

## 摘 要

灾害应急遥感任务驱动的遥感影像数据检索是提高灾害应急响应的重要手段之一。灾害应急遥感影像应用案例知识是实现灾害应急遥感任务驱动的遥感影像数据检索的基础。灾害应急遥感应用学术文本摘要中含有大量灾害应急遥感影像应用案例知识，然而学术文本摘要是一种非结构化的数据，需要借助自然语言处理等技术进行知识抽取。文本语料库标注了文本中的命名实体和关系等内容，是自然语言处理模型训练、学习的重要数据资源，因此，构建标注语料库对获取灾害应急遥感影像应用案例知识具有重要意义。本文根据遥感在灾害应急中的应用，构建了灾害应急遥感应用案例特征模型；通过对灾害应急遥感应用学术文本摘要特征分析，建立了学术文本标注体系，设计了灾害应急遥感影像应用案例知识标注规范；以学术文本摘要为数据源标注了语料库。本文的工作主要包含三个部分：

### （1） 遥感影像应用学术文本标注体系设计

基于遥感在灾害应急中的应用，分析灾害应急遥感应用案例要素组成、要素关系及案例间的关系，建立了灾害应急遥感应用案例特征模型；通过对灾害应急遥感应用学术文本摘要特征分析，设计了遥感影像应用学术文本标注体系。

### （2） 设计灾害应急遥感影像应用案例知识标注规范

灾害应急遥感应用学术文本摘要中的命名实体以及实体间的关系表达复杂多样。为了规范语料库标注，针对灾害应急遥感任务实体、遥感数据实体、方法实体、效果实体、时间实体、空间实体，设计了实体标注规范；对于实体间关系，设计了句子级和篇章级的关系标注规范。

### （3） 构建遥感影像应用案例知识标注语料库

针对灾害应急遥感影像应用领域没有公开标注语料库的问题。基于本文设计的知识标注规范，对学术文本摘要进行标注；在标注过程中，采用深度学习模型对语料库进行验证，优化标注过程；最终构建了由 2000 篇学术文本摘要标注而成的语料库，语料库中包含实体共计 31575 例，实体间关系共计 15424 例。

**关键词：**灾害应急；遥感影像应用案例；知识库；语料库；标注

## ABSTRACT

Task-oriented remote sensing image data retrieval for disaster response is one of the key methods to improve emergency response capabilities. Knowledge of disaster emergency remote sensing image application cases is the foundation for achieving remote sensing image data retrieval driven by disaster emergency remote sensing tasks. The academic text summary of disaster emergency remote sensing application contains a large number of disaster emergency remote sensing image application case knowledge. However, the academic text summary is a kind of unstructured data, which needs to use Natural language processing and other technologies to extract knowledge. Text corpora are important data resources for Natural language processing model training and learning, which annotate named entities and relationships in the text. Therefore, building an annotated corpus is of great significance for acquiring disaster emergency remote sensing image application case knowledge. This article constructs a feature model for disaster emergency remote sensing application cases based on the application of remote sensing in disaster emergency response; By analyzing the characteristics of academic text abstracts for disaster emergency remote sensing applications, an academic text annotation system was established, and a knowledge annotation standard for disaster emergency remote sensing image application cases was designed; A corpus was annotated using academic text abstracts as the data source.

This work mainly consists of three parts:

(1) Design of Academic Text Annotation System for Remote Sensing Image Applications

Based on the application of remote sensing in disaster emergency, this paper analyzes the composition and relationship of elements in disaster emergency remote sensing application cases, and establishes a feature model for disaster emergency remote sensing application cases; A remote sensing image application academic text annotation system was designed by analyzing the characteristics of academic text abstracts for disaster emergency remote sensing applications.

(2) Specification for knowledge labeling of disaster emergency remote sensing image application cases

The named entities and the relationships between entities in the academic text abstract of disaster emergency remote sensing applications are complex and diverse. In order to standardize corpus annotation, entity annotation specifications have been designed for disaster emergency remote sensing task entities, remote sensing data entities, method entities, effect entities, time entities, and spatial entities; For the relationship between entities, we have designed sentence level and discourse level relationship annotation standards.

(3) Constructing Knowledge Annotation Corpus of Remote Sensing Image Application Cases

There is no publicly annotated corpus in the field of disaster emergency remote sensing image application. Based on the knowledge annotation specification designed in this article, annotate academic text abstracts; During the annotation process, a deep learning model is used to validate the corpus and optimize the annotation process; Finally, a corpus consisting of 2000 academic text abstracts was constructed, which included a total of 31575 entities and 15424 relationships between entities.

Key words: Disaster emergency response; Remote sensing image application cases; knowledge base; corpus; annotation

## 目 录

第 1 章 绪论	1
1.1 研究背景与意义	1
1.2 国内外研究现状	2
1.2.1 命名实体抽取方法	2
1.2.2 命名实体关系抽取方法	3
1.2.3 标注语料库	5
1.3 研究目标与内容	7
1.3.1 研究目标	7
1.3.2 研究内容	7
1.4 技术路线	7
1.5 本文的主要工作	8
第 2 章 灾害应急遥感影像应用案例知识建模	10
2.1 遥感灾害应急应用	10
2.1.1 自然灾害分类	10
2.1.2 遥感在灾害应急中的应用	11
2.2 灾害应急遥感应用案例特征模型	12
2.2.1 灾害应急遥感应用案例要素组成	12
2.2.2 灾害应急遥感应用案例要素关系	13
2.2.3 灾害应急遥感应用案例间关系	15
2.3 灾害应急遥感应用学术文本摘要特征分析	16
2.3.1 命名实体分类及描述特征	16
2.3.2 命名实体关系及描述特征	19
2.4 遥感影像应用学术文本标注体系设计	23
2.5 本章小结	26

第 3 章 灾害应急遥感影像应用案例知识标注规范 .....	27
3.1 命名实体标注规范 .....	27
3.1.1 灾害应急遥感任务实体标注规范 .....	27
3.1.2 遥感数据实体标注规范 .....	29
3.1.3 方法实体标注规范 .....	30
3.1.4 效果实体标注规范 .....	33
3.1.5 时间实体标注规范 .....	35
3.1.6 空间实体标注规范 .....	35
3.2 句子级关系标注规范 .....	36
3.2.1 遥感数据实体间的关系标注规范 .....	36
3.2.2 方法实体间的关系标注规范 .....	39
3.2.3 灾害应急遥感任务和遥感数据实体间关系标注规范 .....	40
3.2.4 效果实体间关系标注规范 .....	40
3.2.5 方法和效果实体间关系标注规范 .....	41
3.2.6 方法和灾害应急遥感任务实体间关系标注规范 .....	41
3.2.7 遥感数据和方法实体间关系标注规范 .....	42
3.2.8 不同灾害应急遥感任务实体间关系标注规范 .....	43
3.2.9 灾害应急遥感任务和时间实体间关系标注规范 .....	44
3.2.10 灾害应急遥感任务和空间实体间关系标注规范 .....	44
3.3 篇章级关系标注规范 .....	45
3.3.1 方法指代关系标注规范 .....	45
3.3.2 数据指代关系标注规范 .....	46
3.4 本章小节 .....	46
第 4 章 遥感影像应用案例知识语料库构建与分析 .....	47
4.1 实体关系语料库构建 .....	47
4.1.1 原始数据获取及预处理 .....	48
4.1.2 语料标注 .....	50
4.1.3 标注流程 .....	51

4.2 基于深度学习实体关系抽取.....	52
4.2.1 模型介绍 .....	52
4.2.2 指标选取 .....	56
4.2.3 实验数据预处理 .....	57
4.2.4 实验平台及参数设置 .....	59
4.2.5 实验结果分析 .....	60
4.2.6 样本分析 .....	64
4.3 语料库分析.....	68
4.4 本章小结.....	74
第 5 章 结论与展望.....	76
5.1 结论.....	76
5.2 展望.....	77
致    谢.....	78
参考文献 .....	79

## 一、学位论文独创性声明

本人声明所呈交的学位论文是本人在导师指导下进行的研究工作及取得的研究成果。据我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得南昌大学或其他教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示谢意。

学位论文作者签名（手写）：陈明亮

签字日期：2023 年 6 月 8 日

## 二、学位论文版权使用授权书

本学位论文作者完全了解南昌大学有关保留、使用学位论文的规定，同意学校有权保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。本人授权南昌大学可以将学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存、汇编本学位论文。同时授权北京万方数据股份有限公司和中国学术期刊（光盘版）电子杂志社将本学位论文收录到《中国学位论文全文数据库》和《中国优秀博硕士学位论文全文数据库》中全文发表，并通过网络向社会公众提供信息服务，同意按“章程”规定享受相关权益。

学位论文作者签名（手写）：陈明亮

导师签名（手写）：李铭

签字日期：2023 年 6 月 8 日

签字日期：2023 年 6 月 8 日

论文题目	灾害应急遥感影像应用案例知识标注语料库构建研究				
姓名	陈明亮	学号	411016620054	论文级别	博士 <input type="checkbox"/> 硕士 <input checked="" type="checkbox"/>
院/系/所	信息工程学院	专业	通信工程(含宽带网络、移动通信等)		
E_mail					
备注:					

公开 保密（向校学位办申请获批准为“保密”，\_\_\_\_\_年\_\_\_月后公开）

## 第1章 绪论

### 1.1 研究背景与意义

遥感数据是灾害应急、军事农业、城市规划等许多领域决策的重要资源，也是地理空间科学领域的大数据<sup>[1]</sup>。随着对地观测技术的蓬勃发展，遥感数据资源愈加丰富，数据规模不断增长。与此同时，海量的遥感数据资源的爆炸性增长使得搜索和发现成为一个重大的挑战。面对这样的挑战，有学者提出了以任务为驱动的遥感数据搜索方式，实践表明，以灾害应急遥感任务为驱动的遥感数据搜索可以明显提高搜索的效率<sup>[2]</sup>。例如为应急响应、损失评估、灾后恢复重建等。

