

文章编号: 1671-251X(2023)05-0046-06

DOI: [10.13272/j.issn.1671-251x.2023020005](https://doi.org/10.13272/j.issn.1671-251x.2023020005)

煤矿综采设备故障知识图谱构建

蔡安江, 张妍, 任志刚

(西安建筑科技大学 机电工程学院, 陕西 西安 710000)

摘要: 现有煤矿综采设备故障诊断方法缺乏对综采设备历史故障数据的系统化管理及应用, 针对该问题, 引入知识图谱技术对综采设备故障数据进行系统化管理。采用自顶而下的方法对综采设备故障知识进行本体构建, 将综采设备故障知识归纳为故障位置、故障现象、故障原因、处理方法4类, 并进行规范化命名; 采用通用的命名实体标注方法 BIOES 对综采设备故障知识进行人工标注; 将双向长短句记忆(BiLSTM)和条件随机场(CRF)相结合, 构建 BiLSTM-CRF 模型, 对已标注的综采设备故障知识进行命名实体识别, 并通过人工抽取实体关系, 从而实现故障知识抽取; 结合 BiLSTM-CRF 模型的实体识别结果和人工抽取的实体关系, 使用 Neo4j 图数据库存储综采设备故障知识, 构建综采设备故障知识图谱。实验结果表明, 相较于 BiLSTM 模型和 BiLSTM-Attention 模型, BiLSTM-CRF 模型精确率显著提高, 为 87%, F_1 值也有一定幅度上升, 为 69%。综采设备故障知识图谱的构建可为大规模、多域综采设备故障数据的有效分析、管理及应用提供支持。

关键词: 煤矿综采设备; 故障诊断; 知识图谱; 知识抽取; BiLSTM; CRF

中图分类号: TD632 文献标志码: A

Fault knowledge graph construction for coal mine fully mechanized mining equipment

CAI Anjiang, ZHANG Yan, REN Zhigang

(School of Mechanical and Electrical Engineering, Xi'an University of Architecture and Technology,
Xi'an 710000, China)

Abstract: The existing fault diagnosis methods for coal mine fully mechanized mining equipment lack systematic management and application of historical fault data of fully mechanized mining equipment. In response to this problem, knowledge graph technology is introduced to systematically manage the fault data of fully mechanized mining equipment. The top-down approach is used to construct the ontology of fully mechanized mining equipment fault knowledge. The knowledge of fully mechanized mining equipment fault is classified into four categories: fault location, fault phenomenon, fault cause, and treatment method. And the naming of the knowledge is standardized. The universal naming entity annotation method BIOES is used to manually annotate the fault knowledge of fully mechanized mining equipment. By combining bi-directional long short-term memory (BiLSTM) and conditional random field (CRF), the BiLSTM-CRF model is constructed. The marked fault knowledge of fully mechanized mining equipment is identified by the named entity, and the fault knowledge extraction is realized by manually extracting entity relationships. Combining the entity recognition results of the BiLSTM-CRF model with the manually extracted entity relationships, a Neo4j graph database is used to store the fault knowledge of fully mechanized mining equipment. A fault knowledge graph of fully mechanized mining equipment is constructed. The experimental results show that compared to the BiLSTM model and BiLSTM-Attention model, the accuracy of the BiLSTM-CRF model is significantly improved, reaching 87%. The F_1 value

收稿日期: 2023-02-01; 修回日期: 2023-04-28; 责任编辑: 胡娟。

基金项目: 工信部物联网集成创新与融合应用项目(2018-470); 榆林市科技计划项目(CXY-2022-172)。

作者简介: 蔡安江(1965—), 男, 安徽舒城人, 教授, 博士研究生导师, 研究方向为人工智能及智能制造, E-mail: cai_aj@163.com。

引用格式: 蔡安江, 张妍, 任志刚. 煤矿综采设备故障知识图谱构建[J]. 工矿自动化, 2023, 49(5): 46-51.

CAI Anjiang, ZHANG Yan, REN Zhigang. Fault knowledge graph construction for coal mine fully mechanized mining equipment[J]. Journal of Mine Automation, 2023, 49(5): 46-51.



扫码移动阅读

also has a certain increase, reaching 69%. The construction of fully mechanized mining equipment fault knowledge graph can provide support for the effective analysis, management, and application of large-scale and multi-domain fully mechanized mining equipment fault data.

