

基于功能缺陷文本的电力系统二次设备 智能诊断与辅助决策

戴宇欣¹, 张俊¹, 季知祥², 刘明忠³, 高天露¹, 郑永康³, 姚良忠¹

(1. 武汉大学 电气与自动化学院, 湖北 武汉 430072; 2. 中国电力科学研究院有限公司, 北京 100192;

3. 国网四川省电力公司电力科学研究院, 四川 成都 610041)

摘要:利用电力系统二次设备功能缺陷文本数据,建立了基于双向长短时记忆网络与条件随机场(BiLSTM-CRF)模型的文本信息抽取模型。在此基础上,为了进一步将数据中蕴含的知识价值应用到电力系统生产、管理过程中,构建了电力系统二次设备功能缺陷知识图谱,将各类数据间所含语义信息融入各类实体间的关系约束,建立了基于BiLSTM-CRF模型与知识图谱的二次设备功能缺陷智能诊断与辅助决策平台。该平台可依据缺陷设备类型与缺陷现象快速诊断设备的缺陷部位及原因,并推荐合理的解决措施。算例分析结果表明,相较于传统的命名实体识别算法、BiLSTM-softmax以及Seq2Seq-Attention模型,所采用BiLSTM-CRF模型的精确率、召回率、 F_1 值这3项评估指标均有较大提升,所建平台能很好地挖掘、应用电力文本数据知识与价值,为电力系统二次设备功能缺陷处理提供有益参考。

关键词:电力系统;二次设备;信息抽取;知识应用;知识图谱;BiLSTM-CRF;智能诊断;辅助决策

中图分类号:TM 761

文献标志码:A

DOI:10.16081/j.epae.202106006

0 引言

电力系统中积累了大量反映电力设备功能缺陷及处理情况的文本数据,这些数据蕴含了电力设备以往发生的缺陷情况及相应的有效解决措施。长期以来,这些数据未能被有效利用,往往闲置于数据系统中。另外,电力设备功能缺陷情况繁杂,处理缺陷的过程在很大程度上依赖于运维检修人员的专业知识与经验。若能将历史积累的功能缺陷文本加以组织和利用,将各省电力系统中二次设备发生的缺陷及处理情况集成在一个平台中,建立缺陷处理案例库,有助于不同区域运维检修人员进行查询、学习、借鉴,更有效地进行相互间的经验交流,从而更高效地处理运维检修人员自身尚未遇到过但历史数据中存在可借鉴的案例的缺陷情况,也有利于新运维检修人员快速掌握设备运维检修工作,提高自身业务水平,这对电力设备功能缺陷处理工作具有重要的指导意义^[1]。

近年来,由于深度学习等智能算法兴起,自然语言处理技术也取得重要进展。国内外学者对以深度学习为代表的机器学习算法与自然语言处理技术在电力系统中的应用进行了广泛的探讨。文献[2-4]将机器学习算法分别应用于电力系统的不同领域,达到了较好的效果。对于电力文本数据的处理(如简单的文本分类任务),一些传统机器学习算法(如隐马尔可夫模型^[5]、支持向量机^[6]、卷积神经网络^[7]等)均取得了很好的效果。然而,简单的文本分类难

以充分利用电力文本数据蕴含的价值。相较而言,对文本进行信息抽取、知识挖掘更具有实际应用意义。文献[8]针对电力缺陷文本,定义了电力语义框架和语义槽,以提取缺陷信息。该语义框架在处理相对规整的短文本数据方面具有较好的效果,但不便于处理蕴含信息较复杂的长文本数据。文献[9]基于依存句法分析技术,构建了用于电力设备缺陷文本和缺陷分类标准文本的依存句法树,完成了从文本到缺陷信息的辨识。但是,基于规则的方法对各类电力文本进行挖掘需要一定程度上的人为干预。同时,所设计的规则和具体的业务场景具有强相关性,迁移能力较弱。文献[10]引入机器学习算法,对电力服务问答语料集中命名实体进行识别与提取,达到了较好的效果。

知识图谱因其本身具有的图结构特性,可对蕴含语义信息及具有明确逻辑关系的文本数据进行组织与管理^[11]。文献[12]针对电力设备等相关数据建立了知识图谱,在一定程度上提高了设备查询及辨识的工作效率。文献[13]基于知识图谱技术对电力设备缺陷文本进行了组织与利用,但对于实体、属性抽取任务,过程较为简化。

综上所述,为了解决电力文本数据难以有效挖掘与应用的问题,本文首先基于双向长短时记忆网络与条件随机场 BiLSTM-CRF (Bi-directional Long Short-Term Memory and Conditional Random Field) 模型完成缺陷文本信息抽取工作;然后,构建电力系统二次设备知识图谱本体模型,并结合所规整数据,完成知识图谱的构建;最后,基于 BiLSTM-CRF 模型

与知识图谱建立电力系统二次设备缺陷智能诊断与辅助决策平台,并通过算例验证了所建平台的可行性与有效性。

1 电力系统二次设备缺陷记录文本描述

电力系统二次设备是对电力系统内一次设备进行监察、测量、控制、保护、调节的辅助设备,包括安全自动装置、电子式互感器等。保障二次设备的安全可靠运行,在其发生故障时能够进行及时有效的处理,对整个电力系统的安全性、可靠性有着较重要的意义。

在电力系统的运行管理中,记录了二次设备发生缺陷的详细数据,包括“缺陷设备”、“缺陷等级”、“缺陷部位”、“保护类别”、“责任单位”、“保护是否退出”、“缺陷原因”、“原因及处理情况”等数据。一条历史记录数据见附录中表A1。其中,“缺陷设备”、“缺陷部位”、“缺陷原因”等数据为规整的名词性短语,可经简单处理后作为后续知识图谱构建的输入数据。在取得的数据中,通过实体消歧等数据处理工作,并经专家确认后,可整理得到缺陷设备(28种)、缺陷部位(121种)、缺陷原因(48种)标准库,具体见附录中表A2。