遥感影像应用案例知识图谱是任务驱动的遥感数据搜索的重要支撑。自谷歌首次提出“知识图谱”概念以来，知识图谱取得快速发展，成为人工智能领域的重要分支之一。知识图谱是通过有向图的方式表达实体、概念及其相互之间语义关系的数据组织形式，其中节点代表实体或者概念，边代表实体/概念的属性或者彼此之间的关系，从本质上讲，知识图谱是一种语义网络<sup>[3]</sup>。灾害应急遥感影像应用案例知识图谱是一种结构化描述时间、空间、遥感数据、灾害应急遥感任务、方法和效果等实体及其关系的语义网络。构建灾害应急遥感影像应用案例知识图谱是灾害应急遥感任务驱动的遥感影像数据搜索的基础性、支撑性工作，将决定遥感影像数据搜索的查询效率以及正确率等重要指标<sup>[4]</sup>。

近年来，随着自然语言处理技术的快速发展，人们可以对非结构化的文本数据进行分词、词性标注、句法分析等处理，为命名实体抽取、关系抽取、情感分析等 NLP 任务带来极大便利。语料库是以计算机为工具建立、存储和使用的语言素材集合<sup>[5]</sup>。通过对语料库进行分析和挖掘，我们可以提取出其中的规律和特征，并将这些信息应用到自然语言处理技术中<sup>[6]</sup>。灾害应急遥感影像应用案例知识的获取也需要标注语料库作为支撑；因此，如何构建有效的灾害应急遥感影像应用案例知识标注语料库是目前亟待解决的问题。

鉴于此，本文研究灾害应急遥感影像应用案例知识标注语料库。首先综合遥感在灾害应急中的应用，分析灾害应急遥感应用案例要素组成、要素关系及案例间的关系构建灾害应急遥感应用案例特征模型；接着对灾害应急遥感应用学术



文本摘要特征分析,构建遥感影像应用学术文本标注体系,然后设计灾害应急遥感影像应用案例知识标注规范;使用网络爬虫技术从开放学术图谱中获取原始数据,经过数据清洗后得到灾害应急遥感应用学术文本摘要数据;使用设计的知识标注规范对学术文本摘要数据进行标注构建语料库。采用基于深度学习的方法验证标注语料库,优化语料库标注过程,为高质量语料库构建提供一种思路,也为灾害应急遥感影像应用案例知识的挖掘和分析研究提供支撑。

## 1.2 国内外研究现状

### 1.2.1 命名实体抽取方法

在自然语言处理领域,命名实体识别是信息抽取的第一个关键环节<sup>[7]</sup>。可以将非结构化灾害应急遥感应用学术摘要数据中的时间实体、空间实体、灾害应急遥感任务实体、遥感数据实体、方法实体以及效果实体抽取出来并加以归类,进而组织成半结构化或结构化的信息,再利用其它技术对文本实现分析和理解目的。命名实体识别技术广泛应用于信息检索、自动问答系统等领域<sup>[8]</sup>,目前主流的命名实体抽取方法分为三类:基于规则和词典的方法、基于传统机器学习方法、基于深度学习方法。

#### 1) 基于规则和词典的方法

基于规则指的是预先制定一系列与实体相匹配的规则,然后将该规则应用到待抽取文本当中;基于词典指的是人为事先维护一个词典,该词典需要包含待识别的实体,然后用词典中的实体和待抽取文本做字符匹配,得到的结果就是要抽取的实体信息<sup>[9]</sup>。例如 Keretna 等提出了一种使用基于词典和基于规则的混合方法从非结构化的医学文本中提取名称,取得了比较好的结果<sup>[10]</sup>。

由国内外研究现状可知,基于规则和词典的方法较为简单,在很多场景下也能取得较好的识别结果。然而,为了保证实体识别准确率,基于规则和词典的方法需要不断更新规则模板以及词典当中的实体名称,在灾害应急遥感影像应用领域存在着一定的局限性。

#### 2) 基于统计机器学习的方法

随着各个领域文本数据量不断增加,对命名实体识别的研究方向从基于规则和词典的研究方向逐渐转向基于统计机器学习的方法<sup>[11]</sup>。比较经典的统计机

器学习方法有隐马尔可夫模型(HMM)、支持向量机 SVM、条件随机场 CRF 等。Fei 等使用支持向量机 SVM 和 CRF 混合方法对文本中生物术语做命名实体识别,在 GENIA 语料库中 macro-F1 达到了 91.67%<sup>[12]</sup>。Li 等使用 CRF 模型做生物医学命名实体识别,在 JNLPBA2004 数据集上进行实验,F 值达到了 74.31%<sup>[13]</sup>。

由国内外研究现状可知,基于传统机器学习的方法虽然也能取得较好的识别效果,但是需要构造复杂的特征工程,在灾害应急遥感影像应用领域同样存在着一定的局限性。

### 3) 基于深度学习的方法

随着深度学习算法的快速发展,对命名实体识别的研究逐渐从传统的统计机器学习方法到基于深度学习的方法。深度学习算法在各个领域都表现出色,目前已广泛应用于命名实体识别任务中<sup>[14]</sup>。Dong 等提出了一种基于卷积神经网络(CNN)的多类分类方法,用于从电子病历中挖掘命名实体<sup>[15]</sup>。Shen 等提出了一种基于深度神经网络架构的生物医学命名实体识别(Bio-NER)方法,在 GENIA 常规测试语料库中测试,F 值达到了 71.01%<sup>[16]</sup>。

基于深度学习本身具有端到端的特性,不需要进行复杂的特征工程,使得其网络框架具有较强的通用性和高精度度,在命名实体抽取任务上可以取得良好的成绩。然而,需要意识到,取得良好成绩的关键在于丰富而高质量的语料库作为基础支撑。因此,构建适用于命名实体抽取的语料库是至关重要的。只有构建完备充足的语料库,才能为深度学习提供充分的训练数据和优化空间。

## 1.2.2 命名实体关系抽取方法

实体关系抽取(Named Entity Relation Extraction)是信息抽取的一个重要子任务,又被称作关系抽取。它的主要研究方向是从大量文本数据中抽取实体间可能存在的关联信息。实体关系抽取可以帮助实现灾害应急遥感影像应用案例知识的结构化表达,具有重要的应用价值。随着命名实体关系抽取技术的不断发展,关系抽取技术已经取得重要进步,目前关系抽取的研究成果主要应用在机器翻译、知识图谱、自动问答等领域<sup>[17]</sup>。主流的命名实体关系抽取方法主要有:有监督的实体关系抽取方法、半监督的实体关系抽取方法、无监督的实体关系抽取方法、基于深度学习的实体关系抽取方法。

### 1) 有监督的实体关系抽取方法

有监督的实体关系抽取方法可以分为基于核函数和特征向量的方法。基于核函数的实体关系抽取方法,最初用于解决支持向量机(SVM)模型中向量的计算,后面逐渐应用到更多的模型当中。核函数隐式计算两个特征向量映射到高维空间后的内积,巧妙地将非线性问题转化为线性问题来处理。Mooney 等使用核函数的方法从生物医学语料库中提取实体间关系<sup>[18]</sup>。基于特征向量的关系抽取方法需要将文本句子实例采用多维的数字化向量表示,将语义关系相似度的计算转换为欧几里得空间中的向量计算过程<sup>[19]</sup>。姚全珠等利用提出一种基于子树特征的实体关系抽取方法,利用子树挖掘和特征选择得到有效子树,并将其作为特征模板构造特征向量并在中文语料库上取得较好分类效果<sup>[20]</sup>。

从国内外技术的发展现状来看,基于有监督实体关系抽取技术受限于语料库规模。因此,基于有监督的实体关系抽取方法适合于有大量标记数据的语料库,而灾害应急遥感影像应用领域属于专业性领域,目前并没有公开标注的语料库,所以,该方法具有一定的局限性。

### 2) 半监督的实体关系抽取

半监督的实体关系抽取也可以称为弱监督的实体关系抽取方法。半监督学习的思想是使用少量已标注数据集作为初始种子集,通过一种循环学习机制去标注大量未标注数据<sup>[21]</sup>。半监督的实体关系抽取方法有利于可以减轻人工标记的数据样本,节省大量的人力成本,受到了不少学者的青睐。Liu 等提出了一种新的 Web 规模关系提取框架 MultiSnowball 提取多种关系<sup>[22]</sup>。王使用半监督的 Bootstrapping 方法,从大量未标注的自然语言文本中学习用于发现某种特定二元关系的模式,并找出具有此种关系的实体对<sup>[23]</sup>。

从国内外研究现状可知,半监督的实体关系抽取方法相比于有监督实体关系抽取方法,虽然减轻了人工标注数据的工作量,降低了对训练样本的要求,但是同时也会存在获取大量无效样本的缺点,在灾害应急遥感影像应用这样的专业领域中,无效样本会降低关系抽取的效果。

### 3) 无监督的实体关系抽取

无监督机器学习的基本思想是基于分布假设理论-具有相同关系的实体对会有相似或相近的上下文内容,同时这些内容中可以选择出具有代表性的词汇描述该类关系<sup>[24]</sup>。马等使用无监督的实体关系抽取方法在低资源语言老挝语上达到了不错的效果,平均准确率达到了 60.43%<sup>[25]</sup>。秦兵提出面向大规模网络文本的无指导开放式中文实体关系抽取方法,首先使用实体之间的距离限制和关系指

示词的位置限制获取候选关系三元组;保证了 80%以上的微观平均准确率<sup>[26]</sup>。

从国内外目前的研究现状来看,无监督的实体关系抽取不需要标注数据,更无需设定关系种类和类别,一般可依靠文本数据结构等信息发现实体关系。但是无监督的实体关系抽取方法的准确率和召回率都比较低,在灾害应急遥感影像应用案例实体关系抽取中同样具有很大的局限性。

