

# 基于社交媒体大数据的灾害事件态势感知

龚万渊, 王慧颖, 江信禧, 周绮凤

厦门大学航空航天学院自动化系, 福建 厦门 361102

## 摘要

灾害事件态势的准确感知取决于及时、有效地获取承载事件信息的相关数据以及对数据的深入理解和分析。社交媒体大数据蕴含了丰富的事件信息, 但其海量、非结构化、时空敏感等特点为动态复杂的灾害事件态势感知带来巨大的挑战。从社交媒体大数据的角度出发, 首先, 通过构建灾害事件的因果知识图谱, 有效整合社交媒体大数据中的异构信息, 解决其非结构化和时空敏感问题; 其次, 利用大型语言模型及微调技术, 提升对灾害事件演变过程的推理能力, 并通过微调后的生成式预训练模型, 更准确地识别具有针对性和实用性的灾害事件的因果子事件, 有效应对数据海量和信息冗余带来的挑战; 最后, 设计了一个灾害事件态势感知系统, 通过用户与系统的交互, 辅助相关人员快速、全面地理解和分析灾害事件情况。实验结果表明, 该系统在灾害事件相关文本分类任务中, 平均F1分数达到0.891, 显著优于基线模型。在因果关系生成方面, 微调后的生成式预训练模型能够更准确地识别具有针对性和实用性的灾害事件的因果子事件, 有效提升了灾害事件态势感知的准确性和效率。

## 关键词

社交媒体大数据; 因果事件; 灾害信息管理系统

中图分类号: TP181

文献标志码: A

doi:10.11959/j.issn.2096-0271.2025080

## *Research on disaster event situational awareness based on social media big data*

Gong Wanyuan, Wang Huiying, Jiang Xinxi, Zhou Qifeng

Department of Automation, School of Aerospace Engineering, Xiamen University, Xiamen 361102, China

## *Abstract*

Accurate perception of disaster event situational relies on the timely and effective acquisition of relevant data carrying event information, as well as in-depth understanding and analysis of such data. Social media big data contains a wealth of event information. However, its characteristics of being voluminous, unstructured, and spatiotemporally sensitive pose significant challenges for the dynamic and complex awareness of disaster event situational. From the perspective of social media big data, we firstly constructed a causal knowledge graph of disaster events to effectively integrate heterogeneous information from social media big data, addressing the issues of unstructured data and spatiotemporal sensitivity. Secondly, we leveraged large language models and fine-tuning techniques to enhance the reasoning capability of disaster event evolution processes. Moreover, through the fine-tuned generative pre-trained model, we could more accurately identify causal sub-events of disaster events that were targeted and practical, effectively

addressing the challenges brought by the large volume of data and information redundancy. Finally, a disaster event awareness system was designed to assist relevant personnel in quickly and comprehensively understanding and analyzing disaster situations through user-system interaction. Experimental results show that the system achieves an average F1 score of 0.891 in disaster event-related text classification tasks, significantly outperforming baseline models. In terms of causal relationship generation, the fine-tuned generative pre-trained model can more accurately identify targeted and practical causal sub-events of disaster events, effectively improving the accuracy and efficiency of disaster event situational awareness.

### *Key words*

social media big data, causal event, disaster information management system

## 0 引言

灾害事件是人与自然或社会矛盾冲突的一种表现形式，具有较强的破坏作用。数据作为描述事件本身及其与相关事件在时间、空间上发展变化的有效载体，蕴含了十分有价值的信息。如果人们能够利用这些数据优化灾害事件信息的表达，清晰地刻画灾害事件的发展脉络，并进一步推理灾害事件的发展逻辑，对提升灾害事件态势感知能力、辅助人类应对突发的自然灾害行为具有重要意义。

近年来，随着信息技术的飞速发展，社交媒体作为发布灾害事件信息的主要平台，承载了大量灾害事件的文本信息。当灾害事件发生时，微博、新浪网等媒体平台都会及时发布很多与灾害事件相关的新闻和报道。这些新闻数据蕴含了反映灾害事件演化及其影响的相关信息。研究基于社交媒体大数据的灾害事件分析技术，通过个人、社区、企业和政府紧密协作，收集灾情数据，及时共享灾情信息，从而合理协调社会资源，可以帮助人们实现灾前充分准备、灾中敏捷响应以及灾后快速恢复。社交媒体灾害事件大数据在数据产生、

事件演化等方面存在独特的性质。首先，灾害事件的发生往往具有突发性、动态性和不确定性等特点<sup>[1]</sup>。通常情况下，灾害事件信息不会大量存在，其在灾害临近、发生过程中，以及发生后的短时期内爆发式地增多。其次，灾害发展过程中产生的数据具有海量、多源、异构、时空敏感等特点<sup>[2]</sup>。再次，灾害事件及其次生事件之间是一个复杂系统，且每一个灾害事件的演化都有其独特性，缺乏历史可学习数据和经验知识。最后，重大灾害事件因其后果十分严重，往往会引起人们的广泛关注，从而产生不良的社会影响，具有广泛的社会性特征。

