

中文引用格式:李红儒,栾婷婷,邓明月,等.面向危险化学品事故的事理图谱构建[J].中国安全科学学报,2024,34(5):195-203.

英文引用格式:LI Hongru, LUAN Tingting, DENG Mingyue, et al. Development of causal graph for hazardous chemical accidents[J]. China Safety Science Journal, 2024, 34(5): 195-203.

# 面向危险化学品事故的事理图谱构建\*

李红儒<sup>1</sup>, 栾婷婷<sup>\*\*1</sup>副教授, 邓明月<sup>1</sup>, 陈文涛<sup>2</sup>副研究员, 张雪<sup>1</sup>

(1 北京石油化工学院 安全工程学院, 北京 102617; 2 应急管理部 信息研究院, 北京 100029)

中图分类号: X937

文献标志码: A

DOI: 10.16265/j.cnki.issn1003-3033.2024.05.1570

**【摘要】** 为提高危险化学品企业安全管理水平,构建危险化学品事故因果关系事理图谱。首先,依据事故调查报告,提出一种基于改进 CasRel 的实体关系联合抽取模型,该模型通过关系感知双向编码器表征法(R-Bert)和 Span 指针网络提高文本信息的抽取准确性;其次,利用相似度计算方法进行事件泛化,使得图谱更加完整和准确,并将其存储在图数据库 Neo4j 中,以可视化的方式直观展示事件之间的关联;最后,基于构建的事理图谱,设计出与之匹配的问答系统架构,开发危险化学品事故因果关系智能问答系统。研究结果表明:改进的 CasRel 模型计算结果  $F_1$  值为 90.5%,较未改进模型准确率提升约 2%。构建的危险化学品事故事理图谱及智能问答系统,在多个评价指标表现良好,能够清晰揭示事件之间的逻辑关系;所提方法可满足危险化学品事故知识问答需求,协助探究事故发生的规律和潜在风险因素,推断事故发展趋势。

**【关键词】** 危险化学品事故; 事理图谱; 知识抽取; 因果关系; 智能问答系统

## Development of causal graph for hazardous chemical accidents

LI Hongru<sup>1</sup>, LUAN Tingting<sup>1</sup>, DENG Mingyue<sup>1</sup>, CHEN Wentao<sup>2</sup>, ZHANG Xue<sup>1</sup>

(1 School of Safety Engineering, Beijing Institute of Petrochemical Technology, Beijing 102617, China;

2 Information Research Institute, Emergency Management Department, Beijing 100029, China)

**Abstract:** The causality causal graph of hazardous chemical accidents was developed to improve the safety management level of hazardous chemical enterprises. Firstly, based on the accident investigation report, an entity-relationship joint extraction model was proposed through an improved CasRel technique. Furthermore, the proposed model aimed to improve the extraction accuracy of textual information by incorporating the relationship-aware bidirectional encoder representation method (R-Bert) and Span pointer network. Subsequently, similarity calculation methods were used to generalize the events to enhance the graph's comprehensiveness and accuracy. Then, the refined data was stored in the Neo4j graph database visualizing the associations between events. Finally, the corresponding question-answering system was proposed based on the developed causal graph, and then an intelligent question-answering system for the causality of hazardous chemical accidents was proposed. The results indicated that the  $F_1$  value calculated by the improved CasRel model was 90.5%, and the prediction accuracy of the proposed model was 2% higher than that simulated by the original model. The hazardous chemical accidents causal

\* 文章编号:1003-3033(2024)05-0195-09; 收稿日期:2023-11-14; 修稿日期:2024-02-21

\*\* 通信作者:栾婷婷(1982—),女,山东高密人,工学博士,副教授,主要从事风险评估、应急技术与管理等方面的研究。E-mail:luantt@bipt.edu.cn。

graph and intelligent question-answering system performed well in terms of multiple evaluation indexes, clearly revealing the logical relationship between events. Therefore, the proposed model in this study can meet question-answering needs of hazardous chemical accidents, facilitating the exploration of accident patterns and potential risk factors, and enabling accident trend prediction.

**Keywords:** hazardous chemical accidents; casual graph; knowledge extraction; causality; intelligent question-answering system

## 0 引言

中国化工行业在快速发展的同时,频繁发生的危险化学品事故给经济和环境带来了巨大影响,甚至造成严重的人员伤亡<sup>[1]</sup>,因此,有必要进行危险化学品事故分析。

近年来,安全分析研究中引入了知识库的概念,以实现知识建模和风险管理,知识图谱作为知识库的一种网络展现形式,为危化品事故的分析建模提供了新思路。2012年,谷歌提出知识图谱<sup>[2]</sup>,用于描述现实世界中实体或概念及其关系。相关研究表明:知识图谱可应用于危险化学品事故预防与管理<sup>[3-4]</sup>,将领域知识转化为可视化的图谱结构,有助于理解和分析潜在风险。然而,传统知识图谱在研究动态推理问题时具有一定的局限性,迫使研究者利用事理图谱进行动态推理问题研究<sup>[5]</sup>,以解决推理规则方面的不足。以事理图谱为基础的研究旨在深入理解事件之间的因果关系和动态演化,更强调动态关系和发展趋势。目前,国内外学者开始进行事理图谱的研究并取得了一定成果。YANG Congcong等<sup>[6]</sup>和哈工大社会计算与信息检索研究中心团队,均运用时间演化图谱和事理图谱的概念,

