

基于社交网络的群体性事件挖掘和预测

黄细凤, 廖泓舟

(中国电子科技集团公司第十研究所, 四川 成都 610036)

摘要: 文本事件挖掘旨在通过挖掘文本来实现结构化的事件表示, 从而支撑进一步的事件分析和预测工作。文本事件挖掘所需要挖掘的结构化事件信息包括事件类型、参与者、触发词、时间、地点等, 其中每一项信息的挖掘都是一个单独的文本分类或者是信息抽取任务。所以, 文本事件挖掘是一项综合性的自然语言处理任务, 具有较高的工程复杂性。社交网络群体性事件挖掘是指针对社交网络这一特定的信息源, 以及群体性事件这一特定的事件类型所开展的文本事件挖掘工作。由于社交网络和群体性事件的特殊性, 发现针对社交网络群体性事件的挖掘结果可以作为事件预测的直接线索, 因此, 实现了一个基于社交网络的群体性事件挖掘系统。在该系统中, 实现了对文本信息事件发现和分类、参与者抽取、行为抽取、地点抽取和时间抽取等子任务, 共同组成完整的事件要素结构。同时, 根据事件时间信息进行事件预测, 并与实际事件发生情况进行对比以评测事件生成效果和预测准确率。

关键词: 社交网络; 事件挖掘; 群体性事件; 事件预测; 文本挖掘

中图分类号: TP391.1

文献标识码: A

文章编号: 1673-629X(2022)06-0039-06

doi: 10.3969/j.issn.1673-629X.2022.06.007

Unrest Event Mining and Prediction Based on Social Network

HUANG Xi-feng, LIAO Hong-zhou

(The 10th Research Institute of China Electronics Technology
Group Corporation, Chengdu 610036, China)

Abstract: Text event mining aims to realize structured event representation through text mining, so as to support further event analysis and prediction. The structured event information to be mined for text event mining includes event type, participant, trigger word, time and place, etc., among which each information mining is a separate text classification or information extraction task. Therefore, text event mining is a comprehensive natural language processing task with high engineering complexity. Social network group incident mining is a textual event mining work carried out by pointer to the specific information source of social network and the specific event type of group incident. Due to the particularity of social networks and group incidents, we found that the mining results of social network group incidents can be used as the direct clues of event prediction. Therefore, we have implemented a group incident mining system based on social networks. In this system, we realize sub-tasks such as text information event discovery and classification, participant extraction, behavior extraction, location extraction and time extraction, which together constitute a complete event element structure. At the same time, we make event prediction based on the event time information and compare with the actual event occurrence to evaluate the event generation effect and prediction accuracy.

Key words: social network; event mining; unrest events; event prediction; text mining

0 引言

群体性事件^[1]是指由某些社会矛盾引发, 特定群体通过规模性聚集、对社会造成负面影响的群体活动, 因而对社会秩序和社会稳定造成重大负面影响的各种事件。由于群体性事件大多以现有公共次序的对立面所出现, 所以通常无法利用传统的新闻媒介公开进行大规模的组织和传播, 其传播的隐蔽性使得提前预知

相关信息十分困难。近年来随着网络社交媒体的广泛应用, 群体性事件的组织策划者得以利用其强大的信息传播能力和组织动员能力推动群体性事件的发生。这一方面极大地推动了群体性事件在全球许多地区愈演愈烈, 而另一方面, 社交媒体的公开性也使得获取群体性事件的相关线索, 甚至进行提前预测成为了可能。

该文从自然语言处理的角度研究了针对社交网络

收稿日期: 2021-01-18

修回日期: 2021-05-20

基金项目: 国家自然科学基金(62001437)

作者简介: 黄细凤(1983-), 女, 高级工程师, 博士, 从事人工智能和自然语言处理方面的研究。

文本的事件挖掘任务。事件挖掘旨在从文本中挖掘出结构化的事件信息,包括是否包含事件、事件类型、事件参与者、事件发生的时间和地点等。所有这些信息被称为事件要素。通过标准化之后的事件要素可以将事件信息以结构化的形式进行关联,形成事件知识图谱,从而为下一步利用知识图谱相关技术对事件分析处理提供基础。

