

中文引用格式:柳晓昱,庄育锋,赵兴昊,等.基于知识注入的燃气知识双向变换器模型[J].中国安全科学学报,2025,35(3):204-211.

英文引用格式:LIU Xiaoyu,ZHUANG Yufeng,ZHAO Xinghao,et al. Gas knowledge bidirectional encoder representations from transformers model based on knowledge injection[J]. China Safety Science Journal,2025,35(3):204-211.

基于知识注入的燃气知识双向变换器模型*

柳晓昱,庄育锋**教授,赵兴昊,王珂璠,张国开

(北京邮电大学智能工程与自动化学院,北京100876)

中图分类号:X913.4

文献标志码:A

DOI: 10.16265/j.cnki.issn1003-3033.2025.03.0223

基金项目:国家自然科学基金资助(52478123)。

【摘要】为提高燃气管网领域的应急管理水平和燃气知识双向变换器(Gas-kBERT)模型。该模型结合聊天生成预训练转换器(ChatGPT)扩充的燃气管网领域数据,以及构建的中文燃气语理解-三元组(CGLU-Spo)和相关语料库,通过改变模型的掩码(MASK)机制,成功将领域知识注入模型中。考虑到燃气管网领域的专业性和特殊性,Gas-kBERT在不同规模和内容的语料库上进行预训练,并在燃气管网领域的命名实体识别和分类任务上进行微调。结果表明:与通用的双向变换器(BERT)模型相比,Gas-kBERT在燃气管网领域的文本挖掘任务中 F_1 值表现出显著的提升。在命名实体识别任务中, F_1 值提高29.55%;在文本分类任务中, F_1 值提升高达83.33%。由此证明Gas-kBERT模型在燃气管网领域的文本挖掘任务中具有出色的表现。

【关键词】燃气管网;燃气知识双向变换器(Gas-kBERT)模型;自然语言处理(NLP);知识注入;双向变换器(BERT)模型

Gas knowledge bidirectional encoder representations from transformers model based on knowledge injection

LIU Xiaoyu,ZHUANG Yufeng,ZHAO Xinghao,WANG Kefan,ZHANG Guokai

(School of Intelligent Engineering and Automation, Beijing University of Posts and Telecommunications, Beijing 100876, China)

Abstract: In order to enhance emergency management in the field of gas pipeline networks, Gas-kBERT model was proposed. The model incorporated data from the gas pipeline network field expanded by Chat Generative Pre-Trained Transformer, (ChatGPT) and Chinese Gas Language Understanding Subject-Predicate-Object (CGLU-Spo) and related corpora were constructed in this field. By altering the model's masking (MASK) mechanism, domain knowledge was successfully injected into the model. Considering the professionalism and specificity of the gas pipeline network field, Gas-kBERT was pre-trained on various scales and contents of corpora and fine-tuned on named entity recognition and classification tasks within this field. Experimental results demonstrated that, compared to the general BERT model, Gas-kBERT exhibited significant performance improvements in F_1 -score in text mining tasks in the gas pipeline network field. Specifically, in the named entity recognition task, the F_1 -score was increased by 29.55%, and in

* 文章编号:1003-3033(2025)03-0204-08; 收稿日期:2024-10-14; 修稿日期:2024-12-18

** 通信作者:庄育锋(1972—),男,上海人,博士,教授,主要从事智能控制与装备安全等方面的研究。E-mail: zhuangyf@bupt.edu.cn.

the text classification task, the F_1 -score improvement reached up to 83.33%. This study proves that the Gas-kBERT model performs exceptionally well in text mining tasks in the gas pipeline network field.

Keywords: gas pipeline networks; gas knowledge bidirectional encoder representations from transformers (Gas-kBERT) model; natural language processing (NLP); knowledge injection; bidirectional encoder representations from transformers (BERT)

0 引言

燃气作为高效、清洁的能源,在工业化国家中占据重要地位。随着“西气东输”一线工程的成功投产以及川气东输、西气东输等工程的持续推进,我国已构建了庞大且复杂的燃气输配管道系统。但由于存在施工质量、管道使用环境与管道老化等多种因素,近年来燃气管网事故频发,对城市公共安全造成极大威胁,并造成重大经济损失^[1]。针对这一现状,“十四五”规划明确指出,未来10~20年,我国管道储运将处于稳定增长期,其中,管网建设将是发展重点。同时,也提出加强管网安全管理和风险防范的迫切需求。

