

基于社交网络群智感知信息的非常规突发事件描述方法研究

陈海燕¹ 徐 峥^{2,3}

(华东政法大学计算机科学与技术系 上海 201620)¹ (公安部第三研究所物联网中心 上海 201142)²
(清华大学公共安全研究院 北京 100084)³

摘要 群智感知是以近年来兴起的物联网、大数据、云计算等多种技术为基础发展起来的。群智感知就是用城市环境中感知的大数据来解决城市本身所面临的挑战,通过对多源异构数据的整合、分析和挖掘来提取知识和价值,从而提高城市服务的现代化水平。非常规突发事件是一类前兆特征不充分、具有明显的复杂性和潜在次生衍生危害、破坏性严重、采用传统管理方式难以应对处置的罕见重大突发事件。由于构成群智感知网络的基本单元是无所不在的移动智能终端用户,这种新型的体系架构无论是在感知及认知的广度、深度还是在构建的成本与速度上,都是传统技术手段所无法比拟的。对社交网络信息进行语义感知、时空关联等以对非常规突发事件进行描述,并通过实例对该方法进行验证。非常规突发事件的语义信息包括关键词、模式、语句等。时空信息用来描述非常规突发事件的客观时间属性与空间属性。时间属性包括非常规突发事件的发生时间、结束时间,以及重要时间点。空间属性包括非常规突发事件的发生位置、影响位置,位置信息可以是路名建筑物名称,也可以具体到 GIS 信息。多媒体信息包括图像、视频、音频等。

关键词 群智感知,非常规突发事件,语义描述,时空感知

中图分类号 TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2016.5.038

Crowdsourcing Based Description of Urban Emergency Events

CHEN Hai-yan¹ XU Zheng^{2,3}

(Department of Computer Science and Technology, East China University of Political Science and Law, Shanghai 201620, China)¹

(Department of Internet of Things, The Third Research Institute of the Ministry of Public Security, Shanghai 201142, China)²

(Department of Public Security, Tsinghua University, Beijing 100084, China)³

Abstract Crowdsourcing is a process of acquisition, integration, and analysis of big and heterogeneous data generated by a diversity of sources in urban spaces, such as sensors, devices, vehicles, buildings, and human. Especially, nowadays, no countries, no communities, and no person are immune to urban emergency events. Detection about urban emergency events, e. g. , fires, storms, traffic jams, is of great importance to protect the security of humans. Recently, social media feeds are rapidly emerging as a novel platform for providing and dissemination of information that is often geographic. The content from social media often includes references to urban emergency events occurring at, or affecting specific locations. In this paper, in order to describe the real time urban emergency event, the crowdsourcing based model was proposed. Firstly, users of social media are set as the target of crowd-sourcing. Secondly, the semantic, spatial and temporal information from the social media are extracted to detect the real time event. Thirdly, a GIS based annotation of the detected urban emergency event is shown. The proposed method was evaluated with extensive case studies based on real urban emergency events. The results show the accuracy and efficiency of the proposed method.

Keywords Crowdsourcing, Urban emergency event, Semantic description, Spatial temporal

1 引言

群智感知是以近年来兴起的物联网、大数据、云计算等多种技术为基础发展起来的^[1]。随着感知技术和计算环境的成熟,各种大数据在城市中悄然而生^[2]。随着网络通信技术的

迅速发展以及移动智能终端(如智能手机、平板电脑等)的快速普及,智能终端已经普遍携带视频监控、音频、加速传感器等感知设备。部分高端智能终端所能携带的视频设备已经超过部分低端的视频监控设备。智能终端的大量普及,使得构建以人为中心的感知与计算网络(简称“群智感知计算”)成为

到稿日期:2015-03-07 返修日期:2015-06-08 本文受国家自然科学基金项目(06BFX051),国家自然科学基金青年项目(6130202),上海高校选拔培养优秀青年教师科研专项基金(hz05046),上海市自然科学基金青年项目(13ZR1452900),中国博士后基金一等面上项目(2014M560085)资助。

陈海燕(1978—),男,硕士生,讲师,主要研究方向为计算机网络、数据挖掘、人工智能、信息安全,E-mail:tom_chy@163.com;徐 峥(1984—),男,博士,助理研究员,主要研究方向为公共安全、群智感知等。

