

电网故障处置知识图谱构建研究与应用

郭榕¹, 杨群¹, 刘绍翰¹, 李伟^{2,3}, 袁鑫¹, 黄香鸿¹

(1. 南京航空航天大学 计算机科学与技术学院, 江苏省 南京市 211106;

2. 智能电网保护和运行控制国家重点实验室, 江苏省 南京市 211106;

3. 南瑞集团(国网电力科学研究院)有限公司, 江苏省 南京市 211106)

Construction and Application of Power Grid Fault Handling Knowledge Graph

GUO Rong¹, YANG Qun¹, LIU Shaohan¹, LI Wei^{2,3}, YUAN Xin¹, HUANG Xianghong¹

(1. College of Computer Science & Technology, Nanjing University of Aeronautics & Astronautics, Nanjing 211106, Jiangsu Province, China;

2. State Key Laboratory of Smart Grid Protection and Control, Nanjing 211106, Jiangsu Province, China;

3. NARI Group Corporation (State Grid Electric Power Research Institute), Nanjing 211106, Jiangsu Province, China)

ABSTRACT: The fault handling pre-plan of power grid has great instructive significance for the quick emergency disposal when the failures or accidents occur. To assist dispatchers in fault handling and improve the capabilities of power grid's emergency handling and the level of dispatch intelligence, the technology of knowledge graph can be used to extract, represent, and manage fault handling pre-plan. By exploring the fault handling pre-plan of the power grid, this paper proposes a new method to construct the knowledge graph for the fault handling. The method combines both the top-down and bottom-up construction strategies and solves the involved problem of the knowledge extraction for the power domain. Firstly, the scheme layer of the knowledge graph in the top-down style is designed, which defines the knowledge framework, the concept types, and the relationships between the concepts of the knowledge graph. Then, according to the characteristics of the pre-plan text, multiple deep learning models are comprehensively used for knowledge extraction, and build the data layer of the knowledge graph in the bottom-up style. To avoid word segmentation errors, the TextCNN model is used based on character-level vectors to classify the text content of the plan. For the word conflict, the LR-CNN model is applied to identify domain named entities. On the basis of named entity recognition, the BiGRU-Attention model is adopted to extract the relationships between the entities. Finally, the effectiveness of the above-mentioned knowledge extraction method is verified through experiments. The constructed power grid fault handling knowledge graph is visualized and its application in intelligent information retrieval and auxiliary fault diagnosis is analyzed.

KEY WORDS: grid fault handling; knowledge graph; knowledge extraction; deep learning

摘要: 电网故障处置预案对电网事故应急工作的高效、有序进行有着重要的指导意义。利用知识图谱技术对故障处置信息进行知识抽取、表示和管理,并用于辅助调度人员进行故障处置,可有效提升电网应急处理能力与调度智能化水平。以电网故障处置预案文本为研究对象,提出了一种自顶向下和自底向上相结合的电网故障处置知识图谱构建方法,并解决了其中涉及的电力领域知识抽取问题。首先,自顶向下定义知识图谱的知识组织架构、概念类型、概念间关系,形成知识图谱的模式层;之后,针对电网故障处置预案文本的特性,综合使用多种深度学习模型进行知识抽取,自底向上构建知识图谱的数据层;为避免分词错误,使用基于字向量的TextCNN模型对预案文本进行分类;为解决候选词冲突问题,使用LR-CNN模型对电力领域的命名实体进行识别;在命名实体识别的基础上,使用BiGRU-Attention模型对实体间关系进行抽取。之后,通过实验验证了上述知识抽取方法的有效性。最后,对构建的电网故障处置知识图谱进行了可视化并对其在智能信息检索和辅助故障诊断中的应用进行了分析。

关键词: 电网故障处置; 知识图谱; 知识抽取; 深度学习

DOI: 10.13335/j.1000-3673.pst.2021.0065

0 引言

目前,电网故障后的处置调度工作主要依赖于调度人员的主观性决策,由调度人员实时分析故障后电网的状态、参数变化情况,查明故障发生的原因并制定相应的故障处置措施^[1]。这种处置方式需要调度人员反复查阅与记忆大量以非(半)结构化文本形式存在的故障处置信息,如系统稳定要求、故障后运行方式、故障处置要点等。以关键字进行匹配的传统文本检索方法虽然能够提供段落定位功

基金项目: 江苏省重点研发计划项目(BE2019012)。

Project Supported by Key R&D Program of Jiangsu Province (BE2019012).