Key words: coal mine fully mechanized mining equipment; fault diagnosis; knowledge graph; knowledge extraction; BiLSTM; CRF

0 引言

煤矿综采设备长期在潮湿、煤尘环境下运行,且受到巨大冲击载荷和异常振动的影响,导致故障频发,从而直接影响生产^[1-3]。通过采集到的设备状态信息实现故障诊断,是实现煤矿智能化、无人化目标的关键环节。目前,已有学者对综采设备故障诊断问题展开研究。李旭等^[4]利用云平台等技术构建了综采设备群的远程故障诊断系统,提高了综采设备的维护效率,但故障诊断的实现依赖于专家经验。张旭辉等^[5]基于深度迁移学习,结合采煤机运行过程中的振动信号,对采煤机摇臂部的滚动轴承进行故障诊断,一定程度上解决了故障诊断依赖专家经验的问题,但采煤机摇臂部滚动轴承的故障诊断仅为综采工作面的单一设备服务。现有研究都缺乏对综采设备历史故障数据的系统化管理及应用。

近年来,有学者开始研究系统管理中故障知识的结构化和可视化问题,将知识图谱引入电源系统故障诊断、电网故障处理等应用中^[6-8],取得了一定成效。知识图谱技术可以有效解决目前实际生产过程中综采设备故障诊断依赖专家经验且历史故障数据难以有效利用的问题。因此,本文将知识图谱应用到煤矿综采设备故障知识管理中,根据综采设备故障数据特点,采用双向长短期记忆(Bi-directional Long Short-Term Memory, BiLSTM)-条件随机场(Conditional Random Field, CRF)模型进行命名实体抽取,通过Neo4j图数据库对提取的故障知识进行存储,构建综采设备故障知识图谱。

1 综采设备故障知识图谱构建

构建综采设备故障知识图谱的主要步骤:①明确综采工作面设备故障知识构建需求,确定故障知识本体并对故障知识进行预处理。②利用基于文本数据的BiLSTM-CRF模型对综采设备故障原始语料进行命名实体识别,并通过人工抽取实体关系,从而实现故障知识抽取。③利用Neo4j图数据库存储综采设备故障知识,构建综采设备故障知识图谱。

1.1 故障知识本体构建及预处理

故障知识能够快速有效地辅助工作人员分析故

障原因并获取处理方法。综采设备故障知识主要以文本数据为主,通过检修日志、专家经验、技术文档、网络爬取等技术手段获取^[9]。综采设备故障知识结构如图1所示。

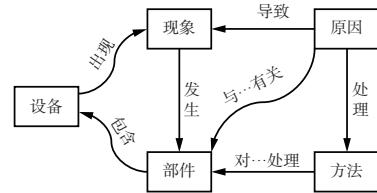


图1 综采设备故障知识结构

Fig. 1 Fault knowledge structure of fully mechanized mining equipment

根据综采设备故障知识图谱的构建需求,采用自顶而下的方法对综采设备故障知识进行本体构建^[10]。将综采设备故障知识归纳为故障位置、故障现象、故障原因、处理方法4类,并进行规范化命名。

1) 故障位置规范化命名。故障位置的规范化命名需要精确定位到某设备。通过整理故障相关资料,故障位置包括采煤机、液压支架、破碎机、转载机、液压泵、刮板输送机、带式输送机。

2) 故障现象规范化命名。选择多个关键词与相应的故障现象进行匹配,从而快速、准确地描述故障现象。例如转载机的传动链条磨损问题可规范化命名为“传动链条磨损”。常见故障现象包括转载机传动链条磨损,液压支架立柱和千斤顶不动作、管路系统故障,乳化液泵运转噪声大、压力过高,带式输送机跑偏、减速机漏油,采煤机只能单向牵引、液压牵引部存在声响,刮板输送机刮板链断链、飘链、跳链或掉链等^[11]。