可能,以实现物理世界和数字世界的完美融合。群智感知就是用城市环境中感知的大数据来解决城市本身所面临的挑战,通过对多源异构数据的整合、分析和挖掘,来提取知识和价值,从而提高城市服务的现代化水平。群智感知计算已成为一种全新的应用模式和发展趋势,正受到学术界的广泛关注。群智感知计算是大量普通用户使用移动智能终端作为基本的感知单元,对周围环境及事物进行信息感知与搜集,并通过互联网或移动通信网络组成动态的信息感知网络,实现大规模的信息交叉感知与汇总。数据处理中心根据用户的需求对所获取的海量感知信息进行选择性过滤与筛选,通过感知融合与深度认知计算,实现从碎片化信息到群智信息的转换,完成大规模的复杂社会事件认知任务。

非常规突发事件(infrequent fatal emergency)是一类前兆特征不充分、具有明显的复杂性和潜在次衍生危害、破坏性严重、采用传统管理方式难以应对处置的罕见重大突发事件^[3]。非常规突发事件的突然性、迫切性、连带性和社交网络的迅速发展,决定了社交网络在非常规突发事件的信息获取及分析中的重要作用^[4]。近年来,社交网络、论坛、博客等新兴媒体的快速发展,使得传统的基于物理传感器感知的非常规突发事件的应急管理机制已经无法适应当前的复杂信息环境。群智感知网络的重要特征预示了它将成为一种独特的、尤其适用于大规模复杂环境及事物的新型感知与认知计算体系架构。由于构成群智感知网络的基本单元是无所不在的移动智能终端用户,无论是在感知及认知的广度、深度,还是在构建的成本与速度上,这种新型的体系架构都是传统技术手段所无法比拟的。例如,大范围自然灾害(如火灾、水灾、雪灾、地震等)、大规模群体事件、违法犯罪等事件的发生往往带有突然性、偶然性或隐蔽性;加上事发现场恶劣、危险与复杂的环境等因素,使得通过传统渠道(如媒体报道)在第一时间获取全面、准确的现场信息成为几乎不可能完成的任务。比较典型的例子有在“波士顿马拉松爆炸案”中推特用户上传案发现场照片,在“上海外滩踩踏事件”中南都周刊记者微博上传实时踩踏视频等。另外,尽管近年来国内各地政府在推动平安城市、建设视频监控基础设施方面做了大量投入,但是由于成本高昂,这些基础设施主要集中于大城市的重要建筑或行人和车辆相对密集的地方,无法实现对所有区域的全空域、无间断覆盖。然而,几乎任何一个重大自然灾害、社会事件或犯罪行为的现场都往往会存在大量的移动智能终端用户,这使得通过群智感知网络实现对事发现场的大规模、低成本感知与认知计算成为可能。

本文主要研究基于群智感知技术的非常规突发事件的语义信息、时空信息、多媒体信息的有效提取方法。主要工作如下:

(1)群智感知信息具有低质量、碎片化、低价值等特性,例如微博用户所发布的微博信息文本较短、用词随意且不规范、噪音歧义较多。因此,群智感知信息的低质量特性给准确挖掘非常规突发事件的语义信息增加了困难。非常规突发事件的语义信息包括关键词、模式、语句等信息。关键词信息指具有代表性的可以描述非常规突发事件的概念词汇,例如“着火”、“火势”等词汇可以描述“火灾”这个突发事件。不同于关键词,模式是由多个关键词所组成的词汇对,模式相较于词汇更能够用来检测非常规突发事件。例如“着火”和“消防车”模

式相较于“着火”能够更准确地检测正在发生的非常规突发事件。语句相较于关键词和模式具有更直接的语义信息,能够更有效地刻画关键词之间的语义关系。

(2)不同于语义信息,时空信息用来描述非常规突发事件的客观时间属性与空间属性。时间属性包括非常规突发事件的发生时间、结束时间以及重要时间点。空间属性包括非常规突发事件的发生位置、影响位置,位置信息可以是路名、建筑物名称,也可以具体到 GIS 信息。由于群智感知信息所具有的低质量、碎片化、低价值等特性,例如微博用户可能在较长的时间之后才发布关于某件突发事件的信息,可能在其他地方发布另一个地方发生的突发事件的信息,本文构建非常规突发事件的大量社交网络传感数据时空关联模型。

(3)不同于语义信息与时空信息,多媒体信息能够更为直观地描述非常规突发事件的实时情况。多媒体信息包括图像、视频、音频等。例如微博用户可能会上传正在发生的非常规突发事件的实时图片或视频,这些多媒体信息可以用来帮助决策者或应急管理部门了解实时情况。本文对这些多媒体信息进行有效提取。