能,但是检索结果缺乏碎片化、缺乏组织,经常出现检索不全、答非所问的情况,容易产生疏忽遗漏,使得故障应急处置工作的效率降低。随着电力系统的快速发展,电网结构和运行模式愈加复杂,故障后的处置难度不断提高,依赖于人工经验的传统调度决策机制越来越难以应对复杂大电网的快速故障分析和故障处置^[2]。电力系统亟需借助智能化技术将非结构化的故障处置文本数据抽取提炼为知识,并将这些知识组织成结构化、可视化的表示形式。在电网故障发生时,帮助调度员快速分析事故原因,全面地掌握故障处理的关键信息,并进行辅助决策,以提高电网的应急处置能力。

知识图谱(knowledge graph)是 Google 公司于 2012 年提出的一种知识表示方法,在本质上是一个结构化语义知识库,它以三元组(头实体,关系,尾实体)的形式对客观世界中的实体及其相互关系进行建模,这些三元组通过共有的实体或属性相互连接,构成网状的知识结构^[3-4]。相较于传统的知识组织、管理方式,知识图谱基于图的数据组织结构支持更高效的数据调取,能够处理复杂多样的关联表示,能够模拟人类思考过程进行语义分析。因此,通过构建电网故障处置的领域知识图谱,以图的形式对故障处置知识进行组织和存储,并将其作为机器理解人类知识的载体,利用计算机进行语义搜索与辅助决策,为故障处置工作提供智能化的信息服务与应用。

目前,关于知识图谱在电力领域中的应用,国内外已经进行了许多的探索。文献[5]为提高电力设备的管理效率,以多源异构的电力设备数据为对象,构建了电力设备知识图谱。文献[6]分析了电网提供能源服务的模式,构建了能源服务知识图谱,集成与能源服务相关的各种领域知识。文献[7]分析了现有的螺栓缺陷分类方法的不足,使用 GGNN 构建栓母对知识图谱,指导栓母对的缺陷分类工作。文献[8]以电力设备缺陷记录文本为基础,构建了电力设备缺陷知识图谱,用于类似故障记录的检索。

然而,在电网故障处置领域,相关的研究还较少,文献[1]提出了电网故障处理知识图谱的应用框架,并对其中关键环节的实施技术进行分析,但并未实际利用电网故障处置数据进行知识图谱的构建。文献[9]构建了线路故障处置知识图谱,但知识抽取的方式依赖于手工与规则模板。

综上所述,虽然知识图谱在电力领域快速发展,但针对电网故障处置的知识图谱构建及其应用研究尚未深入开展。因此,本文以电网故障处置预案文本为对象,设计了领域知识图谱的构建框架。

其中,利用深度学习方法进行知识抽取,以减少对规则模板的需求。将抽取的知识利用 Neo4j 图数据库进行存储和可视化表示,实现一个用于电网故障处置信息智能检索和辅助故障诊断的知识图谱。

1 电网故障处置预案

1.1 电网故障处置预案文本内容

为了提高电网的故障应急处置能力,保证故障应急处置工作高效、有序地进行,最大限度减少事故对社会的影响,电力企业及有关部门针对电网运行中可能发生的故障进行分析,模拟故障后电网运行方式的变化情况,确定预想故障的大致影响范围和电网薄弱环节,通过调度经验针对性地制定故障处置措施,形成文本格式的电网故障处置预案(以下简称为预案)。

预案文本由预案编号、故障名称、稳定要求、故障后运行方式、故障处置要点 5 部分内容组成。其中,故障后方式描述了预想故障发生后电网运行方式的变更情况及原因;处置要点描述了电网恢复正常运行所需采取的处置措施和紧急状态调整;稳定要求描述了故障处置工作需要遵守的稳定规定。

1.2 电网故障处置预案文本特点

相较于日常领域的文本,预案文本具有以下的特点:

1) 预案文本的构成复杂,类型多样。其中既包含了结构化的信息,如预案编号、故障名称、稳定要求,又包含了非结构化信息,如故障后运行方式、故障处置要点。并且故障后运行方式和故障处置要点又可根据描述对象、处置措施进一步细分。

2) 预案文本包含了大量电力领域的专有名词和专业术语,例如厂站设备、电力调度机构、调度指令、设备指标等。一些适用于日常领域的自然语言处理技术,如文本分词、词性标注等,难以直接应用于电网故障处置预案文本中。

3) 预案文本中的实体往往由地区、厂站、电压等级等多个名词嵌套构成,实体的界限模糊。

本文在构建电网故障处置知识图谱时,充分考虑了上述特点,采取有针对性的文本预处理方法,并构建了一套适用于电力领域的知识抽取模型,实现预案文本从传统的基于手工和规则模板的知识抽取方法到基于深度学习的智能化知识抽取方法的转变。

2 电网故障处置知识图谱的构建方法

2.1 构建框架

知识图谱的构建方式通常可分为自顶向下、自

底向上和混合方式 3 种^[10]。自顶向下的构建方式先构建知识图谱的模式(scheme)层, 预先定义知识图谱的知识组织架构, 之后在模式层的指导下利用一系列知识抽取方法从数据源中抽取知识实例添加到知识库中。与之相反, 自底向上的构建方式没有预先定义知识图谱的模式层, 而是从数据源中归纳模式层, 在知识抽取的过程中逐步形成并不断地更新概念与概念间关系的组织架构。两者相结合的混合构建方式则是在初期有预定义的模式层, 同时随着知识抽取的进行, 根据数据源的知识组织结构对模式层进行改进更新, 形成置信度更高的知识组织架构, 使知识图谱更加完备和可靠。