社交媒体大数据的复杂性使灾害事件态势感知面临诸多挑战。例如，海量数据中包含大量与灾害事件无关的信息，导致信息过载；非结构化数据形式增加了信息提取的难度；时空敏感性则要求灾害事件感知系统能够快速响应并准确反映灾害事件的动态变化。为了应对上述挑战，本文提出了一种基于社交媒体大数据的灾害事件态势感知系统。该系统通过构建因果知识图谱，整合灾害事件相关信息，并利用大型语言模型及其微调技术，实现了对灾害事件演变过程的高效推理。此外，所提系统还通过用户交互功能，进一步提升了

用户对复杂灾害事件的理解能力，为决策者提供更及时、全面的灾害情报支持。

## 1 社交媒体大数据分析技术在灾害事件态势感知中的应用

### 1.1 社交媒体大数据在灾害管理中的应用

随着社交媒体的普及，其在灾害管理中的作用日益凸显。在灾害发生时，社交媒体平台成为公众获取和分享信息的重要渠道。研究表明，社交媒体上的实时信息对于提高灾害预警的准确性、加快应急响应速度以及优化资源分配具有显著影响<sup>[3-4]</sup>。例如，通过分析社交媒体上的帖子，赈灾决策团队可以快速识别受灾区域和受灾人群的需求，从而为救援团队提供宝贵的现场信息。此外，社交媒体大数据还可以用来监测灾后受灾区域恢复进程和评估灾情救援效果，为政策制定者提供决策支持等。

### 1.2 因果关系提取技术的发展现状

因果关系提取是自然语言处理领域的重要研究方向，旨在从文本中识别出事件之间的因果联系。在灾害管理领域，因果关系提取技术可以帮助人们分析灾害的发展过程、理解灾害发生的原因，以及预测可能的后果。目前，因果关系提取方法主要包括基于规则的方法<sup>[5-6]</sup>、机器学习方法<sup>[7-8]</sup>和深度学习方法<sup>[9-10]</sup>。基于规则的方法依赖于预定义的模式和语法结构，机器学习方法和深度学习方法通过训练数据学习因果关系的表示。最新的研究趋势是利用大规模语料库和复杂的神经网络模型<sup>[11-12]</sup>来提高因果关系提取的准确性和鲁

棒性。例如，Li等<sup>[7]</sup>提出了一种基于自注意力机制的因果关系提取模型，能够有效处理文本中的复杂因果关系；Khetan等<sup>[11]</sup>提出了一种基于因果BERT（bidirectional encoder representation from transformer）的模型，专用于文本中事件因果关系的检测；Liu等<sup>[9]</sup>提出了一种知识增强的事件因果关系识别方法，通过引入外部知识库（如ConceptNet）来提高因果关系提取的准确性和鲁棒性；Zhu等<sup>[12]</sup>提出了一种跨语言事件因果关系识别的预训练方法，通过在多种语言上进行预训练，提高了模型的泛化能力。

### 1.3 知识图谱在灾害情报分析中的应用

知识图谱作为一种结构化的语义知识表示方法，在灾害情报分析中发挥着重要作用。知识图谱通过构建包含与灾害相关的实体和关系的图谱，帮助人们组织和整合分散的信息，提供灾害事件的全面视图。随着知识图谱技术的不断发展，其在灾害情报分析中的应用前景广阔，有望进一步提升灾害管理的效率和效果。

目前对灾害事件数据构建知识图谱的方法主要包括多模态数据融合方法、灾害链角度分析方法，以及二者相结合的方法。多模态数据融合方法通过整合遥感影像、社交媒体文本和地理信息等多种数据源，实现知识图谱构建<sup>[13]</sup>；灾害链角度分析方法需要分析灾害发展过程中关联的诸多要素，建立统一的灾害链本体语义表达框架<sup>[14]</sup>；二者相结合的方法要求从数据源角度自底向上构建数据层，通过数据获取、知识抽取、融合、存储建立实体间关联关系，且需要从灾害链角度自顶向下构建模式层，通过本体建模形成知识图谱的概念框架<sup>[15]</sup>。

## 2 基于社交媒体大数据的灾害事件态势感知系统

本文构建了一个基于社交媒体大数据的灾害事件态势感知系统，其整体架构如图1所示。首先，该系统从海量社交媒体大数据中识别与特定灾害相关的文本信息，解决社交媒体大数据带来的信息过载问题；然后，从灾害文本信息中进一步提取灾害相关子事件并进行事件解析；最后，通过构建灾害子事件的因果知识图谱，推理事件的演变过程，提高用户对复杂灾害事件的理解，并为决策者提供更及时、全面的灾害情报支持。

