

一种从新闻报道中构建突发事件应急处置过程的方法*

王学贺¹, 李晓磊¹, 成洪豪¹, 赵华²

(1. 菏泽医学专科学校 计算机教研室, 山东 菏泽 274000; 2. 山东科技大学 计算科学与工程学院, 山东 青岛 266590)

摘要: 为了让民众更加清楚地了解突发事件的应急处置过程, 提高政府公信力, 提出并实现了一种从新闻报道中构建突发事件应急处置过程的方法. 将应急措施看作三元组, 构建训练语料, 采用条件随机场(CRF)模型抽取三元组的元素, 然后将应急措施按时间排序得到应急处置过程. 所设计的基于CRF的三元组抽取模型的准确率为99.6%、精确率为93.8%、召回率为76.2%、F1值为84.1%. 同时, 通过对比抽取的应急处置过程和人工构建的应急处置过程可知抽取效果完全达到了实用水平. 所提出的方法能自动准确地生成突发事件应急处置过程, 为突发事件应急处置科学决策提供技术支持.

关键词: 突发事件; 应急处置过程; 条件随机场

DOI: 10.13568/j.cnki.651094.651316.2024.03.16.0001

中图分类号: TP274 **文献标识码:** A **文章编号:** 2096-7675(2024)04-0444-08

引文格式: 王学贺, 李晓磊, 成洪豪, 赵华. 一种从新闻报道中构建突发事件应急处置过程的方法[J]. 新疆大学学报(自然科学版中英文), 2024, 41(4): 444-451.

英文引文格式: WANG Xuehe, LI Xiaolei, CHENG Honghao, ZHAO Hua. A method of constructing emergency response process from news report[J]. Journal of Xinjiang University(Natural Science Edition in Chinese and English), 2024, 41(4): 444-451.

A Method of Constructing Emergency Response Process from News Report

WANG Xuehe¹, LI Xiaolei¹, CHENG Honghao¹, ZHAO Hua²

(1. Division of Computer Science, Heze Medical College, Heze Shandong 274000, China; 2. College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao Shandong 266590, China)

Abstract: In order to let the public know more about the emergency response process and improve the credibility of the government, this paper proposes and realizes a method of constructing the emergency response process from news report. The emergency measures were regarded as triples, the training corpus was constructed, the conditional random field (CRF) model was used to extract the elements of triples, and then the emergency measures were sorted in time order to get the emergency response process. The accuracy of CRF model was 99.6%, the accuracy was 93.8%, the recall was 76.2%, and the F1 value was 84.1%. At the same time, by comparing the extraction result of this paper and the manual extraction result, we can see that the extraction method of this paper achieves the practical level. This method can accurately generate emergency response process, and provide technical support for scientific decision-making in emergency response.

Key words: emergency; emergency response process; conditional random field

0 引言

应急处置过程是指政府和相关机构在突发事件发生时作出的事发应对、事中处置和善后恢复的过程. 及时向民众公布应急处置过程有助于加强民众对政府部门应对突发事件应急处置的了解和监管, 有利于增强政府公信力. 但是面对频发的突发事件, 人工手动生成应急处置过程费时费力.

2016年11月, 国务院办公厅印发《〈关于全面推进政务公开工作的意见〉实施细则》要求^[1], “对涉及群众切身利益、影响市场预期和突发公共事件等重点事项, 要及时发布信息. 对涉及特别重大、重大突发事件的政务

* 收稿日期: 2024-03-16

基金项目: 山东省自然科学基金面上项目“基于定制感知与流程比对的突发事件网络舆情异常定量分析研究”(ZR2021MG038).

作者简介: 王学贺(1976—), 男, 硕士, 教授, 主要从事情感计算、文本挖掘的研究, E-mail: wangxuehe@hzmc.edu.cn.