#### 4) 基于深度学习的实体关系抽取

基于深度学习的有监督的关系抽取方法是近年来关系抽取的研究热点,该方法可以避免经典方法中人工特征选择等步骤,减少并且改善特征抽取过程中的误差积累问题。深度学习模型在实体关系抽取任务中应用广泛,常用的模型有卷积神经网络(CNN)、循环神经网络、脉冲耦合神经网络(PCNN)、长短期记忆网络、Transformer 和 BERT 等<sup>[27]</sup>。根据实体识别及关系分类两个子任务完成的先后顺序不同,基于深度学习的有监督的实体关系抽取方法可以分为流水线(pipeline)方法和联合学习(joint learning)方法。Kumar 等提出了一种基于注意力的深度剩余网络(ResNet)模型,以识别中文电子病历中医疗概念关系,ResNet 模型在人工注释的中文电子病历语料库上获得了 77.80%的 F1 分数<sup>[28]</sup>。

从国内外研究现状来看,基于深度学习强大的表示学习能力,在实体关系抽取任务上,深度学习模型取得了良好的效果。然而,基于深度学习的实体关系抽取离不开标注语料库作为基础资源支撑。因此,构建适用于深度学习实体关系抽取的标注语料库是十分必要的。

### 1.2.3 标注语料库

自 20 世纪 80 年代以来,针对语料库<sup>[29]</sup>的研究逐渐成为学者们的研究热点。语料库是为自然语言处理领域中各种子任务服务的一种资源<sup>[30]</sup>。自然语言处理领域中的各项子任务都需要使用到语料库资源,如命名实体识别、关系抽取、情感分析<sup>[31]</sup>、机器翻译<sup>[32]</sup>等。

目前,国内外比较大型的语料库有人民日报标注语料库<sup>[33]</sup>、北京大学汉语树库<sup>[34]</sup>、Penn Treebank<sup>[35]</sup>语料库等。然而这些都是通用的语料库,并不关注于灾害应急遥感领域。

灾害应急遥感影像应用案例知识由灾害应急遥感任务实体、遥感数据实体、方法实体、效果实体、时间实体、空间实体以及实体间的语义关系组成。在实体

标注方面,杜志强等标注了文献摘要中1034个句子中的模型方法实体,并使用条件随机场(CRF)模型识别标注的方法实体<sup>[36]</sup>。在关系标注方面,张雪英等对中文文本中地理空间关系进行研究,建立了地理空间关系标注语料库<sup>[37]</sup>。在标注体系设计方面,张春菊对中文文本中事件时空与属性信息的语言描述特点和语义结构进行研究,设计了事件时空与属性信息的标注体系<sup>[38]</sup>。在领域语料库构建方面,耿昊天对地理文本中的实体和关系进行研究,构建了地理领域标注语料库<sup>[39]</sup>。

综上所述,针对语料库的标注和构建已经有了广泛的研究。然而,当前在灾害应急遥感影像应用领域,面临着缺乏可获取、公开标注语料库的问题,这一问题亟待解决。此外,如何构建具有实际应用价值的语料库也是当前该领域的一大挑战。

因此,在灾害应急遥感影像应用案例知识语料库的构建上还存在三个问题:

### (1) 缺少遥感影像应用学术文本标注体系

目前关于时间、空间实体标注的研究比较多,而对于灾害应急遥感任务实体、遥感数据实体、方法实体以及效果实体的标注研究还比较缺乏,没有直接可应用的标注体系。如何设计面向灾害应急遥感任务实体、遥感数据实体、方法实体和效果实体的标注体系还有待研究。

### (2) 缺乏灾害应急遥感影像应用案例知识标注规范

灾害应急遥感应用学术文本摘要中的时间实体、空间实体、灾害应急遥感任务实体、遥感数据实体、方法实体、效果实体以及实体间关系的描述复杂多样,即使设计了标注体系,没有统一的标注规范,在标注过程中仍然存在不确定性,如何构建统一的标注规范还需要研究。

### (3) 缺乏遥感影像应用案例知识语料库

灾害应急遥感影像应用案例知识的获取需要语料库的支撑,而在灾害应急遥感影像应用领域,存在缺乏公开可用的语料库的问题;针对遥感影像应用案例知识语料库的构建还需要研究。

## 1.3 研究目标与内容

### 1.3.1 研究目标

针对目前灾害应急遥感影像应用案例知识语料库缺乏的问题，本文建立遥感影像应用学术文本标注体系，设计灾害应急遥感影像应用案例知识标注规范，构建遥感影像应用案例知识语料库，为以灾害应急遥感任务驱动的遥感数据发现提供支撑。

### 1.3.2 研究内容

#### (1) 灾害应急遥感影像应用案例知识建模

基于遥感在灾害应急中的应用，分析灾害应急遥感应用案例要素组成、要素关系及案例间的关系构建灾害应急遥感应用案例特征模型；分析灾害应急遥感应用学术文本摘要特征，设计遥感影像应用学术文本标注体系。

#### (2) 灾害应急遥感影像应用案例知识标注规范

针对灾害应急遥感应用学术文本摘要中实体和关系描述复杂多样的问题，为了规范语料库标注，针对命名实体，设计命名实体标注规范；针对实体间关系，设计句子级和篇章级的关系标注规范。

#### (3) 遥感影像应用案例知识语料库构建与分析

基于设计的灾害应急遥感影像应用案例知识标注规范，对通过数据清洗后得到的灾害应急遥感应用学术文本摘要进行标注；针对标注语料库，采用基于深度学习的方法进行验证，根据反馈的结果发现问题，优化语料库标注过程，并对构建的遥感影像应用案例知识语料库进行分析。

## 1.4 技术路线

首先通过梳理遥感在自然灾害应急当中的应用，通过分析灾害应急遥感应用案例要素组成、要素关系以及案例间关系构建灾害应急遥感应用案例特征模型；通过对灾害应急遥感应用学术文本摘要特征分析，建立遥感影像应用学术文本标注体系，设计灾害应急遥感影像应用案例知识标注规范；运用网络爬虫技术

获取原始数据，通过对原始数据进行清洗，从而获得灾害应急遥感应用学术文本摘要数据；基于设计的知识标注规范，对学术文本摘要数据进行标注；针对标注的遥感影像应用案例知识语料库，采用基于深度学习的方法验证并优化标注语料库，进而构建高质量语料库。技术路线如图 1.1 所示。

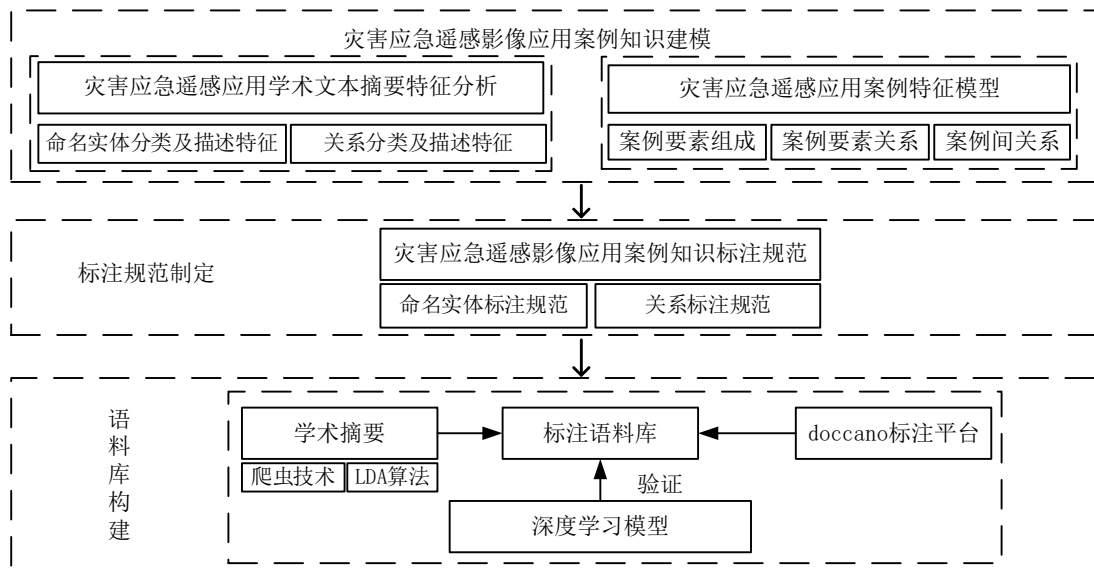


图 1.1 技术路线

## 1.5 本文的主要工作

全文主要分为五章，章节组织结构如下：

第一章节主要介绍了本文的研究背景以及现阶段相关理论与方法在国内外的研究进展，并且对本文的研究目标和研究内容以及技术路线进行阐述。

第二章节主要针对灾害应急遥感影像应用案例知识建模。首先梳理遥感在自然灾害应急中的应用；通过分析灾害应急遥感应用案例要素组成、要素关系以及不同案例间的关系，构建了灾害应急遥感应用案例特征模型；接着对灾害应急遥感应用学术文本摘要特征分析，设计了遥感影像应用学术文本标注体系。

第三章节在基于构建的遥感影像应用学术文本标注体系的基础之上，针对文献摘要中命名实体及关系描述复杂多样的问题，设计了灾害应急遥感影像应用案例知识标注规范。

第四章节在基于构建的灾害应急遥感影像应用案例知识标注规范基础之上，

首先使用网络爬虫技术从开放学术图谱中获取原始数据，接着对原始数据进行清洗，得到灾害应急遥感应用学术文本摘要数据；使用 doccano 标注平台结合实体及关系标注规范标注语料库；采用基于深度学习的模型进行验证并优化标注语料库，最后对标注语料库进行分析。