笔者在对事件挖掘的研究过程中发现,社交网络上所挖掘出的事件,有相当一部分其发布时间会早于事件的实际发生时间,例如在图 1 中所展示出的从 Twitter 中挖掘到的事件信息。这一现象也很容易解释:社交网络作为群体性事件的重要组织工具,理所当然地会发布大量的事件召集信息。也就是说,社交网络上的信息也许可以作为群体性事件预测的指示信息。基于这一观察结果,重点研究利用社交网络挖掘得到的结构化事件信息作为精确事件预测指标的可行性。

首先给出基于社交网络事件挖掘的完整流程,逐一详细描述所采用的处理步骤和方法,包括事件预处理、事件要素抽取、事件要素信息标准化和事件归并等具体流程;接下来会详细测试事件挖掘的整体流程以及各个步骤的性能,给出定量的评测结果;最后探讨基于所挖掘的结构化事件信息进行群体性事件预测的效果,并针对香港地区近年来多发的群体性事件,给出了具体的预测结果;之后讨论与该文相关的其他事件挖掘和预测的最新研究工作,与该文研究工作进行对比

探讨;最终给出结论和未来研究方向。

1 相关研究

事件库是以结构化事件的方式对历史信息进行收集和记录的一种重要工具,事件库的构建需要从粗糙的非结构化数据批量生成结构化的事件数据。相关的学术研究和实际应用经历了手动集成到大规模自动提取和集成的发展过程。Correlates of War event data (COW)^[2]是最早的互联网事件库,采用手动的方式收集和整理了一系列国家间军事争端事件。Global Terrorism Database (GTD)^[3]收集了全球 15 万条恐怖事件的完整数据,是垂直领域上人工构建事件库的优秀成果。Integrated Conflict Early Warning System (ICEWS)^[4]是第一个有着全球影响力的自动事件数据库,它也是针对暴力冲突类事件,收集了超过 2.5 亿条事件数据,并免费向公众开放。由 Google 支持的 Global Dataset of Events, Location, and Tone (GDELT)把这些工作提升到一个新的高度,它监控了全球超过 1.5 万个新闻源,以 15 分钟更新一次的方式实时地向公众发布事件库信息。除此之外,针对特定领域和地区的事件库还包括 News Reader、Event Registry、EMBERS 等。这些现有的事件库大多以新闻或者专业文档(如维基百科)作为事件源,但是目前还没有发现以社交网络作为主要信息源的事件数据库。Tweet 事件挖掘示例如图 1 所示。

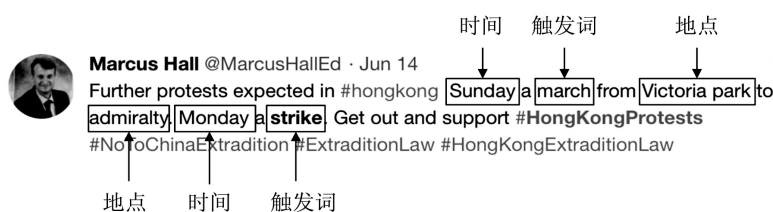


图 1 Tweet 事件挖掘示例

结构化的事件抽取是事件数据库自动化构建的核心技术,可以分为数据驱动和知识驱动两大基本途径。数据驱动的事件抽取采用统计机器学习的方法,通过“标注—学习—抽取”的范式来开展事件抽取工作。与之对应,知识驱动的事件抽取依赖于专家知识,通常采用“模式编写—模式匹配”的方式开展工作。早期的结构化事件抽取工作大多采用知识驱动的技术途径,利用词汇句法模式(Lexical Syntactic Patterns)的匹配来抽取事件^[5]。由于其高性能和有效性,在之后的各个事件抽取工作中也大都作为其中的一个处理步骤^[6]。文中系统中的事件发现部分就是采用这种传统而行之有效的方法。由于早期统计模型的计算能力不足,数据驱动的事件抽取的准确率大多较低,但是其优

