEMI E, et al. Towards designing a knowledge graph-based framework for investigating and preventing crime on online social networks[C]. *Cham: International Conference on e-Democracy*, 2019: 181-195.
- [34] PATEL H, PARASKEVOPOULOS P, RENZ M. GeoTeGra: a system for the creation of knowledge graph based on social network data with geographical and temporal information[C]. *Barcelona: 2018 International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 2018: 617-620.
- [35] 丁晟春, 叶子. 基于知识图谱的企业风险发现[J]. *图书馆论坛*, 2022, 42(2): 129-138.
- [36] 杨阳. 涉警网络舆情主题图谱构建及传播分析——以“女子深夜离奇失踪”为例[J]. *云南警官学院学报*, 2021(4): 90-95.
- [37] 杜亚军, 吴越. 微博知识图谱构建方法研究[J]. *西华大学学报(自然科学版)*, 2015, 34(1): 27-35+89.
- [38] 韩佳伶, 余天池. 基于LDA主题知识图谱的公共安全事件微博舆情实证研究——以“山西农村饭店坍塌事件”为例[J]. *情报探索*, 2021(9): 85-93.
- [39] 王晰巍, 李玥琪, 邱程程, 等. 直播平台用户信息交互行为图谱及特征研究[J]. *图书情报知识*, 2021, 38(04): 15-26+61.
- [40] 黄微, 卢国强, 赵旭. 基于知识图谱的微博主题演变路径研究[J/OL]. (2021-01-14). *情报理论与实践*. <http://kns.cnki.net/kcms/detail/11.1762.G3.20211102.1031.003.html>.
- [41] XIE T, YANG Y, LI Q, et al. Knowledge graph construction for intelligent analysis of social networking user opinion[C]. *Cham: International Conference on e-Business Engineering*, 2019: 236-247.
- [42] 齐金山, 梁循, 李志宇, 等. 大规模复杂信息网络表示学习: 概念、方法与挑战[J]. *计算机学报*, 2018, 41(10): 2394-2420.
- [43] BORDES A, WESTON J, COLLOBERT R. Learning structured embeddings of knowledge bases[C]. *San Francisco: The AAAI Conference on Artificial Intelligence*, 2011: 301-306.
- [44] SOCHER R, CHEN D, MANNING D. Reasoning with neural tensor networks for knowledge base completion[C]. *Lake Tahoe: Advances in Neural Information Processing Systems*, 2013: 926-934.
- [45] YANG B, YIH W, HE X. Embedding entities and relations for learning and inference in knowledge bases[C]. *San Diego: The 3rd International Conference on Learning Representations*, 2015: 1-12.
- [46] QUOC L, MIKOLOV T. Distributed representations of sentences and documents[C]. *Beijing: The 31st International Conference on Machine Learning*, 2014: 1188-1196.
- [47] 刘知远, 孙茂松, 林衍凯, 等. 知识表示学习研究进展[J]. *计算机研究与发展*, 2016, 53(2): 247-261.
- [48] WANG Z, ZHANG J, FENG J. Knowledge graph embedding by translating on hyperplanes[C]. *Quebec: The AAAI Conference on Artificial Intelligence*, 2014: 1112-1119.
- [49] JI G, HE S, XU L. Knowledge graph embedding via dynamic mapping matrix[C]. *Beijing: The 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, 2015: 687-696.
- [50] XIAO H, HUANG M, HAO Y. TransG: a generative mixture model for knowledge graph embedding[C]. *Phoenix: The AAAI Conference on Artificial Intelligence*, 2016: 2659-2665.
- [51] TAN H, AGICHTEIN E, IPEIROTIS P, et al. Trust, but verify: predicting contribution quality for knowledge base construction and curation[C]. *New York: The 7th ACM*

International Conference on Web Search and Data Mining, 2014: 553-562.

[52] FADER A, SODERLAND S, ETZIONI O. Identifying relations for open information extraction[C]. Edinburgh: The Conference on Empirical Methods in Natural Language Processing, 2011: 1535-1545.

[53] 李赵洁. 基于文本的人物画像挖掘技术的研究与应用[D]. 成都: 电子科技大学, 2016.

[54] SZNAJD K, WERON R. A simple model of price formation[J]. International Journal of Modern Physics C, 2002, 13(1): 115-123.

[55] YANG Y, WANG Z, PEI J. Tracking influential individuals in dynamic networks[J]. IEEE Transactions on

Knowledge & Data Engineering, 2017, 29(11): 2615-2628.

[56] FISCHBACH S, ZARZOSA J. Big data on a smaller scale: a social media analytics assignment[J]. *Journal of Education for Business*, 2018, 93(3): 142-148.

