

基于双向长短时记忆网络的地铁应急知识 抽取与推理

叶雨涛^{1,2,3}, 王鹏玲^{1,2,3}, 徐瑞华^{1,2,3}, 肖晓芳^{1,2,3}, 葛健豪^{1,2,3}

(1. 同济大学交通运输工程学院, 上海 201804; 2. 同济大学上海市轨道交通结构耐久与系统安全重点实验室, 上海 201804;
3. 同济大学上海市多网多模式轨道交通协同创新中心, 上海 201804)

摘要: 为解决文本类地铁应急处置流程存在的流程顺序关系不明确、流程执行人员模糊等问题, 提出了基于 BiLSTM-CRF (Bidirectional Long Short-Term Memory-Conditional Random Field) 的地铁应急知识抽取与推理方法。首先, 利用 BiLSTM-CRF 方法对地铁应急处置流程的文本资料进行命名实体识别, 完成文本资料的知识抽取; 其次, 选用 TransD 模型对识别后实体数据进行知识推理, 从而完成以实体和属性对为节点、关系对为边的知识图谱构建; 最后, 利用 Neo4j 图数据库对构建的地铁应急处置流程知识图谱进行了可视化展示和案例分析。研究表明, 基于 BiLSTM-CRF 的知识抽取模型的精确率、召回率和 F1 值均达到了 90% 以上, 且基于 BiLSTM-CRF 的 TransD 模型的推理结果准确率提升了 22.92%, 保证了知识图谱构建的准确性, 可为地铁应急管理提供决策支持。

关键词: 地铁应急处置; 知识图谱; 条件随机场的双向长短时网络; TransD 模型; 知识抽取

中图分类号: U491.1

文献标志码: A

Metro Emergency Knowledge Extraction and Knowledge Reasoning Based on BiLSTM-CRF

YE Yutao^{1,2,3}, WANG Pengling^{1,2,3}, XU Ruihua^{1,2,3},
XIAO Xiaofang^{1,2,3}, GE Jianhao^{1,2,3}

(1. College of Transportation Engineering, Tongji University, Shanghai 201804, China; 2. Shanghai Key Laboratory of Rail Infrastructure Durability and System Safety, Tongji University, Shanghai 201804, China; 3. Shanghai Collaborative Innovation Research Center for Multi-network and Multi-model Rail Transit, Tongji University, Shanghai 201804, China)

Abstract: To address issues such as the unclear sequence

of procedures and ambiguity in the personnel responsible for executing the emergency response procedures in text-based metro emergency response processes, this paper proposes a knowledge extraction and knowledge reasoning method for metro emergency response procedures based on knowledge graph of bidirectional long short-term memory-conditional random field (BiLSTM-CRF). First, the BiLSTM-CRF method is used to identify the named entity of the text data of the metro emergency response process, and complete the knowledge extraction of the text data. Then, the TransD model is selected to conduct knowledge inference on the identified entity data, thereby completing the construction of a knowledge graph with entities and attribute pairs as nodes and relational pairs as edges. Finally, the Neo4j graph database is used to visualize and analyze the knowledge graph of metro emergency response process. The research results show that the precision, recall, and F1 value of the knowledge extraction model based on BiLSTM-CRF have all reached more than 90%, and the accuracy of the inference results of the TransD model based on BiLSTM-CRF has increased by 22.92%, ensuring the accuracy of knowledge graph construction and providing decision support for subway emergency management.

Keywords: metro emergency response; knowledge graph; bidirectional long short-term memory-conditional random field (BiLSTM-CRF); TransD model; knowledge extraction

城市地铁系统的复杂性和其突发事件的不可预测性要求有一个清晰、高效的应急处置流程。现阶段, 大多数地铁运营公司均以普通文档的形式对突

收稿日期: 2023-08-03

基金项目: 国家自然科学基金联合基金(U2368216); 国家自然科学基金青年科学基金(72101184); 上海市自然科学基金(23ZR1467400)

第一作者: 叶雨涛, 博士生, 主要研究方向为轨道交通信息数据挖掘。E-mail: yutaoye@tongji.edu.cn

通信作者: 王鹏玲, 教授, 博士生导师, 工学博士, 主要研究方向为数据挖掘和列车运行控制。

E-mail: pengling_wang@tongji.edu.cn



论文
拓展
介绍

发事件应急处置流程进行了保存,但没有进行系统整理和数字化存储,且文本类应急处置流程存在结构化程度低、不易快速查阅、处置流程不明确等问题,因此,为了满足应急处置准确性、便捷性和普适性的要求,对地铁应急处置流程进行建模分析,对于地铁应急管理具有十分重要的实践和理论意义。

知识图谱可以从多种类型的复杂数据中抽取实体和关系,转化为简单、清晰的“头实体—关系—尾实体”的三元组,以图数据的方式对文本数据进行结构化存储和展示^[1],被广泛应用于地理^[2]、医疗^[3]、图书档案^[4]等领域。现有关于知识图谱在城市交通领域的研究大多集中于知识图谱的构建以及结果预测方面。Zhang等^[5]和Zhou等^[6]利用知识图谱的方法分别实现了交通事故推理和城市道路拥堵区域的预测。朱广宇等^[7]基于构建城市轨道交通突发事件知识图谱,结合关系图卷积神经网络模型提出一种城市轨道交通突发事件演化结果预测方法。此外,知识图谱也被运用于铁路运维^[8]。林海香等^[9]从实际出发,提出了基于建筑信息模型和数据驱动的铁路设备运维多模态知识图谱构建方法。上述研究表明知识图谱在基于交通流量、交通事故等数字型数据的挖掘、处理和应用等领域有着较好的研究前景,虽然已有知识图谱研究基于铁路运营文本数据进行了相关探索,但在地铁应急管理领域的文本类数据应用尚未开展相关研究。因此,文本类地铁应急处置数据因描述多样化,容易导致流程表述不清、执行人员职责模糊等问题,这对构建精准、高效的地铁应急处置流程知识图谱提出了更高要求。