本文第 2 节介绍了群智感知技术的相关工作;第 3 节介绍了如何对社交网络群智感知信息进行语义感知;第 4 节介绍了如何对社交网络群智感知信息进行时空感知;第 5 节针对一个现实环境中的非常规突发事件进行实证研究;最后总结全文。

2 相关工作

在计算机科学领域,与群智感知相关的概念包括群体计算(crowd computing)、参与式感知(participatory sensing)、社群感知(social sensing)、众包等等,它们都以大量用户参与作为基础,旨在将大家的智慧(collective intelligence)利用起来。相比群智感知,群体计算包含的范围更广一些,众包多指任务分发的机制,而参与式感知和社群感知与群智感知在理念上更近,但分别强调的是不同的方面。在群智感知中,完成复杂感知任务的参与者不需要拥有专业技能。与此相反,大量草根用户可以成为中坚力量,通过合理的协作来完成他们单独不可能或者说根本想不到要完成的任务。在国外,利用互联网实现群智计算的比较典型的项目有两个。一个是足不出户寻找外星人(SETI@home)——通过互联网,利用家用个人计算机处理天文数据,共同搜寻地外文明。从 1999 年到 2004 年,繁复的科学计算任务被分解为千百万个小的工作单元,由客户电脑在计算资源闲置时(即屏幕保护运行时)对数据进行处理。任何电脑用户只要能上网就能参与这个项目,奉献计算资源的同时又不影响自己使用。一旦发现有价值的信号,SETI@home 将按照国际天文学联合会的电报发表公告,这是天文学界取得重大发现时公之于众的一种标准方式;而用其屏幕保护程序找到该信号的人将被赋予“合作发现者”的称号。2005 年关闭之前,SETI@home 吸引了 543 万个用户,这些用户的电脑累积工作了 243 万年,分析了大量积压数据。另一个是 Louis Von Anh 开发的 reCAPTCHA。该系统旨在利用网络用户将《纽约时报》古老的报纸存档数字化。reCAPTCHA 在验证的确是正常用户而不是机器在后台操纵的同时,用户对于污染、扭曲文字的识别能力被用来数字化古籍中不能被计算机自动识别的文字。reCAPTCHA 被超过

10 万家网站使用,每天数字化超过 4 千万个单词,结果《纽约时报》所存的 130 年的资料本来需要巨大的时间和人力资源,最终在几个月之内就由网友们完成了,而且是在网友们事前无知、事后惊讶中完成的。2009 年,美国杜克大学的研究者成功地尝试让智能手机持有者通过主动提交的方式实现信号指纹采集,这就突破了专业设备的壁垒,但还是需要有人把这个感知当作额外的任务来做^[5]。根据日常的工作和生活需要,手机用户经常会在各种建筑物内走动,每走到一个位置,手机都可以不用用户操作而自动记录下当前位置的信号指纹,由此收集就变得简单而几乎没有成本了。

近几年无线网络与移动计算有关的几个著名学术会议中,MobiSys 2011 专门开辟了一个会议单元收录众包的相关论文,MobiCom 2012 中至少有 4 篇论文与群智感知有关^[6-9]。将普通用户的移动设备(手机、平板电脑等)作为基本感知单元,通过移动互联网进行有意识或无意识的协作,从而实现感知任务分发与感知数据收集,完成大规模的、复杂的社会感知任务。在国内,最早提出群智感知的概念并且进行相应研究的是清华大学的刘云浩教授。刘云浩教授在 2012 年的中国计算机学会通讯上面首次系统性地提出并解释了群智感知计算的概念^[1]。MobiCom 2012 收录了两篇用群智理念研究室内定位的论文:一篇来自刘云浩教授科研组的杨铮和吴陈沭,该论文根据信号指纹间的时空特性建立指纹与位置的对应关系^[7]。国家自然科学基金委员会针对群智感知问题,专门支持了重点项目群。近期,北京邮电大学的马华东等在 IEEE TON 上发表了关于群智感知网络在环保监测上应用的文章^[10]。西安交通大学的龚怡宏等在 IEEE TKDE 上发表了关于群智感知信息质量控制的文章^[11]。西北工业大学的郭斌等在 IEEE TPDS 上发表了群智感知网络自组织方面的研究成果^[12]。在利用群智感知信息进行语义感知和事件描述方面,目前的工作依然不多。Crooks 等人^[14]把推特看作是一个大的分布式传感器系统,对某次地震进行了时空信息的分析。Longueville 等人^[15]利用推特对火灾进行时空信息的分析。Liu 等人^[16]开发了 MoboQ 系统,利用位置信息进行用户之间的问答。Qu 和 Zhang^[17]利用推特的位置信息进行商业区的金融分析。

