

# 意外爆炸毁伤知识图谱研究

王继民<sup>1</sup>, 姜灿<sup>1</sup>, 韩斌<sup>1</sup>, 王幸<sup>2,3</sup>, 张磊<sup>2,3\*</sup>

(1. 河海大学计算机与软件学院, 江苏 南京 211100;

2. 目标易损性评估全国重点实验室, 河南 洛阳 471023;

3. 军事科学院国防工程研究院工程防护研究所, 河南 洛阳 471023)

**摘要:** 利用具有多源、异构、重叠等特征的爆炸事故调查报告建立意外爆炸毁伤知识图谱, 对进行数据驱动的爆炸评估以及溯源具有重要的作用。针对意外爆炸事故调查数据中存在重叠和嵌套事件的特点, 采用以事件联合抽取为核心的知识图谱构建方法以及爆炸调查报告构建了意外爆炸毁伤知识图谱; 通过余弦相似度在知识图谱中检索类似爆炸事件并采用贝叶斯分类方法进行分类, 较准确地实现了对贝鲁特港口爆炸事故爆炸源物资种类的确定。知识图谱构建结果表明, 在意外爆炸毁伤语料库上的事件分类以及事件元素分类分析, 相较于现有抽取模型, 提出的基于动态掩码的事件联合抽取方法的 F1 值分别提高至少 2% 和 5.4%。溯源分析表明, 基于知识图谱的溯源与传统的人工溯源相比, 速度和准确性都有较大的提高。

**关键词:** 意外爆炸; 知识图谱; 事件抽取; 爆炸溯源

中图分类号: XXXX.X

国标学科代码: XXXXX

文献标识码: A

## Research on the knowledge graph of accidental explosion damage

WANG Jimin<sup>1</sup>, JIANG Can<sup>1</sup>, HAN Bin<sup>1</sup>, WANG Xing<sup>2,3</sup>, ZHANG Lei<sup>2,3\*</sup>

(1. College of Computer Science and Software Engineering, HoHai University, Nanjing 211100, Jiangsu, China

(2. State Key Laboratory of Target Vulnerability Assessment, Luoyang 471023, Henan, China;

(3. Engineering Protection Research Department of Defense Engineering Institute, AMS, PLA, Luoyang 471023, Henan, China.)

**Abstract:** Utilizing accident explosion investigation reports characterized by multi-source, heterogeneous, and overlapping information to construct an accidental explosion damage knowledge graph plays a significant role in enabling data-driven<sup>1</sup> explosion assessment and traceability. To extract the overlapping and nested event commonly present in accidental explosion investigation data, a knowledge graph construction method centered on joint extraction of Chinese events was proposed. This construction process involved four key steps: semi-automatic ontology construction, corpus building, joint extraction of Chinese event based on dynamic mask, and event coreference resolution. During semi-automatic ontology construction, TextRank was employed to compute keyword importance scores, and K-Means clustering was applied to identify core domain terminology, then, domain knowledge was used to analyze the clustering results. 13 top-level ontologies and four types of accidental explosion events were obtained. Subsequently, a corpus for accidental explosion damage was built by annotating texts gathered from online explosion investigation reports. The annotation was performed using a method that combines human effort and prior knowledge. To enhance the capability of recognizing overlapping and nested events, a GlobalPointer layer was integrated into the pre-trained RoBERTa(Robustly optimized Bidirectional Encoder Representations from Transformers approach) to form the RoBERTa-

\*收稿日期: 2025-09-30; 修回日期: 2026-02-02;

基金项目: 国家自然科学基金(12172381);

第一作者: 王继民(1976—), 男, 硕士, 副教授. E-mail: wangjimin@hhu.edu.cn;

通信作者: 张磊(1974—), 男, 博士, 研究员. E-mail: ustczhanglei@163.com。

GPointer model in this paper, and the RoBERTa-GPointer model was trained on the constructed corpus. Finally, extracted entities was aligned based on semantic and syntactic similarity, and the extracted events with same time and location were aligned. The constructed knowledge graph was applied to trace the explosion source of the Beirut port explosion on August 4, 2020. Initially, common explosion source materials were categorized into 11 types based on domain knowledge. The accidental explosion knowledge graph was then vectorized using confidence-optimized embedding. Descriptive texts of the Beirut port explosion phenomena were extracted, and cosine similarity was used to retrieve a set of similar historical explosion cases from constructed knowledge graph. Finally, a Bayesian classifier was applied to predict the type of explosion source material. The traceability analysis correctly identified the explosion source of the Beirut port explosion. The experimental results of knowledge graph construction show that the proposed RoBERTa-GPointer improves the F1 scores for event classification and event element classification on the accidental explosion damage corpus by at least 2% and 5.4%, respectively, compared to existing extraction models. The traceability results demonstrate that the knowledge graph-based approach offers substantial improvements in both speed and accuracy compared to traditional manual traceability methods. Furthermore, it also demonstrates that the constructed knowledge graph for accidental explosions can be adapted to downstream applications related to damage effects.

**Keywords:** accidental explosion; knowledge graph; event extraction; explosion traceability

随国家基础设施建设的发展, 易燃易爆品的生产、运输和使用也更加频繁, 在促进经济、社会发展的同时, 意外爆炸事故也随之增多, 这对国家稳定、经济建设和人民生命财产安全都带来了重大隐患<sup>[1,2]</sup>。意外爆炸往往具有爆炸威力大、爆源物资种类多、事故现场复杂等特点, 使得依赖传统科研“三范式”, 即理论分析、试验研究和数值仿真, 进行爆炸事故分析捉襟见肘。随着机器学习技术的发展, 以及领域数据的积累, 使得“数据驱动”的研究范式成为可能, 即通过收集、分析和挖掘数据, 揭示隐藏的规律。意外爆炸事故调查数据是该领域最常用、信息容量最大的数据, 但这些数据往往具有多源、异构和重复等特征, 从而为数据驱动的毁伤效果评估带来了困难。

知识图谱是一种语义网络, 为知识提供了一个结构化的表示形式, 可以高效地发掘事物间的复杂关联信息<sup>[3,4]</sup>。知识图谱非常适用于进行信息查询和推理, 因此被广泛应用于信息检索、智能问答, 推荐系统<sup>[5]</sup>以及垂直领域<sup>[6]</sup>。通过信息抽取从大量意外爆炸毁伤资料中抽取出实体关系三元组, 并建立意外爆炸毁伤知识图谱, 有利于将分散的、难以利用的意外爆炸数据有效组织起来, 构建出精简的、统一视角的意外爆炸毁伤信息库, 对意外爆炸的评估以及溯源具有重要的作用。刘宝等人<sup>[7]</sup>基于循环神经网络 (recurrent neural network, RNNs) 技术, 抽取化学品的实体与关系, 构建危险评估知识图谱, 提升评估人员获取相关知识和质检信息的效率。李聪等人<sup>[8]</sup>对于燃气管泄漏事故建立燃气管网知识图谱, 研究人员、管道缺陷、管理之间的关系, 并且根据管道缺陷的信息预测防腐缺陷的类型。何允等人<sup>[9]</sup>构建化工事故原因知识图谱, 围绕爆炸、中毒、泄漏、火灾等事故, 分析人员行为、企业决策和物品状态等致因, 建立完整的原因分类体系。

