

基于深度学习的电网故障预案信息抽取研究

石赫¹ 杨群¹ 刘绍翰¹ 李伟^{2,3,4}

1 南京航空航天大学计算机科学与技术学院 南京 211106

2 南瑞集团(国网电力科学研究院)有限公司 南京 211106

3 国电南瑞科技股份有限公司 南京 211106

4 智能电网保护和运行控制国家重点实验室 南京 211106

(657332737@qq.com)

摘要 为对电网调度部门保存的应急预案进行处理,以方便调度人员面临突发事件时能快速检索和匹配预案中的类似事故并借鉴以往经验处理突发事故,需要进行预案的信息抽取,提取其关键信息。传统的电网故障应急预案处理方法的通用性和可扩展性不强,无法有效实现电网故障应急预案的数字化。采用深度学习的方法,对调度紧急预案的内容进行句法分析,得到分析树,在此基础上抽取故障时系统的状态信息和对故障的处置要点,进而将非结构化的预案信息转化成结构化数据存储和管理。采用上述方法可以有效管理电网预案,辅助调度人员完成故障的判定并按预案给出的操作方法进行电网运行管理,从而提高故障处理效率。同时,该方法具有通用性好、扩展性强的优势,可以实现模型的持续改进。

关键词: 应急预案;深度学习;句法分析;分析树;信息抽取

中图分类号 TP391

Study on Information Extraction of Power Grid Fault Emergency Pre-plans Based on Deep Learning

SHI He¹, YANG Qun¹, LIU Shao-han¹ and LI Wei^{2,3,4}

1 College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China

2 NARI Group Corporation (State Grid Electric Power Research Institute), Nanjing 211106, China

3 NARI Technology Co. Ltd., Nanjing 211106, China

4 State Key Laboratory of Smart Grid Protection and Control, Nanjing 211106, China

Abstract The emergency pre-plans are saved by the power grid dispatching department. The power grid dispatching department has formulated it based on the power grid operation and maintenance experiences, which can assist dispatchers to deal with the emergency fault. When an emergency fault occurs in the power grid, in order to effectively deal with the emergency pre-plans, so that dispatchers can learn from previous experience to handle emergencies, it's necessary to extract the key information of the pre-plan. So in this way they can quickly retrieve and match similar accidents in the pre-plans. However, there are many problems of the traditional power grid fault emergency pre-plans processing method. The traditional processing method is not versatile and scalable, and it is unable to effectively digitize the power grid fault emergency pre-plans. This leads to a limited scope of application of the traditional method. In this paper, the deep learning method is used to make up for the shortcomings of traditional processing methods. It is used to analyze the syntax of the sentences for the dispatch emergency pre-plans. Then it will generate a syntactic parser and the parser will give the syntactic parse tree of the sentences. After that, the system state information and disposal points for the emergency pre-plans are extracted, and then the unstructured text information of the emergency pre-plans is transformed into the structured data. Using the deep learning method, the power grid emergency pre-plans can be effectively managed, and the dispatcher can determine the fault type and carry out the operation quickly. This paper also carries out experimental verification for the method. It can be concluded that the method can improve the efficiency of fault processing, while it also has the advantages of good versatility and strong expansibility, and can achieve continuous improvement of the model.

Keywords Emergency pre-plans, Deep learning, Syntactic analysis, Syntactic parse tree, Information extraction

1 引言

电网安全可靠地运行对于国民经济发展具有十分重要的

意义^[1]。然而,电网是个复杂系统,随着电网规模扩大^[2]、拓扑结构日益复杂,电网运行中发生的故障也越来越多。为有效组织电网系统的运行,需要在故障发生时快速形成故障预

基金项目:智能电网保护和运行控制国家重点实验室课题;江苏省重点研发计划(BE2019012)

This work was supported by the Project of "the State Key Laboratory of Smart Grid Protection and Operation Control" and Primary Research & Development Plan of Jiang Province(BE2019012).