综上所述,群智感知仍处在刚刚开始阶段,依然存在低数据质量、缺乏数据有效感知手段、缺乏数据有效组织互联、缺乏数据协同融合应用等问题。尤其在利用群智感知信息进行非常规突发事件分析上仍然属于空白。

3 基于社交网络群智感知信息的非常规突发事件语义感知

非常规突发事件的语义信息包括关键词、模式、语句等信息。关键词信息指具有代表性的可以描述非常规突发事件的概念词汇,例如“着火”、“火势”等词汇可以描述“火灾”这个突发事件。不同于关键词,模式是由多个关键词所组成的词汇对,模式相较于词汇更能用来检测非常规突发事件。例如“着火”和“消防车”模式相较于“着火”能够更准确地检测正在发生的非常规突发事件。语句相较于关键词和模式具有更直接的语义信息,能够更有效地刻画关键词之间的语义关系。

3.1 有效微博信息过滤

本文以微博平台作为提取社交媒体信息的平台。由于微博本身所具有的特性,用户提供的信息往往具有噪音。例如,虽然某条微博信息说明了一个正在发生的火灾,但是其中并不包含任何有意义的位置信息或图像信息,无法判定事件的发生点。因此,需要设定一些规则来对微博信息进行有效性筛选。通过对实际的例子进行观察及总结,本文列出 4 条筛选规则。

规则 1 有效微博信息中具有地理位置信息。地理位置信息反映了事件实际发生的地方,因此具有重要性。

规则 2 有效微博信息中具有签到信息。不同于地理位置信息,签到信息是指用户利用其移动互联网设备如手机进行签到,标记其自身所处的实际位置。

规则 3 有效微博信息中具有多媒体信息。多媒体信息可以反映非常规突发事件的实时情况,因此也具有重要性。

规则 4 有效微博信息应该是原创的微博而非转帖。原创的微博表明了用户在非常规突发事件的现场概率较大。

如图 1 所示,(b)和(c)都不是有效微博。(b)尽管叙述了一场正在发生的火灾,但是并没有设计地理位置信息。(c)虽然具有着火这两个词,但是语义上并不是说火灾相关的事情。(e)是一条有效微博信息,该信息不仅说出了火灾的地址信息,还包含自己的签到信息,并且提供了火灾的现场图片。

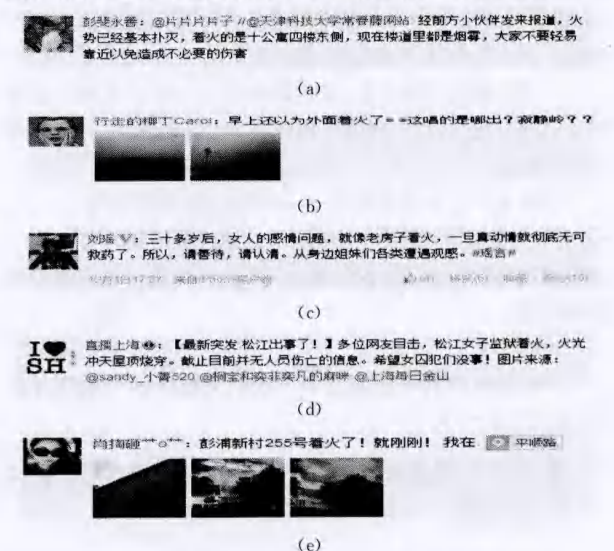


图 1 有效微博信息举例

3.2 关键词提取

关键词信息指具有描述非常规突发事件的概念词汇,例如“消防车”、“消防员”可以用来描述“火灾”。在本文中,关键词的提取流程如下:

1) 选取有效微博信息,即满足规则 1、规则 2、规则 3 或规则 4 的微博信息,其中规则 1 和规则 2 中必须满足一个,规则 4 必须满足。

2) 利用中科院分词软件¹⁾提取微博中的名字,一般来说名字具有实际的语义。对提取出的关键词集合定义如下:

$$K = \{k_1, k_2, \dots, k_{|K|}\} \quad (1)$$

3) 对提取的关键词按照其出现的篇频进行排序,即出现

¹⁾ www.ictclas.org

在多少个微博信息之中,定义为 $df(k_i)$ 。例如关键词“着火”在 10 个微博信息中出现了 4 次,那么 $df(k_i)$ 就为 0.4。

3.3 模式提取

不同于关键词,模式具有更好的描述非常规突发事件的能力,也具有更好的检测非常规突发事件的能力。例如,如果仅是“着火”这个词出现在微博中,可能并不代表有火灾正在发生,但如果是“着火”和“消防车”一起出现在微博中,那么说明正在发生火灾的概率很大。在本文中,模式提取的流程如下:

1) 选取有效微博信息,即满足规则 1、规则 2、规则 3 或规则 4 的微博信息,其中规则 1 和规则 2 中必须满足一个,规则 4 必须满足。

2) 利用频繁模式提取算法^[13]提取模式,在本文中,只考虑二阶的频繁模式,因为微博信息往往较短,二阶以上的频繁模式出现的概率较小。将提取出的模式定义为一个关键词的二元组, $p = \langle k_i, k_j \rangle$ 。

3) 对提取的模式按照其出现的篇频进行排序,即出现在多少个微博信息之中,定义为 $df(p)$ 。例如模式“着火 浓烟”在 10 个微博信息中出现了 2 次,那么 $df(p)$ 就为 0.2。

3.4 语句提取

语句是比关键词与模式具有更高具体含义的语义感知元素。语句可以较为准确地描述关键词之间的语义关系。同样地,利用语句来描述一个非常规突发事件可以更好地为人所理解。在本文中,语句提取的流程如下:

1) 选取有效微博信息,即满足规则 1、规则 2、规则 3 或规则 4 的微博信息,其中规则 1 和规则 2 中必须满足一个,规则 4 必须满足。

2) 利用字符串分割算法,将语句的字符串设定为“.,!?,……”4 个字符串,按照这些结尾字符串提取语句。将提取出的模式定义为一个关键词的集合,其中 m 为语句包含的关键词的数量:

$$s = \{k_1, k_2, \dots, k_m\} \quad (2)$$

3) 对提取的语句按照其中出现的关键词的个数进行排序,定义为 $df(s)$ 。

4 基于社交网络群智感知信息的非常规突发事件时空关联

不同于语义信息,时空信息用来描述非常规突发事件的客观时间属性与空间属性。时间属性包括非常规突发事件的发生时间、结束时间以及重要时间点。空间属性包括非常规突发事件的发生位置、影响位置,位置信息可以是路名、建筑物名称,也可以具体到 GIS 信息。由于群智感知信息所具有的低质量、碎片化、低价值等特性,例如微博用户可能在较长的时间之后才发布关于某件突发事件的信息,可能在其他地方发布另一个地方发生的突发事件的信息,因此非常规突发事件的时空信息的全面感知旨在通过对社交传感器发布的社交传感器数据进行时间与空间信息的全面感知,对非常规突发事件的发生时间及影响地点进行感知。

4.1 空间信息提取

空间信息刻画了非常规突发事件的发生位置和影响范

围,是检测、分析非常规突发事件的一个重要因素。在微博信息中,空间信息分为两种:位置信息和 GIS 信息。

位置信息为用户在微博中所提到的空间信息。如图 2 所示,微博用户在信息中涉及的位置信息为“淮海中路 358 弄”。



图 2 时空信息举例

对非常规突发事件的位置信息的提取流程如下:

1) 选取有效微博信息,即满足规则 1、规则 2、规则 3 或规则 4 的微博信息,其中规则 1 必须满足,规则 4 必须满足。

2) 利用 openstreetmap¹⁾ 构建位置信息库,库中包含了城市中的所有位置信息的字符串。

3) 利用位置信息库抽取微博信息中含有的位置信息,所有的位置信息集合 W 定义为:

$$W = \{w_1, w_2, \dots, w_{|W|}\} \quad (3)$$

GIS 信息为用户的签到信息,主要为智能手机的位置信息,一般表现为用户智能手机的经纬度坐标信息。如图 2 所示,经纬度信息为 121.46 和 31.22。相较于位置信息,经纬度信息可以更为精准地标记出空间信息。但是经纬度信息是用户所在的实际位置,而不一定是非常规突发事件的发生位置。对非常规突发事件的 GIS 信息的提取流程如下:

1) 选取有效微博信息,即满足规则 1、规则 2、规则 3 或规则 4 的微博信息,其中规则 2、规则 4 必须满足。

2) 利用 html 页面解析,将 GIS 元素从微博信息中解析出来。将提取出的 GIS 信息定义为一个经纬度的二元组。

4.2 时间信息提取

微博信息中包含时间信息,时间信息为用户发表微博的时间,一般来说该时间可以认为是与非常规突发事件相关的时间信息。对于一个非常规突发事件来说,有意义的时间信息包含时间的开始时间、时间演化过程中的重要时间点。例如对于一场火灾,火灾的发生时间需要进行有效的提取。同样地,对火灾具有重要影响的时间点也需要提取,例如救火车出现的时间、居民撤离的时间等,这些代表事件演化特征的时间信息都需要提取。对非常规突发事件的开始时间的提取流程如下:

1) 选取有效微博信息,即满足规则 1、规则 2、规则 3 或规则 4 的微博信息,其中规则 1 和规则 2 中必须满足一个,规则 4 必须满足。

2) 利用 html 页面解析,将时间元素从微博信息中解析出来。

3) 按照时间先后顺序对这些时间点进行排序,其中 l 为

1) www.openstreetmap.com

时间节点的数量,定义为:

$$T = \{t_1, t_2, \dots, t_i\} \quad (4)$$

其中, t_1 为非常规突发事件的开始时间。本文所提取的开始时间只是最早在微博平台发布事件相关信息的时间,相比实际的发生时间可能有一定的延迟。

除了开始时间,还有一些重要的时间点需要提取。通过对实际的例子进行观察及总结,本文列出两条筛选规则来筛选重要的事件演化时间点。

规则 5 重要的事件演化时间点具有新的关键词,例如“消防车”这个关键词第一次出现的时间点可以看作是一个重要的时间点。

规则 6 重要的事件演化时间点具有新的地理位置信息,新的地理位置信息可以看作是事件的影响范围的变化。

对非常规突发事件的重要时间点的提取流程如下:

1) 选取有效微博信息,即满足规则 1、规则 2、规则 3 或规则 4 的微博信息,其中规则 1 和规则 2 中必须满足一个,规则 4 必须满足。

2) 如果某条微博符合规则 5 或规则 6,则将其提取出的时间点标记为重要时间点。

4.3 多媒体信息提取

不同于语义信息与时空信息,多媒体信息能够更为直地地描述非常规突发事件的实时情况。多媒体信息包括图像、视频、音频等。例如微博用户可能会上传正在发生的非常规突发事件的实时图片或视频,这些多媒体信息可以用来帮助决策者或应急管理部门了解实时情况。如图 2 所示,用户通过微博上传了 4 张与一场火灾相关的图片,这些图片可以用来了解这次火灾的发生情况。对非常规突发事件的多媒体信息的提取流程如下:

1) 选取有效微博信息,即满足规则 1、规则 2、规则 3 或规则 4 的微博信息,其中规则 1 和规则 2 中必须满足一个,规则 3 和规则 4 必须满足。

2) 利用 html 页面解析,将多媒体信息从微博信息中解析出来,包括图片、视频、音频等。

5 实例验证

本节将利用实际的非常规突发事件的案例来对提出的基于群智感知的非常规突发事件描述方法进行实例验证。本节所用的实例为 2014 年 5 月 29 号 15:20 分左右在广州发生的火灾。火灾是影响城市安全的典型的非常规突发事件,仅在上海从 2014 年 5 月 16 号到 5 月 26 号就大约有 792 起火灾警报。而作为影响面较广的非常规突发事件,微博用户又往往是第一时间感知到火灾发生的目击者。因此利用微博信息的群智感知,可以较快且较准地感知到实时发生的火灾事件。实例验证的流程如下:

1) 首先选取 5 个关键词作为搜索词来对微博平台进行搜索,这 5 个关键词是“着火,火灾,起火,失火,走火”。搜索时间定为 2014 年 5 月 29 号下午 3:00 至晚上 8:00,搜索地点定为广州。一共有 246 条微博作为搜索结果被反馈。