知识图谱的构建采用自然语言处理(Natural Language Processing, NLP)方法对文本资料进行命名实体识别(Named Entity Recognition, NER),以实现实体标注和关系链接^[10]。当前,命名实体识别应用比较广泛的方法有卷积神经网络(Convolutional Neural Network, CNN)、长短时记忆网络(Long Short-Term Memory, LSTM)、双向长短时记忆网络(Bi-directional Long Short-Term Memory, BiLSTM)等。在此基础上,条件随机场(Conditional Random Field, CRF)被引入了BiLSTM模型构成了BiLSTM-CRF模型,进一步提升了命名实体识别的准确度,逐渐成为了命名实体识别的主流方法^[11-12]。目前, BiLSTM-CRF模型在文本类地铁应急处置流程领域尚未有相关应用,但由于模型的双向性,更适用于文本类地铁应急处置流程的专业术语识别和上下文联系紧密的文本处理。因此,本文将BiLSTM-CRF模

型用于地铁应急处置流程文本数据的实体识别。

基于此,需要利用知识推理模型来挖掘潜在的实体关系,用于实现文本类地铁应急处置流程知识图谱的构建。现阶段,应用最为广泛的推理模型有TransE(Translation Embedding)^[13]、TransH(Translation on Hyperplanes)模型^[14]、TransR(Translation by Relation-specific Matrices)模型^[15]、TransD(Translation via Dynamic Mapping Matrix)模型^[16]。上述模型均是基于Word2vec模型扩展而来,基本原理是将词映射为结构化的向量,通过计算向量间的距离来刻画实体之间的语义关联^[17]。其中,TransD模型更适用于多对多的复杂网络结构和关系推理^[18],更适合处理地铁应急处置流程中多主体间的相互联系,因此,选取TransD模型来实现地铁应急处置流程的知识推理。

综上,为解决地铁应急处置流程的文本数据存在结构化程度低、不易快速查阅、处置流程不明确等问题,利用BiLSTM-CRF方法对地铁应急处置流程文本进行实体识别,完成文本类地铁应急处置流程的知识抽取并进行结构化存储。在此基础上,通过知识推理将上述结构化数据以知识图谱的方式实现对文本类地铁应急处置流程的模型构建,以图形方式代替文本信息实现地铁应急处置流程的清晰可视化展示以及流程关系间的深度挖掘。

1 问题描述

1.1 现有应急处置流程案例分析

现阶段大多数地铁运营公司均以普通文档的形式对突发事件应急处置流程进行了保存。但受限于各地铁公司发展现状、中文语言灵活多变等原因,一方面导致了现有的文本类地铁应急处置流程对同一应急事件的描述存在多样化的差异,另一方面也暴露了文本描述的地铁应急处置流程存在流程顺序关系不明确、流程执行人员模糊、流程间联系不紧密等问题,以某地轨道交通火灾应急处置流程为例^[19],具体示例如下所示。

例1 “步骤2、确认为火灾误报警,由车站行车值班员消音处理后向环控调度报告故障;步骤3、确认为真实火灾,且为可控制火灾时,由车站人员现场灭火处理。”

例2 “步骤5、车站行车值班员报行车调度、环控调度、站长、部门安全生产专员、119,视情况拨打120、110;步骤6、值班站长启动乘客疏散应急预案,通知站台岗疏散站台乘客。”

例3 “(处置流程A)步骤4、车站行车值班员报行车调度、环控调度、站长、部门安全生产专员、119,视情况拨打120、110。(处置流程B)步骤3、指令车站将FAS主机切换到自动位置,执行相应的火灾模式。做好引导协助疏散并视情况拨打119、120、110。”

例1显示了在同一处置流程中步骤2和步骤3之间并不是标号的顺序关系,而是2个步骤不可能同时发生的互斥关系,应急处置流程顺序关系不明确导致应急处置人员按照处置流程操作时需要选择判断,一定程度上影响了应急处置效率。例2为车站值班人员的应急处置流程步骤,但其相邻2个步骤的应急处置执行人员发生变动,不易于应急处置操作的开展。例3罗列了2套应急处置流程中的一项步骤,且2个步骤之间存在先后顺序关系,表明2套流程之间存在相应的处置关系,但通常缺失对上述情况的描述,存在流程间联系不紧密的问题。

1.2 基于知识图谱的地铁应急处置流程构建

为解决上述问题,利用BiLSTM-CRF方法对地铁应急处置流程文本进行实体识别,并通过知识推理以知识图谱的方式实现文本信息的数据化存储和流程关系间的深度挖掘。

采用基于BiLSTM-CRF的地铁应急处置流程知识图谱的构建,首先需要收集基础文本资料并明确主体流程与流程间的相互关系。其次,运用BiLSTM-CRF方法将数据按照实体、关系、属性的类别进行命名实体识别,即知识抽取过程。命名实体识别后的数据经知识推理后完成实体对、关系对和属性对存储。最后,选用TransD模型进行知识推理用于明确流程实体之间的相互关系,即以“头实体—关系—尾实体”三元组的形式描述实体及其之间的关系,从而实现以实体和属性对为节点、关系对为边的知识图谱构建。三元组示例如下,文本描述内容为“地铁站台发生了火灾”,基于BiLSTM-CRF方法和TransD模型即可完成“火灾—事件地点—地铁站台”三元组的构建,其中“火灾”为头实体,“地铁站台”为尾实体,“事件地点”为关系,头实体通过关系指向尾实体,表明应急事件“火灾”发生的地点为“地铁站台”。最后,基于Neo4j平台将三元组数据以图结构方式进行展示,从而完成文本类地铁火灾应急处置流程的知识图谱的搭建。