通信作者:杨群(qun.yang@nuaa.edu.cn)

案供调度人员及时处理。

电网调度部门根据以往运行维护经验制定了故障应急预案手册,当故障出现时,调度人员手工查找故障应急预案手册,并根据事件的特征匹配故障类型,找到相应的应急预案。由于上述处理方法效率低下,因此实际应用中常常采用基于规则的模板库来辅助查找。然而传统的基于规则的模板库方法在语义理解和维护方面存在不足,从而上述方法存在如下问题:

(1)实用性不强。一方面,基于规则的模板库采用关键词来刻画故障及其处理方法,对故障现象的描述能力有限;另一方面,故障类型和导致故障的因素众多,并且不同区调对同一种故障的处理方法也不尽相同,模板无法涵盖不同区域、不同类型的复杂故障现象,造成故障发生时难以在模板库中匹配恰当的类型,因此这种方法并不实用。

(2)处理方案维护困难。传统的故障处理方案采用规则描述,形式单一,且无法表达规则之间的逻辑关系,因此规则库不具备泛化能力^[3-4];另一方面,各区调存在实时运行的参数规格不同、故障相同类型但处理预案不同的情况,这些情况均需采用不同的规则表示。这些规则都需要手工添加到模板库中且修改更新也由手工完成,所以规则库的通用性和扩展性均很差。

(3)故障判定效率低。模板中的字段并不能精准地刻画不同区域的各种故障,因此实际运行中规则数量多、查找速度慢,故障判定的效率低^[5-7]。

采用深度学习的方法管理电网运行故障是近年来的一条新思路^[8-14]。过去有工作采用浅层神经网络处理电网故障数据,但浅层神经网络对复杂函数的拟合能力有限,面临复杂问题时其泛化能力受到一定制约,分类精度也不如深度学习^[3-4]。

深度学习是机器学习研究领域的一个分支,是一种模拟人脑机制对数据进行表征、分析和学习的方法。它采用多层神经网络、通过组合低层特征形成更抽象的高层来表示属性类别或特征,发现数据的分布式特征表示,是一种无监督学习。

深度学习通过构建具有多个隐层的机器学习模型,使用海量数据训练并抽取和组合更抽象的特征表示,实现“端到端的”机器学习,提高了分类或预测的准确性。对存在大量高维度数据的电力数据系统来说,采用深度学习的方法具有以下优点:

1)深层神经网络拟合能力强。可以充分利用电力运行大数据,结合神经网络的特征提取和组合能力,实现模型训练,提高分类和拟合的精度。

2)采用深度学习,可以针对各区调预案数据不同的问题,将训练好的模型迁移到不同调度部门,适应不同的应用场合,实现模型的通用性。

3)采用深度学习,可以基于原有知识学习新知识,持续改进模型,实现模型的可扩展性。

4)采用深度学习还可以解决电力工业数据集标注量少的问题,并可以提高模型的稳定性和可泛化性。

本文提出一种基于深度学习的故障信息抽取方法,对已有预案进行句法分析和信息提取,并将结果以结构化^[15-16]文档形式存储和管理,使得预案能在检索时以有效的形式呈现,提高故障处理的效率,同时发挥深度学习的优势,解决规则的通用性和扩展性问题。

2 基于深度学习的电网故障预案研究

基于深度学习的电网故障预案处理方法首先训练句法分析器模型,通过句法分析器分析语句的结构,得到层次结构清晰的语法树,然后根据语法树提取语句成分,得到电网调度所需的关键信息,并采用结构化数据表示。其过程包含数据预处理、模型训练、句法分析和信息抽取,其中句法分析起关键作用,它是电网故障预案有效信息提取的基础。精确句法分析是自然语言处理领域的难点问题,电网领域中的专业词语的含义不同于开放领域的词语,因此采用传统的基于统计或者基于规则的句法分析电网预案得到的语法树精确度不高。下面介绍本文基于深度学习进行电网故障预案信息抽取的具体方法和过程。

2.1 数据预处理

2.1.1 分词

由于汉语和英语的不同,汉语结构比较复杂,主谓宾的位置与英文相反,词与词之间也是相连的,因此,不能完全按照英文的处理方法进行分析。在对一个中文句子进行处理时,首先做的是对句子进行分词。本文采用如下分词算法对电网调度预案进行分词处理。