势在于领域的普适性。近年来随着深度神经网络的大发展,数据驱动的事件抽取逐渐成为了主流,CNN、RNN、LSTM、Transformer 等最新的神经网络架构被应用到事件抽取中,极大地提高了事件抽取的准确性。但是在面对大规模数据的时候,神经网络方法大多存在计算开销过大的劣势,需要与传统方法融合使用。

事件预测是一个还有待研究人员进一步深入探索的领域。随着事件库构建技术的逐渐成熟,已经有不少研究者将下一步的研究目标瞄准了事件预测。这类研究大多以统计模型为基础,通过大量地提取事件的各种潜在特征和前期征兆来构建复杂的神经网络模型,从而实现事件预测^[7-17]。但是这些方法准确率和召回率都极低,不存在实际应用的可能性。这是由于

事件预测这一任务本身所具备的复杂性决定的。与前述的这些方法不同,该文瞄准群体性事件这一特定的事件类型,采用社交网络这一特殊的数据源,用文本挖掘的方式较好地实现了事件预测这一困难任务。

2 事件挖掘与预测

事件挖掘指从已获取的社交网络文本语料中挖掘历史事件,识别事件类型,提取事件要素,归并不同来源的结构化事件表示,最终得到事件库。事件库是一个大规模的结构化知识库,可以导出到知识图谱系统进行可视化查询,检索,分析和浏览。

事件挖掘系统的输入为非结构化的文本信息,输

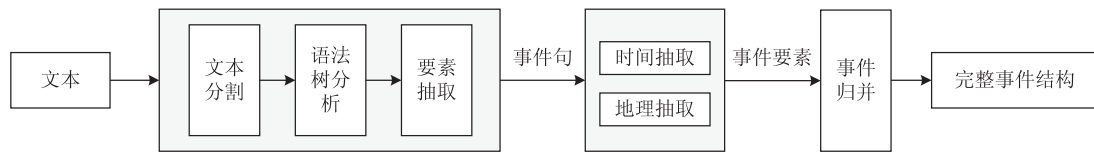


图2 事件挖掘系统工作流程

2.1 文本预处理

在进行事件要素抽取和事件类型判断之前,需要对文本进行预处理,对语料进行切割,以便对句子级别的文本进行事件分析。要素抽取的过程依赖于语法树的建立,使用基于 StanfordNLP^[19]的短语分析工具对句子进行预处理,生成树状结构以对句子中的句法关系进行描述。将语法关系统一表示为类型依赖关系。语法解析以一个完整的句子作为输入,以一颗对应的语法解析树作为输出。使用统计的方法生成语法分析树,如概率上下文无关语法(Probabilistic Context-Free Grammar, PCFG)。

PCFG是一个监督学习方法,需要对PCFG的参数进行估计。其训练数据为一个句子及其解析树,通常把这样的标注数据称为树库(Tree Bank)。PCFG的参数就是某种规则的概率,当给定一个Tree Bank后,统计相应的规则转移的概率。StanfordNLP利用PCFG对输入句子进行分析,分析结果包含句子中单词的基本形式、依存关系、词性、命名实体等。在事件挖掘过程中,依赖其词性标注方法生成词性语法树,并对语法树进行建模分析。

2.2 人物要素抽取与事件类型判断

在文本预处理得到解析树后,对解析树进行触发词判断,发现事件要素和事件类型。

本模块以句子及其解析树作为输入,以三元组(Actor1, Actor2, Type)作为输出。解析核心基于触发词,在进行事件识别时,本模块将解析树的核心动词与事件触发词表进行匹配,从而实现事件的发现以及类型识别。

人物要素抽取与事件类型判断的核心逻辑基于对

出为完整的事件要素五元组 Event: (Actor1, Actor2, Type, Time, Location)。这种五元组的事件表示是国际通用的表示框架,并且在CAMEO (Conflict and Mediation Event Observations)^[18]中得到了明确定义。CAMEO是一个事件编码框架,是为了研究国际争端中的重大事件而优化,同时也包括政治,经济,民生领域的几乎所有重要事件类型。其中的141和144类型即为该文所关注的群体性事件类型。该文的事件挖掘系统将会遵循CAMEO的事件框架定义。