舆情,要快速反应,最迟要在5小时内发布权威信息,在24小时内举行新闻发布会,并根据工作进展情况,持续发布权威信息”。基于此,突发事件一旦发生,事件相关的信息将会在新闻媒体、微博、微信等多种平台中被报道、评论、转发等,其中包含丰富的有关应急处置的新闻报道,这些新闻报道数据为自动抽取应急处置措施提供了可能。然而,从多个单篇新闻报道中抽取到的应急处置措施对于民众来说是一些碎片化的数据,这些碎片化的数据仅仅反映了事件的某个局部^[2],民众从碎片化的数据中只能得到“盲人摸象”的效果,很难清楚掌握整个应急处置过程。本文提出将碎片化的数据以时间为序排列构建应急处置过程,可以帮助人们了解事情完整的发展过程。

为此,本文提出了一种从新闻报道中自动生成突发事件应急处置过程的方法。首先收集关于突发事件的新闻报道,然后从新闻报道中抽取应急处置措施,将应急处置措施按照时间排序形成应急处置过程,最后进行可视化。本文将应急处置措施定义为一个三元组 $EM = \langle t, g, m \rangle$,其中 t 表示应急措施发生的时间, g 表示应急措施的实施者集合, m 表示应急行为集合;而将应急处置过程定义为应急措施的序列,即 $EMP = \langle EM_1, EM_2, \dots, EM_n \rangle$,其中 $EM_i (1 \leq i \leq n)$ 表示一个应急措施,而且 EM_i 的时间小于等于 EM_{i+1} 的时间。

本文构建了一个用于从新闻报道中自动生成突发事件应急处置过程的语料库,并给出了一种基于词性标注的语料库标注方法,该方法可以充分利用词性标注自动生成结果,大大节约了人力,提高了标注效率。

1 相关工作

突发事件发生后,需要各级政府和相关部门、人员及时作出响应,积极应对,最大限度地减少灾害损失。高效的应急响应是减少灾害损失、提高灾害救助效率的重要前提。为提高应急响应效率,应急领域专家积极构建应急响应流程,通过形式化方式描述应急响应执行过程,为应急决策提供技术支撑。现有的应急处置过程生成方法大多是从应急预案中抽取应急处置流程^[3],主要包括以下三种方法。

基于规则的方法。该方法首先对文本进行分词、词性标注、依赖关系和语义角色标注等操作,然后根据规则实现抽取。Ghose等^[4]提出了快速业务流程发现框架和原型工具,可以查询异构信息资源,如企业文档、Web内容等,并快速构建原型模型。Friedrich等^[5]首先对文本进行分词、指代消解,然后实现了一种自动生成流程模型的方法。Schumacher等^[6]扩展了工作流提取框架与控制流提取框架。Yordanova等^[7]基于词性标注和依赖关系识别动作和因果关系,继而实现人类行为模型抽取。Ferreira等^[8]定义了32个抽取规则,实现了一种识别自然语言文本中业务流程元素的方法。Van等^[9]首先从文本约束描述中识别活动及其相互关系,然后提出了一种提取声明性过程模型的自动化方法。Halioui等^[10]利用面向过程的基于实例推理从科技文本中自动提取工作流,并应用于系统发育分析领域。

基于机器学习的方法。流程模型元素在文本中被描述为连续的词序列,因此将流程元素抽取问题看作序列标注问题来解决。Li等^[11]提出了一种基于策略的过程挖掘方法,从策略文档中自动提取过程信息。Scaria等^[12]从文本中自动抽取进程内的事件和事件之间的关系,用图表示过程,图的边描述了时间的或者因果的事件关系,并描述了图的结构性质。Chang等^[13]从菜谱中抽取做菜流程,并实现了一个原型界面,支持可视化查询。倪维健等^[14]针对流程案例描述文本,应用半监督机器学习技术,对过程文本案例抽取任务进行了形式化描述,然后抽象出实体识别、关系识别和活动顺序关系识别3个任务。郭文艳^[15]基于规则和统计学习方法,实现了基于混合方法的应急响应流程抽取方法。

基于深度学习的方法。随着深度学习方法不断地被探索和应用,研究者也对基于深度学习方法的流程抽取进行了研究。其中常用的模型有长短期记忆网络(Long Short-Term Memory, LSTM)和双向长短期记忆网络(Bidirectional Long Short-Term Memory, Bi-LSTM)^[16]。Lu等^[17]提出基于LSTM预测流程中缺失的活动标签。Evermann等^[18]基于递归神经网络模型(Recurrent Neural Network, RNN)构建了荷兰一家金融机构的贷款申请流程。Evermann等^[19]基于LSTM模型引入双层RNN预测下一事件,抽取BPI2012和BPI2013数据中的子流程。Camargo等^[20]采用LSTM网络预测下一事件类型、时间和相关角色方法,完成对事件日志中的业务流程抽取。黄华等^[21]采用Bi-LSTM和注意力机制模型捕获序列信息,从7个真实事件日志中得到差旅费报销流程。袁永旺等^[22]提出基于双层BERT神经网络和融合流程的多视角行为分析方法,预测事件日志中的下一事件得到过程信息。