本文采用最大概率路径算法找出基于词频的最大切分组合对电网故障数据进行分词处理。

首先计算句子 C 的分词方案为 A 的概率,其中 $A = \{a_1, a_2, \dots, a_k\}$:

$$p(A|C) = \prod_{j=1}^k \frac{p(a_j, C)}{p(C)} \quad (1)$$

在切分所得的所有结果中求某个切分方案,使得该切分方案 S 的概率最大,其中 R 为句子 C 的所有切分方案的集合。

$$p(S) = \max(p(R|C)) \quad (2)$$

采用上述分词方法处理电网故障预案时往往出现错误,原因在于上述分词方法是通过查找最大概率路径找出基于词频的最大切分组合,而电网故障领域中的专业词词频低,在计算最大概率路径时会被过滤掉,因此在上述分词方法的基础上采用基于电网故障预案的互信息分词方法进行预案领域分词。

首先定义预案文本中句子中的词 A 、 B 之间的互信息:

$$I(A, B) = \log_2 \frac{P(A, B)}{P(A)P(B)} \quad (3)$$

其中, $P(A, B)$ 为词 A 和 B 在预案中共同出现的概率, $P(A)$ 为词 A 在预案中出现的概率, $P(B)$ 为词 B 在预案中出现的概率。其中词 A 、 B 的长度在计算过程逐步增大,范围在 1—8 之间。(预案中一个词最多包含 8 个字)。

同时本文还建立了预案专用字典,采用基于词典的分词方法辅助上述算法,实现正确分词。

2.1.2 词向量训练

词向量^[17]的训练是为了表达词语与词语之间的语义关系。为充分考虑电网调度预案中词义问题对分析结果的影响,本文使用词向量得到关键词间的语义相似度等信息,把语法层次的分析提升到语义层次,为模型训练做好数据准备。

采用 Google 开源的用于词向量计算的工具 word2vec 进行词向量计算。训练过程如下:首先为文本建立一个词频表,为每一个词生成一个 one-hot 向量,作为模型输入;然后训练

过程中不断调整输入层到隐藏层的参数矩阵,最终得到的矩阵作为每一个词的词向量。

2.2 基于深度学习的句法分析模型

2.2.1 基于概率上下文无关文法的句法分析

传统的句法分析方法是基于规则的分析方法,这种方法缺少灵活性,在处理不确定性或者歧义性问题时会出现问题,这时需要找到一种方法从多种可能的语法树中找出最可能的一棵树,作为句法分析的最终结果。而概率上下文无关文法(PCFG)^[18-19]对句子进行句法分析时,对于每一棵生成的语法树,将其中所有规则的概率的乘积作为语法树的出现概率,通过概率值的大小从多种可能的语法树中找出概率最大的一棵树作为该句子的最佳句法分析树。

PCFG 是一个五元组的形式: $G=(V_N, V_T, S, R, P)$,其中 V_N 是非终结符的集合,在句法分析中表示词性标签。 V_T 是终结符的集合,且 $V_N \cap V_T = \emptyset$,在句法分析中表示句法标签。 S 是非终结符,表示上下文无关文法的开始符号, S 在集合 V_N 中。 R 是规则的集合,形如: $A \rightarrow \alpha$,其中 $A \in V_N, \alpha \in (V_N \cup V_T)^*$, P 表示产生的每个规则的统计概率,即:

$$P: R \rightarrow (0, 1], \text{其中对于任意 } N \in V_N, \sum_{\alpha: N \rightarrow \alpha \in R} P(N \rightarrow \alpha) = 1$$

本文定义句法分析采用的规则如下:

$X \rightarrow Y_1 Y_2$ ①,其中 Y_1 和 Y_2 属于非终结符

$X \rightarrow y$ ②,其中 y 属于非终结符

语句 \rightarrow 句子 终结符 ③

句子 $\rightarrow NP VP$ ④

$VP \rightarrow VV NP$ ⑤

$NP \rightarrow$ 冠词 NN ⑥

$NP \rightarrow$ 专用名词 ⑦

其中,终结符是预案中的词,非终结符为词所对应的词性。通过计算语法树中第 i 个单词到第 j 个单词的概率,得到候选语法树。

计算某个句子的分析树就是找出该句子中概率最大的分析树。公式如下,其中 S_j 表示句子 C 中一个可能的分析树, R 为语法规则集合。

$$p(S_j | C) = \prod_{r \in R} p(r) \quad (4)$$

$$p(C) = \max(P(S_j | C)) \quad (5)$$

采用 PCFG 时需要确保句子的上下文无关、位置无关等多种条件。而对于一个句子而言,同一个概念会用到不同的词,同一个词被不同的人使用,或者在不同的环境下使用时表达的含义也是不同的,这对于句法分析的结果有着很大的影响。因此,我们不能用单纯的概率相乘的方法来分析句法结构,而应多考虑一些词与词之间的语义关系来衡量一个分析树的概率值。

2.2.2 基于循环神经网络的句法分析

针对模型的独立性假设,研究者们也提出了一些改进方法来增加 PCFG 模型的上下文描述能力。随着深度学习的发展,越来越多学者将深度学习方法引入句法分析中,采用深度学习的句法分析方法相较于传统的 PCFG 句法分析方法在上下文描述能力方面有较好的表现,目前已成为新的研究热点。

使用循环神经网络(RNN)进行句法分析的模型如下,其中输入为两个候选孩子的词向量表示,输出为两个节点合并时的语义表示以及新节点的解析决策分数,其中节点的解析决策分数用于衡量整个语法树的得分,是由每个节点上的解

析决策分数之和得到的。模型如图 1 所示:

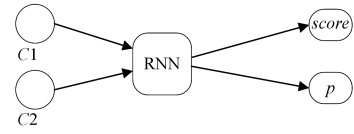


图 1 RNN 进行句法分析的过程

Fig. 1 Syntax analysis process of RNN

其中具体公式如下:

$$score = Up \quad (6)$$

$$p = \tanh\left(W \cdot \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}\right) \quad (7)$$

$\begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$ 表示该节点的孩子结点词向量,树的所有节点上的

权重矩阵 W 参数相同, U 为训练模型综合权重, $score$ 为节点的解析决策分数, p 为两个节点合并时的语义表示。

2.2.3 基于组合向量文法的句法分析

Socher^[20] 在 2013 年提出一种句法分析方法,将 PCFG 与循环神经网络结合,得到组合向量文法(Compositional Vector Grammars, CVG)的句法分析方法。

Socher 考虑到传统的循环神经网络(RNN)模型在训练时没有考虑到关键词之间的词义关系对训练结果产生的影响,从而造成句法分析不准确,因此他提出的 CVG 模型不仅将应用广泛的 PCFG 模型与 RNN 模型相结合,并在 RNN 的基础上提出了结合词义的循环神经网络(Syntactically Untied Recursive Neural Networks, SU-RNN)。SU-RNN 和 RNN 的不同之处在于网络中权重矩阵 W 不再唯一,而是拥有一系列这样的权重矩阵,其值依赖于子节点的句法类别,其个数由 PCFG 规则右部的数量决定。

该模型进行两遍自底向上的搜索:首先将待分析的句子输入到概率上下文无关文法(PCFG)中,经过 PCFG 的训练生成该句子的候选分析树,然后将生成的候选树作为 SU-RNN 的输入,结合关键词的词向量,训练得到分析树的 Score 值,最后选出得分最高的分析树作为最终结果。