第五章节针对全文的研究进行了总结，此外，还对灾害应急遥感影像应用案例知识标注语料库构建研究进行展望。



## 第2章 灾害应急遥感影像应用案例知识建模

灾害应急遥感影像应用案例知识由时间实体、空间实体、灾害应急遥感任务实体、遥感数据实体、方法实体、效果实体及不同实体间的语义关联关系组成。本章节首先分析遥感在自然灾害当中的应用，其次建立灾害应急遥感应用案例特征模型，接着分析灾害应急遥感应用学术文本摘要的特征，最后设计遥感影像应用学术文本标注体系。

### 2.1 遥感灾害应急应用

#### 2.1.1 自然灾害分类

我国学术界对自然灾害知识的分类与组织的研究起步较晚,直到 20 世纪 80 年代末才逐渐有相关学者着力于灾害知识分类方面的研究。在对灾害知识的研究中,尤其对自然灾害种类的探讨最多;彭珂珊按成因将灾害分为了地质灾害、气象灾害、环境污染灾害、火灾、海洋灾害、生物灾害 6 类,又按不同表现方式细分为 44 个种类<sup>[40]</sup>。当然,自然灾害的分类原则和方法尚未统一,还有一些学者按照其他的分类标准对自然灾害进行了分类,如根据灾害波及范围分类、根据灾害持续时间的长短分类,根据灾害发生的先后关系分类、根据地貌类型分类、根据灾害形成和结束速度分类、根据灾害空间分布分类、根据灾害成因和我国灾害管理现状分类等<sup>[41]</sup>。

本文根据中华人民共和国国家标准《自然灾害分类与代码》(GB/T28921-2012),按照自然灾害最基本、最稳定的属性及其存在的逻辑关联进行灾害种类的划分。采用线分类法对自然灾害概念分类,形成灾害本体的概念框架。将自然灾害划分为气象水文灾害、地质地震灾害、海洋灾害、生物灾害和生态环境灾害共 5 大类灾害,简称灾类,灾类下又划分为 39 种灾害,简称灾种,灾种中设其他作为收容类目,涵盖了灾害领域所有可能涉及的灾害概念<sup>[42]</sup>。自然灾害概念层次分类如图 2.1 所示。



图 2.1 自然灾害概念层次分类

### 2.1.2 遥感在灾害应急中的应用

地震、洪水<sup>[43]</sup>、泥石流<sup>[44]</sup>等自然灾害对人类社会的稳定和发展造成巨大的负面影响。但是随着遥感技术<sup>[45]</sup>不断发展，遥感技术在灾害应急中发挥着越来越重要的作用，并在防灾减灾中做出贡献。遥感技术可以为不同类型的自然灾害应急提供从灾前的监测、灾中的应急响应、灾后的恢复重建<sup>[46]</sup>阶段的技术支撑。如图 2.2 所示。

遥感数据广泛应用于不同自然灾害应急任务中，在地震滑坡灾害监测方面，李强等使用高分卫星影像数据对四川九寨沟地震滑坡灾害进行普查<sup>[47]</sup>。在震后评估方面，焦其松等使用 GF-7 号卫星影像数据对青海门源地震导致的地表破裂特征进行解译分析<sup>[48]</sup>。在洪涝灾害过后，进行及时而有效的监测是防范和应对灾害的关键步骤之一，王景旭等采用 GF-3 号影像数据为数据源对郑州市市辖区水体提取信息，对郑州洪涝灾害进行监测<sup>[49]</sup>。在农业旱情灾中监测方面，刘立文等采用 MODIS 遥感影像数据，通过 TVDI 方法获取山西省农业旱情情况，为山西

省农业旱情的监测提供科学依据<sup>[50]</sup>。以上遥感数据在实际自然灾害应急中应用都说明遥感数据在自然灾害的监测<sup>[51]</sup>、评估等方面中存在着巨大的应用潜力。

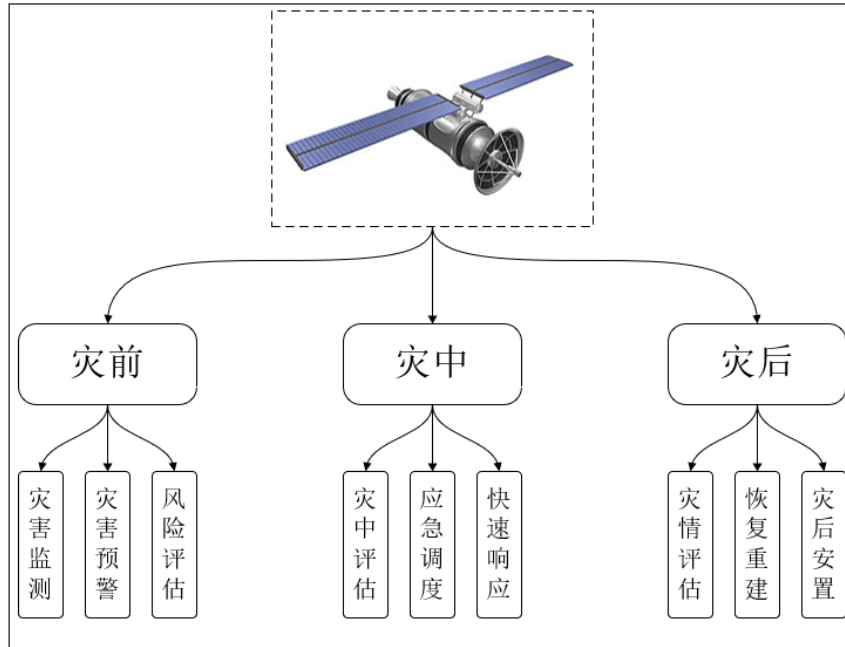


图 2.2 遥感在灾害应急中的应用

## 2.2 灾害应急遥感应用案例特征模型

### 2.2.1 灾害应急遥感应用案例要素组成

根据灾害应急遥感应用案例数据的语言和内容描述特点，对灾害应急遥感应用案例进行深入分析总结，案例的要素主要由灾害应急遥感任务要素、遥感数据要素、方法要素、效果要素、时间要素、空间要素等六类组成，这些要素共同构成了灾害应急遥感应用案例的主体成分，具体分类如下：

#### (1) 灾害应急遥感任务要素

在灾害应急遥感应用案例中灾害应急遥感任务要素根据自然灾害种类的不同可分为：干旱监测、洪涝损失评估、地震灾后评估、台风预警等。

#### (2) 遥感数据要素

灾害应急遥感应用案例中的遥感数据要素可分为：遥感卫星名称、成像模式、传感器、空间分辨率、遥感卫星简写、波段、遥感数据指代词。

(3) 方法要素

灾害应急遥感影像应用案例中方法要素可以分为：图像处理方法、空间分析方法、信息提取方法、模型类方法、分类方法、方法指代词。

(4) 效果要素

灾害应急遥感影像应用案例中效果要素分为定量效果要素和定性效果要素两大类。定量效果要素有：正确率分析、精度分析、误差分析、数值型表示；定性效果要素为定性分析。

(5) 时间要素

(6) 空间要素

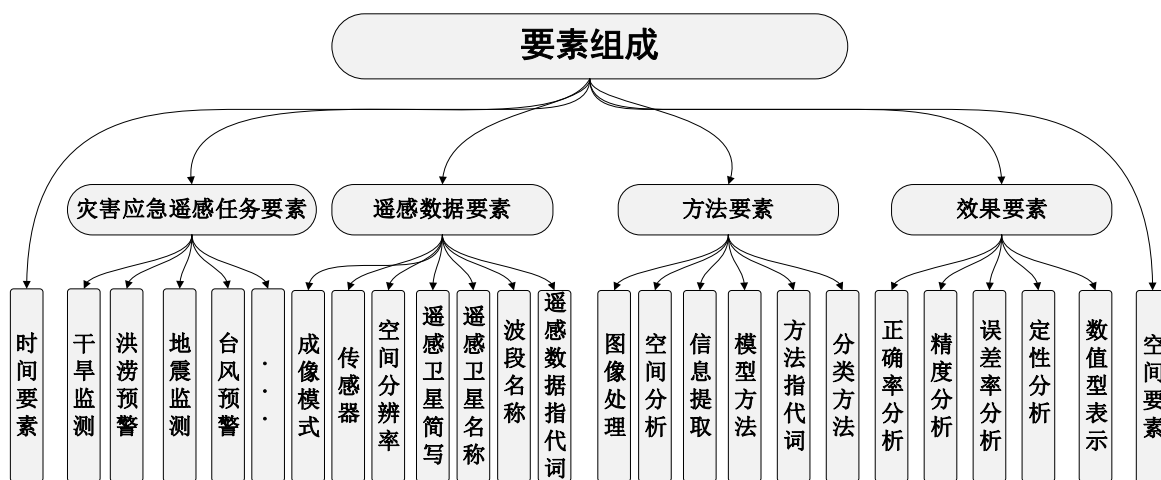


图 2.3 灾害应急遥感应用案例要素组成

### 2.2.2 灾害应急遥感应用案例要素关系

灾害应急遥感应用案例要素关系指的是灾害应急遥感应用案例中要素间的语义关联关系。要素间的关系可分为以下几大类：

- (1) 遥感数据要素间关系：具有传感器、具有分辨率、具有成像模式、具有波段、同义、数据指代。
- (2) 灾害应急遥感任务要素和遥感数据要素间关系：使用数据。
- (3) 方法要素间关系：并列、方法指代。
- (4) 效果要素间关系：具有数值。
- (5) 方法要素和效果要素间关系：具有效果。
- (6) 方法要素和灾害应急遥感任务要素：评估。

- (7) 方法要素和遥感数据要素间关系：使用方法、获取、数据输入。
  - (8) 灾害应急遥感任务要素间关系：灾害关联。
  - (9) 灾害应急遥感任务要素和时间要素间关系：具有时间
  - (10) 灾害应急遥感任务要素和空间要素间关系：具有地点
- 关系描述和例子如表 2.1 所示,使用三元组表达形式<主体要素 1, 关系, 客体要素 2>来描述要素间关系。