由上述方法可知,目前应急处置过程生成方法的研究主要集中于如何在应急预案中抽取预设的应急处置流程,但是对于如何抽取针对某一个突发事件的事后应急处置过程关注得较少。

2 面向应急处置过程生成的语料库构建

目前没有可以公开获取的从新闻报道构建突发事件应急处置过程的语料库。本文根据研究目的进行了语料标记。所使用的原始数据包含两部分,一是上海大学组织整理和标注的中文突发事件语料库(Chinese Emergency Corpus, CEC)^[23];二是从互联网上爬取的关于“山东栖霞矿难”的新闻报道。CEC语料库包含了互联网上关于自然灾害、事故灾难等新闻报道,包括自然灾害、事故灾难、公共卫生事件和社会安全事件,共计332篇,其中地震事件62篇、火灾事件75篇、交通事故85篇、恐怖袭击事件49篇、食物中毒事件61篇。“山东栖霞矿难”事件相关的新闻报道128篇,全部来自中国新闻网。

为了节约人力,本文首先基于分词和词性标注识别初始的时间、机构名和应急行为,然后采用一些规则的方法进行后处理,最后进行人工校对。具体的处理方法介绍如下。

2.1 应急时间的标注

利用分词工具可以识别出大部分的时间表达式,本文选用了中国科学院NLPIR分词工具。但是,NLPIR会将一些时间标记为数词,比如“2008年04月05日09:12”的标注结果为“2008年\t 04月\t 05日\t 09:12\m”,其中09:12被标记为数词。为此,本文对分词工具结果进行人工核对以保证所构建语料库中时间表达式的准确性。同时,分词结果会将一个时间表达式识别为多个时间元素,比如“2008年04月05日”被分解成2008年、04月、05日这三种基本的时间元素。本文选择将连续出现的时间元素合并为一个时间表达式,表1列出了本文采用的15种基本的时间元素。

表 1 基本的时间元素

编号	类别	举例	编号	类别	举例
1	小时时间	1时,12时	9	月时间	十二月份,12月份
2	分钟时间	25分,十二分	10	日时间	十五日,15日
3	秒时间	三十秒,12秒	11	时间点	早上,中午
4	阿拉伯表达式	2012/5/2	12	现在时间	今天,现在
5	世纪时间	21世纪	13	将来时间	明年,后天
6	世纪点	20年代初	14	过去时间	前天,昨夜
7	年时间	2006年	15	星期时间	本周五
8	季度时间	第三季度			

2.2 应急措施实施者的标注

NLPIR可以有效识别部分组织机构名,但是以下几个方面还不能满足本文的研究目的。

1) 错误识别。例如“应急\vn 救援部\n”中,“应急”被错误识别成了动名词,同时“救援部”没有被准确识别为组织机构名,而本文想要的结果是“应急救援部\nt”。

2) 识别粒度较小。例如“山东省指挥救援部”被识别为“山东省\ns 指挥救援部\nt”,而不是本文所需要的“山东省指挥救援部\nt”。

3) 无法加入一些特殊的实施者触发词,如“市长”“省委”这一类名词,而本研究中,将这一类词也作为应急措施的实施者。

为了解决上述问题,本文在基本分词结果的基础上,进行以下两个后处理操作。1) 模式匹配,本文关注各种名词类词语,如名词\n、动名词\vn、地点名词\ns等。将以“局”“部”“所”“队”“委”“厅”结尾作为模式匹配格式对上述名词进行模式匹配,并修改相应的词性。例如将“警察\n局\n”合并成“警察局\nt”,将“指挥部\n”改变为“指挥部\nt”。2) 机构合并,当两个组织机构元素连续出现时,将前者的词性标注去除,并与后者合并,例如“山东省\ns 安全指挥部\nt”被合并成“山东省安全指挥部\nt”。

2.3 应急行为的标注

为了有效标注应急行为,首先定义了应急行为触发词,即表示应急行为发生的核心词,多为动词或名词。首先人工收集种子触发词,如“召开”“召集”等;然后抽取语料中的名词、动词,记为Vword,最后基于Word2Vec计

算Vword中每个词与种子触发词的相似度,当相似度大于一定的阈值时,就将该词标记为触发词.最终,得到了表2所示的应急措施触发词.