描述了新闻和金融领域的事件关系;WANG Hong等<sup>[7]</sup>构建了航空事故因果事件演化图谱;王洁宁等<sup>[8]</sup>提出基于事理图谱的事故分析方法,以识别飞行冲突的诱发事件与关键演化链路;张鹏翔<sup>[9]</sup>采用主题模式匹配+多维字符特征+BiLSTM+CRF的信息抽取方法,有效抽取了事故报告中关键信息,简化了数据分析。现有研究在危险化学品事故分析方面,还鲜见事理图谱模型研究。

鉴于此,笔者拟通过文本分析和预处理,提出改进的CasRel模型,再进行事件泛化处理,构建危险化学品事故事理图谱,并存储于Neo4j,利用该图谱构建智能问答系统,以可视化的方式揭示危险化学品事故的发生及演变特征,以期推断事故发展趋势,强化危化品安全管理。

## 1 危险化学品事故事理图谱的构建

### 1.1 事理图谱构建流程

危险化学品事故事理图谱是为危化品专业领域设计的一个知识图谱,采用自上而下的构造方法,包括数据获取和预处理、建模、提取、存储和应用5个阶段,如图1所示。

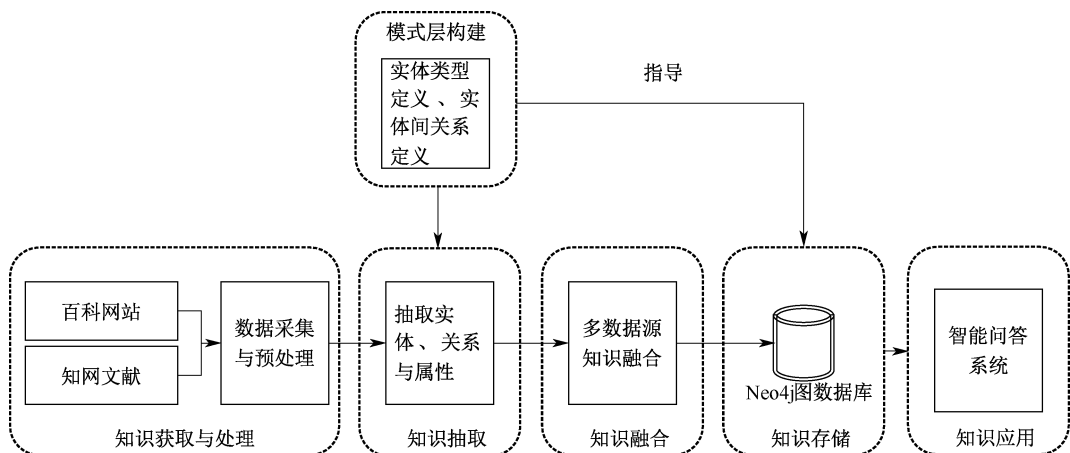


图1 事理图谱构建流程

Fig. 1 Causal graph development process

## 1.2 本体模式设计

为获得逻辑知识体系,需要构建本体,建立模型层的概念模板。本体表达领域概念和关系的规范化形式,是模式层构建的依据<sup>[10]</sup>。事件本体表示模型是特殊的领域本体,以事件为认知单元,提供粒度更

大、动态、完整的结构化知识,能描述事件的各要素及事件间的复杂关系<sup>[11]</sup>。借助本体模式,危险化学品事故实体的关系、属性和类型实现标准化,本体模式设计如图 2 所示。

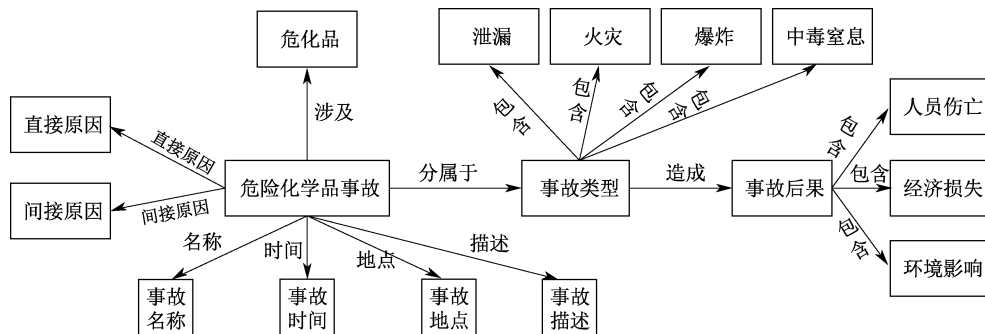


图 2 本体模式设计

Fig. 2 Ontology pattern design

## 1.3 数据采集和处理

从应急管理局官方网站、中国化学品安全协会的事故数据库以及安全管理网等收集 278 篇危险化学品事故调查报告,以此作为数据来源。经过语义分析、切分长语句、分词和去除停用词等文本处理,聚焦事故过程、原因和性质,删除无用数据,生成去噪后并具有具体逻辑语义的危险化学品事故语料库。