地铁应急处置流程知识图谱通过对基础数据的知识抽取和知识推理从而实现数据存储和图谱构建,所述流程如图1所示。其中基础数据包括应急处置预案、应急操作规范、应急处置评价等,然后依

照事件主体、事件等级、事件原因等特征信息分类对基础数据进行知识抽取,同时对相关特征信息进行知识推理,完成概念图谱、案例图谱、逻辑图谱等不同类型的图谱构建。

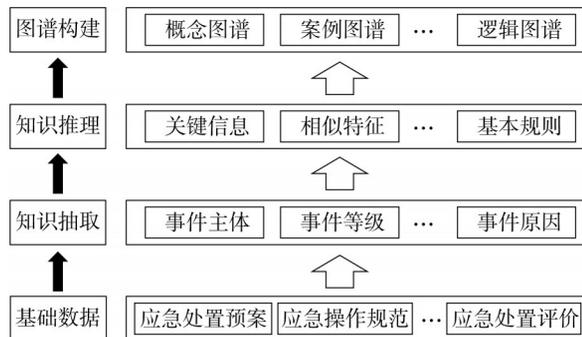


图1 地铁应急处置流程知识图谱构建流程信息

Fig. 1 Knowledge graph construction process information of metro emergency response process

2 基于BiLSTM-CRF的命名实体识别

地铁应急处置流程包含了许多步骤和环节,而BiLSTM-CRF可以同时考虑正向和逆向的上下文信息,可以充分考虑每个步骤与其前后环节之间的相互联系,从而提高命名实体识别的准确性。此外,BiLSTM-CRF可以通过多层的非线性变换去捕捉到地铁应急处置流程中的复杂语义信息,且以往研究表明BiLSTM-CRF在实现量化数据的序列标注方面有很好的效果^[9-10]。因此,选用BiLSTM-CRF进行命名实体识别可以实现地铁应急处置流程基本语义资料的知识抽取。

命名实体识别是基于构建好的文本数据集而展开,其中文本数据集构建需要通过数据收集、数据清洗、实体划分、实体标注等步骤完成,也是搭建知识图谱的首要步骤。首先,将从互联网搜集到的地铁应急处置流程文本案例作为实验的数据集,其中包括上海、兰州、南昌等地铁在站台、轨道、区间、列车等场所的应急处置流程。然后对收集到原始文本数据进行清洗,处理数据中的噪声和错误,如:无关信息、特殊字符、HTML标签等,同时标准化文本数据的格式和结构,以便后续处理。其次是实体划分,将地铁应急处置流程内容赋予其对应的类别。

如表1所示,在地铁应急处置流程中涉及实体有应急事件、事件地点、处置人员、处置组织、应急处置流程等,各实体均有其涉及的相关属性,各实体之间存在顺序、因果、并发、协同、互斥、隶属等关系。

表 1 地铁应急处置流程中实体、关系、属性信息

Tab. 1 Entity, relationship, and attribute information in the metro emergency response process

相关实体	涉及属性	存在关系
应急事件	事件原因、事件等级等	顺序关系、因果关系、并发关系、协同关系等
事件地点	地点名称	并发关系、协同关系等
处置人员	人员名称,所属组织等	顺序关系、并发关系、协同关系、隶属关系等
处置组织	组织名称	顺序关系、并发关系、协同关系、隶属关系等
应急处置流程	流程名称、流程内容、流程进度等	顺序关系、因果关系、并发关系、协同关系、互斥关系等

基于此,对文本数据采用BMEO标注方法进行实体标注,其中B表示实体的开头,M表示实体的中间部分,E表示实体的结尾,O表示非实体数据。根据地铁应急处置流程知识图谱构建需求定义相关标注符号如表2所示。以“车站安全员接到火灾报警信息后,立即报告车控室”这一事件为例,基本语义资料中利用BMEO标注结果示例如表3所示。

表 2 实体标注定义示例

Tab. 2 Example of entity annotation definition

实体名称	标注符号
地铁应急事件	OBJE(Object)
应急处置流程	PROCE(Process)
事件地点	PLACE(PLACE)
事件原因	CAUSE(CAUSE)
处置人员	NAME(NAME)
处置组织	ORG(Organization)

表 3 BMEO标注结果示例

Tab. 3 Example of BMEO labeling results

索引编号	标注	索引编号	标注
1	B-NAME	12	O
2	M-NAME	13	O
3	M-NAME	14	O
4	M-NAME	15	O
5	E-NAME	16	O
6	O	17	O
7	O	18	O
8	B-OBJE	19	O
9	E-OBJE	20	B-ORG
10	O	21	M-ORG
11	O	22	E-ORG

根据上述实体划分和实体标注的方法,将每一个地铁应急处置流程文本标注为原文序列、索引序列、命名实体序列、关系序列,组成一个序列集合 $\{s_0, s_1, s_2, s_3\}$ 。对上述序列进行编码得到具有相同纬度、长度为样本字符的长度、由特征向量组合而成的文本词序列向量。在此基础上,将文本词序列向量引入BiLSTM-CRF中,完成命名实体标注用于地铁应急处置流程图谱的构建,BiLSTM-CRF流程见图2,图中字母上箭头表示神经网络的传播方向。BiLSTM-CRF主要包括输入层、前向LSTM层和后向LSTM层(即BiLSTM层)、输出层、CRF层。