2) 利用 4 条规则来筛选这些微博,一共有 21 条微博提供了空间信息和多媒体信息,部分微博信息如图 3 所示。



图 3 部分涉及广州火灾的微博

3) 提取语义信息,包括关键词、模式、语句,部分提取的语义信息如表 1 所列。从中可以看出,整个火灾引起了浓烟和异味,产生了较大的影响和破坏。

表 1 部分提取的语义信息

关键词	烟	味	警察	鼻子
模式		烟味		味鼻子
语句	整条中山路 布满浓烟和异味		烟弥漫在空中,还 夹杂着刺鼻的气味	

4) 提取时空信息,包括开始时间、重要时间点、位置信息、多媒体信息,提取的时空图谱如图 4 所示。深色原点所示的位置为火灾的实际发生点。包括开始时间点,一共有 6 个时间点具有新的空间信息或关键词信息,每个时间点的位置都被标注在地图当中。从这张火灾事件的时空图谱中可以看出该火灾的地理影响范围以及重要的时间点。



图 4 广州火灾时间空间示意图

结束语 群智感知是近年来兴起的以物联网、大数据、云计算等多种技术为基础发展起来的。群智感知就是用城市环境中感知的大数据来解决城市本身所面临的挑战,通过对多源异构数据的整合、分析和挖掘来提取知识和价值,从而提高城市服务的现代化水平。非常规突发事件是一类前兆特征不充分、具有明显的复杂性和潜在次生衍生危害、破坏性严重、采用传统管理方式难以应对处置的罕见重大突发事件。由于构成群智感知网络的基本单元是无所不在的移动智能终端用户,这种新型的体系架构无论是在感知及认知的广度、深度,或在构建的成本与速度上,都是传统手段所无法比拟的。本文通过对社交网络信息进行语义感知、时空关联等对非常规突发事件进行描述,并通过实例对方法进行验证。非常规突发事件的语义信息包括关键词、模式、语句等信息。时空信息用来描述非常规突发事件的客观时间属性与空间属性。

(下转第 218 页)

杨鹏,孙波.改进的描述逻辑框架 Rough-SHOIN[J].计算机工程与应用,2012,48(4):23-26

- [13] Pawlak Z. Rough sets [J]. International Journal of Computer and Information Sciences, 1982, 11(8): 341-356
- [14] Yao Yi-yu, Chen Yao-hua. Rough set Approximations in formal concept analysis [C] // Proc of 2004 Annual Meeting of the North American Fuzzy Information Processing Society. 2004: 73-78
- [15] Li Jin-hai, Mei Chang-lin, Xu Wei-hua, et al. Concept learning via granular computing: A cognitive viewpoint [J]. Information Sciences, 2015, 298: 447-467
- [16] Ganter B, Wille R. Formal Concept Analysis [M]. New York: Mathematical Foundation, Springer-Verlag, 1999
- [17] Wille R. Restructuring lattice theory: an approach based on hierarchies of concept [M] // Formal Concept Analysis: Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2009: 314-339
- [18] Li Jin-jin, Li Ke-dian, Wu Yan-hua. The approximation concept of concept lattices [J]. Journal of Nanjing University (Natural Sciences), 2013, 49(2): 244-249 (in Chinese)
李进金, 李克典, 吴燕华. 概念格上的近似概念 [J]. 南京大学学报, 2013, 49(2): 244-249
- [19] Hu Ke-yun, Lu Yu-chang, Shi Chun-yi. Advanced in concept lattice and its application [J]. Journal of Tsinghua University (Sci-

ence and Technology), 2000, 40(9): 77-81 (in Chinese)

胡可云, 陆玉昌, 石纯一. 概念格及其应用进展 [J]. 清华大学学报 (自然科学版), 2000, 40(9): 77-81

- [20] Shao Ming-wen, Liu Min, Zhang Wen-xiu. Set approximations in fuzzy formal concept analysis [J]. Fuzzy Sets and Systems, 2007 (158): 2627-2640
- [21] Shi Lian, Sun Ji-gui. Description Logic Survey [J]. Computer Science, 2006, 33(1): 194-197 (in Chinese)
石莲, 孙吉贵. 描述逻辑综述 [J]. 计算机科学, 2006, 33(1): 194-197
- [22] Yan Hong-can, Yan Hong-tu, Liu Bao-xiang. Research on application of Tableau Algorithm in rough-logical knowledge reasoning [J]. Journal of Guizhou Normal University (Natural Science), 2013, 31(1): 40-43 (in Chinese)
阎红灿, 闫宏图, 刘保相. Tableau 算法在粗逻辑知识推理中的应用 [J]. 贵州师范大学学报 (自然科学版), 2013, 31(1): 40-43
- [23] Mei Jing, Lin Zuo-quan. From ALC to SHOQ(D): A Survey of Tableau Algorithm for Description Logic [J]. Computer Science, 2005, 32(3): 1-11 (in Chinese)
梅婧, 林作铨. 从 ALC 到 SHOQ(D): 描述逻辑及其 Tableau 算法 [J]. 计算机科学, 2005, 32(3): 1-11
- [24] Sattler U. A concept language extended with different kinds of transitive roles [C] // 20th Deutsche Jahrestagung für Künstliche Intelligenz Berlin. Springer-Verlag. 1996, 1137: 333-345