由于预案文本的核心构成要素相对固定, 但核心要素可进一步细分为类型多样的非结构化信息。因此, 本文采用自顶向下与自底向上相结合的方式构建电网故障处置知识图谱。首先, 通过分析预案文本的内容, 采用自顶向下的方式设计知识图谱的模式层。之后, 在模式层的指导下, 采用自底向上的方式构建数据层, 针对预案文本的特点设计合适的抽取方法, 进行实体、关系和属性 3 个知识要素的抽取, 形成一系列高质量的事实表达, 通过知识图谱的底层存储方式, 映射到相关概念节点当中。其构建流程如图 1 所示。

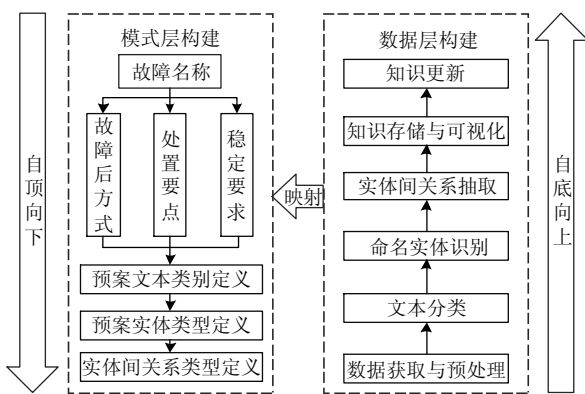


图 1 电网故障处置知识图谱构建流程
Fig. 1 Construction process of power grid fault handling knowledge graph

2.2 模式层构建

模式层是知识图谱的知识组织架构, 是对领域内实体、实体间关系以及属性进行描述的数据模型。本文在电力领域专家的帮助下对预案文本内容进行详细分析, 提炼出电力领域里有意义的概念类型与相关的属性, 以及概念之间关系, 从而形成领域知识体系。

如图 2 所示, 电网故障处置知识图谱的模式层由故障名称、系统稳定要求、故障后运行方式、故障处置要点 4 个核心要素以及它们之间的相互

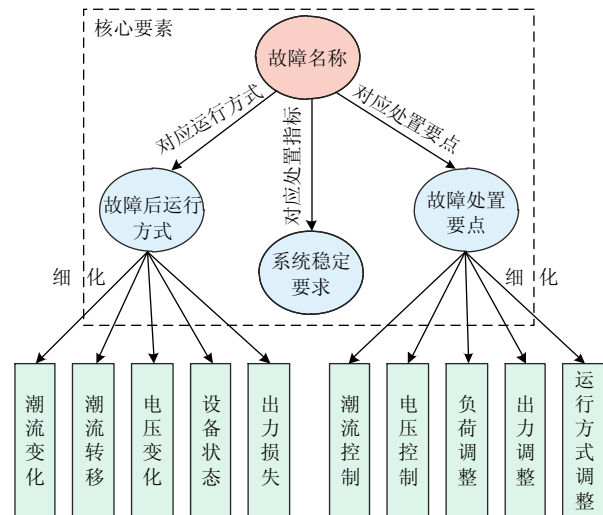


图 2 电网故障处置知识图谱模式层
Fig. 2 Scheme layer of power grid fault handling knowledge graph

关系构成。

2.3 数据层构建

数据层的构建主要分为知识抽取、知识融合与知识更新 3 个步骤。其中, 知识抽取是在模式层知识组织架构的指导下, 通过一系列知识抽取方法从非(半)结构化数据中获取实体、实体间关系以及属性等结构化知识; 知识融合是对知识抽取所得到的实体进行实体消歧和共指消解处理; 知识更新则是在知识图谱应用的过程中, 对其中知识的质量与时效性进行评估, 并结合知识的发展进行更新和修正。

目前, 电力领域的知识抽取主要采用规则模板或领域字典的方法^[11-12]。这种方法需要在领域专家的帮助下构建并维护一个领域字典, 之后通过手工编写大量规则模板进行知识抽取。然而规则模板的适用范围有限, 难以适应复杂的语言环境和形式多变的实际应用需求。为解决上述问题, 本文针对预案文本的特点, 提出了一套基于深度学习的知识抽取方法, 具体包括: 文本分类、命名实体识别、实体间关系抽取。

2.3.1 基于字向量 TextCNN 的预案文本分类

文本分类(text classification)是将预案文本按预先定义类别进行自动分类标记。如表 1 所示, 预案文本中的“故障后运行方式”常分为潮流变化信息、潮流转移信息、电压变化信息、设备状态变化信息、出力损失信息 5 类情况, 描述故障后电网运行方式的变化情况; “故障处置要点”也可细分为: 潮流控制、电压控制、负荷控制、出力调整、运行方式调整 5 类, 描述针对预想故障应采取的具体处置措施。对预案文本进行准确、高效地分类, 是知识抽取的一个重要内容。