3) 故障原因规范化命名。故障现象和故障原因存在一对一或一对多的关系,可采用故障原因1、故障原因2等进行规范化命名。

4) 处理方法规范化命名。一种故障原因与一种解决方法相对应,在规范化命名时应尽可能简洁明了,如“更换密封圈”“清洁周围杂物”“更换液压管”。

实体是构建综采设备故障诊断知识图谱最基本元素,实体的完整性和准确性将直接对综采设备故障知识库的整体质量产生影响。知识抽取是指从多

种来源、多种结构类型的数据中提取需要的故障知识,形成结构化数据,以便存储到对应的知识图谱中^[12]。知识抽取主要包括命名实体识别及实体关系抽取。实体关系是通过人工抽取的,因此本文重点介绍命名实体识别。

在进行命名实体识别前,采用通用的命名实体标注方法 BIOES 对综采设备故障知识进行人工标注^[13]。BIOES 方法中, B 为综采设备故障实体起始位置(Begin), I 为综采设备故障实体内部(Inside), O 为非综采设备故障实体(Outside), E 为综采设备故障实体尾部(End), S 为单独一个综采设备故障实体(Single)。

1.2 命名实体识别模型构建

命名实体识别是知识抽取过程中最重要的一步,一般方法是通过深度学习技术^[14],从不同数据源中自动或半自动地抽取实体。传统的实体抽取技术主要包括 CRF 和 BiLSTM。BiLSTM 采用梯度下降法进行权值和阈值更新,易使参数陷入局部最优^[15]。CRF 可对特征进行归一化,求得全局最优解^[16]。因此,本文将 BiLSTM 和 CRF 相结合,采用基于文本数据的 BiLSTM-CRF 模型对已标注的综采设备故障知识进行实体识别,该模型由字嵌入层、BiLSTM 层、CRF 层 3 个部分构成。

1.2.1 字嵌入层

字嵌入层的主要作用是将综采设备故障相关资料转换成模型可识别并能够进行计算的形式。对已经完成人工标注的综采设备故障语料进行简单预训练。建立综采设备故障语料库的词汇特征表,将每个字映射为低维稠密向量。综采设备故障数据语料中的故障语句,在词汇表中均可查找到与其相对应的字向量,将对应的字向量纵向拼接,可形成综采设备故障矩阵,该矩阵为后续模型的输入。

1.2.2 BiLSTM 层

长短期记忆(Long Short-Term Memory, LSTM)网络是一种时间循环神经网络,通过有选择性地记忆综采设备故障信息,可以充分利用综采设备故障样本数据记录的绝大部分信息,从而快速准确地进行综采设备故障诊断^[17-18]。LSTM 网络结构如图 2 所示。

上一刻综采设备故障输出 h_{t-1} 与当前 t 时刻综采设备故障样本输入 x_t 同时决定细胞状态 C_{t-1} 中需要被保留的综采设备故障信息。

$$f_t = \sigma(W_f(h_{t-1}, x_t) + b_f) \quad (1)$$

式中: f_t 为遗忘程度; σ 为 Sigmoid 激活函数; W_f 为遗

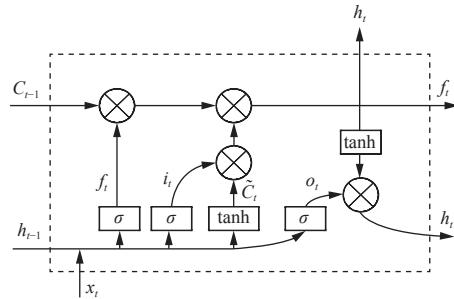


图 2 LSTM 网络结构

Fig. 2 Network structure of LSTM

忘门权重; b_f 为遗忘门偏置。

细胞状态 C_{t-1} 的主要作用是存储前一时刻的综采设备故障记忆数据。

$$C_t = f_t C_{t-1} + i_t \tilde{C}_t \quad (2)$$

$$i_t = \sigma(W_i(h_{t-1}, x_t) + b_i) \quad (3)$$

$$\tilde{C}_t = \tanh(W_c(h_{t-1}, x_t) + b_c) \quad (4)$$

式中: C_t 为当前细胞状态; i_t 为输入门对综采设备故障新数据的更新程度; \tilde{C}_t 为当前新状态数据的候选值; W_i 为输入门权重; b_i 为故障数据提取过程中的偏置; W_c 为故障数据提取过程中的权重; \tanh 为激活函数。