### 2.1 灾害相关文本识别模块

灾害相关文本识别模块负责从海量社交媒体文本中识别出与灾害相关的信息，

主要包括文本预处理和基于预训练语言模型的分类模型，判断文本是否与特定灾害相关，如图2所示。

在预处理步骤中，识别模块首先从原始文本去除噪声内容。本文引入3种预处理设置来探讨它们分类性能的影响。

- 去除 URL、日期、提及、符号、表情符号、标点符号、数字、话题标签和停止词。

- 只删除 URL、日期和提及。

- 保留所有原创文字内容。

本文将以上的3种预处理设置分别命名为 RoBERTa-1、RoBERTa-2 以及 RoBERTa-3。

在分类模型对社交媒体文本进行分类时，本文使用规定的微调过程对其进行微调。如图2所示，首先，识别模块使用 RoBERTa 对社交媒体文本进行编码，并从最后一层获取令牌[CLS]的隐藏状态作为文本表示；接着，将其输入神经网络线性

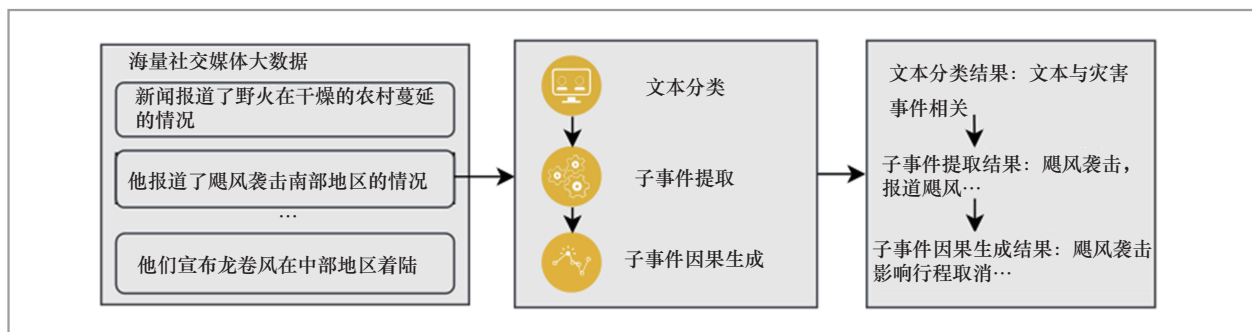


图1 基于社交媒体大数据的灾害事件态势感知系统的整体架构

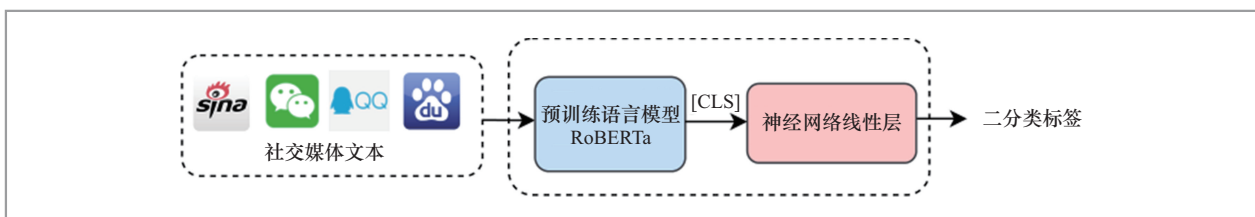


图2 分类模型进行社交媒体文本分类

层以输出预测的标签（二值分类，即相关或不相关）。

以 CrisisBench 数据集<sup>[16]</sup>为例，分类模型在该数据集上的评估结果见表1。适当的预处理可以提升模型性能，但过度预处理可能会丢失关键信息。

表1 基线模型和3种预处理设置下的分类模型在CrisisBench数据集上的评估结果

方法	精度	召回率	F1分数
基线模型	0.883	0.884	0.883
RoBERTa-1	0.886	0.886	0.886
RoBERTa-2	<b>0.891</b>	<b>0.892</b>	<b>0.891</b>
RoBERTa-3	0.889	0.890	0.889

## 2.2 子事件提取与解析

灾害事件感知系统通过3步流程来提取灾害子事件，如图3所示。其中，*P*表示指示代词，*V*表示动词，*N*表示名词，*subj*表示主语与谓语构成主谓结构，*obj*表示宾语与宾语构成动宾结构。ARG0、ARG1是AMR (abstract meaning representation) 解析中使用 Amrlib 产生的边。ARG0表示边相连的两个节点构成动宾结构，ARG1表示边相连的两个节点构成主谓结构。

第一步，对社交媒体中的文本进行解析，以识别可能与灾害相关的子事件。这个过程类似于寻找句子中的关键动作和对

象，如“洪水淹没了村庄”中的“洪水”和“淹没”。

第二步，去除重复或不太重要的子事件，保留能够提供有关灾害重要信息的子事件，帮助用户更加准确地理解灾害的主要信息。为了实现这一目标，系统运用了两种文本解析技术：依赖项解析<sup>[17]</sup>和AMR解析<sup>[18]</sup>。这两种技术可以从复杂的文本中提取出重要的事件片段，并将它们转换成更易于分析的结构。例如，系统能够识别出“hitting”实际上是“hit”的变形，从而确保用户得到准确的事件描述。