表 1 文本类别及示例
Table 1 Text categories and examples

预案文本类别	示例
潮流变化信息	XX-XX 双线潮流越限
潮流转移信息	潮流转移到 XX3 台主变
电压变化信息	XX1000kV 电压略有下降
设备状态信息	XX 3 期 2 台 100 万机组跳闸
出力损失信息	损失出力 232 万
潮流控制	控制 XX-XXX 潮流断面在限额内
电压控制	XX 监控抬高 XX 地区 500kV 电压水平
负荷调整	转移 220kV XX 变部分负荷至 XX 供区
出力调整	XX 省调减 XX 机端出力至 270 万
运行方式调整	XX 分区与 XX220kV 改合环运行

依靠人工进行文本分类的方法不仅效率低而且对人员的专业知识有很高的要求,本文参考 Yoon Kim^[13]提出的 TextCNN 模型,构建适用于预案文本的文本分类模型。针对通用领域分词工具不能很好地对预案文本进行分词的情况,使用字向量代替词向量对文本进行 embedding 表示,避免了分词错误带来的影响。

2.3.2 基于 LR-CNN 的命名实体识别

命名实体识别(named entity recognition, NER)是对文本中具有特定含义的实体进行边界确定和类别识别,本文定义的命名实体类别如表 2 所示。电网故障处置预案属于特定领域的文本,因此在构建命名实体识别模型时应充分考虑文本的特点。

表 2 命名实体类别及示例
Table 2 Named entity categories and examples

实体类别	实体说明	实体示例
调度机构、厂站设备	电力调度机构、发电厂、变电站、电力设备、输电线路等	XX 网调、XX 抽蓄机组
指标	电力指标	潮流、出力
动作	描述处置操作、指标状态变化	抬高、越限
数量	描述指标量或设备数的数值	73 万、2 台
程度	描述处置操作的紧急程度、指标变化的剧烈程度	紧急、严重
属性	描述电压等级、设备容量	220kV、200 万

通过分析,预案文本的结构紧凑,实体成分复杂、边界难以界定,候选词冲突的现象突出。如图 3 所示,在“山西省调控制芝堰—信安线潮流不越限”一句中,“山西省调控制”存在 6 个候选词:“山西”、“省调”、“山西省”、“山西省调”、“调控”、“控制”。而“芝堰—信安线”又由“芝堰”、“信安”两个地名实体嵌套而成。

针对上述问题,本文使用引入 Lexicon Rethinking 机制的 CNN 网络(LR-CNN)^[14]实现电网故障处置领域的命名实体识别并解决候选词冲突问题。模型结构如图 4 所示。

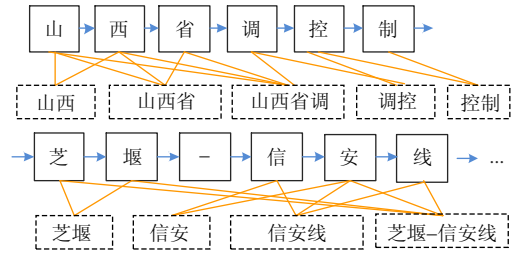


图 3 候选词冲突示例

Fig. 3 Example of word conflict

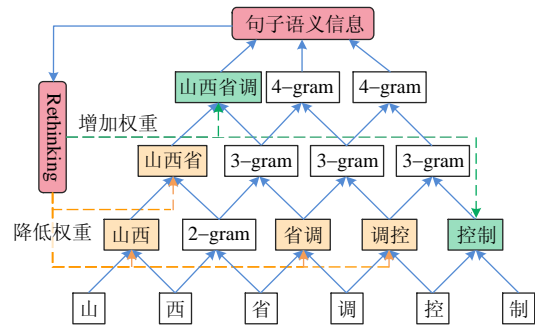


图 4 LR-CNN 模型结构

Fig. 4 Structure of LR-CNN model

LR-CNN 模型使用堆叠的窗口大小为 2 的 CNN 对句子的字符、候选词特征进行提取,并使用注意力机制模块合并字符和候选词信息。由于 CNN 的层次结构,低层的 CNN 无法引用高层 CNN 中的信息对错误的候选词的权重进行调整,不能很好地处理候选词冲突,因此通过引入 Rethinking 机制^[15],向每个 CNN 层添加反馈层,使用高层的字词信息来调整低层注意模块的权重,通过降低错误候选词的权重解决候选词冲突问题。

2.3.3 基于 BiGRU-Attention 的实体间关系抽取

实体间关系抽取(named entity relation extraction, NRE)是在命名实体识别的基础上判断实体间是否存在预定义的关系,从而构成一系列三元组知识。本文定义的实体间关系类别如表 3 所示。