信息技术的发展为燃气管网安全管理提供新的方法,如风险评价^[2]、可视化技术^[3]、3D视觉技术^[4]等。然而,燃气管网领域的文本数据,如记录、安全检查和经验教训等,尚未得到充分利用。这些非结构化的信息采用手动分析文本非常耗时^[5],需要通过自动化手段进行整理和筛选。自然语言处理(Natural Language Processing, NLP)技术的兴起,为其提供了新的解决方案^[6],通过迁移学习方法从多样化数据源中提取有价值信息^[7-9]。随着NLP的发展,预训练语言模型经历了从早期word embedding^[10]到现代上下文感知预训练模型的演变^[11-12]。其中,双向变换器(Bidirectional Encoder Representations from Transformers, BERT)模型以其掩码语言模型和下一句预测任务展示了强大的泛化能力。但当涉及专业术语和特定词汇时,通用预训练语言模型就难以准确理解特定专业领域内的文本^[13]。许多研究尝试将外部知识融入以BERT为代表的预训练语言模型中^[14-17],但目前在燃气管网领域缺乏针对该领域的定制化预训练模型。考虑到燃气管网领域的专业性和复杂性,定制化预训练模型在事故分析、风险评估和管理决策等方面具有巨大应用潜力,有望成为管理人员处理大量文档和案例的智能工具,以提高工作效率和决策质量。

鉴于此,笔者提出燃气知识BERT(Gas-knowledge, Gas-kBERT)模型,以解决通用模型对特

殊术语理解困难的问题,为验证Gas-kBERT模型的有效性,进行命名实体识别(Named Entity Recognition, NER)和文本分类(Classification, CLASS)2个任务的试验,以期能够在相关领域实现技术创新和成果突破。

1 Gas-kBERT模型构建

1.1 Gas-kBERT模型结构

Gas-kBERT是一种自我监督的预训练模型,旨在更好地学习燃气管网领域知识,但是不同于BERT的输入策略。首先,使用聊天生成预训练转换器(Chat Generative Pre-Trained Transformer, ChatGPT)工具生成额外文本数据,以提升模型训练效果。其次,基于大量燃气管网相关语料生成三元组,进一步丰富文本内容^[18]。然后,使用2种不同的掩码机制将燃气管网领域知识注入模型中。最后,将输入文档分割成片段作为正样本,并从其他样本中随机抽取句子作为负样本,用于下一句预测。模型结构如图1所示。

1.2 文本增强与知识注入

1.2.1 文本增强

ChatGPT经过大规模数据训练,并通过大量人类反馈进行强化训练,使模型具有人类语言的自然性。对NLP模型的泛化性能、对抗攻击、干扰波动有很好的提升作用。

Gas-kBERT首先利用ChatGPT工具将训练样本中的每个句子重述为多个语义相近但概念不同的样本,增强后的文本可有效用于下游模型的训练,显著提高模型的泛化能力和准确性,为后续的任务提供更好的支持。其次,通过构建燃气管网领域三元组图谱—中文燃气语言理解—三元组(Chinese Gas Language Understanding Subject-Predicate-Object, CGLU-Spo)扩充领域专业术语和行业词汇。通过mask技术掩盖三元组中的实体词,并通过注入三元组中的关系属性知识来深入学习这些实体词的含义,从而增强模型对燃气管网领域的理解能力,这一过程如图2所示。最后CGLU-Spo图谱基于现有