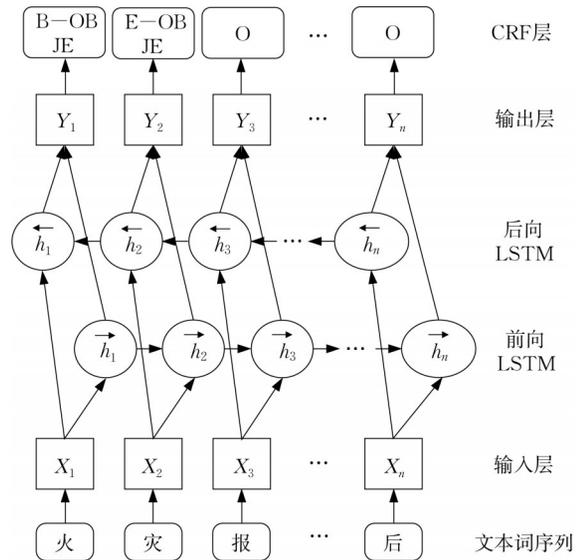


图 2 BiLSTM-CRF命名实体识别流程

Fig. 2 Flowchart of BiLSTM-CRF named entity recognition

(1) 输入层是由文本词序列向量 $\{x_1, x_2, x_3, \dots, x_n\}$ 经Word2vec工具转化为词向量序列 $\{X_1, X_2, X_3, \dots, X_n\}$ 而来,Word2vec词向量工具将预处理的词序列转换为低维且计算机可处理的词向量,并含有初始语义信息^[15]。

(2) BiLSTM层(即前向LSTM层和后向LSTM层)对输入的词向量进行语义信息建模,并进行前后向拼接后输出。在该层中,每一个 h_n 即为一个LSTM,LSTM神经元结构如图3所示。

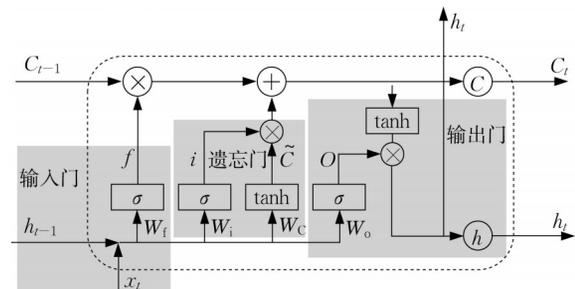


图 3 LSTM神经元结构

Fig. 3 Structure of LSTM neuron

图3实现过程如下,对于 t 时刻,输入的内容包括 t 时刻的输入向量 X_t (即输入层的词向量),上一时

刻的状态信息 C_{t-1} 和隐藏层输出信息 h_{t-1} , σ 为 sigmoid 函数, \tanh 为激活函数, 完整计算步骤为

$$f_t = \sigma(W_{i,x}x_t + W_{i,h}h_{t-1} + b_i) \quad (1)$$

$$i_t = \sigma(W_{i,x}x_t + W_{i,h}h_{t-1} + b_i) \quad (2)$$

$$\tilde{C}_t = \tanh(W_{c,x}x_t + W_{c,h}h_{t-1} + b_c) \quad (3)$$

$$C_t = i_t \tilde{C}_t + f_t C_{t-1} \quad (4)$$

$$o_t = \sigma(W_{o,x}x_t + W_{o,h}h_{t-1} + b_o) \quad (5)$$

$$h_t = o_t \tanh(C_t) \quad (6)$$

式中: f_t , i_t 和 o_t 分别为遗忘门限、输入门限和输出门限的输出; $W_{i,x}$, $W_{i,h}$ 和 $W_{o,x}$ 为 x_t 在 3 个门限的输出权重; $W_{f,h}$, $W_{i,h}$ 和 $W_{o,h}$ 为 h_{t-1} 在 3 个门限的输出权重; $W_{c,x}$ 和 $W_{c,h}$ 分别为实时序列特征 \tilde{C}_t 的权重; b_f , b_i , b_o 和 b_c 分别为计算 3 个门限和实时序列特征时的偏差量。

(3) 输出层就是将前向 LSTM 和后向 LSTM 输出信息拼接, 作为 BiLSTM 的输出, 即输入词向量的相对应的标注概率分布。

$$Y_n = [\vec{h}_n; \overleftarrow{h}_n] \quad (7)$$

(4) CRF 层通过转移概率矩阵用于修正 BiLSTM 输出, 最后得到文本序列的最优标记。假设对于文本序列 $\{x_1, x_2, \dots, x_n\}$ 的预测标签序列为 $\{y_1, y_2, \dots, y_n\}$, 它的最佳得分计算步骤为

$$f(x, y) = \sum_{i \in n} (A_{y_i, y_{i+1}} + K_{i, y_i}) \quad (8)$$

式中: $A_{y_i, y_{i+1}}$ 为标签 y_i 之后是标签 y_{i+1} 的概率, 即转移概率矩阵; K_{i, y_i} 为第 i 个字符映射到第 y_i 个标签上的概率, 即 BiLSTM 的输出概率矩阵。

最后, CRF 层会将得分最高的标注序列输出, 从而实现命名实体识别完成知识抽取, 从而实现从文本数据抽取实体的过程。

3 基于 TransD 的三元组关系搭建

TransD 模型通过计算实体间(头实体和尾实体)之间的相似性, 从而实现三元组的关系搭建, 即实体关系抽取过程。TransD 模型将知识图谱中的实体编码为关系空间中的低维嵌入向量, 从而实现实体在向量空间中的映射, 而每个实体关系对都被视为嵌入空间中的平移。

如图 4 所示, 在 TransD 模型中每个实体或者关系都由 2 个向量表示, 一个向量 (p, r, q) 用于描述头尾实体 (p, q) 两者关系 r 的实际意义, 另外一个为投