“原因及处理情况”为短文本数据,由一线运维检修人员消除缺陷后填写,其描述较为口语化,没有固定的格式与结构,如:“×年×月×日,×时×分,西地站220 kV榆西线2号过压远切装置报开入、开出插件异常,经核实插件已损坏。汇报省调后退出2号过压远切装置。×月×日厂家已将插件寄到西地站,×月×日厂家到现场将插件更换完毕,2号过压远切装置投入运行”。同时,此类文本数据也包含了重要的缺陷信息,如上述短文本数据中包含的缺陷现象为“开入、开出插件异常”,解决措施为“现场将插件更换完毕”等。这类历史数据对后续运维检修人员的工作往往有较大的指导与借鉴意义。

在缺陷文本记录数据中,也记载了设备家族性缺陷情况。如“原因及处理情况”中某条数据记载“通信插件家族性缺陷,×月×日厂家人员来版本升级”,通过溯源该设备批次与责任单位,一线运维检修人员在处理同批次设备时可特别注意该家族性缺陷问题。

因此,基于知识图谱与自然语言处理技术对电力系统二次设备缺陷记录文本数据进行挖掘与应用,构建二次设备功能缺陷智能诊断与辅助决策平台,对电力系统二次设备功能缺陷处理工作具有较大的参考意义,也便于运维检修人员之间的经验交流和自身业务水平的提升。

2 二次设备缺陷智能诊断与辅助决策平台

针对电力系统二次设备缺陷文本数据的挖掘、应用问题,本文构建了基于BiLSTM-CRF模型与知识图谱的二次设备功能缺陷智能诊断与辅助决策平台,其架构如图1所示,包括数据层、本体层、业务层、应用层。其中,数据层基于二次设备的历史缺陷文本数据,引入命名实体识别NER(Named Entity Recognition)模型以完成信息抽取工作,从而为下一步构建知识图谱提供数据支撑;本体层基于Protégé本体构建工具完成实体间的关系构建,以建立二次设备功能缺陷知识图谱本体模型;业务层基于本体层构建的本体模型与数据层提供的数据构建二次设备功能缺陷知识图谱,并提供面向二次设备功能缺陷的信息查询和知识推理业务,以支撑应用层的智能诊断和辅助决策功能。

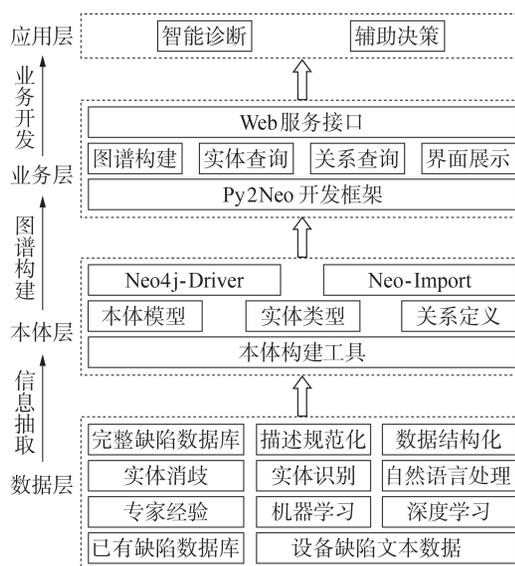


图1 基于BiLSTM-CRF模型与知识图谱的二次设备功能缺陷智能诊断与辅助决策平台

Fig.1 Intelligent diagnosis and auxiliary decision platform of secondary equipment functional defect based on BiLSTM-CRF model and knowledge graph

2.1 NER

在数据层中处理“原因及处理情况”这一类缺陷文本数据时,本文引入NER中的BiLSTM-CRF算法。NER是信息抽取和信息检索中的一项重要任务,其目的是在文本中定位命名实体的边界并将其分类到预定义集合的过程。目前,NER方法主要分为基于规则、字典和在线知识库的方法以及基于智能算法的方法2类^[14]。近年来,深度学习技术已被广泛应用于各种自然语言处理任务中。相比于传统的NER模型,基于深度学习的NER方法无需人工设计规则或特征,易从原始输入文本中提取隐含的语义特征^[15]。

本文以“缺陷现象”为例,虽然一些表征“缺陷现

象”的短语并不是一个名词或者名词性短语,不符合常规的实体定义,但就任务而言,如在“运行灯灭,重启后正常”等文本数据中,提取“运行灯灭”,符合NER模型的基本输入/输出特征,且在取得的缺陷文本数据中,其“缺陷现象”均可用原文本中的某一子序列(或多个子序列,可视为多个实体)加以概括,因此可将从缺陷文本中提取“缺陷现象”信息视作NER任务,并具体针对“缺陷现象”引入深度学习中用于NER模型的BiLSTM-CRF算法模型,对其进行信息抽取。

2.2 知识图谱

知识图谱是Google公司为了支撑其语义搜索而建立的知识库,包含实体、概念及其之间的复杂语义关系^[11]。在完成数据规整工作后,各类数据之间蕴含着丰富的语义关系,例如“缺陷部位”是“缺陷设备”的组成部件,“缺陷现象”是“缺陷设备”发生缺陷的具体表现等。传统的数据存储管理方式是将数据以表格形式存储在数据库中,这样容易丢失数据之间的语义关系,阻碍对数据知识与价值的应用,考虑到知识图谱的技术优势,本文构建了电力系统二次设备功能缺陷知识图谱。

当前,知识图谱主要分为2类:①面向大众的通用领域知识图谱,如卡内基梅隆大学构建的NELL(Never-Ending Language Learner)系统^[16]、复旦大学研发的CN_DBPedia^[17]等;②面向垂直行业的专业领域知识图谱,如医疗领域的“百度医疗大脑”、电商领域的“京东大脑”等。

然而,面向电力系统二次设备的知识图谱尚缺乏相关的研究与探讨。针对电力系统二次设备建立行业型知识图谱,可将各类数据之间丰富的语义信息融入知识图谱中各类实体间的关系约束中,并可基于业务需求,结合专业知识,设定特定的路径约束条件,以在各类应用场景中深度挖掘、应用数据。本文将所建知识图谱应用于二次设备功能缺陷智能诊断与辅助决策这2个具体业务场景。