式(2)中, $f_t C_{t-1}$ 为当前需要被遗忘的综采设备故障数据, $i_t \tilde{C}_t$ 为目前需要被存储的综采设备故障数据。通过上述处理之后,所有需要的综采设备故障信息将被存储在记忆模块中,而后通过输出门输出综采设备故障数据。

$$o_t = \sigma(W_o(h_{t-1}, x_t) + b_o) \quad (5)$$

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

式中: o_t 为输出门的输出; W_o 为输出门权重; h_t 为隐藏层的输出; b_o 为输出门偏置。

LSTM 神经网络只能学习到前向数据,而双向的 LSTM 结构有利于充分利用综采设备故障知识的上下文信息数据,即利用前向和后向 2 个 LSTM 网络对序列进行建模。在处理序列任务时,同时考虑前面和后面的信息,从而更好地捕捉序列中的长期依赖关系和上下文信息,对文本序列的时序特征进行有效学习。BiLSTM 网络结构如图 3 所示, Z_t 为输出数据。

首先将综采设备故障资料经过输入层输入,通过连接 LSTM 前传层与后传层的 BiLSTM 网络层对综采设备故障样本数据信息进行正序和逆序特征提取,将得到的 2 个故障向量进行拼接,形成新的故障向量,传入故障矩阵输出层,最后输出综采设备故障信息。

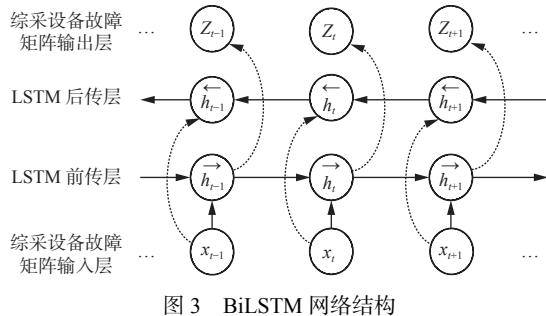


Fig. 3 Network structure of BiLSTM

1.2.3 CRF 层

CRF 层对 BiLSTM 层输出的综采设备故障融合特征数据进行解码,通过特征函数得到各序列的转移概率,用 Viterbi 算法计算分数最高的序列标签^[19]。

假设数据输入包含 5 个单词 w_0-w_4 , w_0, w_1 为故障位置, w_3 为故障现象, w_2, w_4 为非实体信息。经过 CRF 层后,输出各单词对应的预测标签,如图 4 所示。对应 w_0-w_4 的输出分别为 B-Local(1.3), I-Local(0.4), O(0.1), B-Phenomenon(0.4), O(0.8)。B-Local 为综采设备故障位置实体起始位置, I-Local 为综采设备故障位置实体内部, B-Phenomenon 为综采设备故障现象实体起始位置, I-Phenomenon 为综采设备故障现象实体内部。

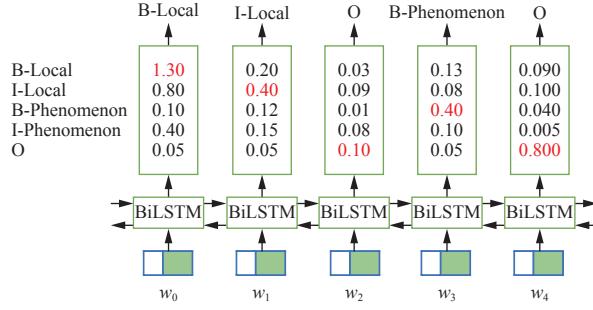


图 4 BiLSTM-CRF 模型样例分析

Fig. 4 Sample analysis of BiLSTM-CRF model

1.3 故障知识存储与查询

综采设备故障知识库属于垂直知识图谱,主要应用于煤矿井下综采工作面设备故障诊断,整体数据量较大。因此,将综采设备故障知识存储在 Neo4j 图数据库中^[20],以便于后续分析。在 Windows 10 环境下建立综采设备故障知识库,通过人机交互界面操作,实现故障知识图谱与用户之间的交流,从而完成故障诊断。综采工作面设备故障语料中每个实体为 1 个节点,每种关系为 1 条边,用来连接头实体和尾实体。结合 BiLSTM-CRF 模型的实体识别结果和人工抽取的实体关系,使用 Neo4j 图数据库构建综采设备故障知识图谱。综采设备故障知识图谱中共计 1 384 个实体节点,2 327 种关系类型。为清晰展示图谱内容,选取部分综采设备故障知识图谱进