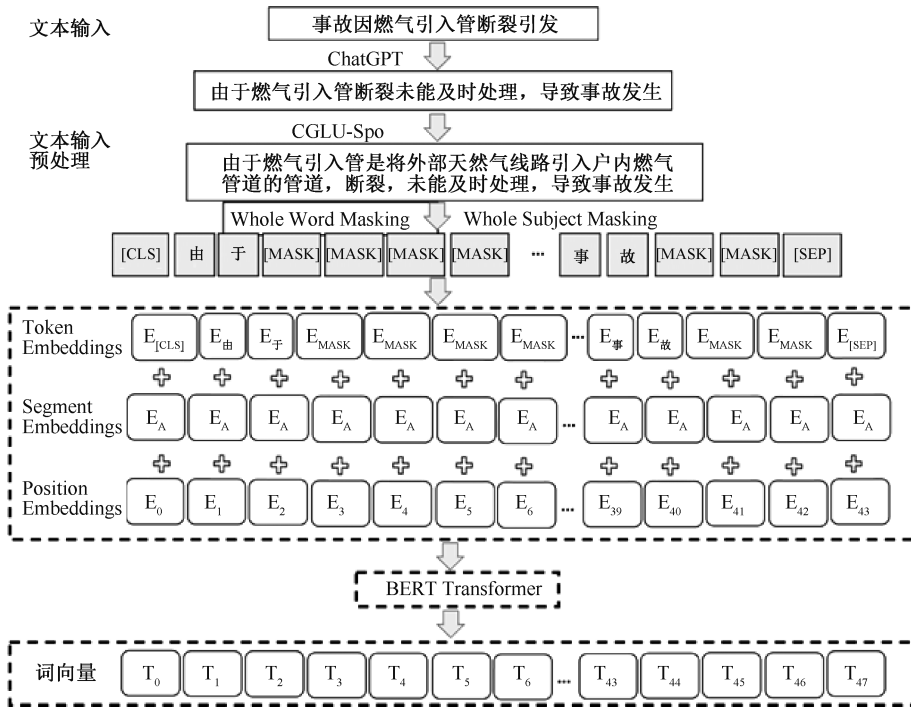


图1 Gas-kBERT 模型结构

Fig. 1 Gas-kBERT model structure diagram

燃气管网知识库,通过工具标注三元组信息,获得 3 990 条三元组,这些三元组结合预训练使用的全部数据,构建一个覆盖面广泛的三元组语料。

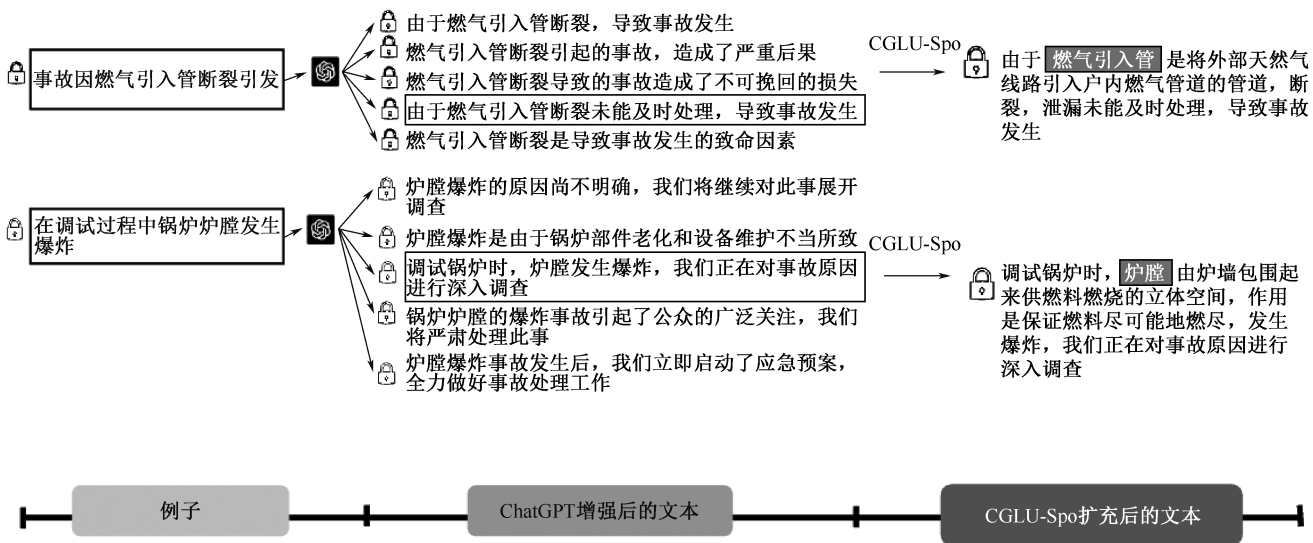


图2 CGLU-spo 扩充后的文本

Fig. 2 Expanded Text from CGLU-spo

1.2.2 知识注入—掩码机制

该模型主要采用 WWM 和 WEM 等 2 种掩码机制(图 3),这 2 种方法有效增强了模型对输入文本的理解能力。

1) WWM。此机制不再仅遮蔽单个标记,而是选择遮蔽整个单词。这种策略在处理依赖单词上下

文信息的任务时尤为有效,相比于单个标记的遮蔽, WWM 能够保留单词的完整语义和上下文信息,从而显著提升模型对文本内容的理解能力,并在相关任务中提升模型性能。

2) WEM。与 BERT 遮蔽机制不同,此机制是一种确定性的策略,利用燃气管网领域特定实体的知

识来增强模型。具体而言,利用预先构建的燃气管网领域三元组语料库,其中,包含关键实体,如“燃气引入管”等;模型在处理文本时,会识别句子中是否包含这些三元组中的实体,并进行遮蔽操作。通过这种方式能够有效地将领域相关的实体和专业知