表 2 应急措施触发词

参与	协助	排查	教育	主持	关闭	指导	询问	侦察	指出	宣传	做好	调拨	奔赴
寻找	组织	指挥	召集	处置	运送	启动	支持	设置	开展	接警	赶赴	保护	提出
救助	调度	披露	支援	敦促	明确	完成	加强	赶往	通知	召开	带领	抵达	协调
介入	联合	发出	调集	调查	发布	报告	打通	计划	成立	采取	发现	要求	集结

3 基于CRF的应急处置过程生成方法

3.1 基于CRF的元素抽取

CRF是Lafferty等^[24]于2001年提出的一种序列建模框架.这种模型的建立是为了学习映射函数 $X_i \rightarrow Y_i$,以此获得输出的标签最大化,即通过输入的特征向量 $X = X_1, X_2, \dots, X_n$ 预测对应的标签序列 $Y = Y_1, Y_2, \dots, Y_n$.假定CRF的输入和输出都属于线性链,且其中的 $P(X|Y)$ 符合马尔科夫性质,则 $P(X|Y)$ 可以表示为:

$$P(X|Y) = \frac{1}{Z(X)} \exp(\sum_{i,k} \tau_k t_k(Y_{i-1}, Y_i, X, i) + \sum_{i,l} \mu_l s_l(Y_i, X, i)) \tag{1}$$

$$Z(X) = \sum_Y \exp(\sum_{i,k} \tau_k t_k(Y_{i-1}, Y_i, X, i) + \sum_{i,l} \mu_l s_l(Y_i, X, i)) \tag{2}$$

其中: $t_k(Y_{i-1}, Y_i, X, i)$ 和 $s_l(Y_i, X, i)$ 分别表示在给定输入序列 X 的前提下,序列 Y 在 $i-1$ 处的值转化为 i 的概率以及其对应值在位置 i 处的概率. $t_k(Y_{i-1}, Y_i, X, i)$ 为转移特征函数; $s_l(Y_i, X, i)$ 为状态特征函数.而 τ_k 和 μ_l 则表示响应函数的权重值.

本文采用“BME0”标注规范对数据进行标注,B、M、E分别表示命名实体的开始字符、中间字符和结尾字符,而O表示非命名实体.本文关注的命名实体分为三类:应急时间、应急措施实施者和应急行为.

为了验证不同特征在识别过程中的性能,本文设定了三种不同的特征模板,如表3所示.值得说明的是,在识别了触发词以后,本文将应急行为触发词开始到触发词所在分句结束的部分识别为应急行为.

表 3 CRF特征模板

模板1	模板2	模板3
# Unigram	# Unigram	# Unigram
U00:%x[-2,0]	U00:%x[-2,0]	U00:%x[-3,0]
U01:%x[-1,0]	U01:%x[-1,0]	U01:%x[-2,0]
U02:%x[0,0]	U02:%x[0,0]	U02:%x[-1,0]
U03:%x[1,0]	U03:%x[1,0]	U03:%x[0,0]
U04:%x[2,0]	U04:%x[2,0]	U04:%x[1,0]
U05:%x[-2,0]/%x[-1,0]/%x[0,0]	U05:%x[-2,0]/%x[-1,0]/%x[0,0]	U05:%x[2,0]
U06:%x[-1,0]/%x[0,0]/%x[1,0]	U06:%x[-1,0]/%x[0,0]/%x[1,0]	U06:%x[3,0]
U07:%x[0,0]/%x[1,0]/%x[2,0]	U07:%x[0,0]/%x[1,0]/%x[2,0]	U07:%x[-3,0]/%x[-2,0]/%x[-1,0]/%x[0,0]
	U08:%x[-1,0]/%x[0,0]	U08:%x[-2,0]/%x[-1,0]/%x[0,0]/%x[1,0]
# Bigram	U09:%x[0,0]/%x[1,0]	U09:%x[-1,0]/%x[0,0]/%x[1,0]/%x[2,0]
B		U10:%x[0,0]/%x[1,0]/%x[2,0]/%x[3,0]
	# Bigram	U11:%x[-2,0]/%x[-1,0]/%x[0,0]
	B	U12:%x[-1,0]/%x[0,0]/%x[1,0]
		U13:%x[0,0]/%x[1,0]/%x[2,0]
		U14:%x[-1,0]/%x[0,0]
		U15:%x[0,0]/%x[1,0]
		# Bigram
		B

3.2 应急处置过程生成算法

经过CRF识别应急时间、应急措施实施者和应急行为后,采用算法1实现应急处置过程的自动生成.值得说明的是,根据突发事件新闻报道内容的分析,在新闻报道关于应急过程的描述中,应急措施实施者和应急行为经常是紧密关联且同时出现在一个句子中,因为这两者是应对紧急状况或危机时不可或缺的两个组成部分.为此算法1中只有当一个句子同时包含应急措施实施者和应急行为时才对其进行抽取应急处置过程.