3 基于BiLSTM-CRF模型的电力系统二次设备功能缺陷文本信息抽取

电力系统二次设备功能缺陷文本由一线运维检修人员填报,难免会出现专业术语表述口语化且不符合常规语法等现象,这增加了文本数据的复杂性及挖掘其潜在知识与价值的难度。数据层的主要任务就是对上述数据进行信息抽取与规范化,以为后续知识图谱的构建提供数据基础。其中,如第1节所述,“缺陷设备”、“缺陷部位”、“缺陷原因”这3类数据为较为规整的名词性短语,可经简单处理后作为后续构建知识图谱模型的输入数据。而关于“缺陷现象”和“解决措施”这2类缺陷信息,其以短文本

数据形式存储在电力系统中,如“断路器失灵保护装置(LFP-923A)自检出错,OP灯熄灭。检修人员到现场检查,更换电源插件后,保护装置恢复正常”,这些文本数据包含某次具体缺陷发生后的缺陷现象以及相应的解决措施,但因其是非结构化的文本数据,需基于自然语言处理技术进行信息抽取,以便得到较为简短、规范化的描述。

因此,本文将从缺陷文本中抽取“缺陷现象”视作NER任务,并引入BiLSTM-CRF算法(对“解决措施”可进行同样的处理,不再赘述)。首先,使用BIO(Begin-Inside-Outside)标注法进行数据标注,其中B表示“缺陷现象”序列的开始,I表示该字符属于“缺陷现象”序列,O表示该字符不属于“缺陷现象”序列;然后,基于BiLSTM-CRF算法进行模型的训练与测试,从缺陷文本中提取“缺陷现象”信息。

基于BiLSTM-CRF的NER模型是目前NER领域使用较为广泛的算法模型。BiLSTM-CRF模型可结合双向长短时记忆(BiLSTM)网络提取文本特征以及CRF处理序列标注任务的优势,完成NER任务^[18]。

其中,长短时记忆(LSTM)网络模型是循环神经网络RNN(Recurrent Neural Network)的变体。LSTM网络模型具有包含遗忘门、输入门、输出门的特殊结构,可解决传统RNN模型不能解决的、处理长序列中存在的梯度爆炸/梯度消失问题^[19]。LSTM网络基本单元示意图见附录中图A1。各状态量的计算公式如式(1)所示。

$$\begin{cases} f_t = \sigma(W_f[h_{t-1}; x_t] + b_f) \\ i_t = \sigma(W_i[h_{t-1}; x_t] + b_i) \\ o_t = \sigma(W_o[h_{t-1}; x_t] + b_o) \\ \tilde{s}_t = \tanh(W_s[h_{t-1}; x_t] + b_s) \\ s_t = f_t \odot s_{t-1} + i_t \odot \tilde{s}_t \\ h_t = o_t \odot \tanh(s_t) \end{cases} \quad (1)$$

其中, $\sigma(\cdot)$ 、 $\tanh(\cdot)$ 为激活函数; h_t 、 s_t 和 h_{t-1} 、 s_{t-1} 分别为当前 t 时刻和 $t-1$ 时刻输出的单元状态、隐藏状态; x_t 为当前 t 时刻的输入; f_t 、 o_t 分别为当前 t 时刻遗忘门、输出门的状态; i_t 、 \tilde{s}_t 为当前 t 时刻输入门的状态; W_f 、 W_i 、 W_o 、 W_s 和 b_f 、 b_i 、 b_o 、 b_s 分别为相应的矩阵权重和偏置项; \odot 表示元素按位相乘。

BiLSTM网络模型是在LSTM网络模型的基础上使用双向架构,使得模型能更加灵活地双向编码序列,这样的模型架构更贴合自然语言特征,可进行上下文特征学习,从而提高自然语言处理任务中的模型性能。

CRF模型用于解决给定输入随机变量 $X = \{X_1, X_2, \dots, X_n\}$ 的情况下,预测输出随机变量 $Y = \{Y_1, Y_2, \dots, Y_n\}$ 的条件分布 $P(Y|X)$ 问题。而线性链

CRF 则主要用于序列标注任务,该模型考虑了与 X_i 对应的标签 Y_i 和上下文标签的约束关系,取得了较好的效果。

BiLSTM-CRF 模型架构如图 2 所示,其主要由输入层、分布式表示层、BiLSTM 网络层(编码器)、CRF 层(解码器)以及输出层构成。在分布式表示层中,由于深度学习模型无法接受非数值型数据作为输入,需要将输入的语句表示为一组向量。在 BiLSTM 网络层与 CRF 层则完成对原始数据的特征提取,最终输出得到目标序列。