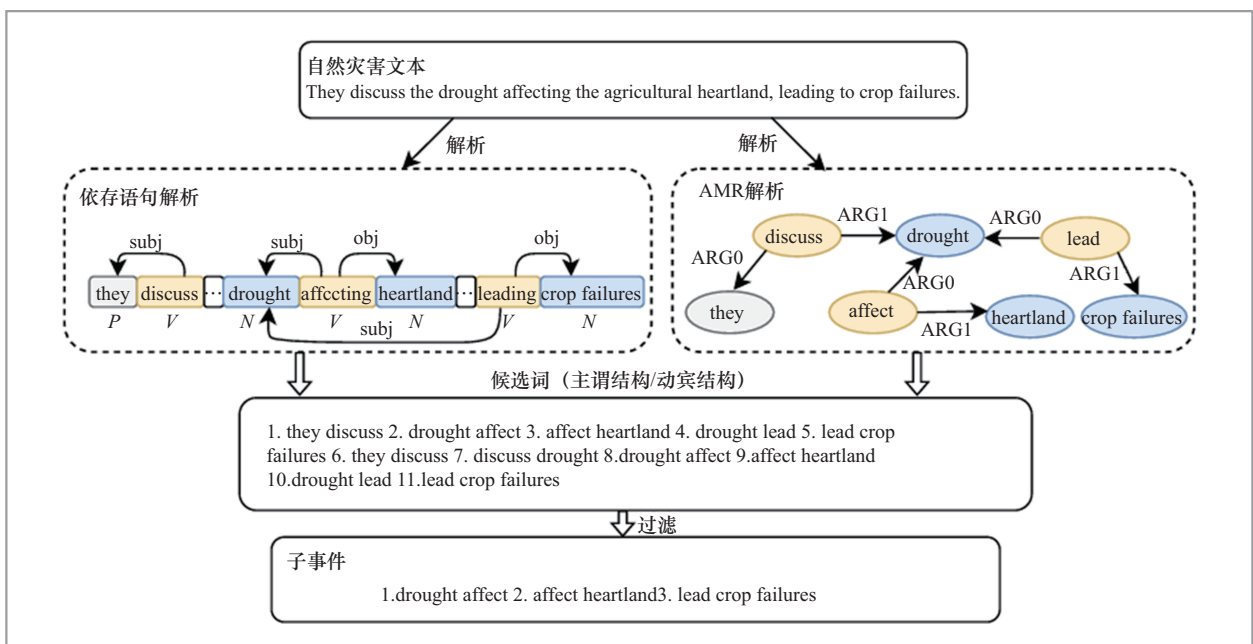


图3 灾害子事件提取流程

第三步，系统需要确保提取的子事件是独一无二的，并且具有实际意义。这需要排除包含常见但无关紧要词汇的子事件，或不符合寻找目标的特定结构（如名词—动词对）的子事件。

通过以上3个步骤，灾害事件感知系统能够自动提取出对灾害事件态势感知最有帮助的信息。

### 2.3 因果生成框架

知识图谱作为一种结构化的语义知识表示方法<sup>[19]</sup>，能够有效整合分散的信息，提供全面的事件视图。此外，其可以清晰地刻画灾害事件的因果关系，为态势感知和决策支持提供依据。目前，知识图谱在灾害情报分析中的应用逐渐增多，但仍面临数据异构性、动态性等挑战。本文设计了一个灾害子事件的因果生成框架，并构建了针对灾害事件的因果知识图谱。

灾害子事件的因果生成框架如图4所示。首先，因果生成框架从因果语料库 CausalBank<sup>[7]</sup>中筛选与灾害相关的文本对，覆盖多种灾害类型；接着，提取子事件与因果化标注，使用前述文本分类和子事件提取与解析的方法，从筛选完后的因果语料库 CausalBank 中构建灾害子事件的因果事件图；最后，进行知识图谱构建和模型微调，将因果对整合到知识图谱中，形成包含灾害事件因果关系的知识图谱，并通过知识图谱对 GPT 模型进行微调，使其能够生成准确的因果子事件。

CausalBank 是一个大型句子级因果语料库，包含 3.14 亿个文本对，存储方式为 <文本 A, 关系, 文本 B>。其中，大多数的关系可以分为因果两种模式：原因和结果（见表2）。因果关系模式为原因时，文本对可以表示为 <文本 A, 原因, 文本 B>，表示 B 是 A 的原因；因果关系模式为结果时，文本对可以表示为 <文本 A, 结果, 文