行可视化展示,如图 5 所示。蓝色节点为故障位置,黄色节点为该故障位置出现的故障现象,节点以现象为边相互连接。

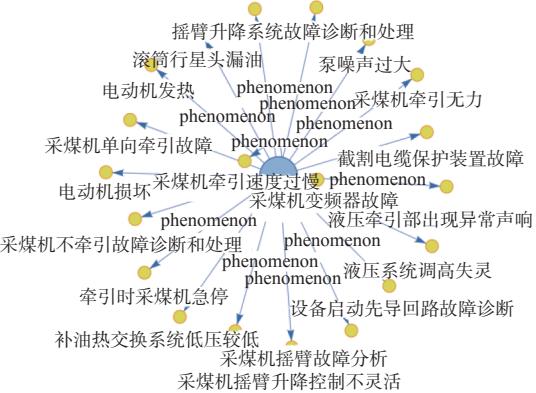


图 5 综采设备故障知识图谱

Fig. 5 Fault knowledge graph of fully mechanized mining equipment

根据综采设备故障知识图谱,可设计知识问答系统,如图 6 所示。在知识问答界面输入问题,可得到相应的解决方案。



图 6 知识问答界面

Fig. 6 Knowledge Q&A interface

2 实验分析

为了评估 BiLSTM-CRF 模型的准确性和有效性,采用对比模型 BiLSTM、添加注意力机制的 BiLSTM-Attention^[21]进行实验分析。由于目前在综采设备故障方面没有公开数据集,结合专家经验、技术文档,采用网络爬取、矿山实地调研等方式获取综采设备故障相关文献资料作为实验语料。考虑到不同类型文献的格式、书写内容、书写方式不同,随机从文献中抽取 30 000 个句子,按照 7:3 划分训练集和测试集,并通过手动分词构建训练语料。采用精确率、召回率和 F_1 值作为评价指标评判实体识别模型的性能。实验结果见表 1。可看出 BiLSTM 模型和 BiLSTM-Attention 模型精确率分别为 69%, 62%, BiLSTM-CRF

模型的精确率显著提高,为87%, F_1 值也有一定幅度上升,为69%。各模型在训练与验证过程中的精确率变化如图7所示。相较于BiLSTM和BiLSTM-Attention,BiLSTM-CRF的精确率显著提升,且整体波动幅度较小。