实现事件挖掘的整体流程如图2所示,其核心步骤包括:文本预处理、事件类型判断、要素抽取、地理位置抽取及标准化、时间抽取及标准化等。

语法树的解析。其接收句子的句法分析树作为输入,并且将分析树存储为链接节点的树形结构,其中每个节点都是一个短语。语法树在数据结构中是一个简单的树形结构,每个短语或单词都是其自己的节点,并带有指向父节点的指针以及子节点集。每个句子都由多个部分组成,一个组成部分可以是单个单词或整个短语(构成要素的组成部分),每个组成部分都有特定的句法角色。对语法树进行迭代处理并将树中的单词类型转化为短语类型,单词被组合在一起直到形成一个名词或动词短语。动词短语包含了当前句子的大部分含义,并解释了句子中相关名词以及动词的交互过程。若短语可以产生符合CAMEO触发词字典的正确交互关系,则将该句判断为事件句,并输出事件基础三元组。

句法树的递归过程,可以抽象为以下步骤:

- (1)使用 Parse 类将句法解析读取到内存中;
- (2)识别名词短语中的参与者编码;
- (3)根据字典确定动词短语;
- (4)识别动词如何与其他词性短语进行交互;
- (5)使用 CAMEO 词库进行判断;
- (6)返回事件编码和参与者编码。

2.3 地点要素抽取

在得到事件句及其所对应的要素三元组(Actor1, Actor2, Type)后,需要进行地点要素抽取。由于文本中地理位置信息描述的多样性,该文采用三段式的处理方式:首先对文本中所有的地理位置信息进行抽取;然后对所抽取的地理信息进行分类,判断其是否描述了目标事件的地理位置;最后采用 GeoNames^[20]数据库进行地理位置信息标准化。

(1) 地理位置信息抽取。

使用 spaCy 命名实体识别框架进行地理位置识别,抽取事件地点,选择事件句实体抽取结果中所有标注为 GEO 的实体作为地理位置要素候选集。

(2) 事件地理位置信息识别。

地理位置信息抽取后,针对地理位置候选集,选择最可能的地理位置作为事件地理要素。使用在 Keras 中实现的分类神经网络,并接受了带有 Prodigy 标签的新注释数据进行训练,以推断每个地名的正确国家和地名词典条目。

(3) 地理位置信息标准化。

在得到事件的地理位置信息后,使用 GeoNames 库提取地名的坐标和位置代码。根据事件地理要素,使用 Elasticsearch 快速检索 GeoNames 地理数据库得到地理位置信息和编码。

2.4 时间要素抽取

事件的时间要素抽取的流程与地点要素抽取的流程一致,也存在着抽取、识别和标准化三个阶段,其中抽取和识别的阶段处理流程与地点要素抽取是一致的,这里不再赘述。时间要素抽取的难点在于时间要素的标准化,由于时间表达方式更加多样化,因此不能采用查表的方式进行时间要素的标准化。

时间表达式归一化的目标是计算机能够理解它们得到的时间信息,比如,在一个文本里出现 1962.3.1,而另一篇出现的是 1962/3/1 或 3/1/1962,计算机应该能够知道它们说的都是同一时间。

在庞大的互联网上,时间信息的表示是多种多样的,导致对文本信息的理解非常困难,如果将时间信息统一到同一种表示方式,将会对机器的自动内容识别有很大帮助,同时对信息检索、文摘生成等领域都发挥重要作用。具体的,在时间抽取任务中,使用 SUTIME 库^[21]实现事件时间抽取,产生事件时间的归一化表达。

2.5 事件归并

事件归并是指将生成的对所抽取的多个事件进行归并。同一事件在不同数据来源会获取多个不同角度的事件描述,事件归并将多个不同来源的同一事件的不同描述归并为一个统一的事件描述,从而避免系统的二义性,提高事件挖掘和预测的准确率。

事件归并依赖于事件要素抽取结果。一个事件可以描述为一个五元组 (Actor1, Actor2, Type, Time, Location)。不同来源的新闻可能对同一事件有不同的描述,事件描述元组相同的事件为同一事件。若不同来源的新闻描述了同一事件,引入事件提及数作为考量,表示同一事件在不同新闻源的提及数,作为衡量事件重要程度的指标。