目前中文领域尚未建立成熟的意外爆炸毁伤知识图谱, 相关研究工作都没有深入研究意外爆炸的爆炸源、毁伤元、以及对目标的毁伤效应等, 尚无法支撑对意外爆炸事故的毁伤评估。如表 1 所示, 意外爆炸毁伤领域除爆炸事件外, 还存在包括引燃、毁伤事件等其他类型的事件, 这些事件构成了嵌套和重叠事件。根据本文作者的统计, 54%的意外爆炸事件中存在嵌套事件, 32%的事件中存在重叠事件, 从而造成触发词抽取和事件类型检测结果的误差, 影响事件的最终抽取效果。

表 1 意外爆炸毁伤事件嵌套和重叠示例

Table 1 Example of nested and overlapping events in accidental explosion damage

问题	例子	说明
重叠事件	爆炸导致小轿车燃烧	事件类型: 爆炸事件、毁伤事件
嵌套事件	爆炸抛出的碎石砸毁了附近的房屋	事件类型: 爆炸事件、毁伤事件

综上, 针对意外爆炸毁伤效应文本资料的特点, 本文提出了构建意外爆炸毁伤知识图谱的方法, 并开展了意外爆炸事故溯源的下游应用分析。首先针对意外爆炸非结构化文本资料的特点, 提出了集成半自动本体构建、基于动态掩码的中文事件联合抽取以及事件共指融合等算法的知识图谱构建方法。并将其与现有抽取模型进行实验对比。然后基于意外爆炸毁伤知识图谱, 对其中的爆炸源、毁伤元、毁伤效应等之间的关联和规律进行挖掘和分析, 实现了爆炸溯源。本文研究为爆炸事故毁伤评估提供了一种新的技术途径。

## 1 意外爆炸毁伤知识图谱构建

意外爆炸毁伤知识图谱(Accidental Explosion Damage Knowledge Graph, AEDKG)构建流程如图 1 所示, 主要包括 4 个步骤: 意外爆炸毁伤领域本体的半自动构建、意外爆炸毁伤语料库的构建、基于动态掩码的中文事件联合抽取以及事件共指融合。

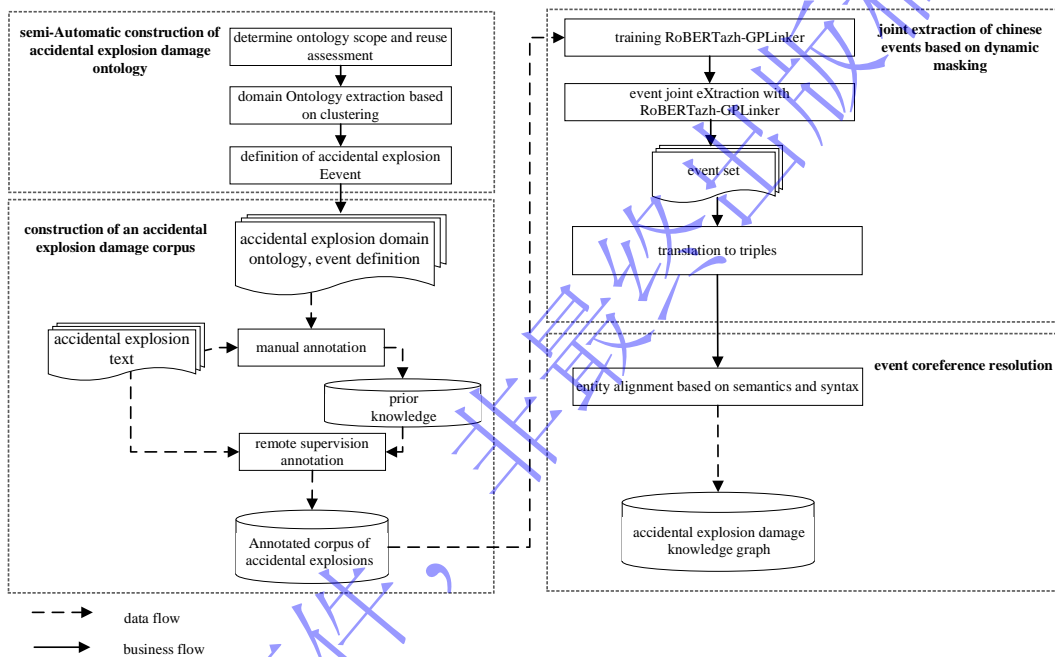


图 1 意外爆炸毁伤知识图谱构建流程

Fig.1 Construction process of knowledge graph for accidental explosion damage

### 1.1 意外爆炸毁伤领域本体半自动构建

本体是共享概念的模型明确的形式化规范说明<sup>[10]</sup>, 是构建知识图谱的“蓝图”, 其主要用于定义知识图谱的概念/类、关系/属性和约束等。鉴于传统的本体构建方法主要采用人工方式自顶向下的方法, 其过分依赖于构建者所掌握的专家知识、缺乏对本领域文本中概念的分析 and 归纳。本文提出基于 TextRank<sup>[11]</sup>和 K-Means<sup>[12]</sup>的半自动本体构建法, 以提高意外爆炸领域本体构建工作的效率和质量。半自动本体构建过程包括: **确定本体范围与复用评估**、基于聚类的领域本体提取以及事件定义三个步骤。

#### 1.1.1 确定本体范围与复用评估

本体构建聚焦于意外爆炸毁伤领域, 研究内容涵盖爆炸源的起爆方式、毁伤元的形成与作用过程。由于国内外在意外爆炸毁伤领域的知识图谱研究尚属空白, 现有军事、化工等相近领域的本体在术语体系与研究重点上差异较大, 难以满足本研究需求, 因此不能复用现有本体, 需从头构建领域专属本体结构。

#### 1.1.2 基于聚类的领域本体提取

为了全面提取意外爆炸毁伤领域的核心术语，本文采用三种方式进行关键词收集。一是从权威机构或出版社发布的专业调查报告、工具书或标准文件中的术语附录中收集；二是采用聚类算法从领域文本中自动筛选关键词；三是由领域专家对自动提取与聚类结果进行补充，完善遗漏的重要术语。

采用聚类从意外爆炸毁伤领域文本中自动提取关键词包括两个步骤。第一步，在预处理的语料上使用 TextRank 算法计算出关键词的得分。假设语料库有词语集合  $V$ ， $V_i$  表示第  $i$  个词，根据公式(1)计算每个词语的重要性得分。

$$Z(V_i) = (1-d) + d \cdot \sum_{V_j \in I(V_i)} \frac{k_{ji}}{\sum_{V_k \in O(V_j)} k_{jk}} Z(V_j) \quad (1)$$

式中： $Z(V_i)$ 为词语  $V_i$  的重要性得分； $d$  为阻尼系数，通常设为 0.85。 $I(V_i)$ 表示指向节点  $V_i$  的所有词语。 $O(V_j)$ 表示  $V_j$  出发指向的所有词语集合； $k_{ji}$  为节点  $V_j$  到  $V_i$  的边权重，默认为 1。使用 Word2vec<sup>[13]</sup> 将这些关键词转为语义向量，对向量使用 K-Means 方法进行聚类。