影向量 (p_s, r_s, q_s) , 用于构建映射矩阵^[16]。此外, 三元组中头实体和尾实体大部分情况下具有不同实际意义, 因此, 对头尾实体选取了不同的映射矩阵进行转换, 其中, 头实体映射矩阵 M_{rp} 和尾实体映射矩阵 M_{rq} 定义为

$$M_{rp} = r_s p_s^T + I \quad (9)$$

$$M_{rq} = r_s q_s^T + I \quad (10)$$

式中: I 为单位矩阵。上述公式表明映射矩阵是由实体和关系共同决定的, 因此投影向量可以表示为

$$p_{\perp} = M_{rp} p \quad (11)$$

$$q_{\perp} = M_{rq} q \quad (12)$$

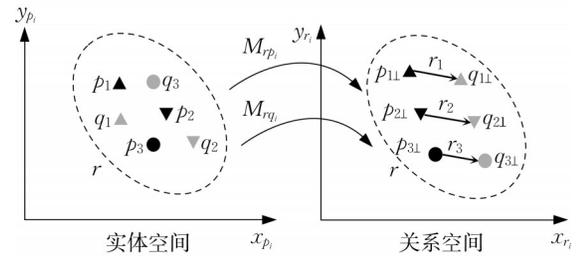


图 4 TransD 模型原理

Fig. 4 Schematic diagram of TransD model

TransD 模型得分函数定义为

$$f_r(p, q) = -\|p_{\perp} + r - q_{\perp}\|_2^2 \quad (13)$$

式中: $\|\cdot\|_2$ 是 L2 范数, 表示向量各元素的平方和然后求平方根, 用于描述空间上 2 个向量矩阵的直线距离。

为了防止模型过拟合, 对式(12)涉及的向量设置了相关约束, 为

$$\|H\|_2 \leq 1, H = \{p, r, q, p_{\perp}, q_{\perp}\} \quad (14)$$

对于 TransD 模型而言, 正确三元组期望得分越高越好, 错误三元组期望得分越低越好, 因此 TransD 模型的损失函数如式(15)所示:

$$L_{\text{loss}} = \sum_{(p, r, q) \in B} \sum_{(p', r, q') \in B'} [\gamma + f_r(p', q') - f_r(p, q)] \quad (15)$$

式中: B 为正三元组集合, 即正确三元组集合; B' 为负三元组集合, 即头实体或者尾实体是错误的三元组集合。因此, 负三元组 (p', r, q') 是基于正三元组实体对 (p, r, q) 挑选出来的, 而 p' 和 q' 是分别以 $tpp/(tpp+hpq)$ 和 $hpq/(tpp+hpq)$ 的概率替换而得, 其中, tpp 表示每个头实体对应的尾实体数量的平均值, hpq 表示每个尾实体对应头实体数量的平均值; γ 为超参数, 用于均衡正负三元组之间的间隔。

根据上述 TransD 模型, 即可实现在给定的

(p, r) 或者 (r, q) 情况下补充对应的 q 或 p ,从而完成知识推理的过程,实现了三元组的构建,所有三元组组合在一起即是知识图谱。

4 实验结果与分析

4.1 基于 BiLSTM-CRF 的命名实体识别

以地铁火灾应急为例,选取上述构建文本数据集中地铁火灾应急处置流程的文本数据进行命名实体识别。实验在 NVIDIA GeForce RTX3060 Laptop GPU 上进行,编程语言采用 Python 3.6,实验所用的深度学习框架为 PyTorch 1.5,同时使用了 Word2vec 0.3.1 作为词嵌入工具。实验选取其中 50% 数据,总计 1 184 份段落文本、182 576 字节数,作为命名实体识别的训练集,再选取 25% 数据,总计 592 份段落文本、92 123 字节数据作为验证集,剩余 25% 数据,总计 592 份段落文本、97 684 字节数据作为测试集。基于此,选取了精确率(Precision, P)、召回率(Recall, R)、F1 值(F1-score, F_1)来评估命名实体识别性能指标,上述 3 个指标被广泛应用于信息检索和统计学分类领域,常用于评价结果的质量,即表征命名实体识别的准确性。其中,精确率 P 和召回率 R 大于等于 90% 时,表明模型的全面性和准确性均具有较好的表现, F_1 值综合考虑精确率和召回率,当 F_1 值达到 80% 以上时,说明模型具备较好的性能。各计算式分别为

$$P = \frac{T_p}{T_p + F_p} \times 100\% \quad (16)$$

$$R = \frac{T_p}{T_p + F_n} \times 100\% \quad (17)$$

$$F_1 = \frac{2PR}{P+R} \times 100\% \quad (18)$$

式中: T_p 为正样本预测为真数量; F_p 为负样本预测为真数量; F_n 为正样本预测为假数量。

根据上述评价指标公式和数据划分,用训练集和验证集进行模型训练和验证,并将训练好的模型在测试集上进行预测,计算结果如表 4 所示。由表中数据可知,该模型在测试集上的精确率、召回率和 F_1 值均大于 90%,证明 BiLSTM-CRF 在该数据集上有着良好的表现,可以对地铁火灾应急处置流程文本数据实现较好的实体识别效果,可进行知识抽取后完成知识图谱搭建。

4.2 TransD 模型的三元组构建

基于实体识别的三元组数据集,使用 TransD 模

表 4 命名实体识别结果

Tab. 4 Results of named entity recognition

数据集	$P/\%$	$R/\%$	$F_1/\%$
训练集	89.45	88.59	88.35
验证集	90.72	90.58	90.19
测试集	91.74	90.23	91.47

型进行关系推理,并选取 TransE、TransH 推理模型进行推理结果对比分析。3 种模型均选用随机梯度下降法更新嵌入向量,模型学习率设置为 0.01,实体和关系的向量维度设定为 100,正负三元组之间的最小间隔设置为 1,迭代次数为 500^[20]。为了对比 3 种模型的推理结果,选用了平均排名(Mean Rank, M_R)、平均倒数排名(Mean Reciprocal Rank, M_{RR})和预测结果前 m 项准确率(Prediction results accuracy of top m , $P_{Hits@m}$)作为评价指标,各评价指标计算式为

$$M_R = \frac{1}{|D|} \sum_{i \in D} r_i \quad (19)$$

$$M_{RR} = \frac{1}{|D|} \sum_{i \in D} \frac{1}{r_i} \quad (20)$$

$$P_{Hits@m} = \frac{1}{|D|} \sum_{i \in D} \prod (r_i \leq m), m = \{1, 3, 10\} \quad (21)$$