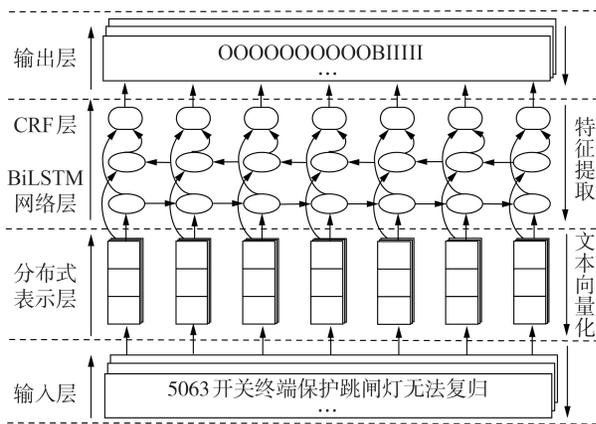


图 2 BiLSTM-CRF 模型架构
Fig.2 Architecture of BiLSTM-CRF model

4 电力系统二次设备功能缺陷知识图谱构建与知识应用:智能诊断与辅助决策

基于当前知识图谱的一般构建流程,结合实际应用的需求,本节在图 1 所示本体层、业务层、应用层中构建与应用二次设备功能缺陷知识图谱。

(1)本体层。本体为知识图谱数据库的抽象概念,本文利用 Protégé 本体构建工具建立二次设备知识图谱的本体模型,如图 3 所示。

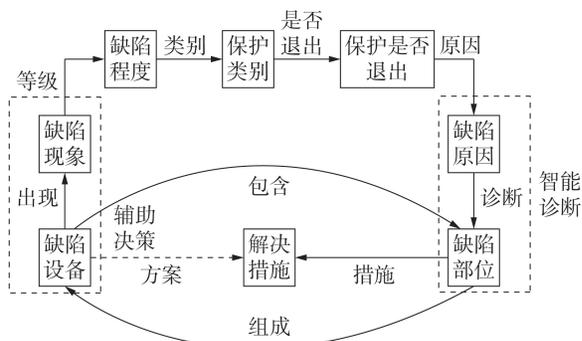


图 3 二次设备知识图谱的本体模型
Fig.3 Ontology model for knowledge graph of secondary equipment

根据数据层的规整数据和应用需求可确定实体类别包含缺陷设备、缺陷现象、缺陷原因、缺陷部位、解决措施 5 类核心实体,以及缺陷程度、保护是否退

出、保护类别 3 类相关实体。进一步地,定义实体间的关系由出现、原因、诊断、措施、等级、是否退出及类别构成。

(2)业务层。对于业务层而言,其基于本体层所构建的本体模型,结合数据层的规整数据,填充具体实体、属性及关系。即:基于图 3 所示本体模型,结合第 3 节中规整的缺陷文本数据,进行本体模型实例化,保护装置本体及安全自动装置的知识图谱子图分别如附录中图 A2 和图 A3 所示。其中,考虑到在处理设备缺陷时不同设备之间具有差异性与独立性,本文针对各二次设备建立了相对独立的子图;否则,各设备之间的缺陷现象、缺陷部位等实体及属性将存在多对多的复杂映射关系,这会导致将针对某一设备的缺陷部位、缺陷原因与解决措施作为智能诊断与辅助决策结果推荐给另一设备。针对各类设备建立子图,大幅降低了知识图谱的结构复杂性,提高了在后续业务功能实现中模型的检索与推理能力。此外,图 A2、A3 中各实体之间的虚线含义为 2 类实体可能存在的相应关系,需要根据实际数据而定。

(3)应用层。在应用层中,本文选用开源 Web 应用框架 Django 作为前端页面开发框架,在框架内集成已训练完成的 BiLSTM-CRF 模型及已构建完成的知识图谱,搭建二次设备功能缺陷智能诊断与辅助决策平台。平台推理过程如图 4 所示。基于该平台,当发生新的缺陷时,通过输入缺陷设备+现象描述,首先进入 BiLSTM-CRF 模型完成缺陷现象信息抽取,再结合所建知识图谱,通过知识推理,返回历史数据中该设备发生相同缺陷现象对应的缺陷部位以及缺陷原因,完成智能诊断功能;进一步地,基于历史数据,推荐合理的解决措施以完成辅助决策功能。在实际应用中,本文会返回历史数据中出现频率最高的 5 条诊断及决策选项,以供运维检修人员选择。

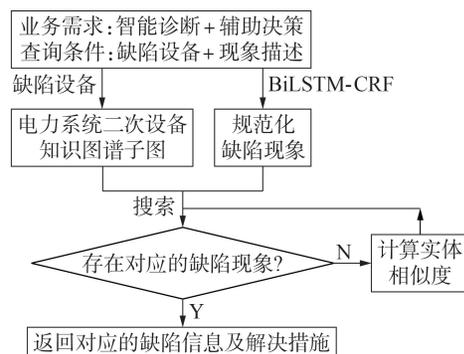


图 4 平台推理过程
Fig.4 Inference process of platform

在将平台部署到某省电网实际应用时,该平台还考虑了以下 2 类具体的特殊情况。

(1)输入现象描述后经过实体识别得到的缺陷现象结合缺陷设备未能在历史数据中找到相对应的案例,如“监控通知,××变××线 rcs-931bm 保护打出通道中断信号,后于××:××自行复归。复归后测量保护背板及光电转换装置收发电平,均在合格范围内”,经实体识别所得的结果为“通道中断信号”,案例库中为“通道中断”。在这种情况下,本文基于 word2vec 技术完成各实体的向量化表达,并计算该现象与各实体间的余弦相似度,以寻找最相似的缺陷现象,推荐该现象对应的缺陷原因、缺陷部位及解决措施。如通过计算“通道中断信号”与各实体间的余弦相似度,并得到与其最相似的现象——“通道中断”,从而推荐“保护装置本体”在发生“通道中断”时对应的缺陷原因、缺陷部位和解决措施。

(2)在实际情况中可能存在设备几个部位均存在异常的情况,即在2.1节提及的现象描述中可能存在多个实体需要识别的情况,具体数据如附录中表A3所示。在这种情况下,当具体进行智能诊断与辅助决策时,本平台会分别依据“缺陷设备”与“缺陷现象1”以及“缺陷设备”与“缺陷现象2”推荐相应的信息。

5 算例分析

本文基于近十年来多省电力公司在对二次设备运维检修时积累的功能缺陷处理文本,经过数据清洗后得到7200条历史数据,搭建基于BiLSTM-CRF模型的电力系统二次设备缺陷现象信息抽取模型,将上述数据以8:2的比例划分为训练集与测试集,以分别进行模型的训练与测试;并基于上述数据构建电力系统二次设备功能缺陷知识图谱,从而建立基于BiLSTM-CRF模型与知识图谱的电力系统二次设备功能缺陷智能诊断与辅助决策平台。所建平台的开发环境配置见附录中表A4。