第二步，利用专家知识拆分、合并得到的关键词聚类结果，抽象出本体的层级。最终将意外爆炸本体分为爆炸事件、毁伤元、爆炸源、环境、目标种类、起爆方式、点火方式、爆炸当量、目标响应、次生灾害、国际规范、爆炸产物和反应过程等 13 个一级分类。进一步根据专家知识在此基础上划分二级本体，其中毁伤元包括冲击波、破片、燃烧、成坑、地冲击以及震动；目标种类包括建筑、设施以及人；目标响应包括压力、距离、加速度、速度、变形以及人员伤亡。

### 1.1.3 事件定义

将意外爆炸毁伤过程中发生的事件分成四类，事件类型和触发词示例如表 2 所示。

表 2 意外爆炸毁伤领域中事件类型、触发词、论元角色和论元

Table 2 Event types, trigger words, ontological roles and ontological elements in the field of accidental explosive damage

事件类型	触发词	论元角色	论元
爆炸事件	爆炸、爆燃、炸、闪爆、爆裂	事件名称	天津大爆炸、清水河大爆炸
		时间	凌晨 5 点 30 分、2017 年 8 月
		地点	工厂、车间
		爆炸源	化学制剂类、火药类、粉尘、混合物
		环境	潮湿、高温、雷击、阵雨、风、封闭
		TNT 当量	1000t、0.05t
		次生灾害	二次爆炸、水污染等。
		毁伤元	冲击波、破片、火焰、深坑等
		目标	建筑、设施和人
		损失金额	人民币 1000 多万元
引爆事件	点燃、引燃、点着、起火、爆燃	爆炸源	化学制剂、火药类、粉尘、混合物等
		点火方式	明火、阴燃、火星、摩擦引热、热表面
反应事件	氧化、硝化	爆炸源	化学制剂、火药类、粉尘、混合物等
		产物	氮气、水蒸气、一氧化碳、二氧化碳、氨气
		毁伤元	冲击波、破片、火焰、炸坑等
毁伤事件	砸毁、冲击、砸、震碎、死亡、受伤、震飞、燃烧	目标	建筑、设施和人
		响应程度	飞出 100m、位移 7m
		毁伤描述	重度变形、坑深 8m
		环境	潮湿、高温、雷击、阵雨、风、封闭

爆炸事件与其他三类事件具有一定的因果关系，爆炸事件的触发词可能也是引燃事件的触发词，事件间触发词存在共享，因此意外爆炸毁伤领域存在事件重叠的情况。

## 1.2 意外爆炸毁伤语料库构建

对非结构化文本数据利用领域知识进行清洗，剔除异常数据，并进行分词和分句等操作。根据本体中定义的 13 种一级实体类型对意外爆炸文本进行标注，构建意外爆炸毁伤语料库。其中关系标注采用人工标注和利用先验知识标注两种方式，首先人工标注单个句子中出现的事件触发词和论元，然后将人工标注得到的三元组作为先验知识采用远程监督的方式增加标注样本量。经过以上步骤，得到了标注的意外爆炸毁伤语料库。为方便读取采用 JSON 格式存储，用于训练、验证和测试抽取模型。

## 1.3 基于动态掩码的事件联合抽取

本文在 RoBERTa (Robustly optimized Bidirectional Encoder Representations from Transformers approach) 模型中加入 GlobalPointer 层形成 RoBERTa-GPointer 模型，来增强对重叠事件和嵌套事件的抽取和识别能力，模型结构如图 2 所示，主要由输入层、编码层、GlobalPointer 层、事件抽取层和输出层组成。

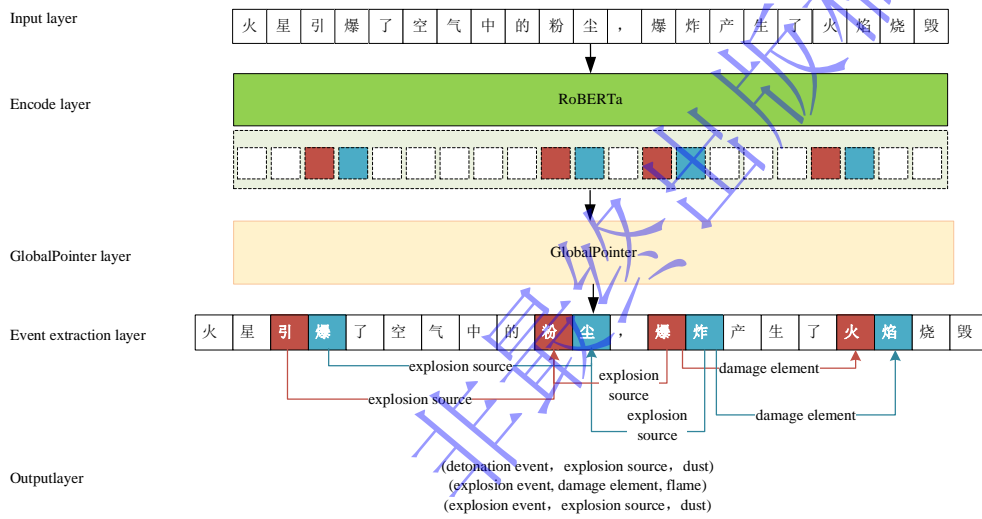


图 2 RoBERTa-GPointer 模型结构

Fig.2 Structure of RoBERTa-GPointer

(1) 输入层。将标记的语料传入模型，将标记的触发词、事件类型、论元和论元角色作为关系抽取任务最终的真实值。模型的输入为中文文本，以字符为单位将待抽取文本切分，得到一个文本序列  $Q_{input}=\{w_1, w_2, \dots, w_n\}$ ，其中， $w_i$  是待抽取文本中第  $i$  个字符。

(2) 编码层。采用中文预训练模型 RoBERTa<sup>[14]</sup> 作为编码层，输入为  $Q_{input}$ ，然后经过 12 层 Transformer 后得到最终的向量表示  $\{h_1, h_2, \dots, h_n\}$ 。编码具体计算公式如(2)和(3)所示。

$$h_0 = E_i + E_p + E_s \quad (2)$$

$$h_\alpha = Trans(h_{\alpha-1}), \alpha \in [1, n] \quad (3)$$

式中： $h_0$  为初始的向量， $E_i$ 、 $E_p$  和  $E_s$  分别表示输入的词嵌入、位置嵌入以及段嵌入向量。 $h_\alpha$  为输入序列  $Q_{input}$  经过  $\alpha$  层 Transformer 编码后的向量。 $n$  代表 Transformer 的层数。

(3) GlobalPointer 层<sup>[15]</sup>。将输出层学习到的字符向量按其在序列中的次序两两拼接，得到 token-pair 的向量表示，并构建多头 token-pair 关系矩阵，用于处理事件抽取中触发词和论元重叠问题。