CVG 的具体流程如图 2 所示。

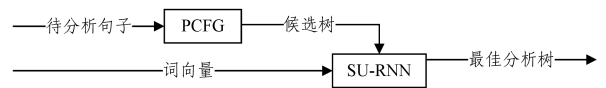


图 2 CVG 的流程

Fig. 2 Flow of CVG

2.3 句法分析模型训练

分析树 Score 值的计算过程如下。

结合语义自底向上计算每一个节点的语义表示,计算公式为:

$$p = f\left(W^{A,B} \begin{pmatrix} a \\ b \end{pmatrix}\right) \quad (8)$$

其中, $W^{A,B}$ 表示该矩阵依赖于孩子节点的词性, a, b 表示该节点的孩子节点词向量,式中 $f = \tanh$ 。

之后计算每一个节点的解析决策分数:

$$s(p) = U \cdot p \quad (9)$$

其中, U 为训练模型综合权重。

最后对所有节点进行求和得到最终语法树的得分。具体过程如图 3 所示。

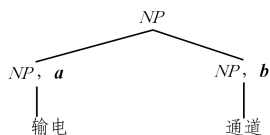


图3 语法树中某一节点及其子节点

Fig. 3 Node in a syntax tree and its children

其中 a, b 为“输电”和“通道”的词向量,计算最上端 NP 节点的语义表示如式(8)所示:

$$p = f\left(W^{NP, NP} \begin{pmatrix} a \\ b \end{pmatrix}\right) \quad (10)$$

之后利用式(9)自底向上计算每个节点的解析决策分数。

2.4 句法分析结果示例

使用训练好的句法分析器进行电网故障预案句法分析,得到的语法树如图4所示。

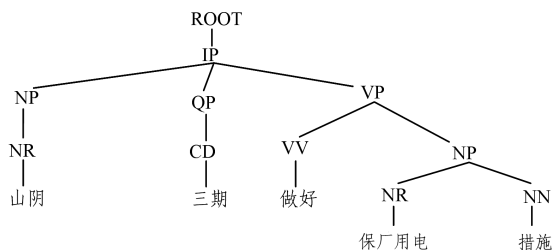


图4 一个语法树的例子

Fig. 4 Example of a syntax tree

由图4可以看出,该句子的核心词是“做好”,在此基础上再进一步找出与“做好”有关系的主语、宾语、定语等,抽取句子的关键信息并表示成结构化表,形成故障处理预案。

3 模型评估

本文使用某电力公司保存的某地区部分的45196条电网故障调度预案作为模型训练数据集,数据集按照约7:3的比例划分为训练集和测试集,具体数据集统计如表1所列。

表1 数据集

Table 1 Data set

分类	数据集	数据量/条
训练数据集	电网故障调度预案	3000
测试数据集	电网故障调度预案	1196

本文使用一个常用的评价指标 F1 值作为模型评估结果。F1 值是召回率和准确率的调和平均数,实验结果如表2所列。

表2 3种算法的数据比较

Table 2 Data comparison of three methods

Algorithm	LP/%	LR/%	F1/%
PCFG	74.6	74.1	74.34
CVG(RNN)	77.9	75.92	76.91
CVG(SU-RNN)	79.2	77.83	78.51