3 实验结果与分析

3.1 实验数据

为了分析通过社交网络进行事件挖掘的有效性,特别是其中的群体性事件的可挖掘性和可预测性,爬取中国香港地区,从 2019 年 1 月 1 日到 2020 年 6 月 30 日的所有 Twitter 推文,共计 370 万条文本数据,然后对其进行了基本的过滤,仅选取其中的英文数据作为实验数据,共计 1 302 801 条。需要指出的是,由于访问量限制过于苛刻,并没有采用 Twitter 所提供的 API,而是利用了搜索接口进行数据获取。不同的数据获取方式所得的数据量可能会有一定的出入,但是作为验证性实验数据笔者认为是足够的。

为了验证社交网络时间挖掘的真实性,使用全球语言数据库^[22] (Global Database of Events, Language, and Tone, GDELT) 作为时间挖掘的对比结果。GDELT 是一个最初由乔治敦大学建立,并被 Google 资助和运行的全球最大的事件数据库。它通过监控全球超过 100 种语言的上万个新闻源来进行事件获取,分析并得到结构化的事件信息。和该系统一样, GDELT 也遵循 CAMEO 事件框架,所以两个系统的事件挖掘结果能够相互印证。虽然 GDELT 也存在着事件遗漏的情况,但仍然足以给该系统提供一个可行的评估依据。

3.2 事件挖掘实验

(1) 评测指标。

在所收集的 130 万条的实验数据之上开展事件挖掘实验,通过准确率、召回率和 F 值三个指标对该系统进行评价。具体如下:

准确率是指系统挖掘到事件的准确性。这里的判断依据是挖掘结果是否符合文本信息,以及是否符合真实情况,称之为字面准确率,表示为 $Precision_{literally}$ 。由于缺乏一个标准事件数据集,所以这里采用 GDELT 作为参考依据,这种准确率也可以称为 GDELT 准确率,表示为 $Precision_{GDELT}$ 。对于字面准确率,采用抽样人工评价的方式进行测试,随机抽取 1 000 条系统挖掘的事件结果,人工对比源文本,判断挖掘结果是否准确。对于 GDELT 准确率,则是比较该系统所挖掘的结果在 GDELT 数据库中是否存在,如果存在则认为准确的。

召回率是指实际发生的事件中被系统挖掘到的数量,同样分为字面召回率 $Recall_{literally}$ 和 GDELT 召回率 $Recall_{GDELT}$ 。字面召回率的统计也是人工进行,随机抽取测试数据中的语料,抽取 1 000 条事件数据,与系统挖掘的结果进行对比判断是否被系统召回。对于 GDELT 召回率,则是比较 GDELT 数据库中的特定地点和特定类型的事件是否被该系统所挖掘到,如果被

挖掘到则认为是成功召回的。

F 值是准确率和召回率的调和平均值,其计算公式为:

$$F - score = \frac{(1 - \beta^2) precision \times recall}{\beta^2 \cdot (precision + recall)}$$

该文取 $\beta = 1$,即为 F-1 值。

(2)实验结果和分析。

表 1 给出了针对上述实验数据的测试结果,其中

表 1 测试结果

事件	Events	P _{literal}	R _{literal}	F _{literal}	P _{GDELT}	R _{GDELT}	F _{GDELT}
所有事件	1 532	0.78	0.82	0.80	0.76	0.01	0.02
群体性事件	932	0.81	0.85	0.83	0.80	0.58	0.67

事件挖掘结果同样保持了不错的准确率,基本能达到预期设想。就字面准确率而言,该系统的事件挖掘子系统准确率能达到约 80%。对于 GDELT 准确率而言,相比字面准确率都有了一定程度的下降,但是差距并不明显。这两种之间的比值可以认为是 GDELT 事件收录的召回率,这一定程度上说明 GDELT 事件集的召回率相当的高。

该系统事件挖掘的字面召回率也保存了不错的水平,达到了 80% 以上。但是,GDELT 召回率则相当的低。这是由于 GDELT 监控了大量的新闻源,能够获取到非常丰富的事件信息,其信息丰富程度远高于社交网络。另一方面,社交网络作为事件挖掘的信息源能够达到约 58% 的召回率,这意味着在群体性事件挖掘方面可以在一定程度上作为 GDELT 的补充。