(4) 事件抽取层。用于进一步实现事件抽取，包括触发词识别和分类、事件元素检测和识别论元角色事件信息。

(5) 输出层。为事件抽取层解码后的结果，得到模型抽取的触发词、事件类型、论元和论元角色等信息。将其转为 (事件类型, 论元角色, 论元) 的三元组形式，便于之后意外爆炸毁伤知识图谱的构建。

在事件抽取任务中，事件分类和事件元素分类为多目标分类任务，损失函数如公式(4)所示。

$$L = \log(1 + \sum_{j \in A} e^{-s_j}) + \log(1 + \sum_{i \in B} e^{s_i}) \quad (4)$$

式中： $A$  是正类别集合， $B$  是负类别集合， $s_i$  为类别  $i$  的得分。

#### 1.4 事件共指融合

常出现同一意外爆炸事件被多篇报告描述的情况，为避免信息冗余并构建更加完整、一致的事件信息库，需要对抽取出的事件进行共指融合。主要包括基于语义语法相似度的实体对齐和意外爆炸事件共指融合。

##### 1.4.1 基于语义语法相似度的实体对齐

实体对齐<sup>[6]</sup>是事件共指融合的基础。首先利用 BERT (Bidirectional Encoder Representations from Transformers) 模型得到实体在句中的向量表示，计算实体之间的余弦相似度，进而得到两实体的语义相似度，接着利用 NTLK (Natural Language Toolkit) 的词法分析判断实体在各自句子中的词性。如果两实体词性相同且语义相似度大于阈值，则认为两个实体在语义、语法上都相似，两个实体实质上为同一实体，将其对齐。具体对齐过程如表 3 所示。

表 3 基于语义和语法相似度的实体对齐算法

Table 3 Entity alignment algorithm based on semantic and syntactic similarity

<b>输入:</b> 待对齐的实体 $ent1$ 和 $ent2$ ，其上下文 $sen1$ 和 $sen2$ ，相似度阈值 $\tau$
<b>输出:</b> $ent1$ 和 $ent2$ 对齐结果
1. 将所有抽取出的实体加入分词词典 $w\_dict$
2. 按照 $w\_dict$ 对 $sen1$ 和 $sen2$ 进行分词
3. 预训练 BERT 模型计算 $ent1$ 在 $sen1$ 中的表示 $vec1$ ， $ent2$ 在 $sen2$ 中的表示 $vec2$
4. //1.语义相似度计算
5. $vec1$ 和 $vec2$ 扁平化处理
6. if $vec1.shape[0] > vec2.shape[0]$ :
7. $vec2$ 进行零填充操作
8. else
9. $vec1$ 进行零填充操作
10. 计算 $Sim(ent1, ent2) = 1 - \cos(vec2, vec1)$
11. //2.语法相似度计算
12. 利用 NTLK 库词法分析判断 $ent1$ 在 $sen1$ 中的词性 $flag1$ ， $ent2$ 在 $sen2$ 中的词性 $flag2$
13. if $flag1 \neq flag2$ :
14. $Pos(ent1, ent2) = 1$
15. else
16. $Pos(ent1, ent2) = 0$
17. if $Pos(ent1, ent2) * Sim(ent1, ent2) < \tau$ :
18.     return
19. return $len(ent1) > len(ent2) ? ent1 : ent2$

在表 3 的算法中，1-10 行计算实体的语义相似度，利用 BERT 模型得到实体在句中的向量表示，通过零填充统一向量长度、计算实体之间的余弦相似度，进而得到两实体的语义相似度  $Sim$ ；12-16 行判断实体在语法上是否相似，利用 NTLK 的词法分析判断实体在各自句子中的词性，如果实体词性不同则语法  $Pos$  不相似，两个实体不对齐；17-19 行表示如果两实体语法相似且语义相似度大于阈值，则认为两个实体在语义、语法上都相似，那么两个实体实质上为同一实体，将其对齐，返回信息更全面的实体作为最终对齐后的结果。表 4 给出意外爆炸毁伤领域的基于语义语法相似度实体对齐的示例，下划线词为需要判断是否对齐的实体。可以看出第一个例子中两实体语义相似度高、且词性相同，因

此应该对齐；第二个例子中，虽然两实体字面上完全一样，但词性不同、在句子中的含义也不同，因此不能对齐两个实体。

**表 4 基于语义和语法进行实体对齐的示例**

Table 4 Example of entity alignment based on semantics and syntax

序号	样例	语法	语义相似度	是否对齐
1	电路老化产生火星点燃 <u>氧气和空气的混合物</u> 。	名词	0.78	是
	<u>空气氧气混合物</u> 被点燃，导致 A 地发生爆炸。	名词		
2	导致这次 <u>爆炸</u> 发生的原因是设施老化	名词	1	否
	10 月 11 日在 A 工厂 B 楼锅炉突然 <u>爆炸</u> 了。	动词		

#### 1.4.2 意外爆炸事件共指融合

根据表 3 的算法判断事件论元实体是否相同，如果爆炸发生的时间和地点等实体相同，则将该爆炸事件进行融合，对所有不同的事件元素进行保留，相同的事件元素仅保留一个；若爆炸事件对应不同的爆炸时间和地点，则两个爆炸事件不进行事件融合。此外，由于爆炸事件、引燃事件、反应事件和毁伤事件间存在共享论元，需要对不同事件类型间的论元实体对齐，整合一场意外爆炸完整的爆炸毁伤过程。通过这种方式将不同报道来源，实质相同的事件合并为一个，得到关于某特定事件全面的、精简的信息，实现了多源异构爆炸信息的融合，完成意外爆炸毁伤知识图谱的构建。

#### 1.5 知识图谱构建实例

下面以广西维尼纶集团有限责任公司 2008 年 8 月 26 日的爆炸事故为例构建意外爆炸毁伤知识图谱。首先使用训练的 RoBERTa-GPLinker 模型从三篇来自互联网的针对该爆炸事故的报告中抽取事件的触发词和论元，将抽取的事件信息转化为三元组列表，表 5 至表 7 分别为从三篇报告中提取的三元组信息；然后对三元组的实体对齐，保留不同的事件元素，相同的事件元素仅保留一个。未进行对齐和融合前，三篇调查报告共包含 27 个三元组，经过对齐和融合后，这一爆炸事故的图谱共包含 23 个关系三元组，消除了冗余的信息，图 3 展示了最终生成的知识图谱信息。

**表 5 从报告 1 中提取的三元组**

Table 5 Triples extracted from Report 1

实体	关系	实体
爆炸	EXTIME_IS	2008 年 8 月 26 日凌晨
爆炸	EXLOC_IS	广西维尼纶集团公司
爆炸	EXSOUR_IS	有机车间
爆炸	EXTIME_IS	2008 年 8 月 26 日 6 时 45 分
爆炸	EXLOC_IS	广西维尼纶集团有限责任公司
爆炸	EXSOUR_IS	乙炔气体