这些报告以 txt 格式存储,每个文件包含一个案例。由于文本数据缺乏完整句子结构和明确因果词汇,数据处理颇具挑战,数据处理如图 3 所示。

```

{"outputs":{"annotation":{"T":["",""],"type":"T","name":"事件原因","value":"芜湖融汇化工有限公司","start":29,"end":39,"attributes":{"id":1},"type":"T","name":"事件原因","value":"对液氯槽车充装液氯过程","start":56,"end":67,"attributes":{"id":2},"type":"T","name":"事件原因","value":"发生氯气泄漏","start":69,"end":75,"attributes":{"id":3},"type":"T","name":"事件结果","value":"相邻企业 19 人受伤住院","start":78,"end":89,"attributes":{"id":4},"type":"T","name":"事件结果","value":"直接经济损失 48 万元","start":90,"end":100,"attributes":{"id":5},"type":"T","name":"事件结果","value":"对液氯槽车充装液氯过程","start":56,"end":67,"attributes":{"id":6},"type":"T","name":"事件结果","value":"发生氯气泄漏","start":69,"end":75,"attributes":{"id":7}},"E":[""],"R":["",""],"name":"间接原因","from":1,"to":6,"arg1":"Arg2","arg2":"Arg2","name":"间接原因","from":2,"to":7,"arg1":"Arg1","arg2":"Arg2","name":"直接原因","from":3,"to":4,"arg1":"Arg1","arg2":"Arg2","name":"直接原因","from":3,"to":5,"arg1":"Arg1","arg2":"Arg2"},"A":[""],"time_labeled":1670985851087,"labeled":true,"content":"2020 年 8 月 29 日 17 时 28 分许,位于芜湖市鸠江区境内的芜湖融汇化工有限公司(以下简称“融汇化工”)液氯工段在对液氯槽车充装液氯过程中,发生氯气泄漏,造成相邻企业 19 人受伤住院、直接经济损失 48 万元。"}

```

图 3 数据处理

Fig. 3 Data processing

在保留报告事故直接和间接原因的基础上,通过数据预处理,获得 15 400 份危险化学品事故相关数据,将数据分配到不同样本中,数据集统计见表 1。

## 1.4 知识抽取

知识抽取是将原始数据转化为符合设计本体的知识三元组的过程,是图谱构建的重要环节,主要任

表 1 数据集统计

Table 1 Dataset statistics

数据集	训练集	验证集	测试集
数据集/条	10 280	2 560	2 560
三元组数量/条	56 930	10 990	10 690

务是提取事件实体和关系。

基于改进的 CasRel 联合抽取模型的整体结构如下:

1) 目标编码器模型。使用 R-Bert 架构,基于预训练双向编码器表征法(Bidirectional Encoder Representations from Transformers, BERT)模型,将关键文本数据转换为表征语义特征的特征向量。通过该模型,将句子的隐含语义输出分为标签、第 1 个实体的隐含信息和第 2 个实体的隐含信息,并选择性提供隐含语义输出的 2 个实体的特征。

首先,输入序列的嵌入层为:

$$\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\} \quad (1)$$

式中: $\mathbf{X}$  为输入序列嵌入向量的集合; $\mathbf{x}_i$  为输入序列中第  $i$  个位置词的向量。

其次,输入序列的编码层,得到编码表示:

$$\mathbf{H} = \Omega(\mathbf{X}) = \{\mathbf{h}_i\}_{i=1}^n \quad (2)$$

式中: $\mathbf{H}$  为每个词  $\mathbf{x}_i$  对应的隐含语义向量; $\Omega$  为 Bert 模型的输出。

2) 高级客体关系标记解码器。用于识别与头实体在特定关系上对应的尾实体,特定关系是直接关系或间接关系。解码过程中使用 Span 指针网络方法,在起始位置  $\mathbf{p}_i^s$  和结束位置  $\mathbf{p}_i^e$  中标注所有客体的位置信息,并通过 2 个 Linear 层进行解码操作,

解决实体覆盖的情况。

客体位置向量为:

$$\{p_j = (p_j^s, p_j^e)\}_{j=1}^m \quad (3)$$

式中: $j$ 为实体的索引,用于标识客体位置向量序号; $m$ 为客体的数量。

关系标记解码器的输出向量: $\{r_j\}_{j=1}^m$ 。

3) 低级主体标记解码器。类似于高级关系解码器,用于识别表征目标事件原因的实主体。同样使用 Span 指针网络方法,在起始位置  $q_k^s$  和结束位置  $q_k^e$  序列中标注所有主体的位置信息。

主体位置向量为:

$$\{q_k = (q_k^s, q_k^e)\}_{k=1}^f \quad (4)$$

式中: $k$ 为实体索引,用于标识主体位置向量序号; $f$ 为主体的数量。

主体标记解码器的输出向量: $\{s_k\}_{k=1}^f$ 。

4) 三元组预测。将目标编码器模型的输出  $h_i$ 、高级客体关系标记解码器的输出  $r_j$  和低级主体标记解码器的输出  $s_k$  进行拼接,形成拼接向量  $z_g$ 。将其输入到全连接层,通过权重矩阵  $U$  和偏置向量  $b$  进行线性变换和非线性变换,得到输出向量  $o_g$ 。最后,使用 softmax 分类器将输出向量  $o_g$  转换为关系分类的概率向量  $y_g$ 。

拼接向量:

$$z_g = [h_i, r_j, s_k] \quad (5)$$

全连接层输出向量:

$$o_g = Uz_g + b \quad (6)$$

最终输出向量:

$$y_g = \text{softmax}(o_g) = \frac{\exp(o_g)}{\sum_{g=1}^m \exp(o_g)} \quad (7)$$

在训练 CasRel 联合抽取模型时,首先,通过目标编码器模型处理预设训练样本得到共享参数;然后,分别使用这些参数中的头实体和尾实体样本训练低级主体标记解码器和高级客体关系标记解码器;最后,利用训练得到头实体位置信息和多层 Linear 下的尾实体信息优化模型损失,得到优化后的模型。损失函数包括头实体、关系数和头尾实体的联合模型的损失函数。