表 3 实体间关系类别及示例
Table 3 Entity relation categories and examples

实体 A 类别	实体 B 类别	关系类别
调度机构、厂站设备	调度机构、厂站设备	A 包含 B/无关系
调度机构、厂站设备	动作	A 动作为 B/无关系
调度机构、厂站设备	指标	A 电力指标为 B/无关系
指标	数量	A 指标量为 B/无关系
指标	动作	A 指标状态为 B/无关系
动作	程度	A 动作程度为 B/无关系
调度机构、厂站设备/ 指标	属性	A 属性 B/无关系
调度机构、厂站设备	数量	A 设备数量为 B/无关系

本文在 BiLSTM-Attention^[16]模型的基础上,使用参数量更小的双向门控循环(BiGRU)结构以提高模型训练速度。BiGRU-Attention 模型结构如图 5 所示。

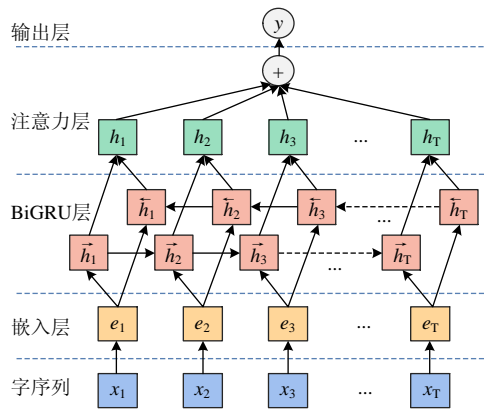


图5 BiGRU-Attention 模型结构

Fig. 5 Structure of BiGRU-Attention model

BiGRU-Attention 模型在 BiGRU 的基础上引入注意力机制，找到对关系分类起重要作用的字，学习得到一个权重，通过赋予这些字更高的权重以提高他们的重要性，从而提高关系抽取的准确率。

2.3.4 知识融合

经过知识抽取得的知识需要经过知识融合进行实体消歧和共指消解处理。其中，实体消歧指的是对可能存在多种含义的实体(例如“《三国演义》”可能指中国的四大名著，也可能指影视剧)进行区分；共指消解是指将具有相同含义和指代的名词和代词在知识图谱中进行合并。由于预案文本属于电力领域文本，实体词义仅限于电力领域，并且电力行业有明确的术语规范，实体歧义的问题基本不存在。然而，在预案文本存在较多的共指问题，这主要是由名词缺省所造成的，例如“申请加出太丹江、富春江电厂出力”一句中，“太丹江”实际上指的是“太丹江电厂”。因此，需要对这些缺省名词进行补全：首先编写正则表达式找出存在名词缺省的句子与实体的结尾，正则表达式为“([\u4e00-\u9fa5])(、)?(或)?(与)?(和)?([\u4e00-\u9fa5])+”，之后根据 NER 的结果确定缺省实体的边界后再编写规则自动补全实体。

2.3.5 知识更新

目前，电力系统的快速发展，电网结构和运行模式愈加复杂，电力设备、处置预案均在不断发展。因此，在知识图谱构建之后需要持续地更新，以保证其中知识的有效性。电网故障处置知识图谱的更新分为模式层和数据层两方面：模式层的更新是指新增的预案中出现当前模式层中不存在的概念(例如新的处置要点类别)时，则根据新的概念对模式层中的文本类型、实体类型、关系类型进行更新；数据层的更新则是当新产生的预案没有产生新的概念时，采用增量更新的方式^[3]，对新增的预案进行

知识抽取与知识融合后添加到原有的图谱中。除此之外，数据层的更新也包含对其中知识的质量与有效性进行评估，及时删除失效知识，这部分工作的工作需要专业调度人员的协助。

3 算例分析

3.1 算例情况

为了验证本文所构建的信息抽取模型的有效性，以某地区的 1035 份电网故障处置预案文本为实验对象，选取其中“故障后运行方式”和“故障处置要点”共 3900 条，经过长句切分后得到共 9875 条，经过人工标注后得到用于训练各信息抽取模型的数据集。各数据集均按 8:1:1 的比例随机划分为训练集、验证集和测试集用于模型的训练和测试。

3.2 预案文本分类实验

3.2.1 实验设计与评价指标

为了对比本文构建的基于深度学习的文本分类模型与传统机器学习方法的分类效果，选取朴素贝叶斯、邻近算法(KNN)、支持向量机(SVM)作为对照进行实验，上述算法的实现基于 Python 的 sklearn 工具包。

为了验证本文使用字向量代替词向量进行 embedding 的有效性，在模型其他参数保持不变的情况下，使用词向量对句子进行 embedding，词向量采用随机初始化的方式在 TextCNN 模型中训练。句子的分词处理使用 Python 的 jieba 工具包完成。