5.1 电力系统二次设备功能缺陷文本的信息抽取

为了在二次设备功能缺陷文本中抽取缺陷现象信息,本文构建了基于BiLSTM-CRF的信息抽取模型。BiLSTM-CRF模型及对照组的相关超参数设置如表1所示。表中, λ 为 L_2 正则化系数;Optimizer表示模型求解优化器(Adam为自适应矩估计算法,SGD为随机梯度下降算法); $N_{\text{Hidden_dim}}$ 为BiLSTM网络隐藏层单元数量; N_{Epoch} 为训练轮数; $N_{\text{Batch_size}}$ 为每一批次训练的样本数量; $\gamma_{\text{Learning_rate}}$ 为模型学习率; $N_{\text{Embedding_dim}}$ 为分布式表示层中将字转换为数值向量的向量维度; γ_{Dropout} 为模型中神经元的随机丢弃率。上述参数均在不同的层面影响模型的最终效果。

本文选取式(2)~(4)所示的3个指标(精确率 P 、召回率 R 、 F_1 值)评估模型性能,4个对照组模型性能结果如表2所示。

表1 模型超参数设定

Table 1 Hyperparameter setting of model

超参数	取值			
	对照组1	对照组2	对照组3	对照组4
λ	0.0001	0.001	0	0.0001
Optimizer	Adam	Adam	Adam	SGD
$N_{\text{Hidden_dim}}$	300	300	300	300
N_{Epoch}	30	30	30	30
$N_{\text{Batch_size}}$	32	32	32	32
$\gamma_{\text{Learning_rate}}$	0.001	0.001	0.001	0.001
$N_{\text{Embedding_dim}}$	100	100	100	100
γ_{Dropout}	0.4	0.4	0.4	0.4

$$P = \frac{T_p}{T_p + F_p} T_p \quad (2)$$

$$R = \frac{T_p}{T_p + F_N} \quad (3)$$

$$F_1 = \frac{2PR}{P + R} \quad (4)$$

其中, T_p 为模型正确识别缺陷现象的样本数量; F_p 为能识别出缺陷现象但与标签不一致的样本数量; F_N 为应该能识别出缺陷现象但没有识别出的样本数量。

表2 对照组模型性能

Table 2 Model performance of control groups

对照组	精确率 / %		召回率 / %		F_1 值	
	训练集	测试集	训练集	测试集	训练集	测试集
1	90.04	87.12	90.38	86.46	0.9021	0.8679
2	83.80	81.63	85.95	82.71	0.8486	0.8217
3	91.61	84.82	91.82	84.58	0.9171	0.8470
4	44.60	48.40	23.72	6.84	0.3097	0.1199

由表2可知,对照组1的模型性能在3个评估指标中均优于其他3个对照组模型。其中,相比于对照组1,对照组2将 λ 增大10倍,对照组3将 λ 设置为0,即损失函数中没有正则化项。经计算,在模型训练结束后,对照组2的模型损失函数值稳定在1.36附近,对照组3的模型损失函数值稳定在0.45附近,结合表2所示结果可知,对照组2模型欠拟合,对照组3模型过拟合。相比于对照组1,对照组4将求解优化器由Adam更改为SGD,表2所示结果表明Adam的求解结果更优。根据上述模型参数调优过程结果,本文最终选取最优参数为对照组1对应参数。

比较本文所提BiLSTM-CRF模型(对照组1)与传统CRF算法、BiLSTM-softmax及自然语言生成模型中的Seq2Seq-Attention模型在测试集的精确率、召回率、 F_1 值评估指标,结果如表3所示。

由表3可知,本文所提BiLSTM-CRF模型在精确率(87.12%)、召回率(86.46%)、 F_1 值(0.8679)方面均优于CRF、BiLSTM-softmax、Seq2Seq-Attention模型。此外,文献[10]在利用CRF算法对电力服务问答语

表 3 不同模型的测试性能对比
Table 3 Test performance comparison among different models

模型	精确率 / %	召回率 / %	F_1 值
CRF	70.01	70.42	0.7042
BiLSTM-softmax	78.71	71.88	0.7514
Seq2Seq-Attention	71.53	—	—
BiLSTM-CRF	87.12	86.46	0.8679

料数据集中位置、故障、解决方案 3 类实体进行识别与提取时,达到了 90% 以上的精确率。然而,如 2.1 节所述,本文中所抽取的缺陷现象并不是典型的命名实体,而是较为复杂的短语类文本,抽取难度更大,因此本文模型的精确率低于文献[10]中模型。事实上,在本文的对比模型中,考虑了文献[10]中所提 CRF 模型,该模型在本文所使用的数据集上的测试精确率仅为 70.01%,这也验证了 BiLSTM-CRF 模型的优越性。

5.2 电力系统二次设备功能缺陷的智能诊断与辅助决策

完成文本信息抽取工作后,结合所建知识图谱,本文建立了二次设备功能缺陷智能诊断与辅助决策平台,下面以 2 个实例详细展示该平台智能诊断与辅助决策功能的实现过程。输入检索条件如表 4 所示。

表 4 输入检索条件
Table 4 Enter retrieval conditions

输入检索条件	属性	输入内容
1	缺陷设备	保护装置本体
	现象描述	1号主变失灵装置运行灯灭
2	缺陷设备	二次回路及辅助继电器 (含对时回路和通信回路)
	现象描述	海里 4363 线第二套线路保护 RCS931A 通信中断

当输入检索条件 1 时,经已训练完成的 BiLSTM-CRF 模型,抽取缺陷现象为“运行灯灭”,然后结合推理条件:缺陷设备为“保护装置本体”,缺陷现象为“运行灯灭”,依据所建知识图谱,平台依据历史数据最终返回 5 条可能的诊断结果与解决措施,如附录中图 A4(a)所示。当输入检索条件 2 时,智能诊断与辅助决策结果如附录中图 A4(b)所示。如图 A4 所示,该平台针对具体缺陷现象,依据历史实际数据推荐以往的解决措施方案,具有较大的可信度与参考意义。在实际应用中,由运维检修人员参考所提供的建议,选择执行具体的某一条措施,并将结果反馈给平台,以便在后续工作中优化平台性能。

可以看到,图 A4(a)与(b)在诊断出的可能缺陷原因中均有“内部通信中断”这一项。本文在第 4 节的业务层构建流程中提出针对不同缺陷设备建立相

对独立的子图方法后,平台在推理时会针对“保护装置本体”对应的缺陷原因为“内部通信中断”的解决措施——“更换 CPU 和 MON 板”展示在图 A4(a),而不是图 A4(b)所示的“更换网线”或“重新做水晶头”。