为提高分类准确性和稳定性,采用聚焦损失 (Focal Loss) 函数<sup>[12]</sup>。传统的交叉熵损失函数中,易分样本和难分样本对损失的贡献是均等的。然而,在实际应用中,往往存在大量的易分样本,这会导致模型训练过程中被这些样本多主导,从而忽视更有信息量的难分样本。Focal Loss 通过引入调制

因子  $\alpha \in [0, 1]$  和参数  $\gamma \in [0, 5]$  减少易分样本的损失,并增加难分样本的损失,提高分类准确性和稳定性。其计算式如下:

$$L(y_g) = -\alpha(1 - y_g)^\gamma \ln(y_g) \quad (8)$$

式中  $L$  为聚焦损失函数。

易分类样本预测概率  $y_g$  相对较高,调制因子接近 0,减少样本的总损失。相反,难分类样本预测概率  $y_g$  相对较低,调制因子可能接近 1,增大了样本的总损失,以更聚焦于难分类样本,提高整体分类的准确度和稳定性。

改进 CasRel 联合抽取模型结构如图 4 所示,R-Bert 架构如图 5 所示,Span 解码架构如图 6 所示。

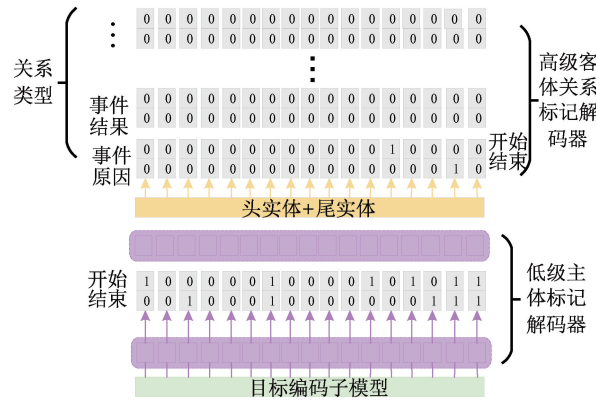


图 4 改进 CasRel 模型结构

Fig. 4 Improved CasRel model structure

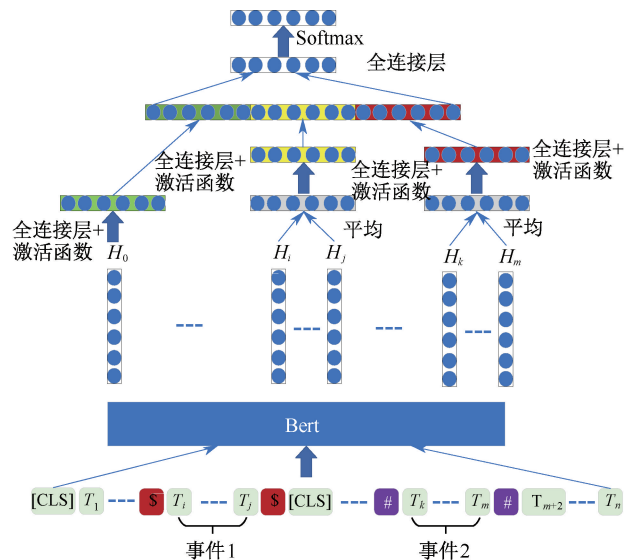


图 5 R-bert 模型

Fig. 5 R-bert model

### 1.5 模型抽取步骤

1) 输入句子到预训练的 BERT 模型,通过 Embedding 层得到词向量和位置编码。再经过双

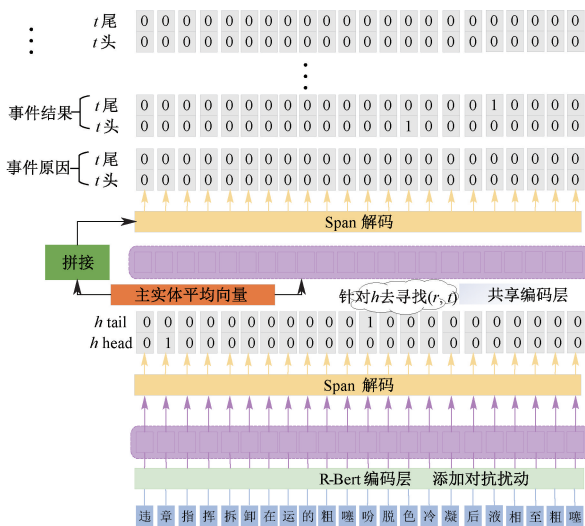


图 6 Span 解码框架模型

Fig. 6 Span decoding framework model

Transformer 层得到单词特征向量。

2) 使用首尾指针标注方法抽取主体  $\lambda$ 。对于每个检测到的主体和所有预定义的关系类型  $\zeta$ , 预测相应的客体  $\xi$ 。

3) 通过搭建级联结构的指针网络, 解码输出关系三元组  $\langle \lambda, \zeta, \xi \rangle$  [13]。

使用改进的 CasRel 联合抽取模型从事件报告中提取实体和关系数据, 并将结果存为 json 格式的危化品事故信息三元组, 见表 2, 为构建事理图谱做基础准备。

### 1.6 事件泛化

在处理批量文档时, 合并多个事件三元组集合, 利用相似度计算方法将相同事件归为同一事件族。同时, 将大量的事件三元组泛化为一组相似的事件族, 并计算权重, 以便进行更好地组织、分析与预测事件。