模型在各类别文本上的分类性能以精确率(precision)、召回率(recall)、F1 值^[17]为评价指标进行评估，模型的整体评价指标采用 macro-F1 值进行评估，macro-F1 值是各类文本 F1 值的算术平均。

3.2.2 TextCNN 模型参数设置

本文构建的基于字向量的 TextCNN 模型的参数设置如表 4 所示。

表 4 TextCNN 模型参数设置	
Table 4 Parameters setting of TextCNN model	
模型参数	模型参数取值
字嵌入维度	300
CNN 卷积核尺寸	2, 3, 4
Dropout 随机失活率	0.5
训练批数(epoch)	10
学习率	0.001
各尺寸卷积核数量(channels)	256

3.2.3 文本分类实验结果与分析

各分类模型的 macro-F1 值如表 5 所示，其中预训练的字向量由文献[18]提供。由表 5 可见，本文

构建的基于字向量的 TextCNN 模型取得最好的分类效果, 优于各类基于传统机器学习的方法, 比其中表现最好的 SVM 方法提高了 7.4% 的 macro-F1 值。

此外, 通过对比基于词向量的 TextCNN 模型可以发现, 使用通用领域分词工具所带来的分词错误会极大地影响模型的性能。可见本文使用的基于字向量的嵌入方式能避免分词错误对文本分类造成的影响, 从而更好地从预案文本中学习句子的语义特征, 提高分类效果。

表 5 文本分类实验结果
Table 5 Results of text classification

模型	macro-F1
朴素贝叶斯	0.8252
KNN	0.8126
SVM	0.9112
TextCNN + 词向量(随机初始化)	0.8562
TextCNN + 字向量(随机初始化)	0.9693
TextCNN + 字向量(预训练)	0.9852

3.3 预案命名实体识别实验

为了验证本文提出的基于 LR-CNN 模型的预案命名实体识别方法的有效性, 选择隐马尔可夫模型(HMM)、条件随机(CRF)、BiLSTM-CRF^[19]模型、IDCNN-CRF^[20]模型和 Lattice LSTM^[21]模型作为对照模型。其中, BiLSTM-CRF 模型、IDCNN-CRF 模型和 Lattice LSTM 模型是中文命名实体任务常用的深度学习模型。

模型的性能指标同样使用精确率、召回率、F1 值进行评价, 其中样本是否被正确预测的评估方式是以实体为单位而不是以单个标签。此外, 将模型的预测耗时作为模型的速度指标。

3.3.1 TextCNN 模型参数设置

本文所构建的 LR-CNN 模型的参数设置如表 6 所示。

表 6 LR-CNN 模型参数设置
Table 6 Parameters setting of LR-CNN model

模型参数	模型参数取值
嵌入方式	50 维字向量+50 维词向量
CNN 层数	4
Dropout 随机失活率	0.5
训练批数(epoch)	10
学习率	0.0015
学习率衰减率	0.05
优化器	Adamax
每层卷积核数量(channels)	128

3.3.2 命名实体识别实验结果与分析

各命名实体识别模型的准确率(P)、召回率(R)、F1 值和预测速度如表 7 所示, 各深度学习模型的训练曲线如图 6 所示。

表 7 命名实体识别实验结果
Table 7 Results of named entity recognition

模型	准确率(P)	召回率(R)	F1 值	预测 100 句用时/s
HMM	0.7420	0.8700	0.8009	—
CRF	0.8677	0.8506	0.8590	—
BiLSTM-CRF	0.9219	0.9386	0.9302	1.84
IDCNN-CRF	0.9258	0.9360	0.9309	1.35
Lattice LSTM	0.9367	0.9579	0.9471	3.88
LR CNN	0.9404	0.9658	0.9529	2.08

由表 7 和图 6 可见, 本文构建的 LR-CNN 模型在各指标上都取得了最好的性能, 而 Lattice LSTM 模型的识别性能和本文模型相近。原因是 BiLSTM-CRF 模型、IDCNN-CRF 模型以字符作为输入, 没有利用相邻字之间存在的语义信息, 无法解决候选词冲突问题。LR-CNN 模型和 Lattice LSTM 模型在字符输入的基础上使用了候选词信息, 能够较好地解决候选词冲突问题。

此外, 通过对比 4 种深度学习模型预测 100 条预案速度可以得出, 基于 CNN 的 LR-CNN 比基于序列的 Lattice LSTM 具有更高的计算效率。