(上接第 213 页)

时间属性包括非常规突发事件的发生时间、结束时间以及重要时间点。空间属性包括非常规突发事件的发生位置、影响位置, 位置信息可以是路名、建筑物名称, 也可以具体到 GIS 信息。多媒体信息包括图像、视频、音频等。

参 考 文 献

- [1] 刘云浩. 群智感知计算 [J]. 中国计算机学会通讯, 2012, 8(10): 38-42
- [2] Zheng Y. Tutorial on Location-Based Social Networks [C] // Proceedings of the 21st International World Wide Web Conference. 2012
- [3] Fan Wei-cheng. Advisement and Suggestion to Scientific Problems of Emergency for Public Incidents [J]. Bulletin of National Natural Science Foundation of China, 2007, 21(2) (in Chinese)
范维澄. 国家突发公共事件应急管理科学问题的思考和建议 [J]. 中国科学基金, 2007, 21(2)
- [4] 范维澄. 我国应急平台建设现状分析 [OL]. <http://news.tsinghua.edu.cn>
- [5] Biagioni J, Gerlich T, Merrifield T, et al. EasyTracker: automatic transit tracking, mapping, and arrival time prediction using smartphones [M] // SenSys'11. 2011
- [6] Yan Bo, Chen Guan-ling. AppJoy: personalized mobile application discovery [C] // Proc. ACM MOBISYS. 2011
- [7] Yang Zheng, Wu Chen-shu, Liu Yun-hao. Locating in Fingerprint Space: Wireless Indoor Localization with Little Human Intervention [C] // ACM MobiCom 2012. Istanbul, Turkey, 2012: 22-26
- [8] Zhou Peng-fei, Zheng Yuan-qing, Li Mo. How Long to Wait?: Predicting Bus Arrival Time with Mobile Phone based Participatory Sensing [C] // Proceedings of ACM MobiSys. 2012
- [9] Wu Chen-shu, Yang Zheng, Liu Yun-hao, et al. WILL: Wireless Indoor Localization without Site Survey [J]. IEEE Transactions on Parallel and Distributed Systems (TPDS), 2013, 24(4): 839-848
- [10] Zhao Dong, Ma Hua-dong, et al. COUPON: A Cooperative Framework for Building Sensing Maps in Mobile Opportunistic Networks [J]. IEEE Trans. Parallel Distrib. Syst., 2015, 26(2): 392-402
- [11] Guo Zhen, Zhang Zhong-fei, Zhu Sheng-huo, et al. A Two-Level Topic Model Towards Knowledge Discovery from Citation Networks [J]. IEEE Trans. Knowl. Data Eng., 2014, 26(4): 780-794
- [12] Guo Bin, Yu Zhi-wen, Zhang Da-qing, et al. Toward a Group-Aware Smartphone Sensing System [J]. IEEE Pervasive Computing, 2014, 13(4): 80-88
- [13] Han Jia-wei, Pei Jian. Mining Frequent Patterns by Pattern-Growth: Methodology and Implications [J]. SIGKDD Explorations, 2000, 2(2): 14-20
- [14] Crooks A, Croitoru A, Stefanidis A, et al. Earthquake: Twitter as a Distributed Sensor System [J]. Transaction in GIS, 2013, 17(1): 124-147
- [15] Longueville B, Smith R, Luraschi G. OMG, from here I can see the flames, a use case of mining location based social networks to acquire spatio-temporal data on forest fires [C] // Proceedings of the International Workshop on Location-Based Social Networks, 2009: 73-80
- [16] Liu Y, Alexandrova T, Nakajima T. Using Stranger as Sensors: Temporal and Geo-sensitive Question Answering via Social Media [C] // Proceedings of the 22th International World Wide Web Conference. 2013: 803-813
- [17] Qu Y, Zhang J. Trade Area Analysis using User Generated Mobile Location Data [C] // Proceedings of the 22th International World Wide Web Conference. 2013: 1053-1063