算法 1 应急处置过程生成算法

输入: 关于突发事件的新闻报道集合 $ES = \{s_1, s_2, \dots, s_n\}$, s_i 表示一篇新闻报道

输出: 突发事件的应急处置过程

- 1) For $i = 1$ to n
- 2) 将 s_i 进行分句, 得到分句集合 $s_i = \{c_{i1}, c_{i2}, \dots, c_{im}\}$;
- 3) 识别 s_i 中时间 t 、应急措施实施者 g 、应急行为 m ;
- 4) For $j = 1$ to m
- 5) If c_{ij} 中含有应急措施实施者和应急行为
- 6) 假设 c_{ij} 中的应急措施实施者包含 p 个, 记为 $g_{ij} = \{g_{ij1}, g_{ij2}, \dots, g_{ijp}\}$
假设 c_{ij} 中的应急行为包含 q 个, 记为 $m_{ij} = \{m_{ij1}, m_{ij2}, \dots, m_{ijq}\}$
- 7) If c_{ij} 中含有 t_{ij}
- 8) 形成三元组 $\langle t_{ij}, g_{ij}, m_{ij} \rangle$, 将其放入 handle 列表;
- 9) Else
- 10) For $k = j - 1$ to 1
- 11) If c_{ik} 中含有 t_{ik}
- 12) 形成三元组 $\langle t_{ik}, g_{ij}, m_{ij} \rangle$, 将其放入 handle 列表;
- 13) Break
- 14) End If
- 15) End for
- 16) End If
- 17) End for
- 18) End for
- 19) End for
- 20) 对 handle 列表以时间顺序进行排序;
- 21) 输出应急处置过程.

4 实验与案例分析

4.1 实验与结果分析

本文使用上述构建的语料库验证基于新闻报道的应急处置过程构建方法的有效性,分别从各类突发事件新闻报道中随机抽取80%组成训练语料,其余20%作为测试语料.同时采用准确率(Accuracy)、精确率(Precision)、召回率(Recall)、F1作为评测标准.本文方法和其它方法的对比实验结果(F1值)如表4所示.

由表4可知,本文提出的方法取得了较为理想的效果.具体来说,文献[25]、文献[26]、文献[27]和本文方法类似,均采用基本的CRF模型,文献[25]采用制定地名修正规则和人名修正规则与CRF相结合的方法识别中文语句中的地理命名实体.文献[26]基于CRF加入词形词性特征、汉语姓氏特征、人名词典、兼类词特征和双词根人名特征对蒙古文人名进行自动识别研究.文献[27]提出中文症状是由16种构成元素组成的序列,基于CRF模型加入词性特征和文字特征设计了6类CRF特征模板,实现对序列的自动标注.由结果可知,本文所提三种特征模板的性能均较优于文献[25-27]中提出方法的性能,验证了本文所设计特征模板的有效性.

文献[28-29]均采用了预训练模型(RoBERTa或ALBERT)和神经网络(Bi-LSTM或BGRU)提取文本特征,然后输入CRF进行实体抽取,其性能优于本文的模板1和模板3,但是略低于本文设计的模板2.这说明,对于句

式较为规范的文本,通过较好地设计特征模板是可以取得同深度学习模型相媲美的实验结果.同时,由表4可知,模板中特征函数的增加在一定程度上会提高CRF模型的预测性能,但并不是特征函数越多精确度就越高,如表4中模板2和模板3的比较.模板2的实验效果最好,基于模板2的CRF模型的详细评测结果如表5所示.

表 4 实验结果

Table with 5 columns: 方法, 应急时间, 应急措施实施者, 应急行为触发词, 整体. Rows include literature [25-29] and the proposed method (templates 1, 2, 3).