识注入模型中,大幅提升模型在燃气管网相关任务中的表现。

通过引入上述的掩码机制,模型可以学习如何使用属性和关系来识别和理解实体,并将这些知识融合到模型中。

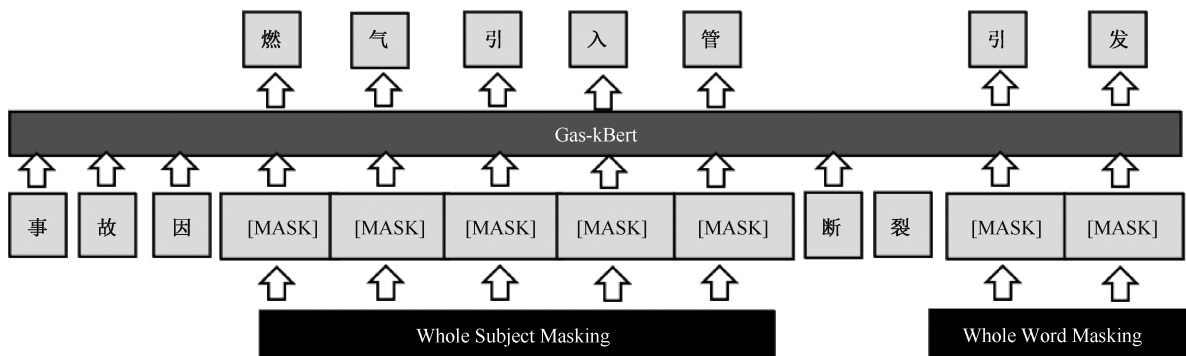


图3 模型的掩码机制

Fig. 3 The masking mechanism of the model

1.3 Gas-kBERT 预训练

BERT 作为通用语言模型,主要在通用领域语料库上进行预训练,但由于缺乏直接支持燃气管网领域的数据库,大量燃气管网领域数据未得到充分利用,使得在燃气管网文本挖掘中因领域差异表现受限。为此,基于 BERT 模型使用燃气管网领域数据预训练模型并做微调,预训练文本见表 1,整体训练流程如图 4 所示。

表 1 预训练语料

Table 1 Pre-training corpus

数据	大小 (原始 文本)	大小 (ChatGPT 增强后)	领域
新闻语料 (News_zh_2016)	—	1.6 G	通用领域
维基百科 (Wiki_zh_2019)	—	1.2 G	通用领域
燃气事故报告 (Gas Accident Report, GAR)	5.1M	20.3 M	燃气管网
燃气标准 (Gas standards, GS)	3.4M	16.9 M	燃气管网
燃气新闻 (Gas News, GN)、燃气应急预案 (Gas Emergency Plan, GEP)等	2.7 M	13.7M	燃气管网