[57] 曹玖新, 闵绘宇, 徐顺, 等. 基于启发式和贪心策略的社交网络影响最大化算法[J]. *东南大学学报(自然科学版)*, 2016, 46(5): 950-956.

[58] 陈红松, 王钢, 张鹏. 基于Hadoop云平台的新浪微博社交网络关键节点挖掘算法[J]. *东南大学学报(自然科学版)*, 2018, 48(4): 590-595.

[59] 梁循, 许媛, 李志宇, 等. 社会网络背景下的企业舆情研究述评与展望[J]. *管理学报*, 2017, 14(6): 925-935.

编辑 何婧

(上接第18页)

[31] SCHULMAN P, KEITH D, SELIGMAN M. Is optimism heritable? a study of twins[J]. *Behavior Research & Therapy*, 1993, 31(6): 569-574.

[32] PAOLILLO A, PLATANIA S, MAGNANO P, et al. Organizational justice, optimism and commitment to change[J]. *Procedia-Social and Behavioral Sciences*, 2015, 191: 1697-1701.

[33] PAPPENHAUSEN C. Managerial optimism and search[J]. *Journal of Business Research*, 2009, 63(7): 716-720.

[34] HMIELESKI K M, BARON R A. Entrepreneurs' optimism and new venture performance: a social cognitive perspective[J]. *Academy of Management Journal*, 2009, 52(3): 473-488.

[35] 刘东, 郑鑫, 周小虎, 等. 创业乐观中介下社交网络对新创企业绩效的影响[J]. *科技进步与对策*, 2016, 33(15): 68-75.

[36] 高山行, 肖振鑫, 高宇. 企业制度资本对新产品开发的影响研究——市场化程度与竞争强度的调节作用[J]. *管理评论*, 2018, 30(9): 110-120.

[37] 曹玉珊, 潘孟, 冉彬双. 管理层权力、市场竞争程度与企业可持续增长——基于国有上市企业的经验研究[J]. *会计之友*, 2018(1): 84-89.

[38] 诸波, 干胜道. 市场竞争程度、经营战略与业绩评价指标选择[J]. *会计研究*, 2015(2): 51-57+94.

[39] 奉小斌. 集群新创企业平行搜索对产品创新绩效的影响: 管理者解释与竞争强度的联合调节效应[J]. *研究与发展管理*, 2016, 28(4): 11-21.

[40] 杨秀芝, 李柏洲. 企业适应能力的内涵及其提升对策研究[J]. *管理世界*, 2007(4): 166-167.

[41] 闫伟宸, 肖星, 王一倩. 国企性质、高管特征和投资效率[J]. *科研管理*, 2020, 41(8): 148-159.

[42] RICHARD O, WU J, MARKOCZY L, et al. Top

management team demographic-faultline strength and strategic change: what role does environmental dynamism play?[J]. *Strategic Management Journal*, 2019, 40(6): 987-1009.

[43] HUANG X, TEOH S H, ZHANG Y. Tone management[J]. *The Accounting Review*, 2013, 89(3): 1083-1113.

[44] HENRY E, LEONE A J. Measuring qualitative information in capital markets research: comparison of alternative methodologies to measure disclosure tone[J]. *The Accounting Review*, 2015, 91(1): 153-178.

[45] LOUGHRAN T, MCDONALD B. When is a liability not a liability? textual analysis, dictionaries, and 10-ks[J]. *The Journal of Finance*, 2011, 66(1): 35-65.

[46] PRICE M K, DORAN J S, PETERSON D R, et al. Earnings conference calls and stock returns: the incremental informativeness of textual tone[J]. *Journal of Banking & Finance*, 2012, 36(4): 992-1011.

[47] 刘志彪, 姜付秀, 卢二坡. 资本结构与产品市场竞争强度[J]. *经济研究*, 2003(7): 60-67+91.

[48] GENTRY R J, SHEN W. The impacts of performance relative to analyst forecasts and analyst coverage on firm R& D intensity[J]. *Strategic Management Journal*, 2013, 34(1): 121-130.

[49] BUSENITZ L W, BARNEY J B. Differences between entrepreneurs and managers in large organizations: biases and heuristics in strategic decision-making[J]. *Journal of Business Venturing*, 1997, 12(1): 9-30.

[50] DIAZ-FERNANDEZ M C, GONZALEZ-RODRIGUEZ M R, SIMONETTI B. Top management teams' demographic characteristics and their influence on strategic change[J]. *Quality & Quantity: International Journal of Methodology*, 2015, 49(2): 1305-1322.

[51] 辛清泉, 郑国坚, 杨德明. 企业集团、政府控制与投资效率[J]. *金融研究*, 2007(10): 123-142.

编辑 何婧