表 5 基于模板2的CRF模型的识别性能

Table with 5 columns: 实体名, 准确率, 精确率, 召回率, F1. Rows include 应急时间, 应急措施实施者, 应急行为触发词, 整体.

4.2 案例分析

我们以一篇测试语料(图1)为例对人工信息抽取和基于本文方法自动抽取进行比对,得到图2、图3.

1月11日晚,国家卫健委获悉山东省烟台市栖霞市五彩龙公司笏山金矿发生爆炸事故,造成22人被困井下.国家卫生健康委高度重视,于1月12日从山东省立医院派出2名国家级医疗专家赶赴当地,指导做好事故紧急医学救援准备.1月19日,国家卫生健康委,带领5名国家级专家赶赴当地,强化救治力量.国家级专家组由北京协和医院、北京天坛医院、北京朝阳医院、北京安定医院的重症医学、营养、神经外科、职业病与中毒医学、心理干预等专业5名专家组成.1月10日,当地卫生健康部门立即启动卫生应急响应,组织开展现场抢救和伤员转运救治准备工作.山东省卫生健康委相关负责同志带领省级医疗专家组,迅速抵达现场,协调指导伤员救治工作.

图 1 测试语料

1月10日 当地卫生健康部门 启动卫生应急响应
1月10日 山东省卫生健康委 带领省级医疗专家组, 迅速抵达现场, 协调指导伤员救治工作
1月12日 国家卫生健康委 派出2名国家级医疗专家赶赴当地, 指导做好事故紧急医学救援准备
1月19日 国家卫生健康委 带领5名国家级专家赶赴当地, 强化救治力量.

图 2 人工信息抽取结果

1月10日 卫生健康部门 启动卫生应急响应, 组织开展现场抢救和伤员转运救治准备工作
1月10日 山东省卫生健康委 负责同志带领省级医疗专家组, 迅速抵达现场, 协调指导伤员救治工作
1月12日 国家卫生健康委 重视, 于1月12日从山东省立医院派出2名国家级医疗专家赶赴当地, 指导做好事故紧急医学救援准备
1月19日 国家卫生健康委 带领5名国家级专家赶赴当地, 强化救治力量

图 3 本文方法自动抽取结果

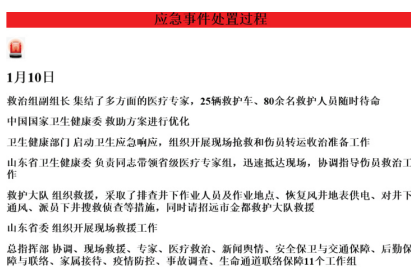


图 4 特定时间的应急处置过程

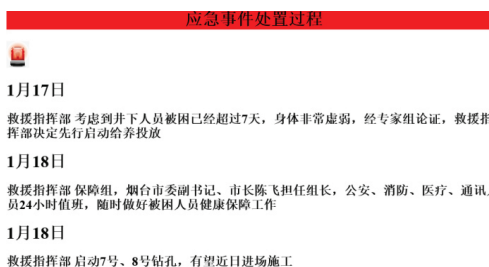


图 5 特定组织机构的应急处置过程

由人工信息抽取和机器信息抽取的内容对比可以看出,机器信息抽取虽然不如人工信息抽取精确,但对需要提取的内容做到了有效提取.同时,本文还实现了抽取特定时间的应急处置过程以及特定组织机构的应急处置过程,如图4显示了1月10日的应急处置过程,而图5显示了“救援指挥部”的应急处置过程.

5 结语

突发事件发生后,从新闻报道中自动生成应急处置过程对于民众了解事件发展及救援过程,提高政府公信力具有重要的作用.本文将应急处置过程生成看作信息抽取技术,在抽取了应急时间、应急措施实施者和应急行为的基础上,探索实现了一种从新闻报道中自动生成应急处置过程的方法.实验结果证明,虽然本文生成的应急过程没有人工生成的过程内容丰富,但已经达到了实用水平,而且节约了大量的人力物力.

下一步研究中将进一步扩大语料库规模,基于更大规模的语料库验证方法的有效性;同时融入更多的语义分析技术,进一步提升应急处置过程生成方法的性能.