式中: D 为所有三元组集合; r_i 为第 i 个三元组的推理排名; $\prod (r_i \leq m)$ 为第 i 个三元组是否满足 m 的排名要求,满足为 1,否则为 0。

根据指标计算公式(19)–(21),考虑 BiLSTM-CRF 对知识图谱构建的影响,分别对 TransD、TransE、TransH 这 3 个模型的推理结果进行展示,具体如表 5 所示。一般 M_R 值越低、 M_{RR} 值和 $P_{Hits@m}$ 值越高,模型推理结果越好。由表可知,无论有无 BiLSTM-CRF 用于知识图谱构建,TransD 模型的推理效果均是最优,各项评价指标均优于 TransE 和 TransH 模型。以 M_{RR} 和 $P_{Hits@10}$ 评价指标为例,采用了 BiLSTM-CRF 的 TransD 模型 M_{RR} 的推理结果比不采用 BiLSTM-CRF 的 TransD 模型 M_{RR} 推理结果提高了 29.09%,而在 $P_{Hits@10}$ 方面,采用了 BiLSTM-CRF 的 TransD 模型推理结果比不采用 BiLSTM-CRF 的 TransD 模型提升了 22.92%。因此,在 BiLSTM-CRF 的基础上运用 TransD 模型进行三元组搭建效果更好,更适用于地铁火灾的应急处置流程知识图谱构建。

4.3 基于 BiLSTM-CRF 的地铁火灾应急处置流程知识图谱

地铁火灾应急处置流程知识图谱,是以 BiLSTM-CRF 的命名实体识别后的实体数据集为

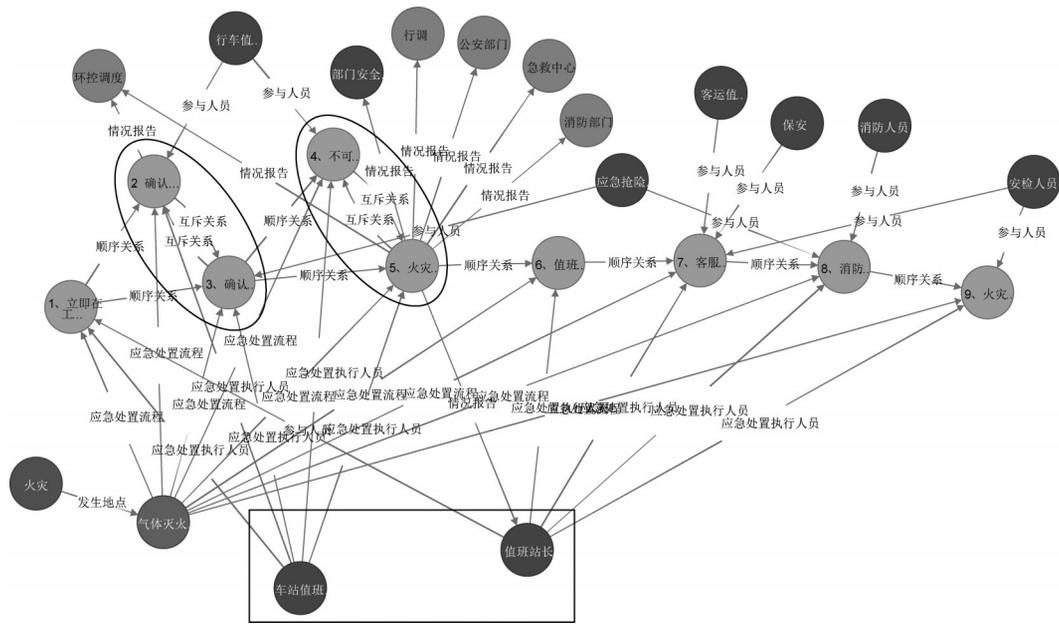


图 6 基于BiLSTM-CRF的地铁气体灭火防护区火灾单一应急处置流程知识图谱

Fig. 6 Knowledge graph of single emergency response process for fire in metro gas fire extinguishing protection zone based on BiLSTM-CRF