表 2.1 要素间关系表

序号	关系类别	关系名称	实例
1	遥感数据要素间关系	具有分辨率	例如: <HJ-1,具有分辨率,30m>
		具有传感器	例如: <GF-1,具有传感器,WFV>
		具有成像模式	例如: <TerraSAR-X,具有成像模式,ScanSAR>
2	灾害应急遥感任务和遥感数据要素间关系	具有波段	例如: <GF-3,具有波段,C band>
		同义	例如: <Gaofen-1,同义,GF-1>
		数据指代	例如: <its,数据指代,HJ-1>
3	方法要素间关系	使用数据	例如: <drought monitoring,使用数据,Terra>
		并列	例如: <SVM,并列,Multispectral transformation>
4	效果要素间关系	方法指代	例如: <This method,方法指代,decision tree>
5	方法和效果要素间关系	具有数值	例如: <accuracy rate,具有数值,50%>
6	方法和灾害应急遥感任务要素间关系	具有效果	例如: <SVM,具有效果, effective>
7	方法和遥感数据要素间关系	评估	例如: < digital elevation model,评估, analyze landslide >
		使用方法	例如: <AQUA,使用方法,decision tree >
		获取	例如: <HJ-1/B,获取,TVDI>

序号	关系类别	关系名称	实例
		数据输入	例如: <CCD,数据输入,Gaussian model >
8	灾害应急遥感任务要素间关系	灾害关联	例如: < earthquake investigation, 灾害关联, landslide investigation>
9	灾害应急遥感任务和时间要素间关系	具有时间	例如<drought monitor,具有时间,July 16 2022>
10	灾害应急遥感任务和空间要素间关系	具有地点	例如<flood monitoring ,具有地点,Ganzhou city>

### 2.2.3 灾害应急遥感应用案例间关系

自然灾害的产生和发展并非彼此独立，相继发生的灾害之间存在着一定的关联。初次灾害容易导致次生灾害的发生，而次生灾害又可能引发更加严重的新的灾害，从而形成一种连锁效应，即灾害链<sup>[52]</sup>。

地震滑坡灾害具有迅速发展、破坏性强的特点，给我国造成巨大财产损失和人员伤亡。地震应急遥感应用是遥感数据在地震灾害上面的实际应用，由于地震灾害的发生通常引发滑坡次生灾害，遥感数据同样应用于监测滑坡次生灾害，通过地震应急遥感任务和滑坡灾害应急遥感任务之间的“灾害链”关联性，可以建立起不同灾害应急遥感应用案例间的联系。

本文在完成构建灾害应急遥感应用案例要素组成和要素关系的基础之上，以 2017 年 8 月 18 日发生在四川九寨沟的地震以及次生灾害滑坡应急遥感案例<sup>[47]</sup>为实例，通过“灾害关联”关系连接“九寨沟地震应急任务”和“九寨沟滑坡应急任务”，构建了不同灾害应急遥感应用案例之间的关系，如图 2.4 所示。

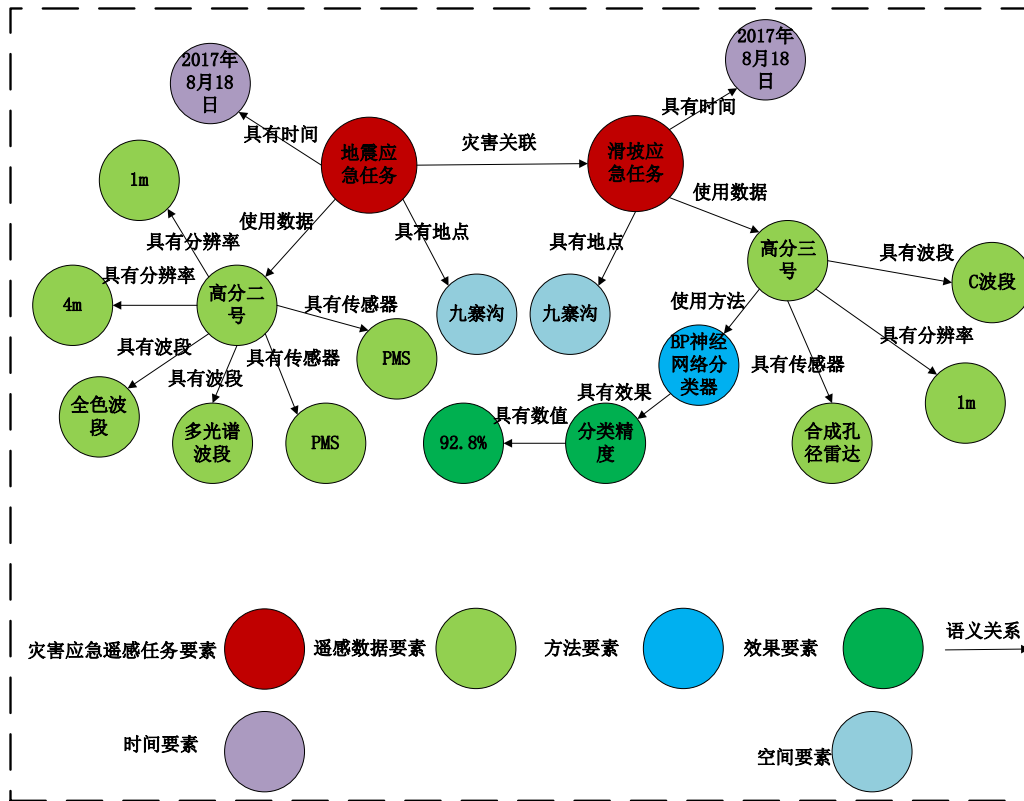


图 2.4 四川九寨沟地震滑坡应急遥感应用

## 2.3 灾害应急遥感应用学术文本摘要特征分析

### 2.3.1 命名实体分类及描述特征

本文在大量阅读灾害遥感应用学术文本摘要的基础之上，根据构建的灾害应急遥感应用案例组成要素，总结了不同种类的命名实体在文献摘要中的描述特征：

#### (1) 灾害应急遥感任务实体

灾害应急遥感任务实体是灾害应急遥感应用学术文本摘要的重要组成部分，本文通过对灾害应急遥感应用学术文本摘要深入分析，总结灾害应急遥感任务实体的描述特征：

##### ✓ 灾害应急遥感任务实体的组成特性

灾害应急遥感任务实体中通常会包含 detection、monitoring、assessment 等特

征词，上下文中会存在自然灾害类别，例如：“detection of flooded areas in high-resolution images”。

✓ 灾害应急任务遥感任务实体的细分粒度

根据自然灾害种类的不同，存在多种灾害应急遥感任务实体。例如“drought monitoring、monitoring and evaluation of flood disaster”等。

(2) 遥感数据实体

遥感数据实体是灾害应急遥感应用学术文本摘要不可或缺的组成成分，贯穿于整个灾害应急遥感应用学术文本摘要，是灾害应急遥感应用学术文本摘要的重要特征之一，本文通过对灾害应急遥感应用学术文本深入分析，总结了遥感数据实体的描述特征：

✓ 遥感数据实体的属性

在遥感数据实体中，传感器(CCD)、分辨率(2m resolution)、成像模式(scanSAR)、波段(L band)等，均为遥感卫星的基本属性。

✓ 遥感数据实体的别名

在灾害应急遥感应用学术文本摘要中，对同一对象会有不同表达，通常分为全称和简写，这也是灾害应急遥感应用学术文本摘要的非常普遍的特征之一。例如“Gao-Fen-3”和“GF-3”。

✓ 遥感数据指代

在灾害应急遥感应用案例中，遥感数据指代词用于指示句内或者句间中的遥感数据实体，通常以“data”结尾的指示性短语。例如“the three temporal data”指示性短语指代前文出现的遥感卫星数据实体。

(3) 方法实体

方法实体也是灾害应急遥感应用学术文本的重要组成部分之一。根据在自然灾害应急过程中的具体任务的不同，对遥感数据的处理也会存在差异，本文基于对灾害应急遥感应用学术文本深入分析的基础之上，总结了方法实体在灾害应急任务过程中应用领域的描述特征：

✓ 方法实体细分粒度

方法实体按照功能的不同，可分为图像处理、空间分析、分类方法，这些方法针对遥感数据实体具有不同的处理功能。

✓ 方法实体的别名

在灾害应急遥感应用学术文本摘要中，方法实体的存在别名的描述，信息提



取方法实体则存在别名的描述特征，例如：Normalized Difference Vegetation Index 和 NDVI 同样都在描述植被归一化指数。

✓ 方法实体的指代

方法指示代词是在灾害应急遥感应用学术文本中对上下文中的方法实体进行指代，通常为指示代词或者短语 that approach、this method 等。

✓ 方法实体组成特性

模型类方法实体的组成中通常会存在 model 特征词结尾，例如：Gaussian model。

(4) 效果实体

在灾害应急遥感应用任务中，采用不同的遥感数据处理方法会产生不同的效果。通过对灾害应急遥感应用学术文本摘要数据深入分析，本文从对方法实体的评价角度总结了效果实体的描述特征：

✓ 效果实体细分粒度

按照对方法实体的评价角度进行划分，效果实体可分为定量评价和定性评价，定性评价效果实体一般是形容词，例如：effective、accurate、practical 等。

✓ 效果实体的组成特性

在灾害应急遥感应用学术文本摘要中，定量效果实体包含正确率、精确率、误差率等定量评价效果实体，正确率在组成上会以“rate、accuracy”，精确率在组成上会以“precision”后缀结尾，误差评价会以“error”结尾。

(5) 时间实体

时间是灾害应急遥感应用学术文本中的重要特征，是灾害事件发生的重要组成，是灾害应急遥感领域不可或缺的元素，本文总结了其在学术文本中的描述特征：

✓ 时间实体的组成特性

在灾害应急遥感应用文本学术摘要中，时间实体在组成上通常包含“年、月、日”。例如“August 21 2022”