1.4 Gas-kBERT 微调任务

通过最小的架构修改, Gas-kBERT 能够胜任多种燃气管网文本挖掘任务。并在 NER 和 CLASS 这

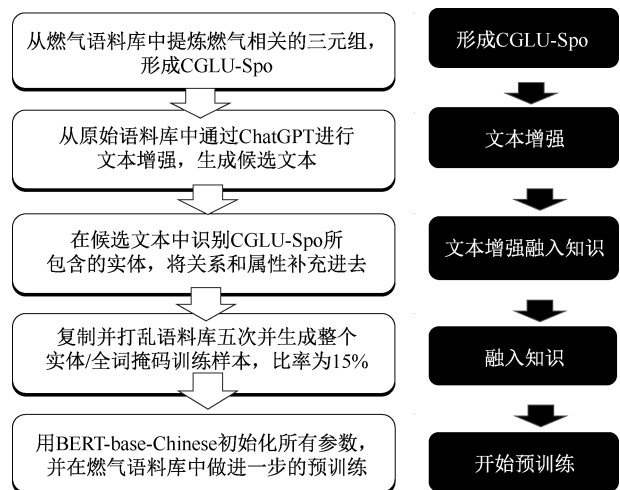


图4 Gas-kBERT 模型整体训练流程

Fig. 4 Gas-kBERT model overall training process

2 个关键任务中进行微调, 评估指标采用精确率 P 、召回率 R 和 F 值。

2 预训练与微调试验

2.1 预训练与微调试验设置

1) 试验数据集获取与处理。通过爬虫等工具收集燃气管网领域的多样化数据, 包括事故报告、标准以及事故新闻等(表 2)。为保证模型训练效果, 数据集经过预处理, 包括去除重复、无效和空文本、分词、标点符号和停用词^[19]。最终获得的预训练和微调任务文本语料见表 2 和表 3。

2) 三元组语料 CGLU-Spo 的部署。为提取结构化知识, 基于现有知识库并选取涉及燃气管网文

表 2 预训练任务语料组合

Table 2 Pre-training task corpus combination

模型	语料组合
BERT	News+Wiki
Gas-kBERT(raw)	GAR+GS+GN+GEP(原始文本)
Gas-kBERTv1.0(GAR+GS+GN+GEP)	GAR+GS+GN+GEP(ChatGPT 增强后文本)
Gas-kBERTv1.1(+GAR+GS+GN+GEP)	News+Wiki+GAR+GS+GN+GEP(ChatGPT 增强后文本)

表 3 微调任务数据集

Table 3 Fine-tuning task dataset

测试任务	数据	训练集	验证集	测试集
命名实体识别	GS	558,458	260,614	111,692
	事故报告	309,782	119,238	2,016
分类	应急预案	1,152	384	128
	事故报告	612	116	84

本的网站,爬取相关词条。借助标注工具,构建 3 990 条燃气管网领域的三元组语料。

3) 模型参数设置。为与 BERT 进行比较,采用相同的模型设置,模型包含 12 层,12 个 self-attention heads,768 维的 hidden size,1 000 000 个 step,每个 batch 包含 256 条序列。learning rate 为 1×10^{-4} $\beta_1 = 0.9$ $\beta_2 = 0.999$ 。learning rate 前 10 000 个 steps 使用 rate warmup,之后开始线性衰减 learning rate。对所有 layer 使用 0.1 概率的 dropout。使用的激活函数为 gelu。

4) 试验环境。试验采用的图形处理器为 NVIDIA TI Tesla T4,代码通过 pycharm 平台进行编程,开发语言 Python,深度学习框架为 Tensorflow1.15.0,torch1.10.2。

5) 对比模型。为验证 Gas-kBERT 模型在下游任务的优越性,引入对照组 BERT 模型,同时为验证预训练数据内容对 Gas-kBERT 模型的影响,引入 Gas-kBERT (raw)、Gas-kBERTv1.0 和 Gas-kBERTv1.1 作为对照组,模型均在相同的试验环境下以相同的方式进行训练。