**表 6 从报告 2 中提取的三元组**

Table 6 Triples extracted from Report 2

实体	关系	实体
爆炸	EXTIME_IS	2008 年 8 月 26 日 6 时 40 分
爆炸	EXTARIS	20 人死亡、60 受伤
爆炸	EXTAR_IS	厂区附近 3km 范围 18 个村屯及工厂职工、家属共 11500 多名群众
大火	EXLOC_IS	广西维尼纶集团有限责任公司所属广西广维化工股份有限公司
爆燃	EXDAM_IS	白雾
爆燃	EXSOUR_IS	乙炔、乙醛、醋酸乙烯等易燃易爆气体
静电火花	IGNIEXSOUR	乙炔、乙醛、醋酸乙烯等易燃易爆气体

爆炸与冲击  
Explosion and Shock Waves

爆炸	EXSOUR_IS	大量可燃液体和乙炔
火源	IGNI_EXSOUR	大量可燃液体和乙炔

**表 7 从报告 3 中提取的三元组**

Table 7 Triples extracted from Report 3

实体	关系	实体
爆炸	EXTIME_IS	2008 年 8 月 26 日 6 时 40 分
爆炸	EXLOC_IS	广西维尼纶集团有限责任公司所属广西广维化工股份有限公司
爆炸	EXTAR_IS	共 11500 多名群众疏散
爆炸	EXTAR_IS	导致 20 人死亡、60 人受伤
爆炸	EXTAR_IS	共 11500 多名群众疏散
爆炸	EXDAM_IS	冲击波
爆炸	EXDAM_IS	大火
冲击波	DAMAGE_TAR	合成、蒸憎、醇解、聚合等工段的部分建筑物和设备、管道
冲击波	DAMAGE_TAR	罐场的储罐
大火	DAMAGE_TAR	罐场的储罐
合成、蒸憎、醇解、 聚合等工段的部分建 筑物和设备、管道	TAREXTENT	震坏
罐场的储罐	TAREXTENT	震坏

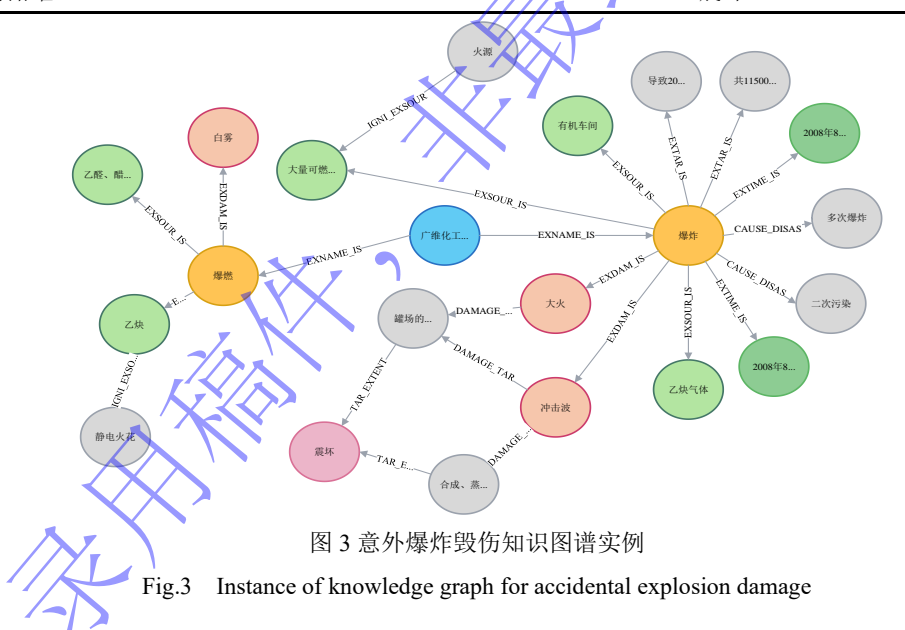


图 3 意外爆炸毁伤知识图谱实例

Fig.3 Instance of knowledge graph for accidental explosion damage

## 2 实验与结果分析

### 2.1 实验数据集

数据集的数据来源包括互联网上爬取的领域内调查报告，以及 NTRL (National Technical Reports Library)、NASA (National Aeronautics and Space Administration)、DHS (Department of Homeland Security) 等权威数据库的科研报告，专业报告，以及从意外爆炸领域的知名专著中提取的意外爆炸事件调查报告。

### 2.2 数据预处理

对数据进行标注，标注格式如图 4 所示。“text”为根据规则生成的待抽取文本，“event\_list”为在文本 text 中出现的所有意外爆炸毁伤领域事件的集合，“event\_type”标记事件的类型，“trigger”标记文本中出现的触发词，“trigger\_start\_index”记录触发词在待抽取文本中出现的位置，“argument\_start\_index”为论元在待抽取文本中出现的位置，“role”为论元角色，“argument”标记文本中出现的论元。对于标记后的数据随机划分训练集、测试集和验证集，比例为 80%、10% 和 10%。

```
{ "text": " Explosion at A city intersection",   "id": "01",
  "event_list": [
    { "event_type": "explosion event", "trigger": "explosion", "trigger_start_index": 7,
      "arguments": [ { "argument_start_index": 0, "role": "location", "argument": "A city", "alias": [] } ]
    }
  ]
}
```

图 4 意外爆炸毁伤事件标注格式

Fig.4 Format of accidental explosion damage event label

### 2.3 评价指标

将事件从非结构化的文本中抽取出来，规定只有在**触发词和论元都正确抽取**的前提下，才能认为当前的事件抽取结果是正确的。采用事件抽取任务通用的精确率 (Precision,  $P$ )，召回率 (Recall,  $R$ ) 以及综合评价指标  $FI$ (F1-Score)对知识图谱的构建结果进行评价。

$$P = \frac{TP}{TP+FP} \times 100\% \quad (5)$$

$$R = \frac{TP}{TP+FN} \times 100\% \quad (6)$$

$$F1 = 2 \times \frac{P \times R}{R+P} \times 100\% \quad (7)$$

式中： $TP$  表示真实例， $FN$  表示反实例， $FP$  表示假实例， $TN$  表示真反例。

### 2.4 实体对齐相似阈值分析

对验证集中的实体进行余弦相似度计算，并进行匹配。选择不同的相似度阈值，分析实体匹配的精确率，如图 5 所示。相似距离在 0.725 以上时，91%的样本匹配正确，因此将实体识别的阈值设为 0.725。当两个实体的相似值在阈值以上时，认为是匹配的，然后通过人工对部分错误实体进行修正。

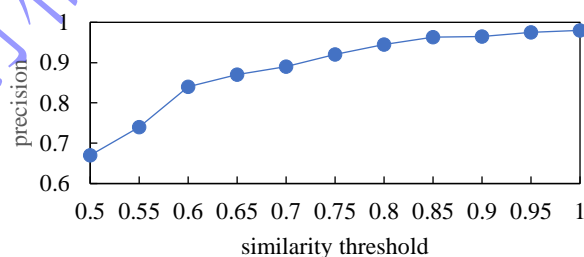


图 5 相似度阈值与实体对齐精确率关系

Fig.5 Precision vs. similarity threshold curve

### 2.5 实验结果分析

为验证提出的模型在事件抽取任务中的性能，在自建的意外爆炸毁伤语料集上分别从事件分类和事件元素分类两个方面进行实验评估。在事件分类实验中，选取了 BERT-BiLSTM-CRF (BERT Bidirectional Long Short-Term Memory Conditional Random Field)、BiLSTM-CRF 和 DMCNN<sup>[17]</sup>(Dynamic Multi-pooling Convolutional Neural Network) 作为对比模型。在事件元素分类任务中，选取了 C-BiLSTM