表 2 知识抽取结果示例

Table 2 Example of knowledge extraction results

事故名称	类别	事故直接原因描述	文本三元组识别结果
三门峡市河南省煤气(集团)有限责任公司义马气化厂“7·19”重大爆炸事故 <sup>[14]</sup>	爆炸	空分装置冷箱内发生泄漏, 直至冷箱板出现裂纹, 富氧液体泄漏至珠光砂中, 使碳钢材质的冷箱构件在低温和压力增高的共同作用下裂纹扩大, 直至冷箱失稳坍塌, 砸裂液氧贮槽, 贮槽内大量液氧迅速外泄气化, 高纯氧遇可燃物发生爆炸, 并引发冷箱中的铝质填料等殉爆	(空分装置冷箱, 发生泄漏, 冷箱板出现裂纹); (空分装置冷箱, 泄漏, 珠光砂中); (富氧液体, 泄漏, 至珠光砂中); (珠光砂, 冷箱构件, 裂纹扩大); (冷箱构件, 低温和压力增高的共同作用, 裂纹扩大); (裂纹扩大, 冷箱失稳坍塌, 砸裂液氧贮槽); (砸裂液氧贮槽, 大量液氧, 外泄气化); (液氧贮槽, 大量液氧, 外泄气化); (高纯氧, 遇可燃物, 发生爆炸); (高纯氧, 爆炸, 引发冷箱中的铝质填料等殉爆); (爆炸, 引发, 冷箱中的铝质填料等殉爆)

事件族是危险化学品相似事件集合, 通过聚类相似事件形成, 并对这些事件的因果关系进行分类和统计。为表达逻辑不确定性, 定义迁移度  $T_r$ 、覆盖度  $C_v$ 、支持度  $B_p$  来描述事件族间的因果关系属性。当 2 个事件族 Q 和 V 中各自包含提取的事件数  $N_Q$  和  $N_V$ , 其原因属于 V, 结果属于 Q 的因果事件为  $N_{QV}$ , 用 V 和 Q 之间的迁移度  $T_r$  衡量它们之间的相互影响<sup>[15]</sup>。

$$T_r = N_{QV} / N_V \quad (9)$$

$$C_v = N_{QV} / N_Q \quad (10)$$

$$B_p = N_{QV} / (N_Q + N_V) \quad (11)$$

与单一指标相比, 多指标能更完整地描述不确定性, 有利于更灵活地制定未来事故原因推断规则。

使用词向量嵌入 (word2vec) 模型, 将文本中单词转化为半结构化向量, 以计算并衡量其相似性。每个单词表示为  $M$  维向量, 通过在语料库中上下文中训练这些向量来捕捉其语义信息。考虑每个单词上下  $2c$  个单词的上下文, 并使用步长  $\eta$  控制模型的学习速度。最终, 获得每个单词的词向量  $w$ , 具体过

程如下:

1) 假设词向量序列  $x_w^\alpha$ , 内部节点参数向量序列  $\theta_\beta^\alpha$ , 训练样本  $\{\text{context}(w)\}$  建立最优二叉树;

$$\theta_\beta^{\alpha+1} = \begin{cases} \theta_\beta^\alpha + \eta(1 - d_\beta^w - \sigma((\theta_\beta^\alpha)^T x_w^\alpha)) x_w^\alpha, & d_\beta^w = 0 \\ \theta_\beta^\alpha - \eta(d_\beta^w - \sigma((\theta_\beta^\alpha)^T x_w^\alpha)) x_w^\alpha, & d_\beta^w = 1 \end{cases} \quad (12)$$

式中:  $\alpha$  为迭代次数;  $d_\beta^w$  为词  $w$  在第  $\beta$  个输出节点上的目标值。这个值通常是一个二进制值 (0 或 1), 用于表示在训练数据中, 词汇  $w$  是否与第  $\beta$  个输出节点相关联。

2) 更新词向量  $x_w^\alpha$ :

$$x_w^{\alpha+1} = x_w^\alpha + \eta \sum_{\beta=2}^{l_w} (1 - d_\beta^w - \sigma((\theta_\beta^\alpha)^T x_w^\alpha)) \theta_\beta^\alpha \quad (13)$$

式中:  $\sigma$  为 sigmoid 函数;  $\eta$  为学习率;  $d_\beta^w$  为 Huffman 编码;  $l_w$  为词 ( $w$ ) 的 Huffman 编码长度。

3) 相似性计算。完成词向量嵌入和模型训练后, 使用事件三元组中的词向量计算它们之间的余

弦相似度,以进行共指消解。通过多次梯度迭代获取事件相似性,用于衡量事件间的关系。

在度量时,将事件元组表示为  $E = (\mu, \rho, \psi)$ ,  $\mu$  为头实体,  $\rho$  为关系,  $\psi$  为尾实体,并计算  $(\mu, \rho, \psi)$  各个论元之间的相似性。

$$A(\mu_\varphi, \mu_\tau) = \frac{\mu_\varphi \cdot \mu_\tau}{|\mu_\varphi| |\mu_\tau|} \quad (14)$$

式中:  $A(\mu_\varphi, \mu_\tau)$  为事件对  $E_\varphi$  和  $E_\tau$  的头实体相似性。

$$\Lambda(\rho_\varphi, \rho_\tau) = \frac{\rho_\varphi \cdot \rho_\tau}{|\rho_\varphi| |\rho_\tau|} \quad (15)$$