2.2 预训练与微调试验结果

2.2.1 命名实体识别任务

燃气管网相关的报告具有很强的非结构化性,通过 NER 技术,可以从这些报告中提取关键信息,如事故发生时间、地点、原因以及性质等。通过对比在不同数据集上进行预训练的模型在 NER 任务上的表现,得出试验结果,具体见表 4。将结论分为 4 类:① Gas-kBERT (包括 Gas-kBERTv1.0 和 Gas-kBERTv1.1) 和 Gas-kBERT(raw) 在所有数据集上的 *F* 分数均优于 BERT,这表明使用燃气管网领域语料进行预训练的模型,在燃气管网领域的文本挖掘任务中表现更为出色。②在 GAR 数据测试任务中 Gas-kBERTv1.0 取得更高的 *F* 值,相较于 Gas-kBERTv1.1 和 BERT 上升 18.94% 和 48.49%。在 GN 数据测试任务中分别提升 18.48% 和 28.15%。这一结果说明,预训练的数据并不是越多越好,而是越有针对性越好。③ Gas-kBERT (raw) 在所有数据集上的 NER 任务测试结果的 *F* 分数均优于 Gas-kBERT,这表明:通过使用 ChatGPT 增强文本能够提升模型的训练效果。④几个模型在数据集上的得分普遍相对较低,原因可能包括:标注数据集有限;模型参数尚未优化调整以达到最佳性能;标注数据集可能存在错误、标签不一致或模糊,这可能影响模型的训练和评估质量。

表 4 燃气管网领域 NER 任务测试结果

Table 4 NER task test results in the gas pipeline network field

数据	评价指标	BERT	Gas-kBERT (raw)	Gas-kBERT v1.0	Gas-kBERT v1.1
		(+news_zh_2016+ wiki_zh_2019)	GAR+GS+GN+GEP (原始文本)	(GAR+GS+GN+GEP) (ChatGPT 增强后文本)	(+news_zh_2016+ wiki_zh_2019+ GAR+GS+GN+GEP) (ChatGPT 增强后文本)
GAR	<i>P</i>	11.11	24.22	55.56	55.56
	<i>R</i>	13.33	35.19	66.67	33.33
	<i>F</i> ₁	12.12	30.68	60.61	41.67
GN	<i>P</i>	30.3	40.02	46.10	40.40
	<i>R</i>	16.67	20.15	14.60	43.01
	<i>F</i> ₁	41.67	50.26	69.82	51.34

2.2.2 分类任务

分类任务是根据燃气管网相关文本对内容类别

进行分类。在这一阶段,采用相同的数据,结果见表 5。从中可以将结果分为 3 类:① Gas-kBERT 在

表 5 燃气管网领域 CLASS 任务结果
Table 5 CLASS task results in the field of gas pipeline network

数据	模型	评价指标	BERT	Gas-kBERT(raw)	Gas-kBERT v1. 0	Gas-kBERT v1. 1
	类型		(+news_zh_2016+ wiki_zh_2019)	GAR+GS+GN+ GEP(原始文本)	(GAR+GS+GN+GEP) (ChatGPT 增强后文本)	(+news_zh_2016+ wiki_zh_2019+ GAR+GS+GN+GEP) (ChatGPT 增强后文本)
GEP	总则	<i>P</i>	42. 42	50. 13	62. 50	60. 87
		<i>R</i>	93. 33	93. 12	100	93. 33
		<i>F</i> ₁	58. 33	70. 23	76. 92	73. 68
	应急组织体系及职责	<i>P</i>	64. 91	40. 29	81. 82	41. 38
		<i>R</i>	94. 87	90. 26	92. 31	30. 77
		<i>F</i> ₁	77. 08	70. 33	86. 75	35. 29
	监测、预警	<i>P</i>	100	50. 21	66. 67	52. 63
		<i>R</i>	8. 33	70. 92	83. 33	83. 33
		<i>F</i> ₁	15. 38	60. 71	74. 07	64. 52
	应急响应	<i>P</i>	45. 95	60. 07	63. 16	66. 67
		<i>R</i>	73. 91	60. 91	52. 17	52. 17
		<i>F</i> ₁	56. 67	56. 34	57. 14	58. 54
	信息报告与发布	<i>P</i>	0	10. 19	25	20
		<i>R</i>	0	8. 28	12. 50	12. 50
		<i>F</i> ₁	0	10. 07	16. 67	15. 38
	善后恢复	<i>P</i>	0	0	0	0
		<i>R</i>	0	0	0	0
		<i>F</i> ₁	0	0	0	0
	保障措施	<i>P</i>	0	40. 43	55. 56	47. 06
		<i>R</i>	0	56. 67	71. 43	57. 14
		<i>F</i> ₁	0	45. 13	62. 50	51. 61
宣传教育、培训和应急演练	<i>P</i>	0	0	0	0	
	<i>R</i>	0	0	0	0	
	<i>F</i> ₁	0	0	0	0	
附则	<i>P</i>	0	80. 27	100	100	
	<i>R</i>	0	70. 19	71. 43	71. 43	
	<i>F</i> ₁	0	79. 62	83. 33	83. 33	
GAR	事故伤亡人员、事故损失	<i>P</i>	86. 21	90. 30	92. 59	92
		<i>R</i>	69. 44	87. 93	92. 59	85. 19
		<i>F</i> ₁	76. 92	80. 51	92. 59	88. 46
	事故原因	<i>P</i>	92. 31	93. 33	95. 65	95. 65
		<i>R</i>	85. 71	95. 25	100	100
		<i>F</i> ₁	88. 89	96. 14	97. 78	97. 78
	性质	<i>P</i>	47. 37	60. 19	100	63. 64
		<i>R</i>	100	100	100	100
		<i>F</i> ₁	64. 29	70. 21	100	77. 78
	事故发生经过	<i>P</i>	60. 00	70. 01	71. 43	100
		<i>R</i>	54. 55	60. 06	62. 50	62. 50
		<i>F</i> ₁	57. 14	62. 15	66. 67	76. 92