(Cascade BiLSTM)、BERT-QA (BERT for Question Answering) 和 DMCNN 进行对比。对比结果如表 8 和表 9 所示。

表 8 事件分类比较

Table 8 Comparison of event classification

模型	$P$	$R$	$F1$
BERT-BiLSTM-CRF	71.4%	49%	57%
BiLSTM-CRF	52.1%	44%	47.7%
DMCNN	66.5%	53%	58%
RoBERTa-GPLinker	69.2%	55%	61.0%

表 8 显示, RoBERTa-GPLinker 在意外爆炸毁伤语料集上具有明显优势, 其在事件分类任务中表现出最好的综合性能  $F1$  值。RoBERTa-GPLinker 的召回率  $R$  相比于 BERT-BiLSTM-CRF 和 BiLSTM-CRF 分别高出约 6% 和 11%, 这主要得益于 RoBERTa-GPLinker 在面对文本中触发词重叠问题时, 仅标注触发词的起始和结束位置, 避免了传统序列标注模型在字符重叠情况下容易出现的遗漏问题。RoBERTa-GPLinker 的精确度  $P$  略低于 BERT-BiLSTM-CRF, 这是由于其将事件抽取建模为关系抽取任务, 触发词识别受到论元抽取误差的影响。与同是联合抽取模型的 DMCNN 相比, RoBERTa-GPLinker 的精确率  $P$  和召回率  $R$  分别提高 2.7% 和 2%。

表 9 事件元素分类比较

Table 9 Comparison of event element classification

模型	$P$	$R$	$F1$
C-BiLSTM	47.3%	46.6%	46.9%
BERT-QA	56.2%	50.7%	53.3%
DMCNN	55.3%	52.2%	53.7%
RoBERTa-GPLinker	63.4%	55.4%	59.1%

表 9 中, RoBERTa-GPLinker 在事件元素分类中的综合性能指标  $F1$  值优于其他模型。相比于 C-BiLSTM, RoBERTa-GPLinker 精确率  $P$  和召回率  $R$  分别提高 16.1% 和 8.8%, 主要原因在于后者基于序列标注, 难以应对重叠和嵌套事件。相较于 BERT-QA, RoBERTa-GPLinker 在精确率  $P$  和召回率  $R$  上提升了 7.2% 和 4.7%, 得益于其采用动态全词掩码策略, 更好地捕捉上下文语义。与 DMCNN 相比, RoBERTa-GPLinker 在精确率  $P$  和召回率  $R$  上分别提升约 8.1% 和 3.2%。

为了分析重叠事件和嵌套事件对模型性能的影响, 将验证集中的文本进行划分成无嵌套和重叠事件、完全嵌套和重叠事件以及混合数据。采用不同模型进行事件类型检测。表 10 为不同数据下, 各模型的事件分类精确率  $P$ 。从表 10 可以看出, 加入嵌套和重叠事件的文本后, 各模型的事件分类精确率  $P$  有所下降。对完全是嵌套和重叠事件的文本中, RoBERTa-GPLinker 比 DMCNN 精确率  $P$  提高了 6.3%。

表 10 不同数据下事件分类精确率比较

Table 10 Comparison of the precision of event type classification under different datasets

模型	混合数据	无嵌套和重叠事件	完全嵌套和重叠事件
BERT-BiLSTM-CRF	71.4%	82.2%	64.7%
BiLSTM-CRF	52.1%	68.3%	43.2%
DMCNN	66.5%	78.6%	60.5%
RoBERTa-GPLinker	69.2%	81.0%	66.8%

### 3 基于知识图谱的意外爆炸溯源应用

#### 3.1 爆炸源类型分析

爆炸溯源是指通过收集、整合和分析多种信息，推理爆炸发生的机制、爆炸源类型等信息的工作。本文应用通过利用爆炸事件现场的物证、痕迹、残余物等进行采集、分析和鉴定，初步推断可能涉及的爆炸源类型。经过专家知识分析，本文将常见的意外爆炸事故的爆源物资类型归类为如表 11 的 11 类。

表 11 爆源类型和说明

Table 11 Type and description of explosion source

编号	爆源类型	说明
A	石油、煤油、汽油、酒精或其他液体易燃易爆品	无毒无异味
B	天然气、煤气、酒精蒸汽、混合气体等气体爆炸	无毒无异味
C	烷、苯、油漆等液态石化产品	有毒有异味
D	烷、苯、烯类固态化工产品	有毒有异味
E	烷、苯、烯类气态化工产品	有毒有异味
F	炸药、火药、烟火药、推进剂等	含 C、H、硝基固态材料
G	金属粉尘	铝、镁、硅等
H	其他粉尘	面粉、木屑等
I	压力容器	物理爆炸
J	静电、雷电、电火花等	物理爆炸
K	其他爆炸源	占比极少、不常见

#### 3.2 基于相似查询和贝叶斯分类的爆炸溯源

基于相似查询和贝叶斯分类的爆炸溯源流程如图 6，主要包括如下步骤：

(1) 置信度优化的图谱向量化表示。使用基于置信度优化的 CompGCN<sup>[18]</sup> (Composition-based Multi-relational Graph Convolutional Networks) 训练意外爆炸毁伤知识图谱，将意外爆炸知识图谱进行向量化表示，得到事件的向量，这也是所有下游应用的基础。同时还要将所有爆炸案例的毁伤现象的主要特征按照表 12 等级进行离散化。

(2) 爆炸特征向量化表示。在步骤 (1) 的基础上行将待溯源的爆炸案例描述向量化。

(3) 相似爆炸案例检索。利用余弦相似度查找相似案例，得到相似案例集合。

(4) 使用贝叶斯分类器<sup>[19]</sup>进行分类。对于相似案例集合中每一场历史上发生的爆炸案例，将其爆炸源所属分类作为标签，爆炸事件中出现的爆炸现象、毁伤描述、目标、目标响应等信息作为贝叶斯算法的输入，然后根据用户所给出的查询条件预测爆炸事件的爆炸源类型。根据贝叶斯准则，在某种特征下每种爆炸源的概率如公式(8)所示。

$$P(c|t) = \frac{P(t|c)P(c)}{P(t)} \quad (8)$$

其中， $c$  为表 11 预先定义的爆炸源类型标签， $t$  代表一场意外爆炸毁伤的特征信息， $P(c)$  为爆炸源类型为  $c$  的爆炸事件的先验概率，表示在相似案例集中  $c$  类别分布的概率； $P(t|c)$  为爆炸源种类为  $c$  时，在案例集中爆炸事件的特征为  $t$  的条件概率； $P(t)$  表示在相似案例集中爆炸特征为  $t$  的概率； $P(c|t)$  为