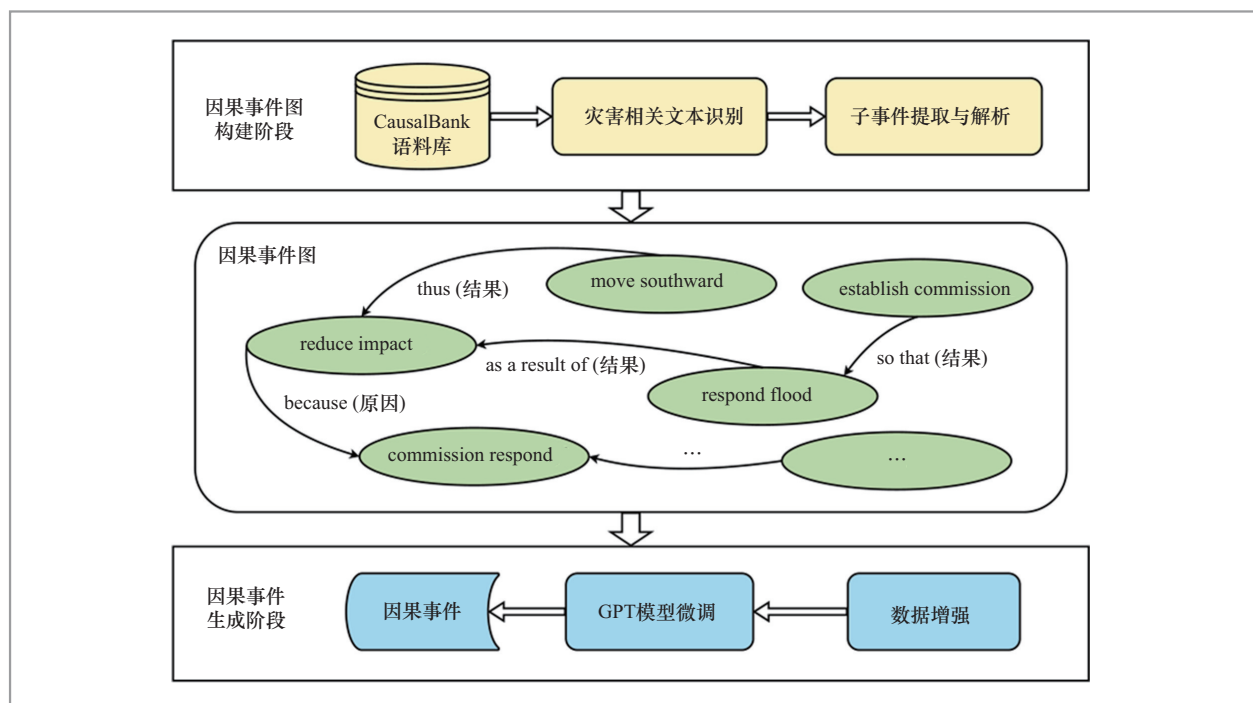


图4 灾害子事件的因果生成框架

表2 因果关系模式与CausalBank数据集中文本对关系的对应

因果关系模式	文本对关系
原因	as, as a consequence of, because, because of, result from, ...
结果	bring about, induce, result in, in order to, thereby, therefore, thus, ...

本 B>，表示 B 是 A 的结果。以文本对为例，首先，本文预处理输入文本 a 和文本 b 的连接([文本 a；文本 b])至文本分类模型中识别是否与灾害相关；然后，使用提取框架分别从文本 a 和文本 b 中提取子事件（若与灾害相关）；最后，将子事件配对成形式<子事件 a，模式，子事件 b>。使用上述方法，本文从 CausalBank 中识别了 70 多万与灾害相关的文本对，并提取了约 20 万个子事件对；对所有子事件对进行整合，构建灾害子事件的因果知识图，图 4 展示了其中一个子图。为了让 GPT 能学习到因果知识图里的灾害知识，本文将 GPT 在因果知识图上进行了微调。

在模型训练阶段，本文通过在训练样本中加入每个子事件对相反模式的描述，进行数据扩充。对于子事件因果对，两种因果关系模式是可以相互转换的，如子事件 b 是子事件 a 的结果，反过来子事件 a 可以是子事件 b 的原因。具体以子事件对<子事件 a，原因，子事件 b>为例，在训练样

本中加入其相反模式<子事件 b，结果，子事件 a>的描述。数据覆盖了多种灾害类型，包括地震（25%）、洪水（20%）、台风（15%）、火灾（10%）、其他灾害（30%）。每种灾害类型的数据均衡分布，以确保模型能够学习到不同灾害的因果关系模式。本文一共进行了 5 轮训练，每轮训练样本数量为 20，采用自适应学习率，初始值为  $1 \times 10^{-5}$ ，并通过学习率调度器进行动态调整。在生成阶段，以因果关系模式为导向词，将查询事件与因果关系模式的拼接输入模型。生成长度设置为 2（即两个单词）。此外，为了生成多样性的结果，将参数温度系数设置为 0.8，将结果个数设置为 5。