多数与燃气管网 CLASS 任务相关的数据集上较 BERT 具有更高的 *F*₁ 分数,可以看出,用于燃气管网语料进行预训练的模型,在燃气管网领域文本挖掘任务中通常表现更好。然而,也注意到在应急组

织体系及职责和应急响应 2 个具体任务上,BERT 的表现优于 Gas-kBERT(raw),但 Gas-kBERT 表现又优于 BERT,这表明:在模型训练过程中,仅注入相关领域文本是不够的,更重要的是要让模型能够

有效地学习和理解这些知识。②Gas-kBERT 较 Gas-kBERT(raw) 具有更高的 F_1 分数,再次证明使用 ChatgPT 增强文本是有效的。③Gas-kBERT v1.1 在所有燃气管网 CLASS 数据集上的训练得分相对于 Gas-kBERT v1.0 较低,可能原因:预训练数据可能过于多样化和杂乱,包括通用和燃气管网领域数据;在不同数据集上可能需要调整不同的训练参数,以获得最佳模型性能。

3 结论

1) Gas-kBERT 模型解决了传统通用模型无法充分识别专业领域文本的问题。试验结果表明: Gas-kBERT 模型在燃气管网文本挖掘任务中明显优

于 BERT 和其他基准模型的效果,尤其是在使用包含(GAR+GS+GN+GEP)数据的 Gas-kBERT v1.0 模型中, F_1 分数提升达到 83.33%。

2) 从知识注入角度思考燃气领域预训练语言模型,通过改变 MASK 策略、注入相关领域三元组、结合 ChatGPT 有效提升了模型在燃气领域任务中的精确率 P 、召回率 R 和 F_1 分数。

3) 文中还存在一些限制,如微调任务相对较少、知识图谱的引入方式简单等。为进一步提升模型的性能和泛化能力,未来的工作中将扩充测试任务数据集、考虑注入的知识在文本中的位置以及探索更多技术手段来优化模型表现。