本文所建平台完成了电力系统二次设备功能缺陷文本从信息抽取到知识应用的全过程,所提技术框架有较好的可实施性与可移植性。最后,该平台的实际应用效果在一定程度上依赖于原始文本数据的质量。例如:图 A4(a)推荐建议中的“更换”一项,并未提及更换何种器件,这是原始数据信息记录不全所导致的问题;图 A4(b)中仅提供了 3 条建议,这是原始数据中针对缺陷设备——“二次回路及辅助继电器(含对时回路和通信回路)”数据量不足所导致的问题。为此,本文在将平台部署到某省电力公司的具体应用时,会根据不断积累的缺陷文本数据,迭代优化平台性能,以不断提高平台对二次设备功能缺陷智能诊断与辅助决策的建议采纳率。

6 结论

(1)本文基于 BiLSTM-CRF 建立了电力系统文本信息抽取模型,引入精确率、召回率、 F_1 值 3 个评估指标,将所提模型与 CRF 等模型进行比较,算例结果表明:相较于 CRF 等模型,本文模型的各项指标均有所提升,可为后续电力系统知识图谱构建提供数据基础,并为电力系统文本信息抽取、规范化处理等任务提供参考。

(2)基于 Neo4j 图数据库与 Django 框架等工具,本文构建了电力系统二次设备功能缺陷知识图谱,当发生新的缺陷时,可依据检索条件完成二次设备功能缺陷智能诊断与辅助决策功能,从而提高电力系统对于二次设备的检修效率,缩短消缺时间,进而保障电网更加安全、可靠地运行。

(3)在对比试验中,Seq2Seq-Attention 事实上是一个序列到序列的“生成”模型,模型输出较为灵活,使得模型具有对文本进行抽象、总结的能力,该类模型是下一步对电力文本深度挖掘的研究重点。

(4)针对设备家族性缺陷问题,若一线运维检修人员发现了某设备缺陷是由设备家族性缺陷导致并将其记录,本平台可进行同批次设备溯源,进行相应处理。但若在处理过程中,未发现该缺陷是由家族性缺陷导致,则需进一步针对同批次设备所链接的知识图谱拓扑进行相应的数据挖掘,如拓扑相似性度量、节点聚类,以识别所建知识图谱中隐藏的设备家族性缺陷,该问题是下一步研究的难点。

下一步的研究工作将着重于自然语言处理技术在电力系统中的应用探索,以提高电力文本数据挖掘

掘的质量与效率。且针对上述问题,基于不断积累的历史数据继续提高BiLSTM-CRF模型的性能,完善所建知识图谱,并探索更高效的搜索算法,使得模型能更加智能、更贴合实际地诊断设备缺陷,为缺陷消除提供更加可靠、合理的建议。

附录见本刊网络版(<http://www.epae.cn>)。

参考文献:

- [1] 杜修明,秦佳峰,郭诗瑶,等. 电力设备典型故障案例的文本挖掘[J]. 高电压技术,2018,44(4):1078-1084.
DU Xiuming, QIN Jiafeng, GUO Shiyao, et al. Text mining of typical defects in power equipment[J]. High Voltage Engineering, 2018, 44(4): 1078-1084.
- [2] 王桂兰,赵洪山,米增强. XGBoost算法在风机主轴轴承故障预测中的应用[J]. 电力自动化设备,2019,39(1):73-77,83.
WANG Guilian, ZHAO Hongshan, MI Zengqiang. Application of XGBoost algorithm in prediction of wind motor main bearing fault[J]. Electric Power Automation Equipment, 2019, 39(1): 73-77, 83.
- [3] 谢小瑜,周俊煌,张勇军. 深度学习在泛在电力物联网中的应用与挑战[J]. 电力自动化设备,2020,40(4):77-87.
XIE Xiaoyu, ZHOU Junhuang, ZHANG Yongjun. Application and challenge of deep learning in Ubiquitous Power Internet of Things[J]. Electric Power Automation Equipment, 2020, 40(4): 77-87.
- [4] 赵文清,沈哲吉,李刚. 基于深度学习的用户异常用电模式检测[J]. 电力自动化设备,2018,38(9):34-38.
ZHAO Wenqing, SHEN Zheji, LI Gang. Anomaly detection for power consumption pattern based on deep learning[J]. Electric Power Automation Equipment, 2018, 38(9): 34-38.
- [5] 邱剑,王慧芳,应高亮,等. 文本信息挖掘技术及其在断路器全寿命状态评价中的应用[J]. 电力系统自动化,2016,40(6):107-112,118.
QIU Jian, WANG Huifang, YING Gaoliang, et al. Text mining technique and application of lifecycle condition assessment for circuit breaker[J]. Automation of Electric Power Systems, 2016, 40(6): 107-112, 118.
- [6] 汪崔洋,江全元,唐雅洁,等. 基于告警信号文本挖掘的电力调度故障诊断[J]. 电力自动化设备,2019,39(4):126-132.
WANG Cuiyang, JIANG Quanyuan, TANG Yajie, et al. Fault diagnosis of power dispatching based on alarm signal text mining[J]. Electric Power Automation Equipment, 2019, 39(4): 126-132.
- [7] 孙国强,沈培锋,赵扬,等. 融合知识库和深度学习的电网监控告警事件智能识别[J]. 电力自动化设备,2020,40(4):40-47.
SUN Guoqiang, SHEN Peifeng, ZHAO Yang, et al. Intelligent recognition of power grid monitoring alarm event combining knowledge base and deep learning[J]. Electric Power Automation Equipment, 2020, 40(4): 40-47.
- [8] 曹靖,陈陆燊,邱剑,等. 基于语义框架的电网缺陷文本挖掘技术及其应用[J]. 电网技术,2017,41(2):637-643.
CAO Jing, CHEN Lushen, QIU Jian, et al. Semantic framework-based defect text mining technique and application in power grid[J]. Power System Technology, 2017, 41(2): 637-643.
- [9] 邵冠宇,王慧芳,吴向宏,等. 基于依存句法分析的电力设备缺陷文本信息精确辨识方法[J]. 电力系统自动化,2020,44(12):178-185.
SHAO Guanyu, WANG Huifang, WU Xianghong, et al. Precise information identification method of power equipment defect text based on dependency parsing technology[J]. Automation of Electric Power Systems, 2020, 44(12): 178-185.
- [10] 杨维,孙德艳,张晓慧,等. 面向电力智能问答系统的命名实体识别算法[J]. 计算机工程与设计,2019,40(12):3625-3630.
YANG Wei, SUN Deyan, ZHANG Xiaohui, et al. Named entity recognition for intelligent answer system in power service[J]. Computer Engineering and Design, 2019, 40(12): 3625-3630.
- [11] 刘峤,李杨,段宏,等. 知识图谱构建技术综述[J]. 计算机研究与发展,2016,53(3):582-600.
LIU Qiao, LI Yang, DUAN Hong, et al. Knowledge graph construction techniques[J]. Journal of Computer Research and Development, 2016, 53(3): 582-600.
- [12] TANG Y C, LIU T T, LIU G Y, et al. Enhancement of power equipment management using knowledge graph[C]//2019 IEEE Innovative Smart Grid Technologies-Asia(ISGT Asia). Chengdu, China: IEEE, 2019: 905-910.
- [13] 刘梓权,王慧芳. 基于知识图谱技术的电力设备缺陷记录检索方法[J]. 电力系统自动化,2018,42(14):158-164.
LIU Ziquan, WANG Huifang. Retrieval method for defect records of power equipment based on knowledge graph technology[J]. Automation of Electric Power Systems, 2018, 42(14): 158-164.
- [14] 刘浏,王东波. 命名实体识别研究综述[J]. 情报学报,2018,37(3):329-340.
LIU Liu, WANG Dongbo. A review on named entity recognition[J]. Journal of the China Society for Scientific and Technical Information, 2018, 37(3): 329-340.
- [15] 张俊遥. 基于深度学习的中文命名实体识别研究[D]. 北京:北京邮电大学,2019.
ZHANG Junyao. Research on Chinese named entity recognition based on deep learning[D]. Beijing: Beijing University of Posts and Telecom, 2019.
- [16] MITCHELL T, COHEN W, HRUSCHKA E, et al. Never-ending learning[J]. Communications of the ACM, 2018, 61(5): 103-115.
- [17] XU B, XU Y, LIANG J Q, et al. CN-DBpedia: a never-ending Chinese knowledge extraction system[C]//International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems. Arras, France: Springer, 2017: 428-438.
- [18] 顾溢. 基于BiLSTM-CRF的复杂中文命名实体识别研究[D]. 南京:南京大学,2019.
GU Yi. Research on complex Chinese named entity recognition based on BiLSTM-CRF[D]. Nanjing: Nanjing University, 2019.
- [19] HUANG Z, WEI X, KAI Y. Bidirectional LSTM-CRF models for sequence tagging[EB/OL]. [2020-07-19]. <https://arxiv.org/pdf/1508.01991.pdf>.

作者简介:



戴宇欣

戴宇欣(1996—),男,安徽合肥人,硕士研究生,主要研究方向为人工智能技术在电力系统中的应用(**E-mail**: yuxindai@whu.edu.cn);

张俊(1981—),男,湖北武汉人,教授,博士研究生导师,博士,主要研究方向为智能系统、人工智能、知识自动化及其在智能电力和能源系统中的应用(**E-mail**: jun.zhang.ee@whu.edu.cn)。

(编辑 陆丹)

Intelligent diagnosis and auxiliary decision of power system secondary equipment based on functional defect text

DAI Yuxin¹,ZHANG Jun¹,JI Zhixiang²,LIU Mingzhong³,GAO Tianlu¹,
ZHENG Yongkang³,YAO Liangzhong¹

(1. School of Electrical Engineering and Automation,Wuhan University,Wuhan 430072,China;

2. China Electric Power Research Institute,Beijing 100192,China;

3. State Grid Sichuan Electric Power Research Institute,Chengdu 610041,China)

Abstract: A text information extraction model based on BiLSTM-CRF(Bi-directional Long Short-Term Memory and Conditional Random Field) model is established by using the functional defect text of power system secondary equipment. On this basis,in order to further apply the knowledge value contained in the data to the production and management process of power system,the knowledge graph for functional defects of power system secondary equipment is constructed,which can integrate the semantic information contained in various types of data into the relationship constraints among various types of entity. And an intelligent diagnosis and auxiliary decision platform for functional defects of secondary equipment based on BiLSTM-CRF model and knowledge graph is established. The platform can quickly diagnose the defective parts and causes of the equipment according to the type and phenomenon of defective equipment,and then recommend reasonable solutions. The numerical example analysis results show that,compared with the traditional named entity recognition algorithm,BiLSTM-Softmax model and Seq2Seq-Attention model,the evaluation indexes of accurate rate,recall rate and F_1 value for BiLSTM-CRF model are greatly improved,and the established platform can well mine and apply the knowledge and value of power text data,providing a useful reference for processing the functional defects of power system secondary equipment.

Key words: electric power systems;secondary equipment;information extraction;knowledge application;knowledge graph;BiLSTM-CRF;intelligent diagnosis;auxiliary decision

(上接第 183 页 continued from page 183)

Imperfect maintenance strategy of transformer based on proportional strength model

ZHAO Hongshan,LI Yuhao

(School of Electrical and Electronic Engineering,North China Electric Power University,Baoding 071003,China)

Abstract: Taking the practical problem of incomplete maintenance into the consideration of transformers' state maintenance strategy,a state maintenance strategy based on proportional strength model is proposed. Firstly,the proportional strength model of transformer is established,and the detailed calculation model of incomplete maintenance parameters is given. The sensitivity analysis of time parameters and maintenance parameters in the calculation process is carried out to determine the reasonable value range and clarify the maintenance effect of incomplete maintenance activities. Then,the state maintenance strategy of transformer is determined. The minimum cost method is used to determine the state maintenance threshold curve of the proportional strength model,and the lower threshold value is added as a restrictive indicator for the state maintenance strategy. Finally,the actual fault statistics and state monitoring data of transformer are used for example analysis. The effectiveness of the improved proportional strength model in the case of incomplete maintenance is analyzed respectively from three aspects of maintenance cost,maintenance frequency and maintenance time,which provides a reliable theoretical basis for transformer's state maintenance decision.