## 2.4 案例分析

表 3 中为两个查询子事件（关闭机场、水位上涨）的因果生成结果，用于研究因果事件图的影响。其中，GPT 表示原始模

表3 不同模型下DSECG对因果生成关系影响的案例研究

子事件查询	生成式预训练模型 GPT		微调后的生成式预训练模型 GPT*	
	原因	结果	原因	结果
关闭机场	当时天气	其他危机	火山喷发	人们受困
	一个不幸的	更多恐惧	地震袭击	关闭度假胜地
	一个恐怖分子	服务中断	飓风袭击	高速公路堵塞
水位上涨	变暖	崩溃	降雨	修筑桥梁
	高温	风险增加	修筑大坝	疏散人群
	地下水增多	更高的级别	冰川融化	洪水发生

型，GPT\*表示在因果事件图上微调的模型。

由表3可知，GPT能够生成常识性答案，得益于在大规模语料库上的预训练；GPT\*能够学习因果事件图中的灾害领域知识并生成结构化的因果事件，其原因是用于训练的子事件对是严格对齐的。例如，在子事件查询“关闭机场”中，GPT生成“一个不幸的”和“一个恐怖分子”的原因，而GPT\*对应生成“地震袭击”和“飓风袭击”。本文认为前者是一种常识，对因果分析没有帮助，而后者可以更好地捕捉事件的原因。本文从子事件查询“水位上升”中也可以发现与之类似的结果，GPT生成“变暖”和“高温”的原因，而GPT\*生成了更具体的“降雨”和“冰川融化”原因。

为了验证构建的灾害子事件因果知识图谱系统的有效性，本文增加了更多灾害类型的生成案例，包括地震、火灾等，并引入人工标注准确率作为定量评估指标。表4是不同灾害类型的因果生成的部分案例及其评估结果。

通过人工标注和对比分析，系统能够有效地表示灾害事件的因果关系，并通过微调后的生成式预训练模型生成准确的因果事件对。本文所提系统在不同灾害类型中的表现均较出色，平均准确率接近

85%，验证了所提系统的有效性和泛化能力。

## 3 系统实现与功能

上述模块集成了一个在线演示系统，其作为一个Web应用程序实现，并由Streamlit工具包提供编程支持。如图5所示，灾害事件态势感知系统有一个直观的用户界面，用于交互式地可视化系统的3个主要功能模块的结果，其由5个步骤构成。

### 3.1 文本分类的可视化

第一步，社交媒体文本输入。当系统接收到输入文本时，其首先只删除URL、日期和提及的预处理方式对文本进行预处理，然后将预处理后的文本提供给RoBERTa-2分类模型。预处理步骤的目的是减少无关信息的干扰，提高模型的分类准确性。

第二步，展示文本分类的结果。分类模型的输出是两个标签（相关或不相关）的概率值，其和为1。这些概率值使用饼状图进行可视化，其中深灰色条表示相关的概率，浅灰色条表示不相关的概率，饼状图的中间显示相关概率值的大小。

### 3.2 子事件提取和因果生成的可视化

如果文本被分类为与灾害相关，系统将进入子事件提取和因果生成的可视化阶段。子事件提取模块首先对文本进行解析，识别出可能与灾害相关的子事件。系统运用依赖项解析和AMR解析技术，从复杂文本中提取出重要的事件片段，并将其转换为更易于分析的结构。接着，在第四步

表4 不同灾害类型的因果生成的部分案例及其评估结果

灾害类型	查询子事件	生成因果事件	人工标注准确率
地震	地震袭击	建筑倒塌	85%
地震	地震袭击	交通中断	80%
火灾	火灾发生	电力中断	90%
火灾	火灾发生	疏散人群	88%
洪水	水位上涨	降雨增加	83%
洪水	水位上涨	冰川融化	82%