式中  $\Lambda(\rho_\varphi, \rho_\tau)$  为事件对  $E_\varphi$  和  $E_\tau$  的关系相似性。

$$\Sigma(\psi_\varphi, \psi_\tau) = \frac{\psi_\varphi \cdot \psi_\tau}{|\psi_\varphi| |\psi_\tau|} \quad (16)$$

式中  $\Sigma(\psi_\varphi, \psi_\tau)$  为事件对  $E_\varphi$  和  $E_\tau$  的尾实体相似性。

$$\Phi(E_\varphi, E_\tau) = \frac{A(\mu_\varphi, \mu_\tau) + \Lambda(\rho_\varphi, \rho_\tau) + \Sigma(\psi_\varphi, \psi_\tau)}{3} \quad (17)$$

式中  $\Phi(E_\varphi, E_\tau)$  为事件对  $E_\varphi$  和  $E_\tau$  的事件相似性。

在实际应用中,为有效表达事件元组之间的相似性,并尽可能全面地涵盖文本内容,常采用相似度阈值法。通过设定一个合理的阈值,筛选出与特定事件有足够相似度的时间元组。在选择相似度阈值时,需要深入分析和试验。考虑到并非所有事件都需要与因果图和事件图相关联,因为有些弱相关的事件图元与核心事件的关联性不强,可能会成为孤立的点。统计阈值设置对相似事件匹配精度的影响,最终对2个事件的相似度设置最小阈值0.6,即  $A(E_\varphi, E_\tau) \geq 0.6$  时才会被消解。相似度计算流程如图7所示。

利用式(14)一式(17)进行事件相似性计算,并进行事件相似度消解,“违规进入有限空间”和“未经允许进入受限空间”的相似度为0.691,以及“在有限空间内违规操作”的相似度为0.517。因此,判定“违规进入有限空间”和“未经允许进入受限空间”为相似事件。“发生氮气窒息”和“氮气窒息中毒”也被判定为相似事件,如图8所示。

通过相似度计算,将复杂事故的因果关系抽象化,构建完整的事理图谱。因果关系抽象化有助于统一事件描述,使得事件更容易被识别和理解。

### 1.7 知识存储

通过 Cypher 的 LOAD CSV 语句将抽取出的危

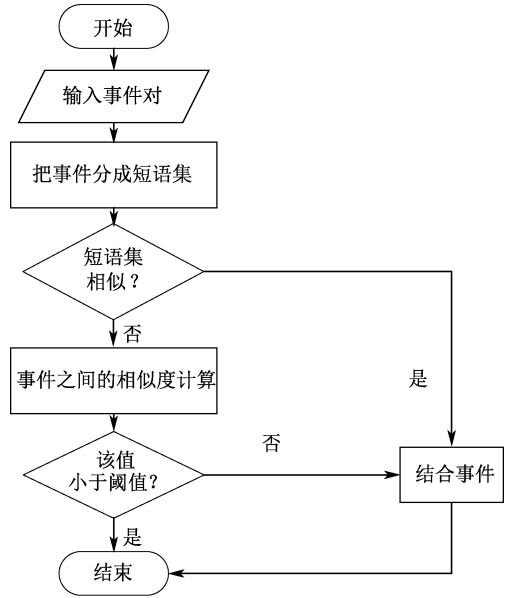


图7 相似度计算流程

Fig.7 Similarity calculation flow chart

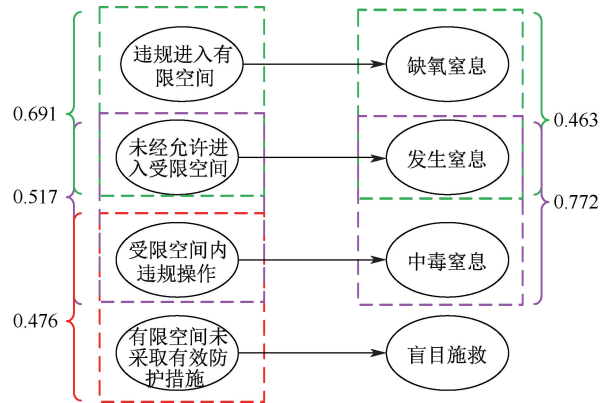


图8 事件相似度

Fig.8 Event similarity

险化学品安全事故数据导入 Neo4j 数据库,所构建的事理图谱如图9所示。

### 1.8 知识应用

系统利用 Neo4j 数据库存储事故因果关系,使用 Python 中的 Django 框架连接前后台,选择图表库 ECharts 实现前端图谱可视化,并构建智能问答系统,如图10所示。

该系统基于自定义的问答模板和 Neo4j 内置的查询语言 Cypher,实现了自然语言处理与图数据库查询的高效结合。用户可以通过自然语言提问,系统会智能解析问题,并自动转换为对应的 Cypher 查询语句,从而快速检索和展示相关的危险化学品事故因果关系。