参考文献

- [1] MONTEL H, VÍLCHEZ JA, ARNALDOS J, et al. Historical analysis of accidents in the transportation of natural gas[J]. *Journal of Hazardous Materials*, 1996, 51 (1/3): 77-92.
- [2] 郭秀军, 刘嘉柠. 智能化燃气管网监测与管理[J]. *网络安全和信息化*, 2024(1): 70-72.
- [3] 李兰. 基于视觉检测与图像处理分析技术的燃气管道检测方法研究[D]. 南京: 东南大学, 2019.
LI Lan. Study on the gas pipeline detection method based on visual inspection and image processing analysis technology[D]. Nanjing: Southeast University, 2019.
- [4] 苟爽. 基于机器视觉的燃气管道内环境感知技术研究[D]. 昆明: 昆明理工大学, 2018.
GOU Shuang. Research on internal environmental perception technology of gas pipeline based on machine vision [D]. Kunming: Kunming University of Science and Technology, 2018.
- [5] 周吉祥, 张浩. 智能管道建设下的城镇燃气管道完整性管理[J]. *煤气与热力*, 2021, 41(2): 32-36, 46.
ZHOU Jixiang, ZHANG Hao. Integrity management of urban gas pipeline under construction of intelligent pipeline[J]. *Gas & Heat*, 2021, 41(2): 32-36, 46.
- [6] 李思洁, 王亚慧, 张子豪. 燃气输配突发事件应急处置的知识图谱构建[J]. *消防科学与技术*, 2022, 41(6): 812-817.
LI Sijie, WANG Yahui, ZHANG Zihao. Construction of knowledge graph for emergency disposal of gas transmission and distribution emergencies[J]. *Fire Science and Technology*, 2022, 41(6): 812-817.
- [7] TIWARY U S, SIDDIQUI T. Natural language processing and information retrieval[M]. New York: Oxford University Press, 2008: 656.
- [8] WANG Chenglong, JIANG Feijun, YANG Hongxia. A hybrid framework for text modeling with convolutional RNN[C]. The 23rd ACM SIGKDD International Conference, 2017; DOI: 10.1145/3097983.3098140.
- [9] QIN Ya, SHEN Guowei, ZHAO Wenbo, et al. A network security entity recognition method based on feature template and CNN-BiLSTM-recognition method based on feature template and CNN-BiLSTM-CRF [J]. *Frontiers of Information Technology and Electronic Engineering*, 2019, 20(6): 13-19.
- [10] RAO N V, KRISHNA B T. Automatic modulation recognition using machine learning techniques: a review[C]//KALYA S, KULKARNI M, SHIVAPRAKASHA K S. *Advances in VLSI, Signal Processing, Power Electronics, IoT, Communication and Embedded Systems*. Singapore: Springer Singapore, 2021: 145-154.
- [11] MA Zhiliang, CAI Shiyao, MAO Na, et al. Construction quality management based on a collaborative system using BIM and indoor positioning[J]. *Automation in Construction*, 2018, 92: 35-45.
- [12] PETERS M, NEUMANN M, IYYER M, et al. Deep contextualized word representations[C]. The 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2018:

2 227-2 237.

- [13] DEVLIN J, CHANG Mingwei, LEE Kenton, et al. BERT: pre-training of deep bidirectional transformers for language understanding[C]. Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2019. DOI:10.48550/arXiv.1810.04805.
- [14] 邓存彬,虞慧群,范贵生,等. 燃气客服热线的中文文本情感分析[J]. 华东理工大学学报:自然科学版,2019,45(1):140-147.
DENG Cunbin, YU Huiqun, FAN Guisheng, et al. Sentiment analysis of Chinese texts for gas customer service hotline [J]. Journal of East China University of Science and Technology: Natural Science Edition, 2019, 45(1): 140-147.
- [15] 王明达,张榜,吴志生,等. 基于强化学习的城镇燃气事故信息抽取方法[J]. 中国安全生产科学技术,2023,19(3):39-45.
WANG Mingda, ZHANG Bang, WU Zhisheng, et al. Information extraction method of urban gas accidents based on reinforcement learning[J]. Journal of Safety Science and Technology, 2023, 19(3): 39-45.
- [16] LEE J H, YOON W J, KIM S D, et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining[J]. Bioinformatics, 2020, 36(4): 1 234-1 240.
- [17] WANG Xuan, ZHANG Xiang, REN Yuhao, et al. Cross-type biomedical named entity recognition with deep multi-task learning[J]. Bioinformatics, 2019, 35(10): 1 745-1 752.
- [18] 宋敦江,杨霖,钟少波. 基于 BERT 的灾害三元组信息抽取优化研究[J]. 中国安全科学学报,2022,32(2): 115-120.
SONG Dunjiang, YANG Lin, ZHONG Shaobo. Research on optimization of disaster triplet information extraction based on BERT[J]. China Safety Science Journal, 2022, 32(2): 115-120.
- [19] 郑彬彬,冯婷婷,王佳贺,等. 基于文本挖掘的城镇燃气事故致因及关联分析[J]. 中国安全科学学报,2023,33(7): 190-195.
ZHENG Binbin, FENG Tingting, WANG Jiahe, et al. Causes and correlation analysis of urban gas accidents based on text mining[J]. China Safety Science Journal, 2023, 33(7): 190-195.



作者简介: 柳晓昱 (2000—),女,天津人,硕士研究生,主要研究方向为文本挖掘与物流信息技术及工程应用。E-mail: ivy@bupt.edu.cn。