(6) 空间实体

空间实体是灾害应急遥感应用学术文本摘要中的重要特征，是灾害事件发生的重要组成，是灾害应急遥感领域不可或缺的元素，本文总结了其在学术文本中的描述特征：

✓ 空间实体的组成特性

在灾害应急遥感应用学术文本摘要中，空间实体在组成上通常包含“city、province、region”等词为后缀。例如“nanchang city、jiangxi province”。

### 2.3.2 命名实体关系及描述特征

通过对灾害应急遥感应用学术文本摘要的特征分析可知，关系连接了灾害应急遥感应用学术文本摘要中的各种命名实体，使得灾害应急遥感应用学术文本摘要的语义表达更加丰富。本文在大量阅读灾害遥感应用学术文本摘要的基础上，根据构建的灾害应急遥感应用案例要素关系，总结了命名实体关系在灾害应急遥感应用学术文本摘要中的描述特征：

(一) 句子级关系及描述特征。

句子级关系以句子为单位,在灾害应急遥感影像应用案例文本数据中涉及的命名实体以及关系必须位于同一句子中。

(1) 遥感数据实体间的关系

遥感数据实体间的关系包含“具有传感器”、“具有分辨率”、“具有波段”、“具有成像模式、同义、数据指代”六种。具体描述特征及关系实例如下：

✓ “具有传感器”关系

遥感卫星和其传感器之间存在这样的语义关联关系，“具有传感器”关系的表达，必然包含遥感卫星和传感器。例如：“A drought monitoring method based on HJ-1A / 1B CCD data”。

✓ “具有分辨率”关系

遥感卫星和分辨率之间存在这样的语义关联关系，通常“具有分辨率”关系的表达，在句中必然包含遥感卫星和分辨率。例如“HJ-1A CCD with 150 m spatial resolution”。环境一号 A 卫星传感器 CCD 的分辨率为 150m。

✓ “具有成像模式”关系

遥感卫星和其成像模式之间存在这样的语义关联关系。“具有成像模式”关系的表达，在句中必然包含遥感卫星和其成像模式。例如“HJ-1A/B also has a hyperspectral imaging mode and a stereo imaging mode”。句中环境一号卫星、高光谱成像模式、立体成像模式可建立语义关联。

✓ “具有波段”关系

遥感卫星和其成像波段之间存在这样的语义关联关系。“具有波段”关系的

表达，在句中必然包含遥感卫星和其成像波段。例如：“The C-band ENVISAT advanced synthetic aperture radar data of vertically transmitted horizontally received VH” ENVISAT 卫星和其成像波段 C 波段有这样的语义关联。

✓ “同义”关系

同义关系的表达见于灾害应急遥感应用学术文本中遥感卫星实体和其简写实体间。例如：“Chinese GaoFen-1 (GF-1) optical satellite has a 16-m resolution”。句中 GaoFen-1 和 GF-1 间存在同义关系。

✓ “数据指代”关系

“数据指代”关系的表达见于遥感数据实体和数据指示代词之间。例如：“RADARSAT-1 has been applied in flood control for its characteristics of all-weather” its 指示代词指代遥感卫星 RADARSAT-1。

(2) 方法实体间的语义关系

方法实体间的关系包含：“方法指代”、“并列”两种；具体关系实例及描述特征如下：

✓ “方法指代”关系

“方法指代”关系的表达见于方法实体和指示代词之间。例如：“the method is principally based on an image segmentation technique.” 句中 “the method” 和 “based on an image segmentation technique” 之间存在“方法指代”关系。

✓ “并列”关系

在灾害应急遥感任务中不同种类方法实体间存在这样的关系。并列关系的表达，在句中必然包含两种或两种以上方法实体。例如：“normalized difference vegetation index (NDVI) values and tasseled cap (TC) transformations of Terra”。句中信息提取方法实体 NDVI 和图像处理方法实体 tasseled cap (TC) transformations 构成并列关系。

(3) 灾害应急遥感任务和遥感数据实体间的语义关系

灾害应急遥感任务和遥感卫星数据实体间存在着关系的表达，具体关系实例及描述特征如下：

✓ “使用数据”关系

在灾害应急遥感任务过程中任务实体和方法实体间存在这样的关系。“使用数据”关系的表达，在句中必然包含灾害应急遥感任务实体以及遥感卫星数据实体。例如：“Estimation of the building damages in Yushu earthquake based on ALOS

SAR data”。句中灾害应急遥感任务“玉树地震建筑物损失估计”和“ALOS”卫星可建立“使用数据”关系。

(4) 效果实体间的语义关系

效果实体间的语义关系存在于不同种类的效果实体之间，具体关系实例及描述特征如下：

✓ “具有数值”关系

“具有数值”关系的表达，在句中必然包含定量效果实体和其数值型描述效果实体。例如：“70% classification precision”句中定量效果实体“classification precision”和数值描述效果实体“70%”存在“具有数值”关系。

(5) 方法和效果实体间的语义关系

方法实体和效果实体在句中存在语义关联关系，具体关系实例及描述特征如下：

✓ “具有效果”关系

“具有效果”关系的表达。句中必然包含方法实体和效果实体。例如“SVM method provides a more accurate analysis”。句中方法实体“SVM”和定性评价效果实体“accurate analysis”存在“具有效果”的关系。

(6) 遥感数据和方法实体间的语义关系

遥感数据和方法实体间的关系包含“获取”、“数据输入”、“使用方法”，具体关系实例及描述特征如下：

✓ “获取”关系

“获取”关系的表达，在句中必然包含信息提取方法实体和遥感卫星数据实体。例如：“NDVI was calculated from HJ-1 CCD data.”信息提取方法实体“NDVI”和“HJ-1”遥感卫星数据实体建立“获取”关系。

✓ “数据输入”关系

数据输入关系的表达，在句中必然包含模型类方法实体和传感器数据实体。例如“The radiative transfer model was used to eliminate difference between OLI sensor and WFV1 camera through the spectral match factor(SMF).”句中模型类方法实体 radiative transfer model 和传感器影像数据 WFV1 间的语义关联。

✓ “使用方法”关系

使用关系的表达，在句中必然包含遥感卫星数据实体和方法实体。例如：“HJ-1B data to monitor ice flood of Yellow River and propose a decision tree method to

extract river ice.”句中遥感卫星数据实体“HJ-1B”和方法实体“decision tree”间存在“使用方法”关系。

(7) 方法和灾害应急遥感任务实体间关系

在灾害应急遥感应用学术文本摘要句子中，方法实体和灾害应急遥感任务实体间存在着关系的表达。具体关系实例及描述特征如下：

✓ “评估”关系

“评估”关系的表达，在句中必然包含模型类方法实体和灾害应急遥感任务实体。例如：“digital elevation model(DEM) were combined to analyze landslide.”句中的模型类方法实体“digital elevation model”和任务实体“analyze landslide”建立“评估”关系。

(8) 灾害应急遥感任务实体间语义关系

由于灾害之间存在“灾害链”的关联，使得不同灾害应急遥感任务实体间也建立起了语义关联。具体关系实例及描述特征如下：

✓ “灾害关联”关系

“灾害关联”关系的表达，在句中必然包含灾害和次生灾害应急遥感任务实体。例如：“earthquake emergency investigation”地震应急调查和“landslide investigation”滑坡勘察之间存在“灾害关联”关系。

(9) 灾害应急遥感任务和时间实体间语义关系

灾害应急遥感应用学术文本摘要具有很强的时间特征，使得灾害应急遥感任务实体和时间实体之间存在语义关联，具体关系实例及描述如下：

✓ “具有时间”关系

“具有时间”关系的表达，在句中必然包含灾害应急遥感任务实体和时间实体，表现的是灾害应急过程的时间特性。

(10) 灾害应急遥感任务和空间实体间语义关系

灾害应急遥感应用学术文本摘要具有很强的地理位置特征，使得灾害应急遥感任务实体和空间实体之间存在语义关联，具体关系实例及描述如下：

✓ “具有地点”关系

“具有地点”关系的表达，在句中必然包含灾害应急遥感任务实体和空间实体，表现的是灾害应急过程的地理位置特性。

(二) 篇章级关系及描述特征。

篇章级关系不同于句子级关系，篇章级关系的表达，在灾害应急遥感应用学

术文本摘要中跨越了不同的句子。可划分为“数据指代”和“方法指代”，具体描述及实例如下：

(1) “数据指代”关系

“数据指代”关系存在于遥感卫星数据和遥感卫星数据指示代词之间，具体描述特征如下：

篇章级的数据指代关系与句子级的数据指代关系不同，篇章级的数据指代关系跨越了不同的句子。例如：一个句子中的遥感数据指示代词“this data”指代另一个句子中的遥感卫星数据“spot-5”。通过这种跨句的遥感卫星数据指代，表达了篇章级的“数据指代”关系。

(2) “方法指代”关系

“方法指代”关系存在于方法实体和方法实体指示代词之间，具体描述特征如下：

篇章级的方法指代关系与句子级的方法指代关系不同，篇章级的方法指代关系同样跨越了不同的句子。例如：一个句子中的方法指示代词“This method”指代另一个句子中的分类方法“logistic regression”。这种跨句的方法实体指代，表达了篇章级的“方法指代”关系。

## 2.4 遥感影像应用学术文本标注体系设计

遥感影像应用学术文本标注体系的设计是遥感影像应用案例知识语料库构建的重要环节，而设计标注体系的目的是为了给灾害应急遥感应用学术文本摘要数据中的实体及关系构建一系列的“标签”。标注体系设计过程遵循以下原则：

(1) 准确性

优质的标注体系对于标注语料库至关重要，为了做到科学而专业，在标注体系设计过程中应该参考领域专家的建议。

(2) 可读性强

由于灾害应急遥感应用学术文本摘要中的实体及关系种类丰富，设计的标记类型必须具备较强的可读性和辨识度。

在遵循遥感影像应用学术文本标注体系原则的基础之上，结合灾害应急遥感应用学术文本摘要中命名实体及关系的描述特征，设计了命名实体及关系的标记类型：

- a) 根据自然灾害种类的不同,建立了不同类型的灾害应急遥感任务实体标记类型,如表 2.2 所示。以 task 开头的标记类型代表灾害应急遥感任务实体。

表 2.2 灾害应急遥感任务实体标记类型

灾害应急遥感任务实体	灾害应急遥感任务实例	标记类型
干旱应急遥感任务	Drought monitoring	taskDrought
洪水应急遥感任务	Flood assessment	taskFlood
地震应急遥感任务	Earthquake damage assessment	taskEarthquake
暴雨应急遥感任务	Detection of rainstorm area	taskRainstorm
风雹应急遥感任务	monitoring hail at regional level	taskHail
...	...	...