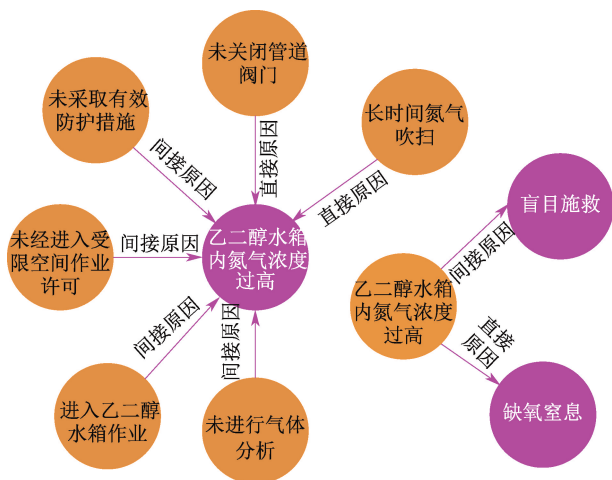


图 9 事理图谱展示案例

Fig. 9 Example of causal graph



图 10 智能问答系统

Fig. 10 Intelligent question-answering system

## 2 改进的 CasRel 模型效能测试与分析

### 2.1 配置与参数调整

在 Linux 操作系统上,使用 Keras 框架和 Tensorflow2.2 后端,通过图形处理器加速模型训练。具体试验参数设置如图 11 所示。

```
self.learning_rate = 1e-5
self.batch_size = 4
self.max_epoch = 100
self.max_len = 256
self.rel_num = 4
self.optimizer = 'adamW' or self.optimizer = 'adan'
self.weight_decay = 0.01
self.teacher_pro = 0.8
self.use_focal = True
self.sub_threshold = 0.5
self.dropout = 0.4
self.attention = "plus"
```

图 11 试验参数设置

Fig. 11 Experimental parameter setting

### 2.2 评估指标

采用评价模型性能的指标分别为:精准率  $P$ 、召

回率  $R$ 、 $F_1$  值。当预测产生的实体—关系三元组与真实的实体—关系三元组完全一致时,作为正确识别的三元组<sup>[16]</sup>。

$$P = \frac{TP}{TP + FP} \times 100\% \quad (18)$$

$$R = \frac{TP}{TP + FN} \times 100\% \quad (19)$$

$$F_1 = 2 \frac{PR}{P + R} \times 100\% \quad (20)$$

式中:TP 为实际情况是正的,且预测为正;FP 为实际情况是负的,预测为正;FN 为实际情况是正的,预测为负。

### 2.3 对比试验

为验证模型性能,与 NovelTagging<sup>[17]</sup>、CopyRE<sup>[18]</sup>、GraphRel<sup>[19]</sup>、ETL-Span 和 CasRel<sup>[20]</sup> 基线模型进行对比试验,结果如图 12 所示。由图 12 可知:文中模型性能更好, $F_1$  值达到 90.5%。

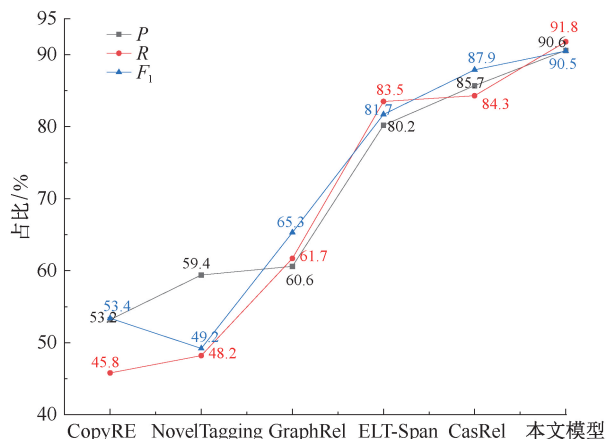


图 12 对比试验结果

Fig. 12 Comparisons between experimental results

### 2.4 消融试验

为深入探索模型各组件对性能的具体贡献,进行消融试验。在试验当中,分别去除 R-Bert 预训练模型和 Span 解码方式,采用更传统的模型进行对比,结果见表 3。试验结果显示,当不使用 R-Bert 模型时,三元组重叠问题的处理能力明显下降,导致整体性能下滑。同样,当移除 Span 指针解码方式时,上下文语义信息的捕获能力受损,进而影响模型的  $F_1$  值。这些发现不仅强调了模型各组件的协同作用,还突显了主体和关系信息在三元组提取中的核心地位。与原始的 CasRel 模型相比,改进模型在保证准确性的基础上,更有效地控制召回率,从而实现整体性能的提升。

表3 消融试验结果

Table 3 Ablation experimental results

模型	$P/\%$	$R/\%$	$F_1/\%$
去掉 R-bert	89.1	88.1	88.3
去掉 Span	90.1	88.7	87.2
改进的 CasRel 模型	90.6	91.8	90.5

### 3 结论

1) 文中构建的危险化学品事故事理图谱及智能问答系统,在多个评价指标上表现优异。该图谱可清晰展现事件之间的复杂逻辑关系,问答系统能够满足危险化学品事故的专业知识问答需求。通过该图谱和系统能探究事故发生的深层次特征,识别潜在的风险因素,合理推断事故的可能发展趋势,为

事故预防和应对提供有力支持。

2) 文中提出的 CasRel 模型试验性能良好,可将精度提高约 5%,召回率提高 6%, $F_1$  得分提高 2%。采用 R-Bert 模型将头部的平均编码向量与输入层向量结合,形成 2 个二分类器,有助于提高头实体抽取的准确性和可靠性。引入 Span 指针网络编码层,可提升模型的泛化能力和抽取效率。

3) 用于事故知识提取的数据源有限,与公共数据库相比,文中开发的数据集相对较小,限制了模型性能的提升,其性能对于支持全自动事故风险分析仍具有挑战性。相对较低的召回率可能导致信息提取过程中关键要素的遗漏,导致在事故分析预测等实际应用中需要人工审核。