后验概率，表示当前发生了新的爆炸事件，并且得知了爆炸事件的信息，判断其爆炸源种类为  $c$  的概率。

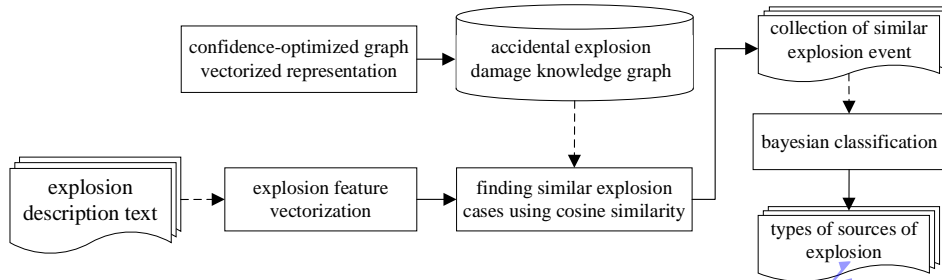


图 6 爆炸溯源流程

Fig.6 Explosion traceability process

表 12 爆炸现象描述及等级划分

Table 12 Description and classification of explosion phenomena

爆炸现象	现象描述	等级划分	等级标号
爆炸成坑	坑深(m)	[0, 0.5)	1
		[0.5,1)	2
		[1,3)	3
		[3,5)	4
		$\geq 5$	5
破片抛掷	破片最大飞行距离(m)	[0,10)	1
		[10,30)	2
		[30,50)	3
		[50,100)	4
		$\geq 100$	5
普通窗玻璃破坏	距离范围(km)	[0,10)	1
		[10,50)	2
		[50,100)	3
		$\geq 100$	4
爆炸火球	高度(m)	[0,10)	1
		[10,30)	2
		[30,50)	3
		[50,100)	4
		$\geq 100$	5
爆源物质等效 TNT 质量	质量(t)	[0,0.5)	1
		[0.5,1)	2
		[1,2)	3
		[2,10)	4
		[10,20)	5
		[20,50)	6
		[50,100)	7
		[100,300)	8

爆 炸 与 冲 击  
Explosion and Shock Waves

爆炸现象	现象描述	等级划分	等级标号
		[300,500)	9
		$\geq 500$	10
烟雾	颜色	白色	1
		黑色	2
		红色或者粉红色	3
气味		无刺激性气味	1
		有刺激性气味	2

在筛选出相似爆炸事件后，对所有事件元素进行统计然后按照表 12 进行处理，将其转为便于贝叶斯算法处理的格式。对应的事件元素采用数值，“-1”代表该元素没有出现，正值整数代表该事件元素出现，且该数值描述该毁伤元的严重程度，数值越大毁伤情况越严重。

### 3.3 溯源案例

以 2020 年 8 月 4 日黎巴嫩贝鲁特港口区爆炸<sup>[20,21]</sup>为例进行溯源。融合相关的报道得到该爆炸的爆炸现象描述包括“剧烈爆炸、升起一团红色烟雾、巨大的冲击波、粉色蘑菇云、港口及港口半径 100 米内的建筑在 1 秒内被夷为平地”。根据爆炸现象描述及其主要特征参数等级分类作为爆炸源的特征，进行相似查询，得到如表 13 的相似历史爆炸案例。

表 13 相似意外爆炸事件

Table 13 Similar accidental explosion events

序号	<i>c</i>	爆炸现象	爆炸现象描述
1	F	普通窗玻璃破坏；爆炸成坑；烟雾	范围 60km；直径约 120 米；红色蘑菇云。
2	D	普通窗玻璃破坏；爆炸成坑	范围 10km；深度 4.57 米。
3	F	普通窗玻璃破坏；爆炸成坑；破片；烟雾和烟	范围 2km；深 2m；飞出 400m；黑色火焰。
4	F	普通窗玻璃破坏；破片；爆炸成坑；烟雾	范围 5km；飞出 100m 外；直径 7m，深约 1m；白色，红褐色烟雾。
5	F	普通窗玻璃破坏；爆炸成坑；烟雾	范围 500m；距离 29m；直径范围 15.6m，深度 1.9m；灰白色蘑菇云。
6	E	普通窗玻璃破坏；烟雾	范围 5km；高约 15m；红色。
7	F	普通窗玻璃破坏；烟雾	约 110 公里；约 150 米高；红色蘑菇云。

将表格 13 处理为贝叶斯算法的输入格式，-1 代表该爆炸没有出现这种毁伤元或者调查没有得到相关的数据，结果如表 14 所示。

经过贝叶斯分类得到该爆炸源为表 11 中的 F 类，即为“炸药、火药、烟火药、推进剂等”。事后多方调查表明，实际爆炸源主要物资为硝酸铵，与基于知识图谱溯源结果吻合<sup>[20,21]</sup>。与传统的人工溯源相比，基于知识图谱的意外爆炸溯源能够同时从海量的爆炸数据中搜索数据，在向量化表示的知识图谱中，相似内容的向量空间分布是靠近的，因此能够通过相似分析发现更多的潜在的相似的爆炸案例，避免人工识别的不准确和遗漏。

表 14 贝叶斯算法输入

Table 14 Input of Bayesian algorithm

编码	<i>c</i>	冲击波	爆炸成坑	烟雾	破片
1	F	4	5	3	-1
2	D	2	4	-1	-1
3	F	1	3	2	5
4	F	1	5	1,3	-1

5	F	1	3	1,2	2
6	E	0	-1	3	-1
7	F	3	-1	3	-1

## 4 结 论

针对意外爆炸毁伤领域的知识图谱构建问题进行研究，提出以事件联合抽取模型为核心的意外爆炸知识图谱构建方法，并在构建的知识图谱上进行了意外爆炸溯源应用分析，得到以下主要结论。

(1) 针对意外爆炸毁伤领域文本中存在事件重叠和嵌套的情况，基于动态掩码的事件联合抽取模型可以较好的提取出事件信息。在意外爆炸毁伤语料库上，相较于现有抽取模型，提出的事件联合抽取方法在事件分类上，精确率  $P$  以及召回率  $R$  分别提高至少 2.7% 以及 2%，因为模型在面对文本中触发词重叠问题时仅标注触发词的起始和结束位置，避免了传统序列标注模型在字符重叠情况下容易出现的遗漏问题。在事件元素分类任务中，综合  $F1$  值、精确率  $P$  和召回率  $R$  指标分别至少提升 5.4%，7.2% 和 3.2%。

(2) 所构建的意外爆炸知识图谱具备良好的下游评估与推理支持能力。在溯源分析中，首先对知识图谱进行向量化表示，随后从贝鲁特港口区爆炸的相关文献报告中抽取爆炸描述，通过向量相似性检索在知识图谱中查找类似爆炸事件，最终利用贝叶斯分类方法准确推断爆炸源种类。基于知识图谱的溯源与传统的人工溯源相比，速度和准确性都有较大的提高，也表明，所建立的意外爆炸知识图谱能够适应毁伤效应的下游分析应用。