- b) 遥感数据实体。针对遥感数据实体的细粒度划分,分别设计了标记类型,如表 2.3 所示。注:以 data 开头的标记类型代表遥感数据实体。

表 2.3 遥感数据实体标记类型

遥感数据实体	遥感数据实例	标记类型
遥感卫星	GF-1、gaofen-1 号	data
遥感卫星简写	GF-1、ALOS	dataAbbre
传感器	WV、CCD	dataSensor
波段	X-band、C-band	dataBand
空间分辨率	2m resolution	dataResolution
遥感数据指代词	this data、that images	dataDemo
成像模式	Wide Swath Mode	dataMode

- c) 方法实体。针对灾害应急遥感应用任务从遥感数据中获取的信息不同,对应的处理方法存在差异,因此,本文对细粒度方法实体分别设计了标记类型,如表 2.4 所示。注:以 met 开头的标记类型代表方法实体。

表 2.4 方法实体标记类型

方法实体	方法实例	标记类型
分类方法	Bayesian Classifier、decision tree	metClassify
图像处理方法	Multispectral transformation、	metImage
信息提取方法	NDVI、PDI、TVDI	metInfor
模型类方法	Gaussian model、geometric model	metModel

第 2 章 灾害应急遥感影像应用案例知识建模

方法实体	方法实例	标记类型
方法指代词	This method 、 approach	metDemo
空间分析方法	network Analysis、	metSpatial

d) 效果实体。根据效果实体的细粒度划分，本文设计了不同效果实体的标记类型，如表 2.5 所示。注以 eft 开头的标记类型代表效果实体。

表 2.5 效果实体标记类型

效果实体	效果实例	标记类型
精度分析	classification precision	Eftprecision
误差分析	error、 Mean square error	Efterror
正确率分析	accuracy、 success rate	Eftaccuracy
数值型表示	70%、 0.2	Eftnumber
定性分析	correctly、 practical、 suitable	Eftqualitative

e) 时间和空间实体，针对时间以及空间实体，本文设计了其标记类型，如表 2.6 所示。

表 2.6 时间和空间实体标记类型

实体	实例	标记类型
时间实体	July 21 2021	time
空间实体	Ganzhou city	loc

f) 关系，针对实体间的关系，本文设计了关系的标记类型。如表 2.7 所示。

表 2.7 关系标记类型

序号	关系类别	关系名称	标记类型
1	遥感数据实体间关系	具有分辨率	hasRes
		具有传感器	hasSensor
		具有成像模式	hasMode
		具有波段	hasBand
		数据指代	dataAnap
2	应急遥感任务和遥感数据实体间关系	同义	equal
		使用数据	useData
3	方法实体间关系	并列	juxtaposit
		方法指代	metAnap



序号	关系类别	关系名称	标记类型
4	效果实体间关系	具有数值	hasValue
5	方法和效果实体间关系	具有效果	hasEft
6	方法和应急遥感任务实体间关系	评估	evaluate
		使用方法	useMethod
7	方法和遥感数据实体间关系	获取	gain
		数据输入	dataInput
8	灾害应急遥感任务实体间关系	灾害关联	disCorrel
9	应急遥感任务和时间实体间关系	具有时间	hasTime
10	应急遥感任务和空间实体间关系	具有地点	hasLoc

## 2.5 本章小结

本章节首先梳理了遥感在自然灾害应急中的应用，通过分析灾害应急遥感应用案例的要素组成、要素关系以及不同灾害应急遥感应用案例之间的关系，构建了灾害应急遥感应用案例特征模型；接着对灾害应急遥感应用学术文本摘要中的命名实体及关系分类，并且分别阐述了其描述特征；最后针对灾害应急遥感任务实体、遥感数据实体、方法实体、效果实体、时间实体以及空间实体这六大类别实体，设计了实体标记类型；针对实体间的 18 种关系设计了关系的标记类型。

## 第3章 灾害应急遥感影像应用案例知识标注规范

本章节在第二章节构建的遥感影像应用学术文本标注体系的基础之上,针对遥感影像应用学术文本中的灾害应急遥感任务实体、遥感数据实体、方法实体、效果实体、时间实体和空间实体等六大类命名实体设计标注规范;其次,针对遥感影像应用学术文本中实体间关系设计句子级和篇章级的标注规范。

### 3.1 命名实体标注规范

由于灾害应急遥感应用学术文本摘要中命名实体的表达复杂多样,为了保证语料库的标注质量,需要对命名实体进行规范标注。由于命名实体具有不同的特征,在标注过程中,需要结合其具体特征进行规范化标注。各类命名实体的特征如表 3.1 所示,表格中 Y 符号代表存在, N 符号代表不存在。

表 3.1 命名实体特征表

特征 命名实体	组成特性 (Y/N)	细分粒度 (Y/N)	别名 (Y/N)	指代 (Y/N)	属性 (Y/N)
灾害应急遥感任务实体	Y	Y	N	N	N
遥感数据实体	N	N	Y	Y	Y
方法实体	Y	Y	Y	Y	N
效果实体	Y	Y	N	N	N
时间实体	Y	N	N	N	N
空间实体	Y	N	N	N	N

#### 3.1.1 灾害应急遥感任务实体标注规范

基于构建的命名实体特征表,通过结合灾害应急遥感任务实体的组成特性、细分粒度特征分别进行规范化标注,具体如下:

- (1) 细分粒度标注

由于自然灾害种类繁多，灾害应急遥感任务实体也可划分为不同种类粒度。针对这样的情况，应对不同类型的灾害应急遥感任务实体分别标注。标注实例如图 3.1 和图 3.2 所示：

The study is aimed to use CCD and IRS data on HJ-1A/B satellite to monitor drought.  
•taskDrought

图 3.1 干旱监测任务实体标注示意图

图 3.1 中“monitor drought”被标注为“taskDrought”标记类型，“monitor drought”两个词整体表达的是对干旱的监测，需要整体标注。

NASA's Aqua satellite are used here to demonstrate the potential  
for improving streamflow forecasts by using remotely sensed  
•taskFlood  
surface soil moisture during a flooding event in northeastern  
Australia (Queensland) during January-February 2004.

图 3.2 洪水流量预测任务实体标注示意图

图 3.2 中“improving streamflow forecasts”被标注为“taskFlood”标记类型，“improving streamflow forecasts”整体表达的是洪涝流量监测，需要整体标注。

以上两例标注示意图，分别描述了干旱监测任务、洪水流量预测任务实体的标注。

## (2) 组成特性标注

在灾害应急遥感应用学术文本摘要中，采用遥感数据进行具体的灾害应急遥感任务的过程中会存在两种及以上由同种自然灾害引发的灾害应急遥感任务实体，通常使用“and”连接词连接。针对这种情况，考虑其组成特性，需要整体标注。如图 3.3 所示：

This project investigated the need for continuous monitoring of levee and boils during  
•taskFlood  
the event of a flood emergency using LiDAR.

图 3.3 同类别灾害应急遥感任务实体标注示意图

图 3.3 中由洪水灾害引发的灾害应急遥感任务实体“continuous monitoring of levee and boils”中存在连接词“and”,为了保证语义的完整性,针对这样的灾害应急遥感任务实体需要整体标注为“taskFlood”。

### 3.1.2 遥感数据实体标注规范

基于构建的命名实体特征矩阵,针对遥感数据实体从别名、属性、指代三个方面设计了标注规范,具体如下:

#### (1) 别名标注

在灾害应急遥感应用学术文本摘要中,对于同一遥感卫星数据在句中会有不同的名称表达形式,这是遥感领域学术文献中比较明显的描述特征;针对这样的情况,在标注时,需要将遥感卫星名称和其简写形式分开标注。标注实例如图 3.4 所示:

The objective of this paper is to evaluate the potential of Gaofen-2 (GF-2) high resolution multispectral sensor (MS) and panchromatic (PAN) imagery on water mapping.

•data      •dataAbbre

图 3.4 遥感卫星数据别名标注示意图

图 3.4 中“GF-2”是遥感卫星“Gaofen-2”的简写形式,针对这样的情况,需要分开标注。

#### (2) 属性标注

在灾害应急遥感应用任务过程中会使用到传感器、空间分辨率、波段、成像模式,而这些都属于遥感卫星具有的基本属性,需要使用不同的标记类型区分标注。标注实例如图 3.5 和图 3.6 所示:

HJ-1 CCD image in detecting landscape change in earthquake areas.

•dataAbbre  
•dataSensor

图 3.5 遥感卫星属性标注示意图



模型类方法实体是指在灾害应急遥感应用任务中对于不能直接进行研究的对象进行处理的方法实体，后面通常跟随 model 特征词，针对这种方法实体需要整体标注。标注实例如图 3.8 所示：

Surface change caused by the earthquake has been automatically detected based on the generalized Gaussian model and KI optimal threshold value change detection  
•metModel  
method.