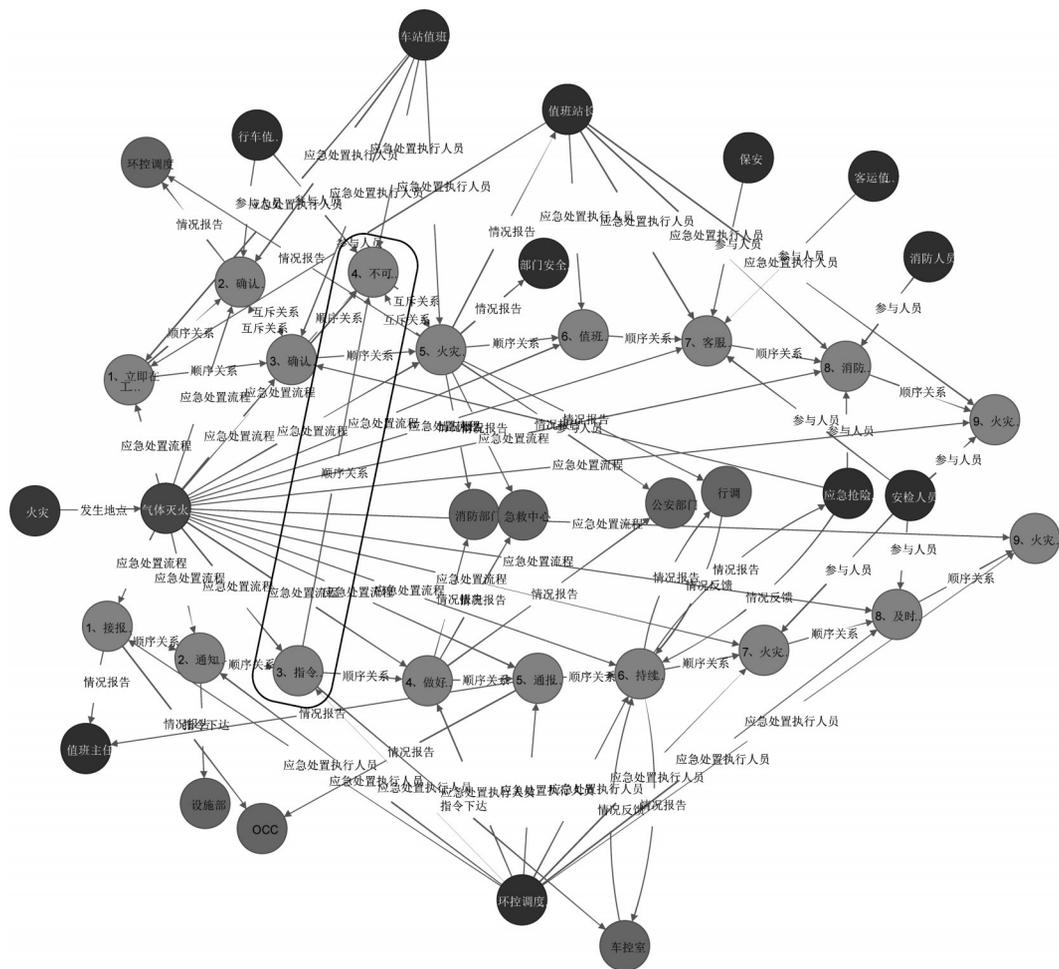


图 7 基于BiLSTM-CRF的地铁气体灭火防护区火灾应急处置流程知识图谱

Fig. 7 Knowledge graph of fire emergency response process in metro gas fire extinguishing protection zone based on BiLSTM-CRF

景下相互独立,彼此之间并无直接关联。根据图 7 中圆角矩形框可知,环控调度人员执行完相应的处置流程 3 之后,车站值班人员才可执行对应的处置流程 4,2 个不同应急处置流程之间存在先后顺序联系。综上,基于 BiLSTM-CRF 的地铁气体灭火防护区应急处置流程知识图谱不仅对同一流程之间的相互关系进行展示,还对同一场景下不同应急处置流程之间的关系也进行了展示,并实现了同一场景下不同应急处置流程的关系推理,确保了应急处置流程的连贯性和有效性。

5.3 不同案例样本量的对比分析

为了进一步验证基于 BiLSTM-CRF 构建的地铁应急处置流程知识图谱的准确性和可靠性,仍以地铁火灾应急事件为例,分别计算了不同案例样本量下的三元组搭建的准确率(Accuracy, A),即搭建正确的结果占总样本的分数,具体计算如式(22)所示,其结果如表 6 所示。

$$A = \frac{T_p + T_n}{T_p + T_n + F_p + F_n} \quad (22)$$

式中: T_n 为负样本预测为假数量。

表 6 三元组搭建准确率

Tab. 6 Accuracy results of triple construction

构建方式	案例样本量	三元组搭建准确率/%
有 BiLSTM-CRF	10	57.51
	20	69.25
	30	82.74
无 BiLSTM-CRF	10	45.83
	20	56.31

由表 6 数据可知,有 BiLSTM-CRF 的三元组搭建准确率优于无 BiLSTM-CRF 的三元组搭建准确率,且随着案例样本数量的增加,三元组搭建准确率也随之上升,当案例样本量为 30 时,三元组搭建准确率为 82.74%,有较好的准确性。因此,数据量越多越能保证构建三元组的准确性,从而提升了所示地铁应急处置流程知识图谱的可靠性,实现了对文本类地铁应急处置流程的优化。

6 结论

基于 BiLSTM-CRF 的地铁应急处置知识抽取与推理方法进行研究,解决目前文本类地铁火灾应急处置流程存在的结构化程度低、没有形成数字化存储、流程顺序关系不明确、流程执行人员模糊等问题,并利用 Neo4j 图数据库实现地铁应急处置流程知识图谱的可视化展示,为地铁应急管理提供辅助

决策支持。实验结果表明,基于 BiLSTM-CRF 的知识抽取模型的精确率、召回率和 F1 值均大于 90%。与此同时,基于 BiLSTM-CRF 推理模型的结果准确率比不采用 BiLSTM-CRF 推理模型提升了 22.92%,保证了知识图谱构建的准确性。此外,基于知识图谱的地铁应急处置流程仍可开展进一步的深入研究,未来可收集更多地铁火灾应急处置流程用于完善知识图谱的构建,并可研究多种地铁应急场景下的应急处置流程知识图谱的构建,从而实现面向地铁应急的智能调度辅助决策系统的构建。

作者贡献声明:

叶雨涛:数据收集、模型构建、论文撰写。

王鹏玲:方法与写作指导。

徐瑞华:写作指导。

肖晓芳:实验结果可视化。

葛健豪:实验结果可视化。

参考文献:

- [1] 张吉祥,张祥森,武长旭,等.知识图谱构建技术综述[J].计算机工程,2022,48(3):23.
ZHANG Jixiang, ZHANG Xiangsen, WU Chhangxu, *et al.* Survey of knowledge graph construction techniques [J]. Computer Engineering, 2022, 48(3): 23.
- [2] 杨晓梅,王志华,刘岳明,等.遥感智能信息处理的发展及技术前景[J].同济大学学报(自然科学版),2023,51(7):1025.
YANG Xiaomei, WANG Zhihua, LIU Yueming, *et al.* Development and technical prospect of remote sensing intelligent information processing [J]. Journal of Tongji University(Natural Science), 2023, 51(7): 1025.
- [3] GONG F, WANG M, WANG H, *et al.* SMR: Medical Knowledge graph embedding for safe medicine recommendation [J]. Big Data Research, 2021, 23:100174.
- [4] 邓君,王阮.口述历史档案资源知识图谱与多维知识发现研究[J].图书情报工作,2022,66(7):4.
DENG Jun, WANG Ruan. Research on knowledge graph and multidimensional knowledge discovery of oral history archives resources [J]. Library and information work, 2022, 66(7): 4.
- [5] ZHANG N, DENG S, CHEN H, *et al.* Structured knowledge base as prior knowledge to improve urban data analysis [J]. ISPRS International Journal of Geo-Information, 2018, 7(7): 264.
- [6] ZHOU G, CHEN F. Urban congestion areas prediction by combining knowledge graph and deep spatio-temporal convolutional neural network [C]// 2019 4th International Conference on Electromechanical Control Technology and Transportation (ICECTT). Guilin: [S.n.]: 2019: 105-108.
- [7] 朱广宇,张萌,裔扬.基于知识图谱的城市轨道交通突发事件演化结果预测[J].电子与信息学报,2023,45(3):9.

- ZHU Guangyu, ZHANG Meng, YI Yang. Prediction of evolution results of urban rail transit emergencies based on knowledge graph [J]. *Journal of Electronics & Information Technology*, 2023, 45(3): 9.
- [8] 李晓健,陈雍君,邱实,等.复杂地区铁路工程建设风险知识图谱的建立与分析方法[J/OL]. [2025-01-20]. <http://kns.cnki.net/kcms/detail/11.2104.u.20240619.1705.002.html>.
- LI Xiaojian, CHEN Yongjun, QIU Shi, *et al.* Establishment and analysis method of risk knowledge graph of railway engineering construction in complex areas [J/OL]. [2025-01-20]. <http://kns.cnki.net/kcms/detail/11.2104.u.20240619.1705.002.html>.
- [9] 林海香,胡娜娜,何乔,等.基于建筑信息模型数据驱动的设备运维多模态知识图谱构建[J]. *同济大学学报(自然科学版)*, 2024, 52(2): 166.
- LIN Haixiang, HU Nana, HE Qiao, *et al.* Construction of a multi-Modal knowledge graph for railway equipment operation and maintenance based on building information model data-Driven approach [J]. *Journal of Tongji University (Natural Science)*, 2024, 52(2): 166.
- [10] DONG X, GABRILOVICH E, HEITZ G, *et al.* Knowledge vault: a web-scale approach to probabilistic knowledge fusion [C]// the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Association for Computing Machinery. New York: Association for Computing Machinery, 2014: 601-610.
- [11] 李新琴,史天运,李平,等.基于文本的高速铁路信号设备故障知识抽取方法研究[J]. *铁道学报*, 2021, 43(3): 92.
- LI Xinqin, SHI Tianyun, LI Ping, *et al.* Research on knowledge extraction method for high-speed railway signal equipment fault based on text [J]. *Journal of the China Railway Society*, 2021, 43(3): 92.
- [12] KE Jia, WANG Weiji, CHEN Xiaojun, *et al.* Medical entity recognition and knowledge map relationship analysis of Chinese EMRs based on improved BiLSTM-CRF [J]. *Computers and Electrical Engineering*, 2023:108.
- [13] BORDES A, USUNIER N, GARCIADURAN A, *et al.* Translating embeddings for modeling multi-relational data [C]// *Neural Information Processing Systems*, New York: Curran Associates Inc., 2013: 2787 - 2795.
- [14] WANG Z, ZHANG J, FENG J, *et al.* Knowledge graph embedding by translating on hyperplanes [C]// *Twenty-Eighth AAAI Conference on Artificial Intelligence*. Quebec: AAAI Press., 2014: 1112-1119.
- [15] LIN Y, LIU Z, SUN M, *et al.* Learning entity and relation embedding for knowledge graph completion [C]// *Twenty-Ninth AAAI Conference on Artificial Intelligence*. Menlo Park: AAAI Press, 2015: 2181-2187.
- [16] JI G, HE S, XU L, *et al.* Knowledge graph embedding via dynamic mapping matrix [C]// *Meeting of the Association for Computational Linguistics & the International Joint Conference on Natural Language Processing, Lecture Notes in Computer Science*. Cham: Springer, 2015: 687-696.
- [17] MIKOLOV T, CHEN K, CORRADO G, *et al.* Efficient estimation of word representations in vector space [J]. *Computer Science*, 2013, 2(26): 1.
- [18] 俞山川,谢耀华,陈晨,等.基于知识图谱推理的风险车辆识别方法研究[J]. *重庆交通大学学报(自然科学版)*, 2022, 41(1): 22.
- YU Shanchuan, XIE Yaohua, CHEN Chen, *et al.* Risky vehicle identification method based on knowledge graph reasoning [J]. *Journal of Chongqing Jiaotong University (Natural Science)*, 2022, 41(1): 22.
- [19] 王丽.兰州轨道交通气体灭火保护区火灾应急处置方式研究[J]. *运输经理世界*, 2022(8): 7.
- WANG Li. Research on fire emergency disposal methods in Lanzhou rail transit gas fire protection area [J]. *World of Transportation Managers*, 2022(8): 7.
- [20] 王会勇,论兵,张晓明,等.基于联合知识表示学习的多模态实体对齐[J]. *控制与决策*, 2020, 35(12): 2855.
- WANG Huiyong, LUN Bing, ZHANG Xiaoming, *et al.* Multi-modal entity alignment based on joint knowledge representation learning [J]. *Control and Decision*, 2020, 35(12): 2855.