3.3 事件预测实验

(1)评价指标。

采用准确率和召回率作为系统的评价指标,同样采用 GDELT 事件数据库作为评价标准。计算预测事件在 GDELT 中是否会发生,记为 Precision_{GDELT};以及 GDELT 中的事件是否会被社交网络预测到,记为 Recall_{GDELT}。

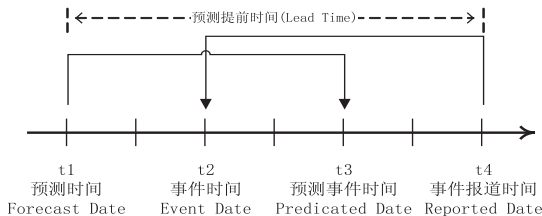


图 3 事件预测提前期计算方法说明

对于事件预测而言,预测提前期(Lead Time)是一个重要的指标。所谓预测提前期是指发出事件预测的日期和事件被报道的日期之差。这里的“发出事件预测的日期”就是该系统通过社交网络信息发现事件的日期;“事件被报道的日期”采用 GDELT 中事件的报到日期为依据。两种相减得到事件预测的预测提前

Events 列给出了挖掘到的事件数量,为 1 532 件。这说明基于社交网络的事件挖掘是非常有效的,确实能够挖掘得到大量的事件。特别要注意的是,这里的事件不仅仅是群体性事件,但是其中的群体性事件数量占有较大比例。在所有的 GDELT 事件中,群体性事件(对应于 CAMEO 的 141 和 144 类型)仅仅占比大约 2% 左右,而社交网络挖掘得到的事件有一大半都是群体性事件。

期,如图 3 所示。

(2)实验结果和分析。

表 2 给出了针对香港地区的社交网络数据群体性事件预测实验结果。

表 2 针对香港地区的社交网络数据群体性事件预测实验结果

	Events	P _{GDELT}	R _{GDELT}	F _{GDELT}
LT=1	385	0.83	0.45	0.62
LT=2...5	301	0.77	0.38	0.52
LT=6...10	68	0.39	0.09	0.15
LT>10	25	0.20	0.03	0.05

其中第一列 LT = n 表明的是不同的事件预测提前期,LT = 1 即是提取一天预测,此时的预测的事件数量最多,准确率和召回率也都较高。随着 LT 数值的增加,能够预测到的事件数量降低,一定程度上意味着大量的群体性事件都并没有经过长时间的准备,而仅仅是在 2 ~ 5 天内短期召集形成的。

根据准确率和召回率进行分析,可以看到基于社交网络的群体性事件预测准确率较高,也是符合预期的:社交网络线索相对明确,一旦出现了社交网络上的群体性事件召集信息,就会有比较大的概率真实发生该事件。另一方面,事件预测召回率在 LT = 1 时达到了 0.45,这意味着有接近一半的群体性事件都可以在社交网络上找到明确线索。还有一点值得注意的是,对于 LT > 10 的情况,准确率和召回率都较低,尚不足以作为可靠的预测线索。

4 结束语

社交网络的群体性事件挖掘是一个自动化的信息抽取任务。该文设计并实现了对社交网络信息进行事件挖掘和事件预测的系统。其中包括事件发现、事件要素抽取、事件分类、地理位置抽取、事件归并等子任务。基于所挖掘到的社交网络群体性事件相对于事件

发生日期有一定提前性的这一发现,探索了将所挖掘到的结构化事件信息用于事件预测的可能性,实验证明这一探索有较好的效果,能够以较高的准确率和召回率提前预测特定地区的群体性事件。

参考文献:

- [1] 于建嵘. 当前我国群体性事件的主要类型及其基本特征[J]. 中国政法大学学报, 2009, 6(11): 114-120.
- [2] DIEHL P F. The correlates of war project: a bibliographic essay, 1961-1991 [J]. International Studies Notes, 1992, 17(2): 21-33.
- [3] LAFREE G. The global terrorism database (GTD) accomplishments and challenges [J]. Perspectives on Terrorism, 2010, 4(1): 24-46.
- [4] BOSCHEE E, LAUTENSCHLAGER J, O' BRIEN S, et al. Integrated crisis early warning system (ICEWS) coded event data [R]. Bethesda: Lockheed Martin, 2015.
- [5] YAKUSHIJI A, TATEISI Y, MIYAO Y, et al. Event extraction from biomedical papers using a full parser [C]//Bio-computing 2001. Hawaii: World Scientific, 2000: 408-419.
- [6] HUNG Sheng-Hao, LIN Chia-Hung, HONG Jen-Shin. Web mining for event-based commonsense knowledge using lexico-syntactic pattern matching and semantic role labeling [J]. Expert Systems with Applications, 2010, 37(1): 341-347.
- [7] QIAO Fengcai, LI Pei, ZHANG Xin, et al. Predicting social unrest events with hidden Markov models using GDELT [J]. Discrete Dynamics in Nature and Society, 2017, 2017: 1-13.
- [8] GALLA D, BURKE J. Predicting social unrest using GDELT [C]//International conference on machine learning and data mining in pattern recognition. New York, NY, USA: Springer, 2018: 103-116.
- [9] WANG Xiuling, CHEN Hao, LI Zhoujun, et al. Unrest news amount prediction with context-aware attention LSTM [C]//Pacific rim international conference on artificial intelligence. Nanjing: Springer, 2018: 369-377.
- [10] PARRISH N H, BUCZAK A L, ZOOK J T, et al. Crystal cube: multidisciplinary approach to disruptive events prediction [C]//International conference on applied human factors and ergonomics. Orlando, Florida, USA: Springer, 2018: 571-581.
- [11] ISLAM M K, AHMED M M, ZAMLI K Z, et al. An online framework for civil unrest prediction using tweet stream based on tweet weight and event diffusion [J]. Journal of Information and Communication Technology, 2019, 19(1): 65-101.
- [12] YUE Ning, ZHAO Liang, CHEN Feng, et al. Spatio-temporal event forecasting and precursor identification [C]//25th ACM SIGKDD international conference on knowledge discovery & data mining. New York: Association for Computing Machinery, 2019: 3237-3238.
- [13] 王莲花, 李筱烨, 高楠. 基于灰色理论的恐怖袭击事件预测与分析 [J]. 数学的实践与认识, 2019, 49(12): 174-182.
- [14] 李良. 突发事件微博舆情的话题发现和热度预测研究 [D]. 西安: 西安理工大学, 2018.
- [15] 王征, 杨茜. 基于情境推演的微博突发事件预测模型研究 [J]. 情报学报, 2017, 36(3): 267-273.
- [16] 陈可, 黄燕霞, 徐选华. 基于突发事件相似度和区间一致性的复杂偏好大群体应急决策方法 [J]. 控制与决策, 2020, 35(9): 2215-2224.
- [17] 张维超. 边境群体性突发事件中的情报获取工作研究 [J]. 武警学院学报, 2020, 36(3): 14-19.
- [18] GERNER D J, JABR R, SCHRODT P A. Conflict and mediation event observations (CAMEO): a new event data framework for the analysis of foreign policy interactions [R]. Storrs Mansfield: International Studies Association, 2002.
- [19] MANNING C D, SURDEANU M, BAUER J, et al. The Stanford CoreNLP natural language processing toolkit [C]//52nd annual meeting of the association for computational linguistics: system demonstrations. Baltimore: Association for Computational Linguistics, 2014: 55-60.
- [20] MARC W, BOUTREUX C. GeoNames [R]. Wollerau, Switzerland: GeoNames Geographical Database, 2011.
- [21] CHANG A X, MANNING C D. SUTIME: a library for recognizing and normalizing time expressions [C]//The eighth international conference on language resources and evaluation. Luxembourg: European Language Resources Association, 2012: 3735-3740.
- [22] LEETARU K, SCHRODT P A. GDELT: global data on events, location, and tone [C]//The 54th ISA annual convention. Storrs Mansfield: International Studies Association, 2013: 673-712.