## 参 考 文 献

- [1] SHAHZAIB A, HARISH G, MUNEEBA K. An industrial disaster emergency decision-making based on China's Tianjin city port explosion under complex probabilistic hesitant fuzzy soft environment[J]. Engineering Applications of Artificial Intelligence, 2023,123:106400. DOI:10.1016/j.engappai.2023.106400.
- [2] GENDING Y, YIHSING D, XIAODONG Y, et al. Holistic case study on the explosion of ammonium nitrate in Tianjin port[J]. Sustainability, 2022,14(6):3429-3429. DOI:10.3390/SU14063429.
- [3] 杭婷婷, 冯钧, 陆佳民. 知识图谱构建技术: 分类、调查和未来方向[J]. 计算机科学, 2021,48(2):175-189. DOI:10.11896/jsjcx.200700010.
- H ANG T T, FENG J, LU J M. Knowledge graph construction techniques: taxonomy, survey and future directions[J]. Compute Science, 2021,48(2):175-189. DOI:10.11896/jsjcx.200700010.
- [4] JI, S., PAN, S., CAMBRIA, E., et al. A survey on knowledge graphs: Representation, acquisition and applications[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 33(2): 494-514. DOI:10.1109/TNNLS.2021.3070843.
- [5] GEVA T, ZAHAVI J. Empirical evaluation of an automated intraday stock recommendation system incorporating both market data and textual news[J]. Decision support systems, 2014, 57: 212-223. DOI: 10.1016/j.dss.2013.09.013.
- [6] 李学亮. 基于深度学习及知识图谱的垂直领域问答研究[D]. 北京:北京邮电大学, 2021: 4-12. DOI:10.26969/d.cnki.gbydu.2021.002355.
- [7] 刘宝, 车礼东, 黄红花, 等. 基于自然语言处理(NLP)技术建立化学品危险评估知识图谱的研究[J]. 计算机与应用化学, 2018,35(7):605-610. DOI:10.16866/j.com.app.chem.201807010.
- LIU B, CHE L D, HUANG H H, et al. Research on the establishment of a chemical hazard assessment knowledge graph based on natural language processing (NLP) technology[J]. Computer and Applied Chemistry, 2018, 35(7): 605-610. DOI:10.16866/j.com.app.chem.201807010.
- [8] 李聪, 徐子烜, 王雨情, 等. 城市燃气管网泄漏事故分析知识图谱构建及应用研究[J]. 中国安全生产科学技术, 2022,18(10):5-12. DOI:10.11731/j.issn.1673-193x.2022.10.001.
- LI C, XU Z X, WANG Y Q, et al. Construction and application of a knowledge graph for urban gas pipeline leakage accident analysis[J]. Journal of Safety Science and Technology, 2022, 18(10): 5-12. DOI:10.11731/j.issn.1673-193x.2022.10.001.
- [9] 何允. 基于化工案例的事故原因知识图谱构建研究[D]. 大连:大连理工大学, 2021:20-33. DOI: 10.26991/d.cnki.gdllu.2021.001795.
- [10] 杜小勇, 李曼, 王珊. 本体学习研究综述[J]. 软件学报, 2006,17(9): 1837-1847. DOI: 10.1360/jos171837.

- DU X Y, LI M, WANG SH. A Survey on Ontology Learning Research[J]. Journal of Software, 2006,17(9): 1837–1847.. DOI: 10.1360/jos171837
- [11] 徐飞, 彭佳佳, 刘军, 等. 基于多特征融合的 TextRank 新闻自动摘要模型[J]. 计算机系统应用, 2023, 32(2): 242-249. DOI:10.15888/j.cnki.csa.008913.
- XU F, PENG J J, LIU J, et al. Automatic News Summarization Model Based on Multi-feature TextRank[J]. Computer Systems & Applications, 2023, 32(2):242–249. DOI:10.15888/j.cnki.csa.008913.
- [12] YANG B, FU X, SIDIROPOULOS ND, et al. Towards k-means-friendly spaces: Simultaneous deep learning and clustering[C]// International conference on machine learning. Sydney, Australia, 2017: 3861-3870. DOI:10.48550/arXiv.1610.04794
- [13] 席笑文, 郭颖, 宋欣娜, 等. 基于 word2vec 与 LDA 主题模型的技术相似性可视化研究[J]. 情报学报, 2021, 40(9): 974-983. DOI: 10.3772/j.issn.1000-0135.2021.09.007.
- XI X W, GUO Y, SONG X N, et al. Research on the Technical Similarity Visualization Based on word2vec and LDA Topic Model[J]. Journal of the China Society for Scientific and Technical Information, 2021, 40(9): 974-983. DOI: 10.3772/j.issn.1000-0135.2021.09.007.
- [14] LIU Y H, OTT M, GOYAL N, et al. RoBERTa: A robustly optimized BERT pretraining approach[C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL 2020). 2020, 1: 3980-3990. DOI:10.48550/arXiv.1907.11692
- [15] SU J, MURTAGHA A, PAN S, et al. Global Pointer: Novel efficient span-based approach for named entity recognition[J]. arXiv preprint arXiv :2208.03054, 2022. DOI: 10.48550/arXiv.2208.03054.
- [16] TRISEDYA B D, QI J, ZHANG R. Entity alignment between knowledge graphs using attribute embeddings[C]// Proceedings of the A AAI conference on artificial intelligence. Honolulu, Hawaii, USA, 2019:297-304. DOI: 10.1609/aaai.v33i01.3301297.
- [17] CHEN Y, XU L, LIU K, et al. Event extraction via dynamic multi-pooling convolutional neural networks[C]// Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing. Beijing, China, 2015:167-176. DOI:10.3115/v1/P15-1017
- [18] VASHISHTH S, SANYAL S, NITIN V, et al. Composition-based Multi-Relational Graph Convolutional Networks[J]. International Conference on Learning Representations, 2020, arXiv preprint arXiv:1911.03082, 2019. DOI:10.48550/arXiv.1911.03082.
- [19] 姜子建, 周荣义, 石云霄, 等. 基于贝叶斯网络的危化品爆炸事故隐患关联与溯源分析[J]. 中国安全科学学报, 2024, 34(06):173-180. DOI:10.16265/j.cnki.issn1003-3033.2024.06.1848.
- JIANG Z J, ZHOU R Y, SHI Y X, et al. Correlation and traceability analysis of hazardous chemical explosion accidents based on Bayesian network [J]. China Safety Science Journal, 2024, 34(6):173-180. DOI:10.16265/j.cnki.issn1003-3033.2024.06.1848.
- [20] KALLAS J, NAPOLITANO R. Understanding critical masonry building attributes shaping vulnerability to blast loads: Data-driven insights from the 2020 Beirut explosion[J]. International Journal of Disaster Risk Reduction, 2024, 110: 104640. DOI:10.1016/j.ijdrr.2024.104640.
- [21] DENNY W J, FARRIMOND G D, GABRIEL S, et al. The influence of obstacle geometric fidelity on blast wave propagation: a reduced-scale case study examining the role of the grain siloes in the 2020 Beirut explosion[J]. Shock Waves, 2025, 35: 673-699. DOI:10.1007/S00193-025-01241-5.