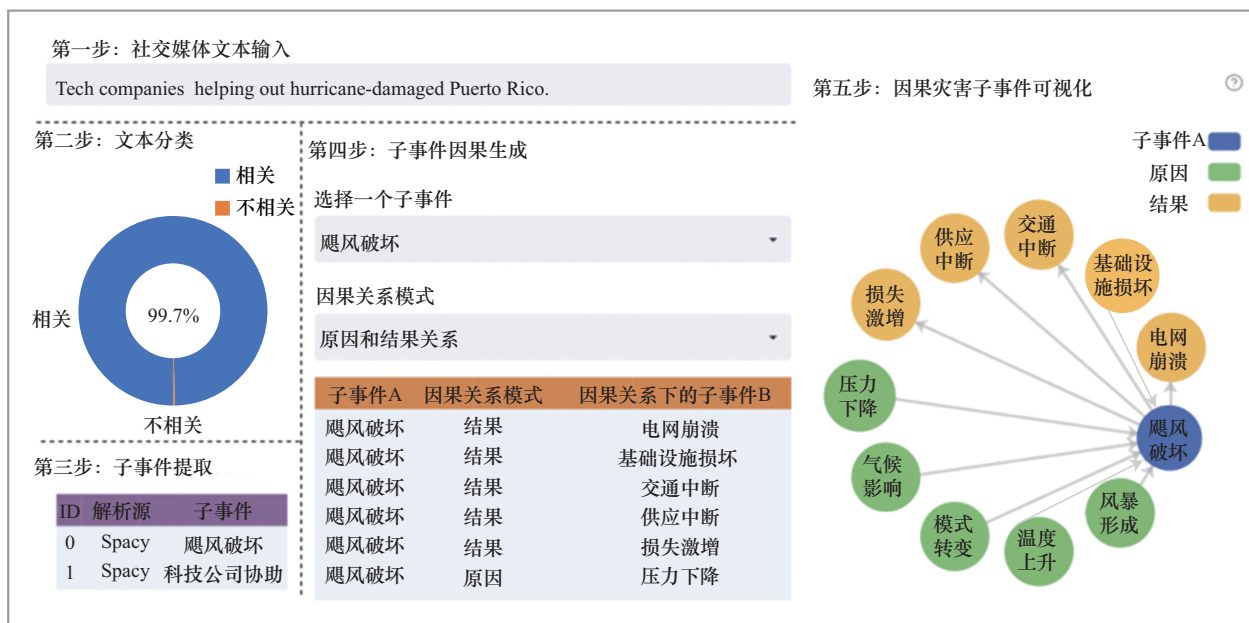


图5 灾害事件态势感知系统用户界面

子事件因果生成步骤中，用户可以选择不同的子事件和不同的因果关系模式选项（“原因”“结果”以及“原因和结果”共3种选项），界面将模型计算出来的因果子事件可视化展示在第五步中。

### 3.3 案例分析

基于以上功能模块，本文进行案例展示并根据系统输出的结果进行分析。首先，系统从社交媒体中提取一个实时社交媒体文本“Tech companies helping out hurricane-damaged Puerto Rico.”，即“科技公司帮助受飓风破坏的波多黎各”。如图5所示，在“飓风破坏”子事件中，系统快速生成风暴形成、温度上升、模式转变、气候影响和压力下降5个原因子事件，以及电网崩溃、基础设施损坏、交通中断、供应中断和损失激增5个结果子事件；接着，让决策者结合原因中已有的信息，得到灾害最有可能发生的结果，结果

子事件可以帮助决策者采取措施解决问题。例如，从获得的原因和结果来看，如果监测到显著的气压下降造成飓风破坏程度大，导致电网崩溃大面积停电、基础设施的损坏影响整个地区的恢复、物资供应的中断影响救援和恢复工作，那么决策者需要考虑优先修复基础设施和恢复供应链，优先修复关键基础设施，如电力、通信和交通系统，以支持救援和恢复工作、迅速调配救援物资和人员，优先保障受影响最严重地区的供应。这些因果关系不仅可以帮助决策者快速了解灾害事件的起因和可能的后果，还为应急响应和资源调配提供重要的参考依据。

## 4 结束语

大数据技术的发展以及社交媒体的普及使人们能够以更加广阔的视角洞察灾害事件的复杂性。事件因果关系发现是实现

事件推理的重要前提。灾害事件的偶发性和独特性使现有大多依赖历史数据的分析方法很难直接推理出灾害事件的潜在发展逻辑和可能产生的危害。基于事件的研究方法（如知识图谱、事理图谱等）能够形成对既定知识逻辑路线的发现与探索，结合因果关系发现技术可以以直观、形象的方式解读复杂事件的发展模式，为灾害事件态势感知提供强而有力的支撑。利用媒体大数据及时、广泛、参与性强的优势，人们可以在时间维度上追踪灾害事件随时间的演化过程，在空间维度上展示某时刻各受灾地点的详细灾情，并根据灾害事件的时空情景自动规划出相应的辅助应急建议措施，提升灾害应急响应的时效性和全面性。尽管本文已取得了一定的研究成果，但仍存在需要进一步改进和完善的地方。例如，当前的模型在处理大规模数据时的准确率还有待提高，且对于不同类型灾害事件的适应性还有待进一步验证。在未来的研究中，笔者将进一步丰富数据来源，引入更多类型的灾害事件数据，构建跨模态灾害事件感知模型，以增强模型对不同灾害类型和灾害数据的适应能力。