### 参考文献

- [1] 刘成军. 炼油化工装置水锤事故分析及处理措施[J]. 石油与天然气化工, 2018, 47(3): 101-107.  
LIU Chengjun. Analysis and treatment measures of water hammer accidents occurring in refining & chemical plants[J]. Chemical Engineering of Oil & Gas, 2018, 47(3): 101-107.
- [2] SINGHAL A. Introducing the knowledge graph: things, not strings[EB/OL]. (2012-05-16). <https://www.blog.google/products/search/introducing-knowledge-graph-things-not/>.
- [3] ZHANG Qi, WEN Yuanqiao, ZHOU Chunhui, et al. Construction of knowledge graphs for maritime dangerous goods[J]. Sustainability, 2019, 11(10): DOI:10.3390/su11102849.
- [4] 龙丹冰, 魏君豪, 杨成. 知识图谱改进的施工行为安全风险与危险位置识别算法[J]. 中国安全科学学报, 2021, 31(12): 10-16.  
LONG Danbing, WEI Junhao, YANG Cheng. An improved algorithm for construction behavior safety risk and hazard location identification based on knowledge graph[J]. China Safety Science Journal, 2021, 31(12): 10-16.
- [5] 刘焕勇, 薛云志. 我们的实践: 事理图谱, 下一代知识图谱[EB/OL]. (2018-12-25). <https://blog.csdn.net/lhy2014/article/details/85247268>.
- [6] YANG Congcong, SHI Xiaodong, WEI Chihping. Discovering event evolution graphs from news corpora[J]. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 2009, 39(4): 850-863.
- [7] WANG Hong, ZHU Han, BAI Yunqing, et al. Construction of causality event evolutionary graph of aviation accident[C]. The 5<sup>th</sup> International Conference on Transportation Information and Safety, 2019: 692-697.
- [8] 王浩宁, 朱妍. 基于事理图谱的飞行冲突事故演化研究[J]. 安全与环境学报, 2023, 23(6): 1961-1969.  
WANG Jiening, ZHU Yan. Research on the evolution of flight conflict accidents based on event logic graph[J]. Journal of Safety and Environment, 2023, 23(6): 1961-1969.
- [9] 张鹏翔. 多维字符特征表示的铁路设备事故信息抽取方法[J]. 中国安全科学学报, 2022, 32(6): 109-114.  
ZHANG Pengxiang. Information extraction method for railway equipment accidents based on multidimensional character feature representation[J]. China Safety Science Journal, 2022, 32(6): 109-114.
- [10] GRUBER T R. Toward principles for the design of ontologies used for knowledge sharing? [J]. International Journal of Human-computer Studies, 1995, 43(5/6): 907-928.
- [11] 李跃艳, 王昊, 邓三鸿, 等. 面向事件本体的医学文本语义关联化研究[J]. 情报学报, 2022, 41(5): 497-511.

- LI Yueyan, WANG Hao, DENG Sanhong, et al. Research on semantic relevance of medical text oriented to event ontology[J]. Journal of the China Society for Scientific and Technical Information, 2022,41(5):497-511.
- [12] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020,42(2): 318-327.
- [13] 拓雨欣,薛涛.融合指针网络与关系嵌入的三元组联合抽取模型[J].计算机应用,2023,43(7):2116-2124.
- TUO Yuxin, XUE Tao. Joint triple extraction model combining pointer network and relational embedding[J]. Journal of Computer Applications, 2023,43(7):2116-2124.
- [14] 海南省应急管理厅.三门峡市河南煤气(集团)有限责任公司义马气化厂“7·19”爆炸事故调查报告[EB/OL]. (2020-08-10). [http://yjgl.hainan.gov.cn/xxgk/sgdcbg/202008/t20200810\\_2831571.html](http://yjgl.hainan.gov.cn/xxgk/sgdcbg/202008/t20200810_2831571.html).
- [15] 葛任贤.基于互联网知识抽取的汽车故障事理图谱构建方法[D].广州:广东技术师范大学,2023.
- GE Renxian. The construction method of automobile fault event-logic graph based on internet knowledge extraction[D]. Guangzhou: Guangdong Polytechnic Normal University, 2023.
- [16] 张维宁,申喜凤,李美婷,等.融合关系标签和位置信息的中文医疗文本因果关系抽取方法研究[J].医学信息学杂志,2024,45(1):21-26.
- ZHANG Weining, SHEN Xifeng, LI Meiting, et al. Study on the method of causality extraction from Chinese medical texts by integrating relational label and location information[J]. Journal of Medical Informatics, 2024,45(1):21-26.
- [17] ZHENG Suncong, WANG Feng, BAO Hongyun, et al. Joint extraction of entities and relations based on a novel tagging scheme[C]. the 55<sup>th</sup> Annual Meeting of the Association for Computational Linguistics, 2017: 1227-1236.
- [18] ZENG Xiangrong, ZENG Daojian, HE Shizhu, et al. Extracting relational facts by an end-to-end neural model with copy mechanism[C]. the 56<sup>th</sup> Annual Meeting of the Association for Computational Linguistics, 2018:506-504.
- [19] FU T J, LI P H, MA Weiyun. GraphRel: modeling text as relational graphs for joint entity and relation extraction[C]. the 57<sup>th</sup> Meeting of the Association for Computational Linguistics, 2019: 1409-1418.
- [20] WEI Zhepei, SU Jianlin, WANG Yue, et al. A novel cascade binary tagging framework for relational triple extraction[C]. the 58<sup>th</sup> Annual Meeting of the Association for Computational Linguistics, 2020: 1476-1488.



**作者简介:** 李红儒 (1996—),女,河北唐山人,硕士研究生,主要研究方向为安全风险评估与应急管理技术。E-mail:2021520204@bipt.edu.cn。