LP表示准确率,LR表示召回率。由上表可知 CVG 模型相对于 PCFG 模型有较大改善,F1 的值由 74.34 提高至 78.51。

图5、图6分别是采用传统方法和深度学习方法进行句法分析的结果。

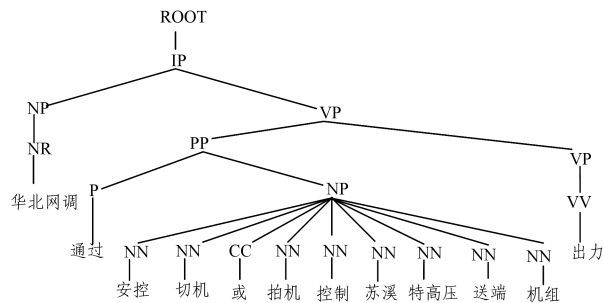


图5 传统的句法分析结果

Fig. 5 Traditional syntactic analysis results

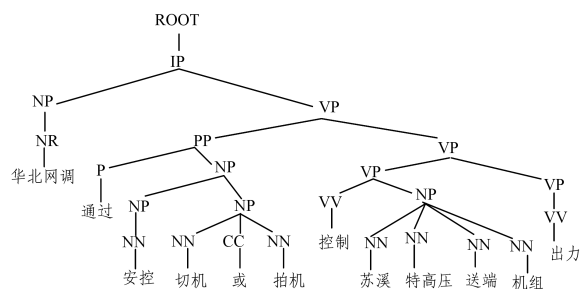


图6 深度学习的句法分析结果

Fig. 6 Deep learning analysis results

由图5、图6可以看出传统方法没有考虑到词语之间上下文关系和词义,导致分析时将“控制”分析成了名词,而采用深度学习之后“控制”被分析成了动词。如果“控制”被分析成名词,之后的信息抽取会忽略“控制”这一关键信息,导致抽取信息不准确。

4 案例分析

下面以某电力公司2015年修订的一个故障处理预案为例,介绍如何进行信息抽取,并通过列表形式对故障所涉及到的稳定要求、故障后方式、处置要点等进行编制。

对“南华剩余单台主变潮流越限,转移95%”进行处理。首先进行语法树分析,由语法树可以得出该句子由3部分构成:动作发出者“南华剩余单台主变潮流”、动作“越限”、程度“转移95%”。从该语法树中抽取出三元组(“南华剩余单台主变潮流,越限,转移95%”)。具体分析结果如图7所示。

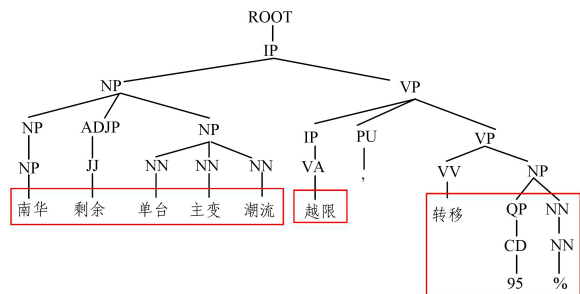


图7 “南华剩余单台主变潮流越限,转移95%”的语法树

Fig. 7 Syntax tree of “Nanhua remaining single main transformer current exceeded, transfer 95%”

对“华北网调抬高高博、上元附近电压水平”进行处理。首先进行语法树分析,由语法树可以得出该句子由四部分构成:动作发出者“华北网调”、动作“抬高”、动作接受者“高博、上元”和程度“电压水平”,从该语法树中抽取出四元组(“华北网调,抬高,高博、上元,电压水平”)。具体分析结果如图8所示。

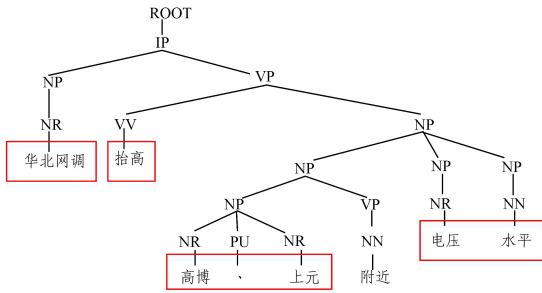


图8 “华北网调抬高高博、上元附近电压水平”的语法树

Fig. 8 Syntax tree of “North China network raising voltage level near Gaobo and Shangyuan”

以上通过语法树抽取出来的信息以表格形式存储,表格的列包含动作发出者、动作、动作接受者、程度,调度人员可通过查表得到需要的故障信息,并根据指示做出相应的处理。

本文对句子结构成分的分析方法可以处理语言中的递归现象,可以分析语句的构成得到句子的主干结构。按照本文方法可以实现预想故障结构化,将故障预案内容提取成计算机可以理解的形式,为自动化处理打下基础。

结束语 随着信息技术的发展及电网规模的变化,深度学习逐渐被应用到电网系统中。采用深度学习技术可以通过逐层训练得到高层故障特征,适合处理高维度电网数据。深度学习方法诊断能力强,并且可以通过参数优化调整,获得良好的泛化能力;同时,它不依赖人工经验,可以消除人工特征选取带来的不确定性。