图 3.8 方法组成特性标注示意图

图 3.8 中的“Gaussian model”属于模型类方法实体，将“Gaussian model”标注为“metModel”。

### (2) 别名标注

信息提取方法实体是指在灾害应急遥感应用任务中从遥感数据中提取相关指数信息的方法实体，例如“归一化差异植被指数、垂直干旱指数”等，这一类方法实体主要通过从遥感卫星影像数据中提取信息，标注实例如图 3.9 所示：

The normalized difference vegetation index (NDVI) was calculated from HJ-1 CCD data.  
•metInfor •metInfor

图 3.9 方法实体别名标注示意图

图 3.9 中的归一化差异植被指数“normalized difference vegetation index”及“NDVI”属于信息提取方法实体，将其标注为“metInfor”。

### (3) 细分粒度标注

图像处理方法实体是指对遥感数据采用诸如“多光谱变换、几何校正等”一类方法进行处理实体。这一类方法实体是对遥感数据进行图像处理的角度进行处理分析。标注实例如图 3.10 所示：

facilitates image enhancement and the numerical comparison of different image takes  
•metImage  
together with data fusion and visualization processes.

图 3.10 图像处理方法实体标注示意图

图 3.10 中的“image enhancement”属于图像处理方法实体，将其标注为“metImage”。

分类方法实体指对遥感数据采用诸如“贝叶斯分类、决策树”等一类方法进行处理实体。这一类方法实体是通过对遥感数据从分类的角度进行处理分析。针对分类方法实体，标注实例如图 3.11 所示：

Polarimetric decomposition-based unsupervised Wishart classification is combined  
•metClassify  
  
with object-based post-classification refinement.  
•metClassify

图 3.11 分类方法实体标注示意图

图 3.11 中的“Wishart classification”和“object-based post-classification”属于分类方法实体，标注为“metClassify”。

空间分析方法实体是指在灾害应急遥感应用任务中使用空间分析类方法实体对遥感数据进行处理实体，诸如“网络分析、叠加分析”等空间分析方法。标注实例如图 3.12 所示：

a simulated image with random bright squares, as building objects, is firstly presented  
  
for analysis of the K-distribution and Getis statistics.  
•metSpatial

图 3.12 空间分析方法实体标注示意图

图 3.12 中的“Getis statistics”属于空间分析方法实体，针对这样的实例，需要将“Getis statistics”标注为“metSpatial”。

#### (4) 指代标注

方法指代实体指的是以“method、approach”等后缀结尾的指示性短语。这一类实体会对句中的方法实体进行指代。针对存在指示性短语对方法实体指代的情况，需要对这种方法指代实体需要进行标注。标注实例如图 3.13 所示：





数值型表示效果实体，体现了正确率评价的定量性。标注实例如图 3.15 所示：

The results of the accuracy assessment showed a success rate of 90.4 and 91.6 % and  
•Eftaccuracy •Eftnumber  
a prediction rate of 89.6 and 91.3 % for FR and ensemble FR and LR models,  
respectively.

图 3.15 正确率评价效果实体标注示意图

图 3.15 中的正确率评价效果实体“success rate”被标注为“Eftaccuracy”标记类型。句中后半部分的数值型实体“90.4 and 91.6 %”需整体标注为“Eftnumber”。

误差评价效果是指采用诸如“Error、Mean Square Error”等评价指标的一类效果实体。这也是灾害应急遥感应用任务中常用的对方法实体的评价方式，实体中常以“Error”后缀结尾。并且句中后半部分会有对数值型实体对误差评价效果进行定量描述。标注实例如图 3.16 所示：

The validation between measured and SVMs derived values indicate a strong positive correlation of 0.88  
and a low value of Root Mean Square Error of 12.62.  
•Efterror

图 3.16 误差评价效果实体标注示意图

图 3.16 中“Root Mean Square Error”均方根误差实体被标注为“Efterror”。

精度评价效果实体也是灾害应急遥感应用任务中常用的对方法实体的评价方式，实体常以“precision”后缀结尾。此外，句中后半部分会有对数值型实体对精度评价效果实体补充。标注实例如图 3.17 所示：

The classification precision of this method reaches 80%.  
•Eftprecision •Eftnumber

图 3.17 精度评价效果实体标注示意图

图 3.17 分类精度实体“classification precision”被标注为“Eftprecision”，数值型实体“80%”使用“Eftnumber”标记类型标注。

### 3.1.5 时间实体标注规范

基于构建的命名实体特征矩阵，可对时间实体的标注进行说明。具体标注如下：

#### (1) 组成特性标注

时间实体是对灾害应急遥感应用学术文本摘要的重要特征之一，通常是以具体的日期为主，实体组成上通常是包含年、月、日。例如：“June 10 2020、2021”等，标注实例如图 3.18 所示：

The present studies deal with the tsunami damage assessment of the 26 December 2004 tsunami that  
•time  
occurred as a result of an Indonesian earthquake (Mw9.3) on Car Nicobar Island.

图 3.18 时间实体标注示意图

如图 3.18 所示，时间实体“26 December 2004”被标注为 time 标记类型。

### 3.1.6 空间实体标注规范

#### (1) 组成特性标注

基于构建的命名实体特征矩阵，可对空间实体的标注进行说明。具体标注如下：

空间实体是对灾害应急遥感应用学术文本摘要重要特征之一，通常是以具体的地点为主，例如：“NanChang city、Hunan Province”等，在实体组成上会包含“city、Province、region”等后缀。标注实例如图 3.19 所示：

We apply high-resolution, X-band, stripmap COSMO-SkyMed data  
to the monitoring of flood events in the Basilicata region.  
•loc

图 3.19 空间实体标注示意图

如图 3.19 所示，空间实体“Basilicata region”被标注为 loc 标记类型。

### 3.2 句子级关系标注规范

句子级的关系标注规范以句子为单位,即句子级关系涉及的命名实体在灾害应急遥感应用学术文本数据中必须包含在同一个句子中。此外,实体间的语义关联关系和实体紧密相关,因此在标注过程中需要考虑到两者之间是否存在语义关系,从而进行实体间关系标注,可由关系存在判别矩阵从而识别实体间是否存在关系,如表 3.2 所示。表格中 Y 符号代表两类实体间存在关系, N 符号代表两类实体间不存在关系

表 3.2 实体间关系表

命名实体 \ 命名实体	灾害应急遥感任务实体	遥感数据实体	方法实体	效果实体	时间实体	空间实体
灾害应急遥感任务实体	Y	Y	Y	N	Y	Y
遥感数据实体	Y	Y	Y	N	N	N
方法实体	Y	Y	Y	Y	N	N
效果实体	N	N	Y	Y	N	N
时间实体	Y	N	N	N	N	N
空间实体	Y	N	N	N	N	N

基于实体间关系表可以得出存在关联关系的实体,从而进行关系的进一步标注。

#### 3.2.1 遥感数据实体间的关系标注规范

本文标注的遥感数据实体间关系集中于灾害应急遥感任务过程中使用的遥感卫星和其它遥感数据实体之间,可细分为“具有分辨率”、“具有传感器”、“具有成像模式”、“具有波段”“同义”“数据指代”六种。具体标注如下:

##### (1) “具有分辨率”关系标注

“具有分辨率”关系是遥感卫星和其属性分辨率间存在的关系。标注实例如图 3.20 所示:

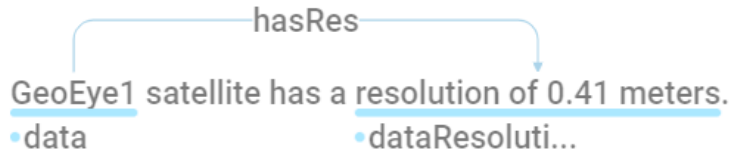


图 3.20 “具有分辨率”关系标注示意图

图 3.20 中的关系由“GeoEye1”遥感卫星主体指向“resolution of 0.41 meters”分辨率客体。

(2) “具有传感器”关系标注

“具有传感器”关系是遥感卫星和其组成部分传感器间存在的关系，标注实例如图 3.21 所示：

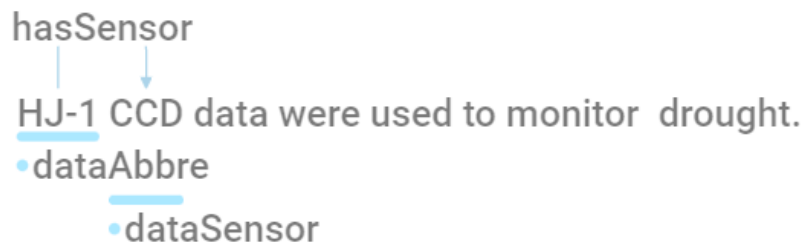


图 3.21 “具有传感器”关系标注示意图

图 3.21 中的关系由“HJ-1”遥感卫星主体指向“CCD”传感器客体。

(3) “具有成像模式”关系标注

“具有成像模式”是遥感卫星和其成像模式间存在的关系。标注实例如图 3.22 所示：

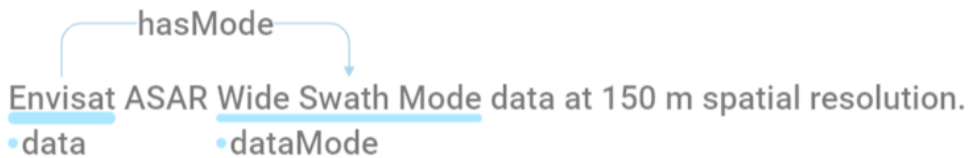


图 3.22 “具有成像模式”关系标注示意图。

图 3.22 中的关系由“Envisat”遥感卫星主体指向“Wide Swath Mode”成像模式客体。

(4) “具有波段”关系标注

“具有波段”关系的表达存在于遥感卫星与其波段之间。标注实例如图 3.23

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/678000072010006030>