参考文献:

- [1] 中华人民共和国国务院办公厅.《关于全面推进政务公开工作的意见》实施细则[EB/OL]. (2016-11-10)[2023-12-20]. https://www.gov.cn/zhengce/content/2016-11/15/content_5132852.htm.
General Office of the State Council of the People's Republic of China. Opinions on comprehensively promote the work of transparency in government affairs[EB/OL]. (2016-11-10)[2023-12-20]. https://www.gov.cn/zhengce/content/2016-11/15/content_5132852.htm. (in Chinese)
- [2] 李风环, 郑德权, 赵铁军. 动态增量式子主题事件演化分析[J]. 计算机研究与发展, 2015, 52(11): 2441-2450.
LI F H, ZHENG D Q, ZHAO T J. Dynamic incremental analysis of sub-topic evolution[J]. Journal of Computer Research and Development, 2015, 52(11): 2441-2450. (in Chinese)
- [3] 张峰, 韩燕波, 陈欣, 等. 基于数字预案的应急处置流程构造方法[J]. 计算机集成制造系统, 2013, 19(8): 1802-1809.
ZHANG F, HAN Y B, CHEN X, et al. Emergency response process construction based on digital plans[J]. Computer Integrated Manufacturing Systems, 2013, 19(8): 1802-1809. (in Chinese)
- [4] GHOSE A, KOLIADIS G, CHUENG A. Rapid business process discovery (R-BPD)[M]//Conceptual Modeling-ER 2007. Berlin, Heidelberg: Springer, 2007: 391-406.
- [5] FRIEDRICH F, MENDLING J, PUHLMANN F. Process model generation from natural language text[C]//International Conference on Advanced Information Systems Engineering. Berlin, Heidelberg: Springer, 2011: 482-496.
- [6] SCHUMACHER P, MINOR M. Extracting control-flow from text[C]//Proceedings of the 2014 IEEE 15th International Conference on Information Reuse and Integration (IEEE IRI 2014). Redwood City, CA, USA. IEEE, 2014: 203-210.
- [7] YORDANOVA K, KIRSTE T. Learning models of human behaviour from textual instructions[C]//Proceedings of the 8th International Conference on Agents and Artificial Intelligence. New York: ACM, 2016: 415-422.
- [8] FERREIRA R C B, SILVA T S, AVILA D T, et al. Recognition of business process elements in natural language texts[C]//International Conference on Enterprise Information Systems. Cham: Springer, 2018: 591-610.
- [9] VAN D H, DICICCIO C, LEOPOLD H, et al. Extracting declarative process models from natural language[C]//International Conference on Advanced Information Systems Engineering. Cham: Springer, 2019: 365-382.
- [10] HALIOUI A, VALTCHEV P, DIALLO A B. Bioinformatic workflow extraction from scientific texts based on word sense disambiguation[J]. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2018, 15(6): 1979-1990.
- [11] LI J X, WANG H J, ZHANG Z, et al. A policy-based process mining framework: Mining business policy texts for discovering process models[J]. Information Systems and E-Business Management, 2010, 8(2): 169-188.
- [12] SCARIA A T, BERANT J, WANG M Q, et al. Learning biological processes with global constraints[C]//Proceeding of the 2013 Conference on Empirical Methods in Natural Language Processing. Seattle, Washington, USA. ACL, 2013: 1710-1720.
- [13] CHANG M, HARE V M, KIM J, et al. Recipe scape: Mining and analyzing diverse processes in cooking recipes[C]//Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems. Denver, CO, USA. ACM, 2017: 1524-1531.
- [14] 倪健, 韦振胜, 曾庆田, 等. 面向自然过程文本的案例信息抽取[J]. 计算机集成制造系统, 2018, 24(7): 1680-1689.
NI W J, WEI Z S, ZENG Q T, et al. Case information extraction from natural procedure text[J]. Computer Integrated Manufacturing Systems, 2018, 24(7): 1680-1689. (in Chinese)