表1 不同模型的实验结果对比

Table 1 Comparison of experimental results of different models

模型	精确率	召回率	F_1	%
BiLSTM	69	67	68	
BiLSTM-Attention	62	63	63	
BiLSTM-CRF	87	69	69	

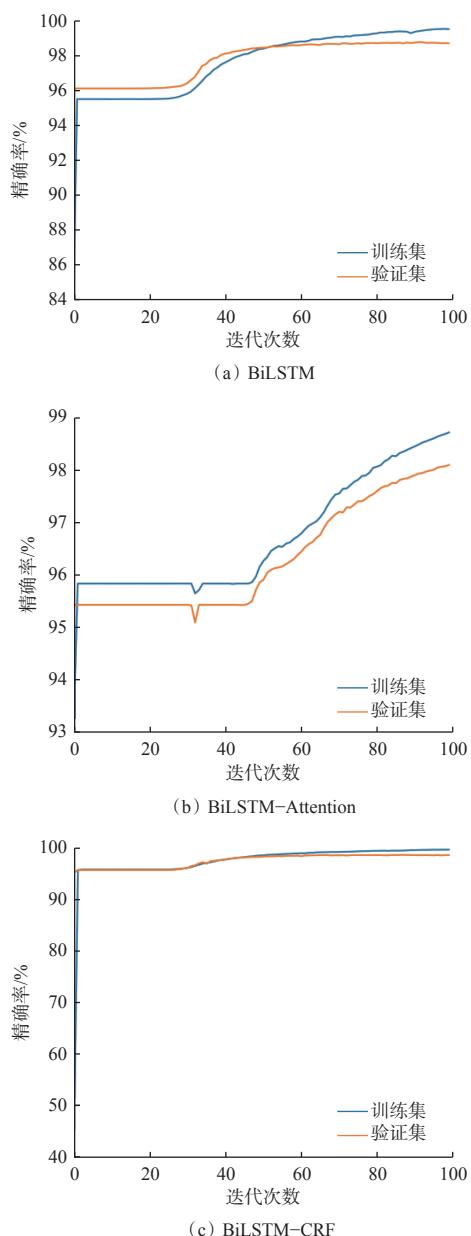


图7 各模型训练、验证过程中精确率变化曲线

Fig. 7 Accuracy curves of each model in training and verification process

3 结语

针对大规模综采设备的故障语料,采用自顶而下的方法对故障原始资料进行分析归类,构建综采设备故障知识本体;综合利用BiLSTM和CRF的优势,构建基于文本数据的BiLSTM-CRF模型,对人工标注的命名实体进行识别;通过Neo4j图数据库对提取的故障知识进行存储,完成综采设备故障知识图谱构建。实验结果表明,相较于BiLSTM模型和BiLSTM-Attention模型,BiLSTM-CRF模型精确率显著提高,为87%, F_1 值也有一定幅度上升,为69%。综采设备故障知识图谱的构建可为大规模、多域综采设备故障数据的有效分析、管理及应用提供支持。

参考文献(References):

- [1] 李梅,杨帅伟,孙振明,等.智慧矿山框架与发展前景研究[J].煤炭科学技术,2017,45(1):121-128,134.
LI Mei, YANG Shuaiwei, SUN Zhenming, et al. Study on framework and development prospects of intelligent mine[J]. Coal Science and Technology, 2017, 45(1): 121-128, 134.
- [2] 王国法,王虹,任怀伟,等.智慧煤矿2025情景目标和发展路径[J].煤炭学报,2018,43(2):295-305.
WANG Guofa, WANG Hong, REN Huawei, et al. 2025 scenarios and development path of intelligent coal mine[J]. Journal of China Coal Society, 2018, 43(2): 295-305.
- [3] 王国法,任世华,庞义辉,等.煤炭工业“十三五”发展成效与“双碳”目标实施路径[J].煤炭科学技术,2021,49(9):1-8.
WANG Guofa, REN Shihua, PANG Yihui, et al. Development achievements of China's coal industry during the 13th Five-Year Plan period and future prospects[J]. Coal Science and Technology, 2021, 49(9): 1-8.
- [4] 李旭,吴雪菲,田野,等.基于云平台的综采设备群远程故障诊断系统[J].工矿自动化,2021,47(7):57-62.
LI Xu, WU Xuefei, TIAN Ye, et al. Remote fault diagnosis system of fully mechanized mining equipment group based on cloud platform[J]. Industry and Mine Automation, 2021, 47(7): 57-62.
- [5] 张旭辉,潘格格,郭欢欢,等.基于深度迁移学习的采煤机摇臂部滚动轴承故障诊断方法[J].煤炭科学技术,2022,50(4):256-263.
ZHANG Xuhui, PAN Gege, GUO Huanhuan, et al. Fault diagnosis method for rolling bearing on shearer arm based on deep transfer learning[J]. Coal Science and Technology, 2022, 50(4): 256-263.
- [6] 聂同攀,曾继炎,程玉杰,等.面向飞机电源系统故障