Key words: electric transformers;state maintenance;imperfect maintenance;proportional strength model

附录

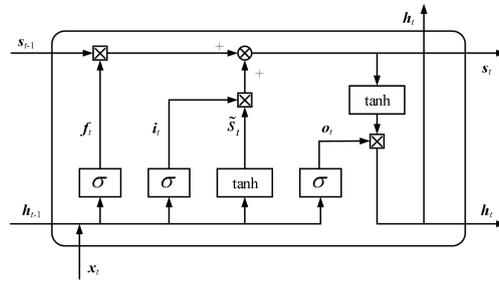


图 A1 LSTM 网络的基本单元
Fig.A1 Basic unit of LSTM network

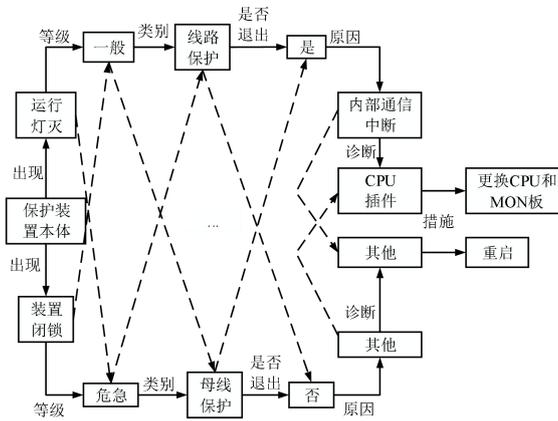


图 A2 保护装置子图
Fig.A2 Subgraph of protection device

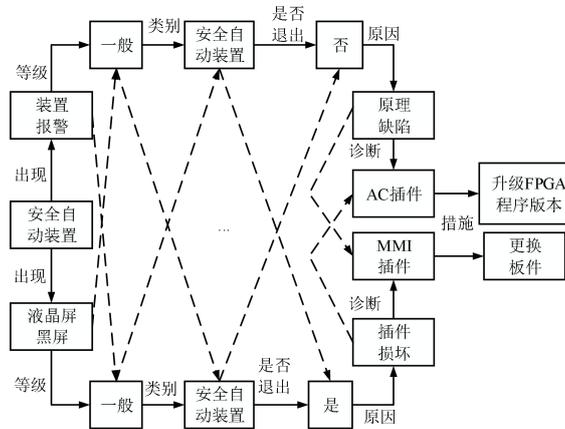


图 A3 安全自动装置子图
Fig.A3 Subgraph of automatic safety device

二次设备缺陷知识图谱	智能诊断与辅助决策	实体及关系查询
主页 智能诊断		
查询条件:		
<input type="button" value="保护装置本体"/>	<input type="button" value="1号主变失灵装置运行灯灭"/>	<input type="button" value="诊断"/>
诊断结果:		
缺陷部位	缺陷原因	解决措施
CPU插件	插件损坏	更换
电源插件	插件损坏	更换电源插件
A/D(VFC)插件	插件损坏	重启
MMI插件	插件损坏	更换管理板
CPU插件	内部通信中断	更换CPU和MON板

(a) 检索条件 1

二次设备缺陷知识图谱	智能诊断与辅助决策	实体及关系查询
主页 智能诊断		
查询条件:		
<input type="button" value="二次回路及辅助继"/>	<input type="button" value="海里4363线第二套线路保"/>	<input type="button" value="诊断"/>
诊断结果:		
缺陷部位	缺陷原因	解决措施
通信回路	内部通信中断	更换网线
通信回路	内部通信中断	重新做水晶头
信号回路	其他	更换电源板

(b) 检索条件 2

图 A4 智能诊断与辅助决策结果展示

Fig.A4 Intelligent diagnosis and auxiliary decision results

表 A1 二次设备缺陷记录

Table A1 Defect recording of secondary equipment

属性	内容	属性	内容
缺陷设备	保护装置本体	责任单位	××××
缺陷等级	严重	保护是否退出	是
缺陷部位	CPU 插件	缺陷原因	装置死机
保护类别	线路保护	原因及处理情况	OP 灯灭, 重启后恢复正常

表 A2 缺陷属性标准库 (部分)

Table A2 Standard library of defect attributes

缺陷属性	数量/种	样例
缺陷设备	28	故障录波器, 操作箱, ...
缺陷部位	121	CPU 插件, 通信插件, ...
缺陷原因	48	操作不当, 光纤折断, ...

表 A3 缺陷记录分析

Table A3 Defect record analysis

缺陷记录	缺陷现象 1	缺陷现象 2
220kV 旁路 290 开关 LFP-902A 超高压线路成套保护装置死机, 液晶无显示。更换内部插件后缺陷消除	装置死机	液晶无显示
# 主变 1 号保护装置现场告警灯亮, 液晶面板显示“召唤配置无应答”。站端申请退出 1 号保护重启后装置恢复正常运行。二次人员对装置通信板件进行升级后缺陷情况消失, 装置正常运行	告警灯亮	液晶面板显示“召唤配置无应答”

表 A4 平台开发环境配置

Table A4 Platform development environment configuration

环境	版本	备注
Python	3.6.5	开发语言
Tensorflow	1.13	深度学习开发框架
Neo4j	3.3.1	图数据库
Django	2.0.3	Web 服务器
Java	1.8	JDK
CPU	—	Intel(R)Core(TM)i5-7200U CPU @2.50GHz
RAM	—	8.00GB