本文采用深度学习的方式对故障预案进行句法分析和提取有效信息,可用于推进电网调度的自动化工作。目前,这种分析方法适用于处理句子主谓宾形式的长距离关系,对于一些细节的信息提取效果不太理想,因此后续工作要解决细节信息提取问题,提取出具体修饰词与主谓宾之间存在的关系,对实体之间关系做进一步分析。

参考文献

- [1] WANG S P, ZHAO D M. Research review and prospect of power grid fault diagnosis[J]. Automation of Electric Power Systems, 2017, 41(19): 164-175.
- [2] ZHANG H B, JIA K, SHI W J, et al. Network fault diagnosis based on information theory and expert system[J]. Journal of Power Systems and Automation, 2017, 29(8): 111-118.
- [3] XIONG G J, SHI D Y, ZHU L. Fuzzy cell fault diagnosis based on radial basis function neural network[J]. Automation of Electric Power Systems, 2014, 38(5): 59-65.
- [4] YUAN P, MAO J L, XIANG F H. Improved grid fault diagnosis based on genetic optimization BP neural network[J]. Journal of Power Systems and Automation, 2017, 29(1): 118-122.
- [5] SHAN X, DAI Z M, ZHANG Z. Research and application of integrated intelligent warning system of intelligent power grid dispatching control system[J]. Automation of Electric Power Systems, 2015, 39(1): 65-72.
- [6] BIAN L, BIAN C Y. Overview of intelligent methods for network fault diagnosis[J]. Force System Protection and Control, 2014(3): 146-153.
- [7] JIANG B, LI Z J. Power grid fault diagnosis based on BP neural network[J]. Energy and Environmental Protection, 2017, 39(3): 246-249.
- [8] LIU G Y. Formalized modeling and decision-making planning of emergency plan[D]. Wuhan: Huazhong University of Science and Technology, 2011.
- [9] LIU A J. Structural research on emergency plan of power industry[D]. Beijing: University of Chinese Academy of Sciences, 2012.
- [10] CAO Y K, HE J W, BAO Z A. Research status and prospect of deep learning in the field of electric power[J]. Journal of Shanghai Electric Power University, 2017, 33(4): 341-345.
- [11] DUAN Y J, LU Y S, ZHANG J, et al. Research status and prospect of deep learning in the field of control[J]. Journal of Automation, 2016(5): 643-654.
- [12] ZHOU Q Y. Syntactic analysis of natural language based on deep learning[D]. Harbin: Harbin Institute of Technology, 2016.
- [13] YAN Y. Research on text representation and classification methods based on deep learning[D]. Beijing: University of Science and Technology Beijing, 2016.
- [14] NIU L Q. Research on text vector representation and modeling based on neural network[D]. Nanjing: Nanjing University, 2016.
- [15] ZHAO X S, XIE B M, ZHANG H W. A power grid fault diagnosis method based on deep learning algorithm[J]. Henan Science and Technology, 2016(23): 53-54.
- [16] JIANG Q, SHEN L, ZHANG W, et al. Research on fault diagnosis method based on deep learning[J]. Computer Simulation, 2018, 35(7): 409-413.
- [17] MIKOLOV T, SUTSKEVER I, CHEN K. Distributed Representations of Words and Phrases and their Compositionality[J]. Advances in Neural Information Processing Systems, arXiv: 1310.4546v1, 2013.
- [18] LIN Y, SHI X D, GUO F. A Chinese syntactic analysis based on probabilistic context-free grammar[J]. Chinese Journal of Information Technology, 2006, 20(2): 1-7.
- [19] CHARNIAK E. Parsing With Context-free Grammar and Word Statistics[C]//Statistical Parsing with a Context-free Grammar and Word Statistics. Proc, 1997.
- [20] SOCHER R, BAUER J, MANNING C D, et al. Parsing with Compositional Vector Grammar[C]//Proceedings of 51st Annual Meeting of the Association for Computational Linguistics(Volume 1: Long papers). 2013: 455-465.



SHI He, born in 1995, postgraduate, is a member of China Computer Federation. Her main research interests include text mining.



YANG Qun, born in 1971, Ph.D, professor, is a member of China Computer Federation. Her main research interests include algorithm design.