- 诊断的知识图谱构建技术及应用[J]. 航空学报, 2022, 43(8): 46-62.
- NIE Tongpan, ZENG Jiyan, CHENG Yujie, et al. Knowledge graph construction technology and its application in aircraft power system fault diagnosis[J]. Acta Aeronautica et Astronautica Sinica, 2022, 43(8): 46-62.
- [7] 林凌云, 陈青, 金磊, 等. 基于知识图谱的变电站告警信息故障知识表示研究与应用[J]. 电力系统保护与控制, 2022, 50(12): 90-99.
- LIN Lingyun, CHEN Qing, JIN Lei, et al. Research and application of substation alarm signal fault knowledge representation based on knowledge graph[J]. Power System Protection and Control, 2022, 50(12): 90-99.
- [8] 侯靖琳, 仇润鹤, 薛季爱, 等. 基于知识图谱嵌入和补全的电梯故障预测[J]. 计算机工程与设计, 2022, 43(1): 224-230.
- HOU Jinglin, QIU Runhe, XUE Ji'ai, et al. Elevator failure prediction based on embedding and completion of knowledge graph[J]. Computer Engineering and Design, 2022, 43(1): 224-230.
- [9] 马红兵. 综采工作面电气设备故障处理分析[J]. 内蒙古石油化工, 2021, 47(1): 74-75.
- MA Hongbing. Analysis of the processing in electrical equipment failure at fully mechanized working face[J]. Inner Mongolia Petrochemical Industry, 2021, 47(1): 74-75.
- [10] 胡芳槐. 基于多种数据源的中文知识图谱构建方法研究[D]. 上海: 华东理工大学, 2015.
- HU Fanghuai. Chinese knowledge graph construction method based on multiple data sources[D]. Shanghai: East China University of Science and Technology, 2015.
- [11] 吴玉龙. 综采工作面煤矿机械设备常见故障研究[J]. 科技创新与应用, 2022, 12(29): 162-164, 168.
- WU Yulong. Research on common faults of coal mining machinery and equipment in fully mechanized working face[J]. Technology Innovation and Application, 2022, 12(29): 162-164, 168.
- [12] 王萌, 王昊奋, 李博涵, 等. 新一代知识图谱关键技术综述[J]. 计算机研究与发展, 2022, 59(9): 1947-1965.
- WANG Meng, WANG Haofen, LI Bohan, et al. Survey on key technologies of new generation knowledge graph[J]. Journal of Computer Research and Development, 2022, 59(9): 1947-1965.
- [13] 卢绍帅, 陈龙, 卢光跃, 等. 面向小样本情感分类任务的弱监督对比学习框架[J]. 计算机研究与发展, 2022, 59(9): 2003-2014.
- LU Shaoshuai, CHEN Long, LU Guangyue, et al. Weakly-supervised contrastive learning framework for few-shot sentiment classification tasks[J]. Journal of Computer Research and Development, 2022, 59(9): 2003-2014.
- [14] TONG Fan, LUO Zheheng, ZHAO Dongsheng. A deep network based integrated model for disease named entity recognition[C]. IEEE International Conference on Bioinformatics and Biomedicine, Kansas, 2017: 618-621.
- [15] 金相臣, 吴子锐, 石敏, 等. 基于BiLSTM的地质片段层位预测方法[J]. 高技术通讯, 2021, 31(6): 607-614.
- JIN Xiangchen, WU Zirui, SHI Min, et al. Geological segment horizon prediction method based on BiLSTM[J]. Chinese High Technology Letters, 2021, 31(6): 607-614.
- [16] LEI Jianbo, TANG Buzhou, LU Xueqin, et al. A comprehensive study of named entity recognition in Chinese clinical text[J]. Journal of the American Medical Informatics Association, 2014, 21(5): 808-814.
- [17] 施海昕, 诸建超, 严骏驰, 等. 基于卷积神经网络和LSTM循环神经网络的客户复购预测方法[J]. 高技术通讯, 2021, 31(7): 713-722.
- SHI Haixin, ZHU Jianchao, YAN Junchi, et al. A prediction method of clients' repurchase based on CNN and LSTM RNN[J]. Chinese High Technology Letters, 2021, 31(7): 713-722.
- [18] 周旭峰, 王醒策, 武仲科, 等. 基于组合RNN网络的EMG信号手势识别[J]. 光学精密工程, 2020, 28(2): 424-442.
- ZHOU Xufeng, WANG Xingce, WU Zhongke, et al. Gesture recognition with EMG signals based on ensemble RNN[J]. Optics and Precision Engineering, 2020, 28(2): 424-442.
- [19] 宋雅文, 杨志豪, 罗凌, 等. 基于字符卷积神经网络的生物医学变异实体识别方法[J]. 中文信息学报, 2021, 35(5): 63-69.
- SONG Yawen, YANG Zhihao, LUO Ling, et al. Biomedical mutation entity recognition method based on character convolution neural network[J]. Journal of Chinese Information Processing, 2021, 35(5): 63-69.
- [20] ANGLES R, GUTIERREZ C. Survey of graph database models[J]. ACM Computing Surveys, 2008, 40(1): 1-39.
- [21] 赵志宏, 李晴, 杨绍普, 等. 基于BiLSTM与注意力机制的剩余使用寿命预测研究[J]. 振动与冲击, 2022, 41(6): 44-50, 196.
- ZHAO Zhihong, LI Qing, YANG Shaopu, et al. Remaining useful life prediction based on BiLSTM and attention mechanism[J]. Journal of Vibration and Shock, 2022, 41(6): 44-50, 196.