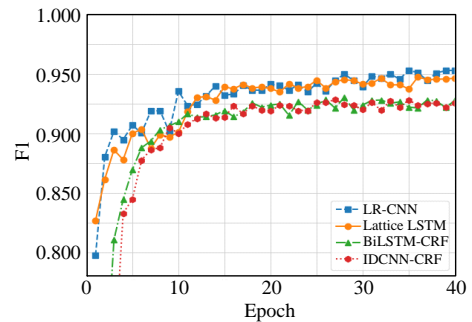


图 6 模型训练曲线

Fig. 6 Model training curve

3.4 实体间关系抽取实验

3.4.1 实验评价指标与参数设置

使用精确率、召回率、F1 值作为模型在各类别关系上的性能指标, 模型整体性能使用 macro-F1 值评价。

使本文所构建的 BiGRU-Attention 模型的参数设置如表 8 所示。

表 8 BiGRU-Attention 模型参数设置
Table 8 Parameter settings of BiGRU-Attention model

模型参数	模型参数取值
嵌入方式	100 维预训练字向量
BiGRU 隐藏层单元数	230
Dropout 随机失活率	0.5
训练批数(epoch)	15
学习率	0.0015

3.4.2 实验结果与分析

表 3 中所定义的各种类型关系的识别结果如表 9 所示。

示。当电网发生故障时，首先由 SCADA/EMS 设备数据采集与监视控制系统返回事故的告警信息；将获得的事事故告警信息进行解析，得到故障后电网运行方式的变化报告；之后利用构建的 TextCNN 文本分类模型对报告内容进行分类，并利用 LR-CNN 模型与 BiGRU-Attention 模型将其

解析为对应的三元组信息；然后利用图数据库 Neo4j 的查询语句 Cypher 查找知识图谱中是否存在故障后方式与解析内容一致的故障类型；若存在则返回故障类型与相应的处置要点信息；若不存在，则由电力调度人员分析故障类型后，查询知识图谱返回对应的处置要点。

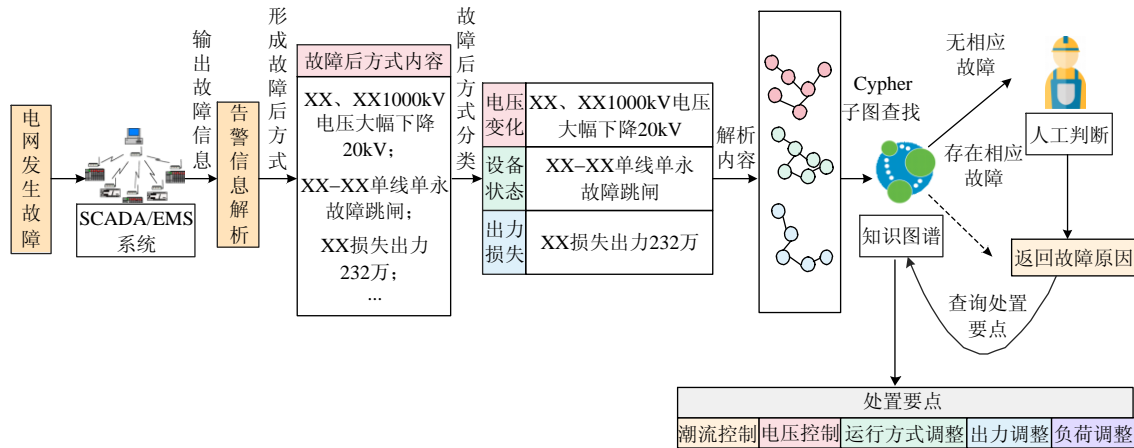


图 8 电网故障处置流程

Fig. 8 Power grid fault handling process

5 结论

本文以电网故障处置预案文本为对象，提出一种自顶向下与自底向上相结合的知识图谱构建方法，并针对预案文本的特点提出了一套基于深度学习的电力领域知识抽取方法：

1) 针对通用领域的分词工具不能很好地对预案文本进行分词的情况，构建了基于字向量的 TextCNN 模型，避免了分词错误带来的影响，实现了预案文本高准确率的自动分类。

2) 针对预案文本实体成分复杂，实体边界难以界定的情况，构建了 LR-CNN 模型进行命名实体识别，通过引入 Lexicon Rethinking 机制，利用高层 CNN 的字词信息对错误低层 CNN 的候选词的权重进行调整，提高了模型识别的准确率。

3) 在命名实体识别结果的基础上，构建 BiGRU-Attention 模型进行关系的抽取，将预案文本转化为三元组信息。

最后，通过 Neo4j 图数据库对本文构建的知识图谱进行了可视化，并对其应用进行了分析。

参考文献

[1] 乔骥, 王新迎, 闵睿, 等. 面向电网调度故障处理的知识图谱框架与关键技术初探[J]. 中国电机工程学报, 2020, 40(18): 5837-5848.
 QIAO Ji, WANG Xinying, MIN Rui, et al. Framework and key technologies of knowledge-graph-based fault handling system in power grid[J]. Proceedings of the CSEE, 2020, 40(18): 5837-5848(in Chinese).

[2] 孙宏斌, 黄天恩, 郭庆来, 等. 面向调度决策的智能机器调度员研制与应用[J]. 电网技术, 2020, 44(1): 1-8.
 SUN Hongbin, HUANG Tianen, GUO Qinglai, et al. Automatic operator for decision-making in dispatch: research and applications[J]. Power System Technology, 2020, 44(1): 1-8(in Chinese).