## 参考文献：

- [1] 王一鸣, 连巧玉, 李群, 等. 《国家自然灾害救助应急预案》修订解读[J]. 中国应急管理, 2024(5): 38-43.  
Wang Y M, Lian Q Y, Li Q, et al. Interpretation of the revision of national emergency plan for natural disaster relief[J]. China Emergency Management, 2024(5): 38-43.
- [2] 李涛, 周绮凤, 郑理, 等. 大数据提升灾难信息管理水平[J]. 大数据, 2015, 1(3): 124-129.  
Li T, Zhou Q F, Zheng L, et al. Big data meets the needs of disaster information management[J]. Big Data Research, 2015, 1(3): 124-129.
- [3] Alam F, Ofli F, Imran M, et al. A Twitter tale of three hurricanes: Harvey, Irma, and Maria[PP]. arXiv preprint, 2018, arXiv: 1805.05144.
- [4] Zhang C, Fan C, Yao W L, et al. Social media for intelligent public information and warning in disasters: an interdisciplinary review[J]. International Journal of Information Management, 2019, 49: 190-207.
- [5] Ittoo A, Bouma G. Extracting explicit and implicit causal relations from sparse, domain-specific texts[M]// Natural Language Processing and Information Systems. Berlin: Springer Berlin Heidelberg, 2011: 52-63.
- [6] Ittoo A, Bouma G. Minimally-supervised learning of domain-specific causal relations using an open-domain corpus as knowledge base[J]. Data & Knowledge Engineering, 2013, 88: 142-163.
- [7] Li Z Y, Ding X, Liu T, et al. Guided generation of cause and effect[C]//Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence. [S.l.:s.n.], 2020: 3629-3636.
- [8] Gao L, Choubey P K, Huang R H. Modeling document-level causal structures for event causal relation identification [C]//Proceedings of the 2019 Conference of the North. Stroudsburg: ACL, 2019: 1808-1817.
- [9] Liu J, Chen Y B, Zhao J. Knowledge enhanced event causality identification with mention masking generalizations[C]//Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence. International Joint Conferences on Artificial Intelligence Organization, 2020: 3608-3614.

- [10] Li Z N, Li Q, Zou X T, et al. Causality extraction based on self-attentive BiLSTM-CRF with transferred embeddings[J]. *Neurocomputing*, 2021, 423: 207-219.
- [11] Khetan V, Ramnani R, Anand M, et al. Causal BERT: language models for causality detection between events expressed in text[M]//*Intelligent Computing*. Cham: Springer International Publishing, 2021: 965-980.
- [12] Zhu E C, Yu Z T, Huang Y X, et al. PTEKC: pre-training with event knowledge of ConceptNet for cross-lingual event causality identification[J]. *International Journal of Machine Learning and Cybernetics*, 2025, 16(3): 1859-1872.
- [13] 沈伟豪, 钟燕飞, 王俊珏, 等. 多模态数据的洪涝灾害知识图谱构建与应用[J]. *武汉大学学报(信息科学版)*, 2023, 48(12): 2009-2018.
- Shen W H, Zhong Y F, Wang J J, et al. Construction and application of flood disaster knowledge graph based on multi-modal data[J]. *Geomatics and Information Science of Wuhan University*, 2023, 48(12): 2009-2018.
- [14] 邱芹军, 吴亮, 马凯, 等. 面向灾害应急响应的地质灾害链知识图谱构建方法[J]. *地球科学*, 2023, 48(5): 1875-1891.
- Qiu Q J, Wu L, Ma K, et al. A knowledge graph construction method for geohazard chain for disaster emergency response[J]. *Earth Science*, 2023, 48(5): 1875-1891.
- [15] 杜志强, 李钰, 张叶廷, 等. 自然灾害应急知识图谱构建方法研究[J]. *武汉大学学报(信息科学版)*, 2020, 45(9): 1344-1355.
- Du Z Q, Li Y, Zhang Y T, et al. Knowledge graph construction method on natural disaster emergency[J]. *Geomatics and Information Science of Wuhan University*, 2020, 45(9): 1344-1355.
- [16] Alam F, Sajjad H, Imran M, et al. CrisisBench: benchmarking crisis-related social media datasets for humanitarian information processing[J]. *Proceedings of the International AAAI Conference on Web and Social Media*, 2021, 15: 923-932.
- [17] Honnibal M, Montani I, Van Landeghem S, et al. spaCy: Industrial-strength natural language processing in python[J]. *Journal of Open Source Software*, 2020, 5(52): 2489.
- [18] Arachie C, Gaur M, Anzaroot S, et al. Unsupervised detection of sub-events in large scale disasters[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, 34(1): 354-361.
- [19] 王文广. 知识图谱推理: 现代的方法与应用[J]. *大数据*, 2021, 7(3): 42-59.
- Wang W G. Knowledge graph reasoning: modern methods and applications[J]. *Big Data Research*, 2021, 7(3): 42-59.

#### 作者简介



龚万渊 (2000-), 男, 厦门大学航空航天学院自动化系硕士生, 主要研究方向为自然语言处理。

## 作者简介



王慧颖（2000-），女，厦门大学航空航天学院自动化系硕士生，主要研究方向为推荐系统。



江信禧（1997-），男，厦门大学航空航天学院自动化系硕士生，主要研究方向为自然语言处理。



周绮凤（1976-），女，厦门大学航空航天学院自动化系教授，主要研究方向为数据挖掘与信息检索。

收稿日期: 2025-01-20

通信作者: 周绮凤, zhouqf@xmu.edu.cn

基金项目: 国家自然科学基金项目(No.62171391)

Foundation Item: The National Natural Science Foundation of China (No.62171391)