- [15] 郭文艳. 基于流程模型抽取的应急预案响应文本质量评估及修正方法研究[D]. 青岛: 山东科技大学, 2020.
GUO W Y. Research on quality evaluation and revision for emergency plan response texts based on process model extraction[D]. Qingdao: Shandong University of Science and Technology, 2020. (in Chinese)
- [16] LAMPLE G, BALLESTEROS M, SUBRAMANIAN S, et al. Neural architectures for named entity recognition[C]//Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg, PA, USA. ACL, 2016: 260-270.
- [17] LU Y, CHEN Q F, POON S K. A deep learning approach for repairing missing activity labels in event logs for process mining[J]. Information, 2022, 13(5): 234.
- [18] EVERMANN J, REHSE J R, FETTKE P. Predicting process behaviour using deep learning[J]. Decision Support Systems, 2017, 100: 129-140.
- [19] EVERMANN J, REHSE J R, FETTKE P. A deep learning approach for predicting process behaviour at runtime[C]//International Conference on Business Process Management. Cham: Springer, 2017: 327-338.
- [20] CAMARGO M, DUMAS M, GONZÁLEZ R. Learning accurate business process simulation models from event logs via automated process discovery and deep learning[C]//International Conference on Advanced Information Systems Engineering. Cham: Springer, 2022: 55-71.
- [21] 黄华, 杨子仪, 李小龙, 等. 基于概念漂移的预测性业务流程监控方法[J/OL]. 计算机应用, (2023-12-22)[2024-02-20]. <https://link.cnki.net/urlid/51.1307.tp.20231221.0952.006>.
HUANG H, YANG Z Y, LI X L, et al. Predictive business process monitoring approach based on concept drift[J/OL]. Journal of Computer Applications, (2023-12-22)[2024-02-20]. <https://link.cnki.net/urlid/51.1307.tp.20231221.0952.006>. (in Chinese)
- [22] 袁永旺, 方贤文, 卢可. 基于深度学习的融合流程多视角行为分析: 预测业务流程监控[J]. 计算机应用研究, 2024, 41(6): 1790-1796.
YUAN Y W, FANG X W, LU K. Multi-perspective behavior analysis of fusion processes based on deep learning: Predictive business process monitoring[J]. Application Research of Computers, 2024, 41(6): 1790-1796. (in Chinese)
- [23] 张亚军, 刘宗田, 李强, 等. 面向事件的中文指代语料库的构建[J]. 上海大学学报(自然科学版), 2018, 24(6): 900-911.
ZHANG Y J, LIU Z T, LI Q, et al. Construction of event-oriented Chinese coreference corpus[J]. Journal of Shanghai University (Natural Science Edition), 2018, 24(6): 900-911. (in Chinese)
- [24] LAFFERTY J D, MCCALLUM A, PEREIRA F C N. Conditional random fields: Probabilistic models for segmenting and labeling sequence data[C]//Proceedings of the Eighteenth International Conference on Machine Learning. New York: ACM, 2001: 282-289.
- [25] 何炎祥, 罗楚威, 胡彬尧. 基于CRF和规则相结合的地理命名实体识别方法[J]. 计算机应用与软件, 2015, 32(1): 179-185+202.
HE Y X, LUO C W, HU B Y. Geographic entity recognition method based on CRF model and rules combination[J]. Computer Applications and Software, 2015, 32(1): 179-185+202. (in Chinese)
- [26] 吴金星, 那顺乌日图, 杨振新. 基于CRF的蒙古文人名自动识别研究[J]. 计算机应用研究, 2016, 33(7): 2014-2017.
WU J X, NASHUNWURITU, YANG Z X. Recognition method of Mongolian person names based on conditional random fields[J]. Application Research of Computers, 2016, 33(7): 2014-2017. (in Chinese)
- [27] 曾露, 高大启, 阮彤, 等. 基于CRF的症状构成分析与标注[J]. 华东理工大学学报(自然科学版), 2018, 44(2): 277-282.
ZENG L, GAO D Q, RUAN T, et al. Analysis and annotation of symptom composition based on CRF[J]. Journal of East China University of Science and Technology(Natural Science Edition), 2018, 44(2): 277-282. (in Chinese)
- [28] 张芳丛, 秦秋莉, 姜勇, 等. 基于RoBERTa-WWM-BiLSTM-CRF的中文电子病历命名实体识别研究[J]. 数据分析与知识发现, 2022, 6(S1): 251-262.
ZHANG F C, QIN Q L, JIANG Y, et al. Research on named entity recognition of Chinese electronic medical records based on RoBERTa-WWM-BiLSTM-CRF[J]. Data Analysis and Knowledge Discovery, 2022, 6(S1): 251-262. (in Chinese)
- [29] 李军怀, 陈苗苗, 王怀军, 等. 基于ALBERT-BGRU-CRF的中文命名实体识别方法[J]. 计算机工程, 2022, 48(6): 89-94+106.
LI J H, CHEN M M, WANG H J, et al. Chinese named entity recognition method based on ALBERT-BGRU-CRF[J]. Computer Engineering, 2022, 48(6): 89-94+106. (in Chinese)