[3] 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(in Chinese).

[4] 徐增林, 盛泳潘, 贺丽荣, 等. 知识图谱技术综述[J]. 电子科技大学学报, 2016, 45(4): 589-606.
 XU Zenglin, SHENG Yongpan, HE Lirong, et al. Review on knowledge graph techniques[J]. Journal of University of Electronic Science and Technology of China, 2016, 45(4): 589-606(in Chinese).

[5] TANG Yachen, LIU Tingting, LIU Guangyi, et al. Enhancement of power equipment management using knowledge graph[C]// Proceedings of the 2019 IEEE Innovative Smart Grid Technologies-Asia (ISGT Asia). Chengdu: IEEE, 2019: 905-910.

[6] CHUN Sejin, JIN Xiongnan, SEO S, et al. Knowledge graph modeling for semantic integration of energy services[C]// Proceedings of the 2018 IEEE International Conference on Big Data and Smart Computing (BigComp). Shanghai: IEEE, 2018: 732-735.

[7] 赵振兵, 段记坤, 孔英会, 等. 基于门控神经网络的栓母对知识图谱构建与应用[J]. 电网技术, 2021, 45(1): 98-106.
 ZHAO Zhenbing, DUAN Jikun, KONG Yinghui, et al. Construction and application of bolt and nut pair knowledge graph based on GGNN[J]. Power System Technology, 2021, 45(1): 98-106(in Chinese).

[8] 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(in Chinese).

- Chinese).
- [9] 余建明, 王小海, 张越, 等. 面向智能调控领域的知识图谱构建与应用[J]. 电力系统保护与控制, 2020, 48(3): 29-35.
YU Jianming, WANG Xiaohai, ZHANG Yue, et al. Construction and application of knowledge graph for intelligent dispatching and control[J]. Power System Protection and Control, 2020, 48(3): 29-35(in Chinese).
- [10] 高海翔, 苗璐, 刘嘉宁, 等. 知识图谱及其在电力系统中的应用研究综述[J]. 广东电力, 2020, 33(9): 66-76.
GAO Haixiang, MIAO Lu, LIU Jianing, et al. Review on knowledge graph and its application in power systems[J]. Guangdong Electric Power, 2020, 33(9): 66-76(in Chinese).
- [11] 曹靖, 陈陆桑, 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(in Chinese).
- [12] 佟佳弘, 武志刚, 管霖, 等. 电力调度文本的自然语言理解与解析技术及应用[J]. 电网技术, 2020, 44(11): 4148-4156.
TONG Jiahong, WU Zhigang, GUAN Lin, et al. Power dispatching text analysis and application based on natural language understanding[J]. Power System Technology, 2020, 44(11): 4148-4156(in Chinese).
- [13] Kim Y. Convolutional neural networks for sentence classification[C]// Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, 2014: 1746-1751.
- [14] GUI Tao, MA Ruotian, ZHANG Qi, et al. CNN-Based Chinese NER with lexicon rethinking[C]//Twenty-Eighth International Joint Conference on Artificial Intelligence. 2019: 4982-4988.
- [15] LI Xin, JIE Zequn, FENG Jiashi, et al. Learning with rethinking: recurrently improving convolutional neural networks through feedback[J]. Pattern Recognition, 2018, 79(1): 183-194.
- [16] ZHOU Peng, SHI Wei, TIAN Jun, et al. Attention-based bidirectional long short-term memory networks for relation classification[C]// Proceedings of the 54th annual meeting of the association for computational linguistics(volume 2: Short papers). 2016: 207-212.
- [17] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016: 30-32.
ZHOU Zhihua. Machine learning[M]. Beijing: Tsinghua University Press, 2016: 30-32(in Chinese).
- [18] LI Shen, ZHAO Zhe, HU Renfen, et al. Analogical reasoning on Chinese morphological and semantic relations[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics(Volume 2: Short Papers), 2018: 138-143.
- [19] HUANG Zhiheng, XU Wei, YU Kai. Bidirectional LSTM-CRF models for sequence tagging[J]. Computer Science, 2015.
- [20] Strubell E, Verga P, Belanger D, et al. Fast and accurate entity recognition with iterated dilated convolutions[C]//Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 2017: 2670-2680.
- [21] ZHANG Yue, YANG Jie. Chinese NER Using Lattice LSTM[C]// Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics(Volume 1: Long Papers), 2018: 1554-1564.



郭榕

在线出版日期: 2021-04-09。

收稿日期: 2021-01-12。

作者简介:

郭榕(1996), 男, 硕士研究生, 研究方向为电网自然语言处理, E-mail: ficus_gr@163.com;

杨群(1971), 女, 通信作者, 副教授, 硕士生导师, 研究方向为深度学习等, E-mail: qun.yang@nuaa.edu.cn。

(责任编辑 王晔)